



A Text for SCI103 and SCI104
Assembled by Nikolaus Sucher

Biology

A Text For
Biology I (SCI103) And Biology II (SCI104)
at Roxbury Community College

Nikolaus Sucher

Contents

Acknowledgements	xxxiv
1 Biology: The Science of Life	1
1.1 Characteristics of Life	3
1.2 Origin of Life	3
1.3 Foundations of Modern Biology	6
1.4 Areas And Levels of Study And Research in Biology	9
1.5 Science	12
1.6 Philosophy of Science	16
1.7 Science And Religion	19
2 Basic General Chemistry For Biology	23
2.1 Matter	23
2.2 Chemical Bonds, Molecules And Compounds	32
2.3 Energy	33
2.4 Ions And Salts	37
2.5 Acids And Bases	37
2.6 Redox	39
2.7 Water	40
3 Basic Organic Chemistry For Biology	43
3.1 Functional groups	44
3.2 Biomolecules	46
3.3 Proteins	48
3.4 Carbohydrates	53
3.5 Lipids	54
3.6 Nucleic Acids	57
5 The Cell Membrane	81
5.1 Structure And Function of The Cell Membrane	86
5.2 Membrane Permeability	88
5.3 The Extracellular Matrix	90
5.4 Cell Junctions	91
6 Bioenergetics	93
6.1 Thermodynamics of Living Organisms	94
6.2 Metabolism	96
6.3 Energy Transformations	100
6.4 Enzymes	102
7 Photosynthesis And Cellular Respiration	111
7.1 The Electron Transport Chain And Chemiosmosis	111
7.2 Photosynthesis	112
7.3 Symbiosis And The Origin of Chloroplasts	117
7.4 Cyanobacteria And The Evolution of Photosynthesis	118
7.5 Cellular Respiration	118

7.6 Fermentation	120
7.7 Anaerobic Respiration	121
8 The Cell Cycle And Cell Division	123
8.1 The Prokaryotic Cell Cycle	123
8.2 Binary Fission	123
8.3 The Eukaryotic Cell Cycle	124
8.4 Chromosomes: The Vectors of Heredity	127
8.5 Mitosis	130
8.6 Regulation of The Eukaryotic Cell Cycle	134
9 Sexual Reproduction And Meiosis	139
9.1 Meiosis	139
9.2 Chromosomal Disorders	144
10 Classical Genetics	149
10.1 Mendelian Inheritance	149
10.2 Mendelian Traits in Humans	155
10.3 Non-Mendelian Inheritance	156
11 Molecular Biology of The Gene	157
11.1 Structure And Function of Deoxyribonucleic Acid	158
11.2 Structure And Function of Ribonucleic Acid	172
11.3 Gene Expression	176
11.4 Techniques of Molecular Genetics	186
12 Regulation of Gene Expression	195
12.1 The <i>lac</i> operon	195
12.2 The <i>trp</i> operon	200
12.3 Regulated stages of gene expression	201
12.4 Post-transcriptional regulation	201
12.5 Regulation of translation	202
13 Bacteria, Archaea And Viruses	203
13.1 Bacteria	204
13.2 Archaea	214
13.3 Viruses	223
14 Protists	235
14.1 Protozoa	239
14.2 Flagellates	242
14.3 Ciliates	243
14.4 Amoebozoa	245
14.5 Algae	246
14.6 Slime Molds	248
15 The Theory of Evolution	251
15.1 History of Evolutionary Thought	252
15.2 The Modern Synthesis	253
15.3 Mechanisms of Evolution	254
15.4 Outcomes of Evolution	259
16 Taxonomy, Binomial Nomenclature And Systematics	263
16.1 The Concept of Species in Biology	264
16.2 Binomial Nomenclature	265
16.3 Taxonomy And Systematics	268
16.4 Kingdoms And Domains	270

16.5 Cladistics	271
17 Fungi	275
17.1 Morphology of Fungi	277
17.2 Growth And Physiology of Fungi	278
17.3 Reproduction of Fungi	278
17.4 Evolution of Fungi	279
17.5 Fungus-like organisms	282
17.6 Ecology of Fungi	282
17.7 Fungi as Pathogens And Parasites	284
17.8 Mycotoxins	284
18 Plants	287
18.1 Embryophytes	289
18.2 Non-vascular Plants	290
18.3 Vascular Plants	291
18.4 Seed Plants	294
18.5 Gymnosperms	294
18.6 Angiosperms	296
18.7 Plant Anatomy	303
18.8 Plant Tissues	303
18.9 Biochemistry of Plants	306
18.10 Plant Secondary Metabolism	307
18.11 Tropisms And Nastic Movements	308
18.12 Plant Reproduction	308
18.13 The Fruit	311
19 Animals	315
19.1 Non-bilaterian animals	320
19.2 Bilaterian animals	331
19.3 Chordata	339
19.4 Tissues, Organs And Organ Systems	357
19.5 Human evolution	361
20 Circulatory Systems	377
20.1 Closed Circulatory System	377
20.2 Open Circulatory Systems	378
20.3 No Circulatory System	378
20.4 The Human Cardiovascular System	378
20.5 The Heart	379
20.6 The Blood Flow	382
20.7 Arteries	384
20.8 Capillaries	384
20.9 Veins	384
20.10 The Systemic Circulation	384
20.11 The Pulmonary Circulation	384
20.12 Cardiovascular Disease	385
20.13 History	385
21 The Lymphatic And Immune Systems	387
21.1 The Lymphatic System	387
21.2 The Immune System	391
21.3 The Innate Immune System	392
21.4 Cellular Components of The Innate Immune System	393
21.5 Inflammation	395

21.6 The Complement System	396
21.7 The Adaptive Immune System	398
21.8 The humoral Immune Response	402
22 Respiratory Systems	405
22.1 The Mammalian Respiratory System	405
22.2 Gas Exchange	407
22.3 Control of Ventilation	409
22.4 The Respiratory System of Birds	411
22.5 The Respiratory System of Reptiles	412
22.6 The Respiratory System of Amphibians	412
22.7 The Respiratory System of Fish	413
22.8 The Respiratory System of Arthropods	414
22.9 The Respiratory System of Molluscs	414
22.10 The Respiratory System of Plants	414
23 Digestive Systems	417
23.1 The Human Digestive System	418
23.2 The Digestive System of Fish	427
23.3 The Digestive System of Annelids	428
24 Osmoregulation And Excretory Systems	429
24.1 Nitrogenous Waste	430
24.2 Excretory Systems	430
24.3 The Human Urinary System	431
25 Nervous Systems	435
25.1 The Cells Of The Nervous System	436
25.2 Comparative Anatomy And Evolution Of Nervous Systems	437
25.3 The Human Brain	442
25.4 Development Of The Nervous System	444
25.5 The Function Of The Nervous System	445
25.6 The Sensory System	445
25.7 The Motor System	446
25.8 Neuronal Signalling	446
25.9 Neurons And Glial Cells	448
25.10 Electrical Basis Of Neuronal Function	453
25.11 Neurotransmission	465
26 Sensation: Receptors, Organs And Systems	473
26.1 Sensory Receptors	474
26.2 The Visual System	476
26.3 The Auditory And Vestibular Systems	486
26.4 The Auditory System	487
26.5 The Vestibular System	490
26.6 The Somatic Sensory System	492
26.7 The Olfactory System	498
26.8 The Gustatory System	502
27 Animal Locomotion And Support Systems	507
27.1 The Skeleton	507
27.2 The Human Musculoskeletal System	509
27.3 The Muscular System	516
28 Endocrine Systems	525
28.1 The Hypothalamus	526

28.2 Hypothalamic-pituitary-thyroid axis	531
28.3 Hypothalamic-pituitary-gonadal axis	531
28.4 Hypothalamic-pituitary-adrenal axis	533
28.5 The Renin-angiotensin System	534
29 Reproductive Systems	537
29.1 The Human Reproductive System	539
29.2 The Female Reproductive System	539
29.3 The Male Reproductive System	539
30 Development And Aging	543
30.1 Animal Development	543
30.2 Genetic Control of Development	545
30.3 Human Embryonic Development	546
30.4 Aging	552

List of Figures

1.1	A Common Eastern Bumble Bee (<i>Bombus impatiens</i>) picking up nectar and pollen from a spear cistle (<i>Cirsium vulgare</i>) on Winter Island in Salem, Massachusetts.	1
1.2	The Miller-Urey experiment ¹ was a chemical experiment that simulated the conditions thought at the time (1952) to be present on the early Earth and tested the chemical origin of life under those conditions. The experiment at the time supported Alexander Oparin's and J. B. S. Haldane's hypothesis that putative conditions on the primitive Earth favoured chemical reactions that synthesized more complex organic compounds from simpler inorganic precursors. Considered to be the classic experiment investigating abiogenesis, it was conducted in 1952 by Stanley Miller, with assistance from Harold Urey, at the University of Chicago and later the University of California, San Diego and published the following year.	5
1.3	Cartoons of a eukaryotic and prokaryotic cell. ²	6
1.4	Tree diagram ³ illustrating the evolutionary relationship of living organisms]	6
1.5	Modern stromatolites in Shark Bay, Western Australia ⁴	8
1.6	Eukaryotes and some examples of their diversity ⁵ – clockwise from top left: Red mason bee, Boletus edulis, chimpanzee, Isotricha intestinalis, Ranunculus asiaticus, and Volvox carteri.	11
1.7	The hierarchy of biological classification's eight major taxonomic ranks from the most specific (top) to the most general (bottom). Intermediate minor rankings are not shown. This diagram uses a 3 Domains / 6 Kingdoms format ⁶	11
1.8	A range of animal behaviours. ⁷	12
1.9	Two scientists working at the National Cancer Institute which is part of the National Institutes of Health of the United States of America. ⁸	12
1.10	At First Solvay Conference (1911), Marie Curie (seated, second from right) confers with Henri Poincaré; standing, fourth from right, is Rutherford; second from right, Einstein; far right, Paul Langevin ⁹	13
1.11	George Washington Carver ¹⁰ (1860s – January 5, 1943) was an American agricultural scientist and inventor who promoted alternative crops to cotton and methods to prevent soil depletion. He was the most prominent black scientist of the early 20th century. Apart from his work to improve the lives of farmers, Carver was also a leader in promoting environmentalism. He received numerous honors for his work, including the Spingarn Medal of the NAACP. In an era of high racial polarization, his fame reached beyond the black community. He was widely recognized and praised in the white community for his many achievements and talents. In 1941, Time magazine dubbed Carver a "Black Leonardo".	13
2.1	An illustration of the helium atom ¹¹ , depicting the nucleus (pink) and the electron cloud distribution (black). The nucleus (upper right) in helium-4 is in reality spherically symmetric and closely resembles the electron cloud, although for more complicated nuclei this is not always the case. The black bar is one angstrom (10^{-10} m or 100 pm).	25

¹<https://commons.wikimedia.org/wiki/File:MUexperiment.png>

²<https://commons.wikimedia.org/wiki/File:Celltypes.svg>

³https://commons.wikimedia.org/wiki/File:Tree_of_Living_Organisms_2.png

⁴https://commons.wikimedia.org/wiki/File:Stromatolites_in_Sharkbay.jpg

⁵https://commons.wikimedia.org/wiki/File:Eukaryota_diversity_2.jpg

⁶https://commons.wikimedia.org/wiki/File:Biological_classification_L_Pengo.svg

⁷https://commons.wikimedia.org/wiki/File:Ethology_diversity_2.jpg

⁸https://commons.wikimedia.org/wiki/File:Researcher_looking_through_microscope.jpg

⁹https://commons.wikimedia.org/wiki/File:1911_Solvay_conference.jpg

¹⁰https://en.wikipedia.org/wiki/George_Washington_Carver

¹¹https://commons.wikimedia.org/wiki/File:Helium_atom_QM.svg

2.2 Niels Bohr's 1913 quantum model of the atom ¹² , which incorporated an explanation of Johannes Rydberg's 1888 formula, Max Planck's 1900 quantum hypothesis, i.e. that atomic energy radiators have discrete energy values ($\epsilon = h\nu$), J. J. Thomson's 1904 plum pudding model, Albert Einstein's 1905 light quanta postulate, and Ernest Rutherford's 1907 discovery of the atomic nucleus. Note that the electron does not travel along the black line when emitting a photon. It "jumps", disappearing from the outer orbit and appearing in the inner one and cannot exist in the space between orbits 2 and 3. The Bohr model of the hydrogen atom ($Z = 1$) or a hydrogen-like ion ($Z > 1$), where the negatively charged electron confined to an atomic shell encircles a small, positively charged atomic nucleus and where an electron jumps between orbits, is accompanied by an emitted or absorbed amount of electromagnetic energy ($h\nu$). The orbits in which the electron may travel are shown as grey circles; their radius increases as n^2 , where n is the principal quantum number. The $3 \rightarrow 2$ transition depicted here produces the first line of the Balmer series, and for hydrogen ($Z = 1$) it results in a photon of wavelength 656 nm (red light).	26
2.3 The five filled atomic orbitals of a neon atom ¹³ separated and arranged in order of increasing energy from left to right, with the last three orbitals being equal in energy. Each orbital holds up to two electrons, which most probably exist in the zones represented by the colored bubbles. Each electron is equally present in both orbital zones, shown here by color only to highlight the different wave phase.	28
2.4 A simple depiction of the periodic table of the elements. ¹⁴	29
2.5 Examples of Lewis dot-style representations of chemical bonds ¹⁵ between carbon (C), hydrogen (H), and oxygen (O). Lewis dot diagrams were an early attempt to describe chemical bonding and are still widely used today.	33
2.6 Formation of an ionic bond. ¹⁶ Sodium and fluorine atoms undergoing a redox reaction to form sodium fluoride. Sodium loses its outer electron to give it a stable electron configuration, and this electron enters the fluorine atom exothermically. The oppositely charged ions – typically a great many of them – are then attracted to each other to form a solid.	33
2.7 A covalent bond between two hydrogen atoms forming a hydrogen molecule. ¹⁷	33
2.8 Basic overview of energy and human life. ¹⁸	34
2.9 Joule's apparatus for measuring the mechanical equivalent of heat. A descending weight attached to a string causes a paddle immersed in water to rotate. ¹⁹	34
2.10 Example of an enzyme-catalysed exothermic reaction. ²⁰	35
2.11 3D diagram showing the pyramidal structure of the hydroxonium ion. ²¹	37
2.12 Lewis structure of the hydroxide ion showing three lone pairs on the oxygen atom ²²	37
2.13 Model of hydrogen bonds (1) between molecules of water. ²³	42
3.1 This diagram ²⁴ shows 5 different structural representations of the organic compound butane. The left-most structure is a bond-line drawing where the hydrogen atoms are removed. The 2nd structure has the hydrogens added depicted—the dark wedged bonds indicate the hydrogen atoms are coming toward the reader, the hashed bonds indicate the atoms are oriented away from the reader, and the solid (plain) bonds indicate the bonds are in the plane of the screen/paper. The middle structure shows the four carbon atoms. The 4th structure is a representation just showing the atoms and bonds without 3-dimensions. The right-most structure is a condensed structure representation of butane.	44
3.2 Biologically important functional groups.	44

¹²https://en.wikipedia.org/wiki/Bohr_model#/media/File:Bohr_atom_model.svg

¹³<https://commons.wikimedia.org/wiki/File:S-p-Orbitals.svg>

¹⁴https://commons.wikimedia.org/wiki/File:Simple_Periodic_Table_Chart-en.svg

¹⁵https://en.wikipedia.org/wiki/Chemical_bond#/media/File:Electron_dot.svg

¹⁶<https://commons.wikimedia.org/wiki/File:NaF.gif>

¹⁷https://commons.wikimedia.org/wiki/File:Covalent_bond_hydrogen.svg

¹⁸https://commons.wikimedia.org/wiki/File:Energy_and_life.svg

¹⁹[https://commons.wikimedia.org/wiki/File:Joule%27s_Apparatus_\(Harper%27s_Scan\).png](https://commons.wikimedia.org/wiki/File:Joule%27s_Apparatus_(Harper%27s_Scan).png)

²⁰https://commons.wikimedia.org/wiki/File:Activation2_updated.svg

²¹<https://commons.wikimedia.org/wiki/File:Hydroxonium-cation.svg>

²²https://commons.wikimedia.org/wiki/File:Hydroxide_lone_pairs-2D.svg

²³https://commons.wikimedia.org/wiki/File:3D_model_hydrogen_bonds_in_water.svg

²⁴https://commons.wikimedia.org/wiki/File:Structural_drawings_of_butane_854px.jpg

3.3	Chemical structure of the peptide bond (bottom) and the three-dimensional structure of a peptide bond between an alanine and an adjacent amino acid (top/inset). The bond itself is made of the CHON elements. ²⁵	48
3.4	The amino-acid sequence, the primary structure of a protein, determines the secondary (α -helix and β -sheet), tertiary and quaternary protein structures ²⁶	51
3.5	The disaccharide sucrose ²⁷	53
3.6	The disaccharide lactose ²⁸	54
3.7	Structures of some common lipids. ²⁹ At the top are cholesterol and oleic acid. The middle structure is a triglyceride composed of oleoyl, stearoyl, and palmitoyl chains attached to a glycerol backbone. At the bottom is the common phospholipid phosphatidylcholine.	55
3.8	Example of an unsaturated fat triglyceride ³⁰ ($C_{55}H_{98}O_6$). Left part: glycerol; right part, from top to bottom: palmitic acid, oleic acid, alpha-linolenic acid.	57
3.9	The structure of the DNA double helix. A section of DNA. The bases lie horizontally between the two spiraling strands. The atoms in the structure are colour-coded by element (based on atomic coordinates of PDB 1bna ³¹ rendered with open source molecular visualization tool PyMol.)	58
5.1	Picture of a molecular dynamics simulation ³² of a cell membrane/protein complex consisting of bovine rhodopsin ³³ incorporated of a phosphatidylcholine (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, POPC) lipid bylayer. POPC and water molecules are depicted as sticks. The lipid layers facing the extracellular and cytoplasmic spaces are shown in white and blue, respectively. Both the extra- and intracellular interfaces are covered with layers of water. The secondary structure of rhodopsin is depicted in rainbow colored cartoon representation. Potassium and chloride ions are shown as spheres (colored in cyan and green, respectively). Image generated from PDB file ³⁴ obtained from the CHARMM-GUI Archive – Protein/Membrane Complex Library ³⁵ using the open source molecular visualization tool PyMol ³⁶ .	81
5.2	Simplified diagram of two types of cell junctions. ³⁷	92
6.1	Structure of adenosine triphosphate (ATP), protonated ³⁸	93
6.2	The Sun ³⁹ is the source of energy for most of life on Earth. It derives its energy mainly from nuclear fusion in its core, converting mass to energy as protons are combined to form helium. This energy is transported to the sun's surface then released into space mainly in the form of radiant (light) energy.	95
6.3	Simplified view of the cellular metabolism ⁴⁰	96
6.4	Simplified diagram of catabolism of proteins, carbohydrates and fats. ⁴¹	99
6.5	Example of an enzyme-catalysed exothermic reaction ⁴²	103
7.1	Schematic of photosynthesis in plants. The carbohydrates produced are stored in or used by the plant which in turn may provide food for heterotrophic organisms such as animals. ⁴³	111
7.2	The electron transport chain ⁴⁴ in the mitochondrion is the site of oxidative phosphorylation in eukaryotes. The NADH and succinate generated in the citric acid cycle are oxidized, providing energy to power ATP synthase.	112

²⁵<https://commons.wikimedia.org/wiki/File:Peptide-Figure-Revised.png>

²⁶(https://commons.wikimedia.org/wiki/File:Protein_structure.png)

²⁷<https://commons.wikimedia.org/wiki/File:Beta-D-Lactose.svg>

²⁸<https://commons.wikimedia.org/wiki/File:Beta-D-Lactose.svg>

²⁹https://commons.wikimedia.org/wiki/File:Common_lipids_lmaps.png

³⁰https://commons.wikimedia.org/wiki/File:Fat_triglyceride_shorthand_formula.PNG

³¹<https://www.rcsb.org/structure/1bna>

³²<https://doi.org/10.1371/journal.pone.0000880>

³³<https://en.wikipedia.org/wiki/Rhodopsin>

³⁴http://www.charmm-gui.org/archive/complex/1gzm_rect_popc.pdb

³⁵<http://www.charmm-gui.org/?doc=archive&lib=complex>

³⁶<https://pymol.org/2/>

³⁷https://commons.wikimedia.org/wiki/File:Cell_junction_simplified_en.svg

³⁸https://commons.wikimedia.org/wiki/File:Adenosintriphosphat_protoniert.svg

³⁹[https://commons.wikimedia.org/wiki/File:Sun_in_February_\(black_version\).jpg](https://commons.wikimedia.org/wiki/File:Sun_in_February_(black_version).jpg)

⁴⁰<https://commons.wikimedia.org/wiki/File:Metabolism.png>

⁴¹https://commons.wikimedia.org/wiki/File:Catabolism_schematic.svg

⁴²https://commons.wikimedia.org/wiki/File:Activation2_updated.svg

⁴³https://commons.wikimedia.org/wiki/File:Photosynthesis_en.svg

⁴⁴https://commons.wikimedia.org/wiki/File:Mitochondrial_electron_transport_chain—Etc4.svg

7.3	Photosynthesis changes sunlight into chemical energy, splits water to liberate O ₂ , and fixes CO ₂ into sugar. ⁴⁵	112
7.4	Absorbance spectra of free chlorophyll ⁴⁶ a (blue) and b (red) in a solvent. The action spectra of chlorophyll molecules are slightly modified in vivo depending on specific pigment-protein interactions.	113
7.5	Chloroplast ultrastructure ⁴⁷ : 1. outer membrane 2. intermembrane space 3. inner membrane (1+2+3: envelope) 4. stroma (aqueous fluid) 5. thylakoid lumen (inside of thylakoid) 6. thylakoid membrane 7. granum (stack of thylakoids) 8. thylakoid (lamella) 9. starch 10. ribosome 11. plastidial DNA 12. plastoglobule (drop of lipids)	114
7.6	The Z scheme ⁴⁸	114
7.7	Light-dependent reactions of photosynthesis at the thylakoid membrane ⁴⁹	115
7.8	Overview of the Calvin cycle and carbon fixation ⁵⁰	116
7.9	Photorespiration ⁵¹ 1. ribulose 1,5-bisphosphate 2. 3-Phosphoglycerate 3. 2-phosphoglycolate 4. glycolate 5. glyoxylate 6. glycine 7. serine 8. hydroxypyruvate 9. glycerate CC Calvin cycle	116
7.10	Overview of C4 carbon fixation ⁵²	117
7.11	A diagram of cellular respiration including glycolysis, Krebs cycle (AKA citric acid cycle), and the electron transport chain ⁵³	118
7.12	Stoichiometry of aerobic respiration ⁵⁴ and most known fermentation types in eucaryotic cell. Numbers in circles indicate counts of carbon atoms in molecules, C6 is glucose C ₆ H ₁₂ O ₆ , C1 carbon dioxide CO ₂ . Mitochondrial outer membrane is omitted.	119
8.1	Many bacteria reproduce through binary fission, which is compared to mitosis and meiosis in this image. ⁵⁵	123
8.2	A diagram of mitosis stages ⁵⁶ . Interphase (G ₂): In this substage, the cell prepares for nuclear division and a protein that makes microtubules for cell division is synthesized. Prophase: The longest stage of mitosis. In this stage the chromosomes become visible and the centrioles separate and move to opposite poles of the cell. Prometaphase: The nuclear envelope disintegrates and microtubules can attach to kinetochores. Chromosomes congress toward the metaphase plate of the cell. Metaphase: In this stage the chromosomes line up across the center of the cell and become connected to the spindle fiber at their centromere. Anaphase: In this stage the sister chromatids separate into individual chromosomes and are pulled apart. Telophase & cytokinesis: Chromosomes decondense and are surrounded by a newly formed nuclear envelope. Cytokinesis typically coincides with and telophase.	127
8.3	The karyotype of a normal human male consisting of 22 pairs of autosomes and 2 sex chromosomes (one x and one y chromosome). ⁵⁷ The karyotype is the characteristic chromosome complement of a eukaryote species.	127
8.4	The major structures in DNA compaction: DNA, the nucleosome, the 10 nm "beads-on-a-string" fibre, the 30 nm fibre and the metaphase chromosome. ⁵⁸	128
8.5	Mitosis in an animal cell (phases ordered counter-clockwise). ⁵⁹	130
8.6	Onion (<i>Allium cepa</i>) root cells in different phases of the cell cycle (drawn by E. B. Wilson ⁶⁰ , 1900). General view of cells in the growing root-tip of the onion, from a longitudinal section, enlarged 800 diameters. a. non-dividing cells, with chromatin-network and deeply stained nucleoli; b. nuclei preparing for division (spireme-stage); c. dividing cells showing mitotic figures; e. pair of daughter-cells shortly after division. ⁶¹	131

⁴⁵https://commons.wikimedia.org/wiki/File:Simple_photosynthesis_overview.svg

⁴⁶https://commons.wikimedia.org/wiki/File:Chlorophyll_ab_spectra-en.svg

⁴⁷<https://commons.wikimedia.org/wiki/File:Chloroplast.svg>

⁴⁸<https://commons.wikimedia.org/wiki/File:Z-scheme.png>

⁴⁹https://commons.wikimedia.org/wiki/File:Thylakoid_membrane_3.svg

⁵⁰<https://commons.wikimedia.org/wiki/File:Calvin-cycle4.svg>

⁵¹<https://commons.wikimedia.org/wiki/File:Photorespiration.svg>

⁵²<https://commons.wikimedia.org/wiki/File:HatchSlackpathway2.svg>

⁵³<https://commons.wikimedia.org/wiki/File:CellRespiration.svg>

⁵⁴https://commons.wikimedia.org/wiki/File:Cellular_respiration.gif

⁵⁵https://commons.wikimedia.org/wiki/File:Three_cell_growth_types.svg

⁵⁶https://commons.wikimedia.org/wiki/File:Mitosis_Stages.svg

⁵⁷https://commons.wikimedia.org/wiki/File:NHGRI_human_male_karyotype.png

⁵⁸https://commons.wikimedia.org/wiki/File:Chromatin_Structures.png

⁵⁹https://commons.wikimedia.org/wiki/File:Animal_cell_cycle-en.svg

⁶⁰https://en.wikipedia.org/wiki/Edmund_Beecher_Wilson

⁶¹<https://commons.wikimedia.org/wiki/File:Wilson1900Fig2.jpg>

8.7 Early prophase: Polar microtubules, shown as green strands, have established a matrix around the currently intact nucleus, with the condensing chromosomes in blue. The red nodules are the centromeres. ⁶²	132
8.8 Early prometaphase: The nuclear membrane has just disassembled, allowing the microtubules to quickly interact with the kinetochores, which assemble on the centromeres of the condensing chromosomes. ⁶³	133
8.9 Metaphase: The centrosomes have moved to the poles of the cell and have established the mitotic spindle. The chromosomes have congressed at the metaphase plate. ⁶⁴	133
8.10 Anaphase: Kinetochore microtubules pull the two sets of chromosomes apart, and lengthening polar microtubules push the halves of the dividing cell further apart, while chromosomes are condensed maximally. ⁶⁵	133
8.11 Telophase: Reversal of prophase and prometaphase events and thus completing the cell cycle. ⁶⁶	134
9.1 Sperm fertilizing an egg. ⁶⁷	139
9.2 In meiosis ⁶⁸ , the chromosome or chromosomes duplicate (during interphase) and homologous chromosomes exchange genetic information (chromosomal crossover) during the first division, called meiosis I. The daughter cells divide again in meiosis II, splitting up sister chromatids to form haploid gametes. Two gametes fuse during fertilization, creating a diploid cell with a complete set of paired chromosomes.	140
9.3 Diagram of the meiotic phases ⁶⁹	141
10.1 Gregor Mendel, the Moravian Augustinian monk who is credited for having founded the modern science of genetics ⁷⁰	149
10.2 Characteristics of pea plants Gregor Mendel used in his inheritance experiments ⁷¹	151
10.3 P-Generation and F1-Generation: The dominant allele for purple-red flower hides the phenotypic effect of the recessive allele for white flowers. F2-Generation: The recessive trait from the P-Generation phenotypically reappears in the individuals that are homozygous with the recessive genetic trait. ⁷²	151
10.4 In the F1 generation all individuals have the same genotype and same phenotype expressing the dominant trait (red). In the F2 generation, the phenotypes show a 3:1 ratio. In the genotype 25% are homozygous with the dominant trait, 50% are heterozygous genetic carriers of the recessive trait, 25% are homozygous with the recessive genetic trait and expressing the recessive character. ⁷³	152
10.5 In <i>Mirabilis jalapa</i> and <i>Antirrhinum majus</i> are examples for intermediate inheritance. As seen in the F1-generation, heterozygous plants have “light pink” flowers—a mix of “red” and “white”. The F2-generation shows a 1:2:1 ratio of red : light pink : white ⁷⁴	152
10.6 Segregation and independent assortment are consistent with the chromosome theory of inheritance. ⁷⁵	153
10.7 Punnett square for homozygous cross.	154
10.8 Punnett square for heterozygous cross.	154
10.9 Monohybrid cross	154
10.10 Dihybrid cross	155
11.1 A cartoon representation of DNA based on atomic coordinates of PDB 1BNA ⁷⁶ , rendered with open source molecular visualization tool PyMol.	158
11.2 The structure of the DNA double helix. A section of DNA. The bases lie horizontally between the two spiraling strands. The atoms in the structure are colour-coded by element (based on atomic coordinates of PDB 1bna ⁷⁷ rendered with open source molecular visualization tool PyMol.)	158

⁶²<https://commons.wikimedia.org/wiki/File:Prophaself.jpg>

⁶³<https://commons.wikimedia.org/wiki/File:Prometaphase.jpg>

⁶⁴<https://commons.wikimedia.org/wiki/File:Metaphaself.jpg>

⁶⁵https://commons.wikimedia.org/wiki/File:Anaphase_IF.jpg

⁶⁶<https://commons.wikimedia.org/wiki/File:Telophaself.jpg>

⁶⁷<https://commons.wikimedia.org/wiki/File:Sperm-egg.jpg>

⁶⁸https://commons.wikimedia.org/wiki/File:Meiosis_Overview_new.svg

⁶⁹https://commons.wikimedia.org/wiki/File:Meiosis_Stages.svg

⁷⁰https://commons.wikimedia.org/wiki/File:Gregor_Mendel.png

⁷¹https://commons.wikimedia.org/wiki/File:Gregor_Mendel_-_characteristics_of_pea_plants_-_english.png

⁷²https://commons.wikimedia.org/wiki/File:Dominant-recessive_inheritance_-_flowers_of_pea_plants.png

⁷³https://commons.wikimedia.org/wiki/File:Dominant-recessive_inheritance_P_-_F1_-_F2.png

⁷⁴https://commons.wikimedia.org/wiki/File:Intermediate_inheritance_P_-_F1_-_F2.png

⁷⁵https://commons.wikimedia.org/wiki/File:Independent_assortment_%26_segregation.svg

⁷⁶<https://www.rcsb.org/structure/1bna>

⁷⁷<https://www.rcsb.org/structure/1bna>

11.3 The structure of the four nucleotides and their base pairing in the DNA double helix. The atoms in the structure are colour-coded by element (based on atomic coordinates of PDB 1bna ⁷⁸ rendered with open source molecular visualization tool PyMol.)	159
11.4 DNA major and minor grooves. PDB 1bna ⁷⁹ rendered with open source molecular visualization tool PyMol.) ¹⁶⁰	
11.5 Top, a GC base pair ⁸⁰ with three hydrogen bonds. Bottom, an AT base pair ⁸¹ with two hydrogen bonds. Non-covalent hydrogen bonds between the pairs are shown as dashed lines. The two types of base pairs form different numbers of hydrogen bonds, AT forming two hydrogen bonds, and GC forming three hydrogen bonds (see figures, right). DNA with high GC-content is more stable than DNA with low GC-content ¹⁶⁰	
11.6 DNA polymerases adds nucleotides to the 3' end of a strand of DNA. If a mismatch is accidentally incorporated, the polymerase is inhibited from further extension. Proofreading removes the mismatched nucleotide and extension continues. ⁸²	163
11.7 Many enzymes are involved in forming the DNA replication fork and DNA polymerization. ⁸³	164
11.8 DNA polymerases adds nucleotides to the 3' end of a strand of DNA. If a mismatch is accidentally incorporated, the polymerase is inhibited from further extension. Proofreading removes the mismatched nucleotide and extension continues. ⁸⁴	165
11.9 Tertiary structure of tRNA (based on atomic coordinates of PDB 1ehz ⁸⁵ rendered with open source molecular visualization tool PyMol.)	174
11.10 RNA polymerase (RNAP) at work. Note the coding and template strands. The resulting RNA is a synthesized from the template strand and identical in sequence to the coding strand. ⁸⁶	178
11.11 Crystal structure of the bacterial 70S ribosome of the bacterium <i>Thermus thermophilus</i> . The 30S (small) ribosomal subunit proteins are colored in green, the 50S (large) subunit proteins are colored in blue, the ribosomal RNA is colored orange. The 30S subunits contains 3 tRNA molecules (based on atomic coordinates of PDB 1JGQ ⁸⁷ and PDB 1GIY ⁸⁸ rendered with open source molecular visualization tool PyMol.) ¹⁸¹	
11.12 Crystal structure of the human 80S ribosome (based on atomic coordinates of PDB 4V6X ⁸⁹ rendered with open source molecular visualization tool PyMol). The 40S (small) ribosomal subunit proteins are shown in lightblue, the 60S (large) subunit proteins in palegreen, the ribosomal RNA in orange.	182
11.13 Diagram showing the translation of mRNA and the synthesis of proteins by a ribosome. ⁹⁰	183
11.14 Diagram of the CRISPR prokaryotic antiviral defense mechanism. ⁹¹	191
12.1 ⁹²	196
12.2 A simulated structural model ⁹³ of a complex between the lac repressor protein (LacI) and a 107-bp-long DNA segment. Two dimeric LacI functional subunits (green + blue and yellow + orange) each bind a DNA operator sequence (top). The mobility of the DNA-binding head groups coupled to the stable body of the body of LacI provide the force for the looping of the DNA (Ville et al. ⁹⁴).	197
12.3 Structure of isopropyl β-D-thiogalactopyranoside (IPTG) ⁹⁵	198
12.4 Analysis of lac regulatory mutants by Lac complementation. ⁹⁶	199
13.1 Scanning electron micrograph of Escherichia coli bacteria. ⁹⁷	203

⁷⁸<https://www.rcsb.org/structure/1bna>

⁷⁹<https://www.rcsb.org/structure/1bna>

⁸⁰https://commons.wikimedia.org/wiki/File:Base_pair_GC.svg

⁸¹https://commons.wikimedia.org/wiki/File:Base_pair_AT.svg

⁸²https://commons.wikimedia.org/wiki/File:DNA_replication_split.svg

⁸³https://commons.wikimedia.org/wiki/File:Eukaryotic_DNA_replication.svg

⁸⁴https://commons.wikimedia.org/wiki/File:DNA_polymerase.svg

⁸⁵<https://www.rcsb.org/structure/1ehz>

⁸⁶https://commons.wikimedia.org/wiki/File:Simple_transcription_elongation1.svg

⁸⁷<https://www.rcsb.org/structure/1jgq>

⁸⁸(<https://www.rcsb.org/structure/1jgq>)

⁸⁹<https://www.rcsb.org/structure/4v6x>

⁹⁰https://commons.wikimedia.org/wiki/File:Ribosome_mRNA_translation_en.svg

⁹¹<https://commons.wikimedia.org/wiki/File:Crispr.png>

⁹²https://commons.wikimedia.org/wiki/File:Lac_operon1.png

⁹³<https://www.modelarchive.org/doi/10.5452/ma-cwr9z>

⁹⁴<https://www.pnas.org/content/102/19/6783>

⁹⁵<https://commons.wikimedia.org/wiki/File:IPTG2.svg>

⁹⁶https://commons.wikimedia.org/wiki/File:Lac_complementation.png

⁹⁷https://commons.wikimedia.org/wiki/File:EscherichiaColi_NIAID.jpg

13.2 Halobacterium sp. strain NRC-1, each cell about 5 μm long ⁹⁸	203
13.3 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the strain of coronavirus that causes coronavirus disease 2019 (COVID-19), the respiratory illness responsible for the COVID-19 pandemic. ⁹⁹	204
13.4 Bacteria display various cell morphologies and arrangements ¹⁰⁰	205
13.5 Helicobacter pylori electron micrograph ¹⁰¹ , showing multiple flagella on the cell surface	207
13.6 A stained preparation of the cell <i>Bacillus subtilis</i> showing endospores as green and the vegetative cell as red. ¹⁰²	207
13.7 Formation of an endospore through the process of sporulation. ¹⁰³	207
13.8 A comparison of the membrane structures of Archaea, Bacteria and Eukarya. Top, an archaeal phospholipid: 1, isoprene chains; 2, ether linkages; 3, L-glycerol moiety; 4, phosphate group. Middle, a bacterial or eukaryotic phospholipid: 5, fatty acid chains; 6, ester linkages; 7, D-glycerol moiety; 8, phosphate group. Bottom: 9, lipid bilayer of bacteria and eukaryotes; 10, lipid monolayer of some archaea. ¹⁰⁴	219
13.9 Structure of tobacco mosaic virus: RNA coiled in a helix of repeating protein sub-units ¹⁰⁵	224
13.10 Structure of an icosahedral cowpea mosaic virus ¹⁰⁶	225
13.11 This negative-stained transmission electron micrograph (TEM) depicts the ultrastructural details of an influenza virus particle, or “virion”. A member of the taxonomic family Orthomyxoviridae, the influenza virus is an enveloped, single-stranded RNA virus. Eight helical capsids are surrounded by the viral envelope. Particle diameter: 80–120 nm ¹⁰⁷	226
13.12 A typical virus replication cycle. ¹⁰⁸	227
13.13 Taxonomy and replication strategies of RNA viruses. ¹⁰⁹ A) Simplified taxonomy of the genome architecture of the RNA viruses described in this review. See main text for used abbreviations. B) (+RNA virus) Infection with a +RNA virus—as exemplified here with a CoV-like virion—releases a single-stranded RNA genome into the cytoplasm (1). (2) Translation of the 5'-terminal open-reading frame of the genome produces the viral replicase. (3) This multi-enzyme complex includes RdRp activity (orange) and associates with intracellular membranes before –RNA synthesis commences. Newly synthesised –RNAs are subsequently used to produce new +RNAs (4), which are typically capped (yellow) and polyadenylated (polyA). (Retrovirus) HIV-1 genomes are packaged as ssRNA in virions. When the ssRNA is released (1) a cDNA copy is synthesised by the RT (2). The RNA is next degraded by the intrinsic RNase H activity in the RT (3) and the single stranded cDNA converted to dsDNA (4). The dsDNA is imported in the nucleus (5) for integration into the host's genetic material. (–RNA virus) (1) As illustrated here with an IAV-like particle, infection with an –RNA virus releases a viral RNA genome that is associated with a viral polymerase (orange) and nucleoprotein (green). (2) In the case of non-segmented –RNA viruses, these complexes support transcription to produce viral mRNAs or cRNAs. (3) Viral mRNAs are next translated and new viral proteins complex with cRNAs to synthesise new vRNAs. (5) The vRNA-containing complexes of some segmented –RNA viruses are imported into the nucleus of the host cell, where (6) the RdRp produces mRNAs or cRNAs. (7) mRNAs are transported to the cytoplasm, while cRNAs are bound by new viral proteins to form cRNPs for –RNA synthesis. (dsRNA virus) Fully duplexed RNA genomes lack cap and polyA elements. (1) The RdRp (orange), therefore, transcribes the viral genome inside the capsid of the virion (blue and red), so viral mRNAs can be (2) released into the cytoplasm as illustrated here with a rotavirus-like virion. In the cytoplasm the mRNA is translated (3) or replicated by newly synthesised viral RdRps (4)]	228
13.14 Structural overview of T2 phage ¹¹⁰	232

⁹⁸<https://commons.wikimedia.org/wiki/File:Halobacteria.jpg>

⁹⁹https://commons.wikimedia.org/wiki/File:SARS-CoV-2_without_background.png

¹⁰⁰https://commons.wikimedia.org/wiki/File:Bacterial_morphology_diagram.svg

¹⁰¹<https://commons.wikimedia.org/wiki/File:EMpylori.jpg>

¹⁰²https://commons.wikimedia.org/wiki/File:Bacillus_subtilis_Spore.jpg

¹⁰³https://commons.wikimedia.org/wiki/File:Endospore_Formation.png

¹⁰⁴https://commons.wikimedia.org/wiki/File:Archaea_membrane.svg

¹⁰⁵https://commons.wikimedia.org/wiki/File:TMV_structure_simple.png

¹⁰⁶<https://commons.wikimedia.org/wiki/File:CowpeaMosaicVirus3D.png>

¹⁰⁷https://commons.wikimedia.org/wiki/File:Influenza_virus_particle_8430_lores.jpg

¹⁰⁸https://commons.wikimedia.org/wiki/File:HepC_replication.png

¹⁰⁹https://commons.wikimedia.org/wiki/File:18_2014_1695_Fig1_HTML.webp

¹¹⁰https://en.wikipedia.org/wiki/Hershey-Chase_experiment#/media/File:T-evenphage.svg

13.15 Diagram showing bacteriophages infecting a bacterial cell (not to scale; bacteriophages are much smaller than bacteria) ¹¹¹	232
14.1 A sampling of protists. ¹¹² red algae (<i>Chondrus crispus</i>); brown algae (Giant Kelp); ciliate (<i>Frontonia</i>); golden algae (<i>Dinobryon</i>); Foraminifera (<i>Radiolaria</i>); parasitic flagellate (<i>Giardia muris</i>); pathogenic amoeba (<i>Acanthamoeba</i>); amoebozoan slime mold (<i>Fuligo septica</i>)	235
14.2 “Monophyletischer Stammbaum der Organismen” from Generelle Morphologie der Organismen (1866) with the three branches Plantae, Protista, Animalia. ¹¹³	235
14.3 Composite image ¹¹⁴ of various Apicomplexan parasites showing Babesia microti in red blood cells (top left), toxoplasma gondii (top right), a septic eugregarine (bottom left), Lankesteria cystodytae (bottom middle), and Plasmodium falciparum in red blood cells (bottom right).	241
14.4 Euglena.	242
14.5 Composite image ¹¹⁵ of various ciliates: Lacrymaria olor, Paramecium bursaria, Coleps, Dileptus, Stentor coeruleus.	243
14.6 Diagram of the ciliate Paramecium ¹¹⁶	244
14.7 Stages of conjugation in Paramecium caudatum ¹¹⁷	244
14.8 A variety of microscopic unicellular and colonial freshwater algae ¹¹⁸	247
14.9 The sexual life cycle of Laminaria, a representative of some 30 different species of brown algae that are commonly called “kelp”. ¹¹⁹	247
14.10 Fuligo septica, a slime mold ¹²⁰ Fuligo septica is a species of plasmodial slime mold, and a member of the Myxomycetes class. It is commonly known as the scrambled egg slime, or flowers of tan because of its peculiar yellowish, bile-colored appearance. Also known as the dog vomit slime mold, it is common with a worldwide distribution, and it is often found on bark mulch in urban areas after heavy rain or excessive watering. Their spores are produced on or in aerial sporangia and are spread by wind.	249
14.11 Dictyostelium Fruiting Body ¹²¹ Dictyostelium is a genus of single- and multi-celled eukaryotic, phagotrophic bacterivores. Though they are Protista and in no way fungal, they traditionally are known as “slime molds”. They are present in most terrestrial ecosystems as a normal and often abundant component of the soil microflora, and play an important role in the maintenance of balanced bacterial populations in soils.	249
14.12 Dictyostelium colony in process of aggregation ¹²²	250
14.13 Pseudoplasmodium or “slug” of a Dictyostelium ¹²³	250
15.1 Biston betularia f. typica, the white-bodied peppered moth. ¹²⁴	258
15.2 Biston betularia f. carbonaria, the black-bodied peppered moth. ¹²⁵	258
15.3 Typica and carbonaria morphs on the same tree. ¹²⁶ The light-coloured typica (below the bark's scar) is nearly invisible on this pollution-free tree, camouflaging it from predators.	259
15.4 Homologous bones in the limbs of tetrapods. The bones of these animals have the same basic structure, but have been adapted for specific uses. ¹²⁷	260
16.1 Biological classification. ¹²⁸	263

¹¹¹https://commons.wikimedia.org/wiki/File:Phage_injecting_its_genome_into_bacteria.svg

¹¹²https://commons.wikimedia.org/wiki/File:Protist_collage_2.jpg

¹¹³https://commons.wikimedia.org/wiki/File:Haeckel_arbol_bn.png

¹¹⁴https://commons.wikimedia.org/wiki/File:Apicomplexa_Composite_Image.png

¹¹⁵https://commons.wikimedia.org/wiki/File:Ciliate_collage.jpg

¹¹⁶https://commons.wikimedia.org/wiki/File:Paramecium_diagram.png

¹¹⁷https://commons.wikimedia.org/wiki/File:Stages_of_ciliate_conjugation.png

¹¹⁸https://commons.wikimedia.org/wiki/File:Водоросли_пресноводного_водоема.jpg

¹¹⁹https://commons.wikimedia.org/wiki/File:Laminaria_Life_Cycle.png

¹²⁰https://commons.wikimedia.org/wiki/File:Fuligo_septica_BI1.JPG

¹²¹https://commons.wikimedia.org/wiki/File:Dictyostelium_Fruiting_Bodies.JPG

¹²²https://commons.wikimedia.org/wiki/File:Dictyostelium_Aggregation.JPG

¹²³https://commons.wikimedia.org/wiki/File:Dictyostelium_Pseudoplasmodium.JPG

¹²⁴<https://commons.wikimedia.org/wiki/File:Biston.betularia.7200.jpg>

¹²⁵<https://commons.wikimedia.org/wiki/File:Biston.betularia.f.carbonaria.7209.jpg>

¹²⁶https://commons.wikimedia.org/wiki/File:Lichte_en_zwarte_versie_berkenspanner.jpg

¹²⁷https://commons.wikimedia.org/wiki/File:Homology_vertebrates-en.svg

¹²⁸https://commons.wikimedia.org/wiki/File:Biological_classification_L_Pengo_vflip.svg

16.2 The tree of life. ¹²⁹	264
16.3 The butterfly genus <i>Heliconius</i> contains many similar species. ¹³⁰	265
16.4 Cladogram of the primates ¹³¹ showing a monophyletic taxon (a clade: the simians or Anthropoidea, in yellow), a paraphyletic taxon (the prosimians, in blue, including the red patch), and a polyphyletic taxon (the nocturnal primates – the lorises and the tarsiers – in red)]	274
17.1 A collage of five fungi (clockwise from top-left): ¹³² a mushroom with a flat, red top with white-spots, and a white stem growing on the ground; a red cup-shaped fungus growing on wood; a stack of green and white moldy bread slices on a plate; a microscopic, spherical grey semitransparent cell, with a smaller spherical cell beside it; a microscopic view of an elongated cellular structure shaped like a microphone, attached to the larger end is a number of smaller roughly circular elements that collectively form a mass around sites.	275
17.2 Fungal hyphae cells ¹³³ 1. Hyphal wall 2. Septum 3. Mitochondrion 4. Vacuole 5. Ergosterol crystal 6. Ribosome 7. Nucleus 8. Endoplasmic reticulum 9. Lipid body 10. Plasma membrane 11. Spitzenkörper 12. Golgi apparatus	277
17.3 Six commonly recognized groups of fungi. ¹³⁴ Zygomycota, or zygote fungi, is a former division or phylum of the kingdom Fungi. The members are now part of two recently proposed phyla.	281
17.4 An environmental isolate of <i>Penicillium</i> hypha conidiophore phialide conidia septa ¹³⁵ 1. hypha 2. conidiophore 3. phialide 4. conidia 5. septa	281
17.5 Nutrient exchanges and communication between a mycorrhizal fungus and plants. ¹³⁶	282
17.6 Fruticose (light green) and crustose (yellow) lichens growing on a tree.	283
17.7 Schematic cross section of foliose lichen: ¹³⁷ a) The cortex is the outer layer of tightly woven fungus filaments (hyphae) b) This photobiont layer has photosynthesizing green algae c) Loosely packed hyphae in the medulla d) A tightly woven lower cortex e) Anchoring hyphae called rhizines where the fungus attaches to the substrate.	283
17.8 Cup-shaped apothecia and thallus of a foliose lichen.	283
18.1 The diversity of plants. ¹³⁸	287
18.2 Reconstruction of Cooksonia, a vascular plant from the Silurian ¹³⁹	290
18.3 <i>Marchantia polymorpha</i> , with antheridial and archegonial stalks. ¹⁴⁰	291
18.4 Hornworts are bryophytes that are believed to be the closest living relatives of the vascular plants. ¹⁴¹	291
18.5 A moss showing both gametophytes (the low, leaf-like forms) and sporophytes (the tall, stalk-like forms).	291
18.6 Microimages of seeds of various plants. ¹⁴² The first row: Poppy, Red pepper, Strawberry, Apple tree, Blackberry, Rice, Carum. Second row: Mustard, Eggplant, Physalis, grapes, raspberries, red rice, Patchouli. The third row: Figs, <i>Lycium barbarum</i> , Beets, Blueberries, Golden Kiwifruit, Rosehip, Basil. The fourth row: Pink pepper, Tomato, Radish, Carrot, Matthiola, Dill, Coriander Fifth row: Black pepper, White cabbage, Napa cabbage, Seabuckthorn, Parsley, Dandelion, <i>Capsella bursa-pastoris</i> . The sixth row: Cauliflower, Radish, Kiwifruit, Grenadilla, Passion fruit, Melissa, <i>Tagetes erecta</i> .	294
18.7 Various gymnosperms Left 1 <i>Welwitschia mirabilis</i> 2 <i>Cycas revoluta</i> 3 <i>Taxus baccata</i> 4 <i>Ginkgo biloba</i> RIGHT 1 <i>Cupressus sempervirens</i> 2 <i>Sequoiadendron giganteum</i> 3 <i>Dammara orientalis</i> 4 <i>Araucaria heterophylla</i> ¹⁴³	295

¹²⁹<https://commons.wikimedia.org/wiki/File:CollapsedtreeLabels-simplified.svg>

¹³⁰https://commons.wikimedia.org/wiki/File:Heliconius_mimicry.png

¹³¹https://commons.wikimedia.org/wiki/File:Monophly,_paraphly,_polyphyly.png

¹³²https://commons.wikimedia.org/wiki/File:Fungi_collage.jpg

¹³³<https://commons.wikimedia.org/wiki/File:HYPHAE.png>

¹³⁴[https://commons.wikimedia.org/wiki/File:02_01_groups_of_Fungi_\(M._Piepenbring\).png](https://commons.wikimedia.org/wiki/File:02_01_groups_of_Fungi_(M._Piepenbring).png)

¹³⁵https://commons.wikimedia.org/wiki/File:Penicillium_labeled_cropped.jpg

¹³⁶https://commons.wikimedia.org/wiki/File:Mycorrhizal_network.svg

¹³⁷https://commons.wikimedia.org/wiki/File:Lichen_cross_section_-_heteromeric_thallus.svg

¹³⁸https://commons.wikimedia.org/wiki/File:Diversity_of_plants_image_version_5.png

¹³⁹https://commons.wikimedia.org/wiki/File:Cooksonia_pertoni.png

¹⁴⁰<https://upload.wikimedia.org/wikipedia/commons/d/dc/Marchantia.jpg>

¹⁴¹[https://commons.wikimedia.org/wiki/File:Hornwort_\(3144429129\).jpg](https://commons.wikimedia.org/wiki/File:Hornwort_(3144429129).jpg)

¹⁴²https://commons.wikimedia.org/wiki/File:Разнообразие_семян.jpg

¹⁴³<https://commons.wikimedia.org/wiki/File:Gymnospermae.jpg>

18.8 Flowers of different families: ¹⁴⁴ St Bernards Lilly (<i>Anthericum liliago</i>): Liliaceae Bermuda Buttercup (<i>Oxalis pes-caprae</i>): Oxalidaceae Oleander (<i>Nerium oleander</i>): Apocynaceae Lantana (<i>Lantana camara</i>): Verbenaceae Scarlet Pimpernel (<i>Anagallis arvensis</i>): Primulaceae Verbascum (<i>Verbascum sinuatum</i>): Scrophulariaceae Common Mallow (<i>Malva sylvestris</i>): Malvaceae Spanish Oyster (<i>Scolymus hispanicus</i>): Asteraceae Stork's Bill (<i>Erodium malacoides</i>): Geraniaceae Bindweed (<i>Convolvulus arvensis</i>): Convolvulaceae Blue Gem (<i>Hebe x franciscana</i>): Plantaginaceae Calla Lily (<i>Zantedeschia aethiopica</i>): Araceae	296
18.9 Anatomy of the flower. ¹⁴⁵	298
18.10 Reproductive parts of Easter Lily (<i>Lilium longiflorum</i>). ¹⁴⁶ 1. Stigma, 2. Style, 3. Stamens, 4. Filament, 5. Petal	298
18.11 A Bee orchid ¹⁴⁷ has evolved over many generations to better mimic a female bee to attract male bees as pollinators.	300
18.12 Scanning electron microscope image (500x magnification) of pollen grains from a variety of common plants: ¹⁴⁸ sunflower (<i>Helianthus annuus</i>), morning glory (<i>Ipomoea purpurea</i>), prairie hollyhock (<i>Sidalcea malviflora</i>), oriental lily (<i>Lilium auratum</i>), evening primrose (<i>Oenothera fruticosa</i>), and castor bean (<i>Ricinus communis</i>). The image is magnified some x500, so the bean shaped grain in the bottom left corner is about 50 µm long. The colors are computer-generated.	302
18.13 Flowers and fruit (capsules) of the ground orchid, Spathoglottis plicata, illustrating an inferior ovary. ¹⁴⁹	312
18.14 Flower of Magnolia × wieseneri showing the many pistils making up the gynoecium in the middle of the flower. ¹⁵⁰	312
18.15 Cross-section through the ovary of Narcissus showing multiple carpels (the female reproductive part of the flower) fused along the placental line where the ovules form ¹⁵¹	312
18.16 The development sequence of a typical drupe, the nectarine (<i>Prunus persica</i>) over a 7.5 month period, from bud formation in early winter to fruit ripening in midsummer. ¹⁵² 1. Bud formation can be observed on new growth on the plant (early winter) 2. Flower buds clearly formed and leaves start to develop (early spring, ≈ 3 months) . 3. Flowers fully develop and are pollinated by wind or insects (early spring, ≈ 3½ months). 4. If successfully pollinated, flowers die back and incipient fruit can be observed; leaves have quickly grown to provide tree with food and energy from photosynthesis (mid-spring, ≈ 4 months). 5. Fruit is well developed and continues to grow (late spring, ≈ 5½ months). 6. Fruit fully ripens to an edible form to encourage spreading of seed contained within by animals (midsummer, ≈ 7½ months)	313
19.1 Diversity of animals. ¹⁵³	315
19.2 Animals are unique in having the ball of cells of the early embryo (1) develop into a hollow ball or blastula (2). ¹⁵⁴	316
19.3 Tiktaalik, ≈375 Ma ¹⁵⁵	317
19.4 Dickinsonia costata from the Ediacaran biota (c. 635–542 MYA) is one of the earliest animal species known. ¹⁵⁶	317
19.5 Anomalocaris canadensis ¹⁵⁷ is one of the many animal species that emerged in the Cambrian explosion, starting some 542 million years ago, and found in the fossil beds of the Burgess shale.	317
19.6 The Placozoa are a basal form of free-living (non-parasitic) multicellular organism. They are the simplest in structure of all animals. Three genera have been found: the classical Trichoplax adhaerens ¹⁵⁸ , Hoilungia hongkongensis, and Polyplacotoma mediterranea.	318

¹⁴⁴https://commons.wikimedia.org/wiki/File:Flower_poster_2.jpg

¹⁴⁵https://commons.wikimedia.org/wiki/File:Mature_flower_diagram.svg

¹⁴⁶https://commons.wikimedia.org/wiki/File:Lilium_Stamens.jpg

¹⁴⁷https://commons.wikimedia.org/wiki/File:Ophrys_apifera_flower1.jpg

¹⁴⁸https://commons.wikimedia.org/wiki/File:Misc_pollen_colorized.jpg

¹⁴⁹https://commons.wikimedia.org/wiki/File:Spathoglottis_flwrs_reduced.jpg

¹⁵⁰https://commons.wikimedia.org/wiki/File:Magnolia_wieseneri_-_labelled_gynoecium.jpg

¹⁵¹https://en.wikipedia.org/wiki/Gynoecium#/media/File:Narcis_zaadhokken.jpg

¹⁵²https://commons.wikimedia.org/wiki/File:Nectarine_Fruit_Development.jpg

¹⁵³https://en.wikipedia.org/wiki/File:Animal_diversity.png

¹⁵⁴<https://commons.wikimedia.org/wiki/File:Blastulation.png>

¹⁵⁵https://commons.wikimedia.org/wiki/File:Tiktaalik_BW.jpg

¹⁵⁶<https://commons.wikimedia.org/wiki/File:DickinsoniaCostata.jpg>

¹⁵⁷<https://commons.wikimedia.org/wiki/File:Anomalocaris2019.jpg>

¹⁵⁸https://commons.wikimedia.org/wiki/File:Trichoplax_adhaerens_photograph.png

19.7 Sponge biodiversity and morphotypes at the lip of a wall site in 60 feet (20 m) of water. Included are the yellow tube sponge, <i>Aplysina fistularis</i> , the purple vase sponge, <i>Niphates digitalis</i> , the red encrusting sponge, <i>Spirastrella coccinea</i> [nl], and the gray rope sponge, <i>Callyspongia</i> sp. ¹⁵⁹	320
19.8 Diagram of the structure of a sponge. ¹⁶⁰	322
19.9 Cells of the protist choanoflagellate clade closely resemble sponge choanocyte cells. Beating of choanocyte flagella draws water through the sponge so that nutrients can be extracted and waste removed. ¹⁶¹	322
19.10 Pelagic (open ocean) ctenophores (a) <i>Beroe ovata</i> , (b) <i>Euplokamis</i> sp., (c) <i>Nepheloctena</i> sp., (d) <i>Bathocyroe fosteri</i> , (e) <i>Mnemiopsis leidyi</i> , and (f) <i>Ocyropsis</i> sp. ¹⁶²	326
19.11 Four examples of Cnidaria: ¹⁶³ A jellyfish <i>Chrysaora melanaster</i> , a gorgonian <i>Annella mollis</i> , a rocky coral <i>Acropora cervicornis</i> , and a sea anemone <i>Nemanthus annamensis</i> .	329
19.12 Diversity of bilaterians. ¹⁶⁴	331
19.13 Idealised bilaterian body plan. ¹⁶⁵ With an elongated body and a direction of movement the animal has head and tail ends. Sense organs and mouth form the basis of the head. Opposed circular and longitudinal muscles enable peristaltic motion.	332
19.14 <i>Xenoturbella japonica</i> , a xenacoelomorph member (xenoturbellids). ¹⁶⁶	332
19.15 The bilaterian gut develops in two ways. In many protostomes, the blastopore develops into the mouth, while in deuterostomes it becomes the anus. ¹⁶⁷	334
19.16 A sea cucumber from Malaysia. ¹⁶⁸	334
19.17 Starfish exhibit a wide range of colours. ¹⁶⁹	334
19.18 Strongylocentrotus purpuratus, a well-armoured sea urchin. ¹⁷⁰	335
19.19 Acorn worm, a hemichordate. ¹⁷¹	338
19.20 A Lancelet (<i>Branchiostoma lanceolatum</i>). ¹⁷²	339
19.21 Gold-mouth sea squirt (<i>Polycarpa aurata</i>), a tunicate. ¹⁷³	339
19.22 A. Lancelet, B. Larval tunicate, C. Adult tunicate ¹⁷⁴ 1. Notochord, 2. Nerve chord, 3. Buccal cirri, 4. Pharynx, 5. Gill slit, 6. Gonad, 7. Gut, 8. V-shaped muscles, 9. Anus, 10. Inhalant siphon, 11. Exhalant siphon, 12. Heart, 13. Stomach, 14. Esophagus, 15. Intestines, 16. Tail, 17. Atrium, 18. Tunic	340
19.23 Actinopterygii – Rayed-Finned Fish ¹⁷⁵	340
19.24 Example of Osteichthyes: Queensland lungfish and West Indian Ocean coelacanth (two Sarcopterygii), Iridescent shark and American black sturgeon (two Actinopterygii). ¹⁷⁶	341
19.25 Diversity of amphibians: ¹⁷⁷ Clockwise from top right: Seymouria, Mexican burrowing caecilian, eastern newt and leaf green tree frog.	341
19.26 Diversity of reptiles: ¹⁷⁸ Clockwise from above left: Green sea turtle (<i>Chelonia mydas</i>), Tuatara (<i>Sphenodon punctatus</i>), Nile crocodile (<i>Crocodylus niloticus</i>), and Sinai agama (<i>Pseudotrapelus sinaitus</i>)	342
19.27 The diversity of birds: ¹⁷⁹ This image shows 17 biological orders of birds (from top, left to right): Musophagiformes, Pelecaniformes, Phaethontiformes, Accipitriformes, Gruiformes, Galliformes, Columbiformes, Apodiformes, Charadriiformes, Casuariiformes, Psittaciformes, Phoenicopteriformes, Sphenisciformes, Pelecaniformes, Suliformes, Passeriformes, Strigiformes, Piciformes.	343

¹⁵⁹https://commons.wikimedia.org/wiki/File:Reef3859_-_Flickr_-_NOAA_Photo_Library.jpg

¹⁶⁰https://commons.wikimedia.org/wiki/File:Sea_sponge_diagram.svg

¹⁶¹https://commons.wikimedia.org/wiki/File:Choanoflagellate_and_choanocyte.png

¹⁶²https://commons.wikimedia.org/wiki/File:Pelagic_ctenophores.png

¹⁶³<https://commons.wikimedia.org/wiki/File:Cnidaria.png>

¹⁶⁴https://commons.wikimedia.org/wiki/File:Animal_diversity_October_2007.jpg

¹⁶⁵https://commons.wikimedia.org/wiki/File:Bilaterian_body_plan.svg

¹⁶⁶https://commons.wikimedia.org/wiki/File:Xenoturbella_japonica.jpg

¹⁶⁷<https://commons.wikimedia.org/wiki/File:Protosdeuterostomes.svg>

¹⁶⁸https://commons.wikimedia.org/wiki/File:Sea_cucumber_at_Pulau_Redang.jpg

¹⁶⁹<https://commons.wikimedia.org/wiki/File:Nerr0878.jpg>

¹⁷⁰https://commons.wikimedia.org/wiki/File:Strongylocentrotus_purpuratus_1.jpg

¹⁷¹[https://commons.wikimedia.org/wiki/File:Eichelwurm_\(cropped\).jpg](https://commons.wikimedia.org/wiki/File:Eichelwurm_(cropped).jpg)

¹⁷²https://commons.wikimedia.org/wiki/File:Branchiostoma_lanceolatum.jpg

¹⁷³https://commons.wikimedia.org/wiki/File:Tunicate_komodo.jpg

¹⁷⁴https://commons.wikimedia.org/wiki/File:Comparison_of_Three_Invertebrate_Chordates.svg

¹⁷⁵<https://commons.wikimedia.org/wiki/File:Actinopterygii-0001.jpg>

¹⁷⁶<https://commons.wikimedia.org/wiki/File:Osteichthyes.jpg>

¹⁷⁷<https://commons.wikimedia.org/wiki/File:Amphibians.png>

¹⁷⁸https://commons.wikimedia.org/wiki/File:Extant_reptilia.jpg

¹⁷⁹https://en.wikipedia.org/wiki/File:Bird_Diversity_2013.png

19.28 Diversity of mammals. ¹⁸⁰	344
19.29 Diversity of arthropods: ¹⁸¹ From left to right and from top to bottom: Kolihapeltis, Stylonurus, scorpion, crab, centipede, butterfly.	345
19.30 A Maine Lobster served in Boston. ¹⁸²	346
19.31 <i>Caenorhabditis elegans</i> , a free-living transparent nematode about 1 mm in length that lives in temperate soil environments. ¹⁸³ In 1963, Sydney Brenner proposed research into <i>C. elegans</i> , primarily in the area of neuronal development. In 1974, he began research into the molecular and developmental biology of <i>C. elegans</i> , which has since been extensively used as a model organism. It was the first multicellular organism to have its whole genome sequenced, and as of 2019, is the only organism to have its connectome (neuronal “wiring diagram”) completed.	346
19.32 A Rotifera (wheel animal). ¹⁸⁴	347
19.33 The turbellarian <i>Pseudoceros dimidiatus</i> . ¹⁸⁵	349
19.34 Schistosoma life cycle. ¹⁸⁶ Schistosoma eggs are eliminated with feces or urine, depending on species (1). Under appropriate conditions the eggs hatch and release miracidia (2), which swim and penetrate specific snail intermediate hosts (3). The stages in the snail include two generations of sporocysts (4) and the production of cercariae (5). Upon release from the snail, the infective cercariae swim, penetrate the skin of the human host (6), and shed their forked tails, becoming schistosomulae (7). The schistosomulae migrate via venous circulation to lungs, then to the heart, and then develop in the liver, exiting the liver via the portal vein system when mature, (8)(9). Male and female adult worms copulate and reside in the mesenteric venules, the location of which varies by species (with some exceptions) (10). For instance, <i>S. japonicum</i> is more frequently found in the superior mesenteric veins draining the small intestine (A), and <i>S. mansoni</i> occurs more often in the inferior mesenteric veins draining the large intestine (B). However, both species can occupy either location and are capable of moving between sites. <i>S. intercalatum</i> and <i>S. guineensis</i> also inhabit the inferior mesenteric plexus but lower in the bowel than <i>S. mansoni</i> . <i>S. haematobium</i> most often inhabits in the vesicular and pelvic venous plexus of the bladder (C), but it can also be found in the rectal venules. The females (size ranges from 7–28 mm, depending on species) deposit eggs in the small venules of the portal and perivesical systems. The eggs are moved progressively toward the lumen of the intestine (<i>S. mansoni</i> , <i>S. japonicum</i> , <i>S. mekongi</i> , <i>S. intercalatum/guineensis</i>) and of the bladder and ureters (<i>S. haematobium</i>), and are eliminated with feces or urine, respectively (1).	351
19.35 Life cycle of the cestode <i>Taenia</i> . ¹⁸⁷ Taeniasis is the infection of humans with the adult tapeworm of <i>Taenia saginata</i> or <i>Taenia solium</i> . Humans are the only definitive hosts for <i>T. saginata</i> and <i>T. solium</i> . Eggs or gravid proglottids are passed with feces (1); the eggs can survive for days to months in the environment. Cattle (<i>T. saginata</i>) and pigs (<i>T. solium</i>) become infected by ingesting vegetation contaminated with eggs or gravid proglottids (2). In the animal's intestine, the oncospheres hatch (3), invade the intestinal wall, and migrate to the striated muscles, where they develop into cysticerci. A cysticercus can survive for several years in the animal. Humans become infected by ingesting raw or undercooked infected meat (4). In the human intestine, the cysticercus develops over 2 months into an adult tapeworm, which can survive for years. The adult tapeworms attach to the small intestine by their scolex (5) and reside in the small intestine (6). Length of adult worms is usually 5 m or less for <i>T. saginata</i> (however it may reach up to 25 m) and 2 to 7 m for <i>T. solium</i> . The adults produce proglottids which mature, become gravid, detach from the tapeworm, and migrate to the anus or are passed in the stool (approximately 6 per day). <i>T. saginata</i> adults usually have 1,000 to 2,000 proglottids, while <i>T. solium</i> adults have an average of 1,000 proglottids. The eggs contained in the gravid proglottids are released after the proglottids are passed with the feces. <i>T. saginata</i> may produce up to 100,000 and <i>T. solium</i> may produce 50,000 eggs per proglottid respectively	352
19.36 <i>Lingula anatina</i> , a brachiopod, from Stradbroke Island, Australia. ¹⁸⁸	353

¹⁸⁰https://en.wikipedia.org/wiki/File:Mammal_Diversity_2011.png

¹⁸¹<https://commons.wikimedia.org/wiki/File:Arthropoda.jpg>

¹⁸²https://commons.wikimedia.org/wiki/File:Lobster_in_Boston.jpg

¹⁸³https://commons.wikimedia.org/wiki/File:Adult_Caenorhabditis_elegans.jpg

¹⁸⁴https://commons.wikimedia.org/wiki/File:Mikrofoto.de-Raedertier_Ptygura_pilula_2.jpg

¹⁸⁵https://commons.wikimedia.org/wiki/File:Pseudoceros_dimidiatus.jpg

¹⁸⁶https://commons.wikimedia.org/wiki/File:Schistosoma_life_cycle.svg

¹⁸⁷https://commons.wikimedia.org/wiki/File:Taenia_solium_Life_cycle.tif

¹⁸⁸<https://commons.wikimedia.org/wiki/File:LingulaanatinaAA.JPG>

19.37 An entoproct: <i>Barentsia discreta</i> . ¹⁸⁹	353
19.38 <i>Cornu aspersum</i> (formerly <i>Helix aspersa</i>) – a common land snail, a gastropod. ¹⁹⁰	353
19.39 The common cuttlefish <i>Sepia officinalis</i> , a cephalopod. ¹⁹¹	354
19.40 <i>Argopecten irradians</i> , the Atlantic bay scallop, a bivalve. ¹⁹²	354
19.41 The earthworm <i>Lumbricus terrestris</i> is a representative of the annelids. ¹⁹³	356
19.42 The leech <i>Hirudo medicinalis</i> sucking blood. ¹⁹⁴ Although blood-letting is used less frequently by doctors, some leech species are regarded as endangered species because they have been over-harvested for this purpose in the last few centuries.	356
19.43 <i>Nereis virens</i> , a polychaete annelid. ¹⁹⁵ These sandworms eat seaweed and microorganisms and can be longer than four feet.	356
19.44 Types of epithelia ¹⁹⁶	358
19.45 Relationship of major animal lineages with indication of how long ago these animals shared a common ancestor. On the left, important organs are shown, which allows us to determine how long ago these may have evolved. ¹⁹⁷	359
19.46 Selection of internal organs in human anatomy. ¹⁹⁸	359
19.47 Anatomy of human skin. ¹⁹⁹	360
19.48 All extant species of hominid. ²⁰⁰ Top) Human, Middle) Common chimpanzee, bonobo, western gorilla & eastern gorilla, Bottom) Bornean, Sumatran & Tapanuli orangutan]	361
19.49 Model of the phylogeny of Hominidae, with adjacent branches of Hominoidea, over the past 20 million years. ²⁰¹	361
19.50 From Thomas Huxley's 1863 Evidence as to Man's Place in Nature, the compared skeletons of apes to humans. ²⁰²	362
19.51 A model of the evolution of the genus <i>Homo</i> ²⁰³ over the last 2 million years (vertical axis). The rapid "Out of Africa" expansion of <i>H. sapiens</i> is indicated at the top of the diagram, with admixture indicated with Neanderthals, Denisovans, and unspecified archaic African hominins. Late survival of robust australopithecines (<i>Paranthropus</i>) alongside <i>Homo</i> until 1.2 Mya is indicated in purple.	365
19.52 Dermoplast reconstruction of a Neanderthal ²⁰⁴	366
19.53 Replica of fossil skull of <i>Homo ergaster</i> (African <i>Homo erectus</i>). Fossil number Khm-Heu 3733 discovered in 1975 in Kenya. ²⁰⁵	368
19.54 Replica of fossil skull of <i>Homo habilis</i> . Fossil number KNM ER 1813, found at Koobi Fora, Kenya. ²⁰⁶	370
19.55 Oldowan-tradition stone chopper. ²⁰⁷	374
19.56 Aurignacian Culture bone tools (needle, points and tools for punching holes), Hayonim Cave, 30000 BP (Before Present). ²⁰⁸ HaYonim Cave (Hebrew: הַיּוֹנִים, מַעֲרָה Me'arat HaYonim, lit. Cave of the Pigeons) is a cave located in a limestone bluff about 250 meters above modern sea level, in the Upper Galilee, Israel.	375
20.1 The systemic circulation and capillary networks shown and also as separate from the pulmonary circulation. ²⁰⁹	378

¹⁸⁹https://commons.wikimedia.org/wiki/File:Barentsa_discreta_suzukokemusi02.jpg

¹⁹⁰<https://commons.wikimedia.org/wiki/File:Snail-wiki-120-Zachi-Evenor.jpg>

¹⁹¹https://commons.wikimedia.org/wiki/File:Washington_DC_Zoo_-_Sepia_officinalis_2.jpg

¹⁹²https://commons.wikimedia.org/wiki/File:Argopecten_irradians.jpg

¹⁹³<https://commons.wikimedia.org/wiki/File:Nerr0328.jpg>

¹⁹⁴https://commons.wikimedia.org/wiki/File:Sucking_leech.jpg

¹⁹⁵https://commons.wikimedia.org/wiki/File:Nereis_virens.jpg

¹⁹⁶https://commons.wikimedia.org/wiki/File:Illu_epithelium.jpg

¹⁹⁷https://commons.wikimedia.org/wiki/File:Origin_of_major_organs_on_the_animal_phylogeny.jpg

¹⁹⁸https://commons.wikimedia.org/wiki/File:Internal_organs.svg

¹⁹⁹<https://commons.wikimedia.org/wiki/File:Skin.jpg>

²⁰⁰[https://commons.wikimedia.org/wiki/File:Hominidae_\(extant_species\).jpg](https://commons.wikimedia.org/wiki/File:Hominidae_(extant_species).jpg)

²⁰¹https://commons.wikimedia.org/wiki/File:Hominoidea_lineage.svg

²⁰²https://commons.wikimedia.org/wiki/File:Huxley_-_Mans_Place_in_Nature.jpg

²⁰³https://en.wikipedia.org/wiki/Human_evolution#/media/File:Homo_lineage_2017update.svg

²⁰⁴https://commons.wikimedia.org/wiki/File:Neandertaler_reconst.jpg

²⁰⁵https://commons.wikimedia.org/wiki/File:Homo_ergaster.jpg

²⁰⁶https://commons.wikimedia.org/wiki/File:Homo_habilis-KNM_ER_1813.jpg

²⁰⁷https://commons.wikimedia.org/wiki/File:Oldowan_tradition_chopper.jpg

²⁰⁸https://commons.wikimedia.org/wiki/File:Aurignacian_Culture_Bone_Tools,_Hayonim_Cave,_30000_BP.jpg

²⁰⁹https://commons.wikimedia.org/wiki/File:2101_Blood_Flow_Through_the_Heart.jpg

20.2 The pulmonary circulation as it passes from the heart. Showing both the pulmonary and bronchial arteries. ²¹⁰	378
20.3 Diagram of the human heart ²¹¹ 1. Superior vena cava 2. 4. Mitral valve 5. Aortic valve 6. Left ventricle 7. Right ventricle 8. Left atrium 9. Right atrium 10. Aorta 11. Pulmonary valve 12. Tricuspid valve 13. Inferior vena cava	380
20.4 Arterial supply to the heart (red), with other areas labelled (blue). ²¹²	382
20.5 Conduction system of the heart. ²¹³	383
20.6 The cardiac cycle as correlated to the EKG ²¹⁴	383
20.7 Atherosclerosis is a condition affecting the circulatory system. If the coronary arteries are affected, angina pectoris may result or at worse a heart attack. ²¹⁵	385
21.1 The human lymphatic system. ²¹⁶	387
21.2 Location and microscopic anatomy of the human thymus ²¹⁷	388
21.3 Location and microscopic anatomy of the human spleen. ²¹⁸	389
21.4 Schematic diagram of a lymph node showing flow of lymph through lymph sinuses ²¹⁹	389
21.5 Lymph capillaries in the tissue spaces. ²²⁰	390
21.6 Nutrients in food are absorbed via intestinal vili (greatly enlarged in the picture) to blood and lymph. Long-chain fatty acids (and other lipids with similar fat solubility like some medicines) are absorbed to the lymph and move in it enveloped inside chylomicrons. They move via the thoracic duct of the lymphatic system and finally enter the blood via the left subclavian vein, thus bypassing the liver's first-pass metabolism completely. ²²¹	391
21.7 A scanning electron microscope image of normal circulating human blood. ²²² One can see red blood cells, several knobby white blood cells including lymphocytes, a monocyte, a neutrophil, and many small disc-shape platelets.	393
21.8 (ref:)	394
21.9 Simplified diagram of the phagocytosis and destruction of a bacterial cell. ²²³	394
21.10 A neutrophil. ²²⁴	395
21.11 An eosinophil. ²²⁵ Red blood cells surround the eosinophil, two platelets at the top left corner.	395
21.12 The complement system ²²⁶ is made up of about 25 proteins that work together to "complement" the action of antibodies in destroying bacteria. Complement proteins circulate in the blood in an inactive form. When the first protein in the complement series is activated— typically by antibody that has locked onto an antigen—it sets in motion a domino effect. Each component takes its turn in a precise chain of steps known as the complement cascade. The end product is a cylinder inserted into—and puncturing a hole in—the cell's wall. With fluids and molecules flowing in and out, the cell swells and bursts.	396
21.13 Overview of the processes involved in the primary immune response ²²⁷	398
21.14 Antigen presentation stimulates T cells to become either "cytotoxic" CD8+ cells or "helper" CD4+ cells. ²²⁸	399

²¹⁰https://commons.wikimedia.org/wiki/File:2119_Pulmonary_Circuit.jpg

²¹¹[https://commons.wikimedia.org/wiki/File:Diagram_of_the_human_heart_\(cropped\).svg](https://commons.wikimedia.org/wiki/File:Diagram_of_the_human_heart_(cropped).svg)

²¹²https://en.wikipedia.org/wiki/Heart#/media/File:Coronary_arteries.svg

²¹³<https://commons.wikimedia.org/wiki/File:ConductionsystemoftheheartwithouttheHeart-en.svg>

²¹⁴https://commons.wikimedia.org/wiki/File:2027_Phases_of_the_Cardiac_Cycle.jpg

²¹⁵https://commons.wikimedia.org/wiki/File:2113ab_Atherosclerosis.jpg

²¹⁶https://commons.wikimedia.org/wiki/File:Blausen_0623_LymphaticSystem_Female.png

²¹⁷https://upload.wikimedia.org/wikipedia/commons/c/cf/Illu_thymus.jpg

²¹⁸https://commons.wikimedia.org/wiki/File:Illu_spleen.jpg

²¹⁹https://commons.wikimedia.org/wiki/File:Schematic_of_lymph_node_showing_lymph_sinuses.svg

²²⁰https://commons.wikimedia.org/wiki/File:2202_Lymphatic_Capillaries_big.png

²²¹https://commons.wikimedia.org/wiki/File:Nutrient_absorbtion_to_blood_and_lymph.png

²²²https://commons.wikimedia.org/wiki/File:SEM_blood_cells.jpg

²²³<https://commons.wikimedia.org/wiki/File:Phagocytosis2.png>

²²⁴<https://commons.wikimedia.org/wiki/File:PBNeutrophil.jpg>

²²⁵https://commons.wikimedia.org/wiki/File:Eosinophil_blood_smear.JPG

²²⁶https://commons.wikimedia.org/wiki/File:Complement_pathway.svg

²²⁷https://commons.wikimedia.org/wiki/File:Primary_immune_response_1.png

²²⁸https://commons.wikimedia.org/wiki/File:Antigen_presentation.svg

21.15 The T lymphocyte activation pathway. ²²⁹ The T lymphocyte activation pathway is triggered when a T cell encounters its cognate antigen, coupled to a MHC molecule, on the surface of an infected cell or a phagocyte. T cells contribute to immune defenses in two major ways: some direct and regulate immune responses; others directly attack infected or cancerous cells.	400
21.16 The B lymphocyte activation pathway. ²³⁰ The B lymphocyte activation pathway. B cells function to protect the host by producing antibodies that identify and neutralize foreign objects like bacteria and viruses.	402
21.17 An antibody is made up of two heavy chains and two light chains. The unique variable region allows an antibody to recognize its matching antigen. ²³¹	403
 22.1 The respiratory system consists of the airways, the lungs, and the respiratory muscles that mediate the movement of air into and out of the body. ²³²	405
22.2 Alveoli and their capillary networks. ²³³	406
22.3 Output of a 'spirometer'. Upward movement of the graph, read from the left, indicates the intake of air; downward movements represent exhalation. ²³⁴	406
22.4 (ref:alive)	408
22.5 A highly diagrammatic illustration of the process of gas exchange in the mammalian lungs, emphasizing the differences between the gas compositions of the ambient air, the alveolar air (light blue) with which the pulmonary capillary blood equilibrates, and the blood gas tensions in the pulmonary arterial (blue blood entering the lung on the left) and venous blood (red blood leaving the lung on the right). All the gas tensions are in kPa. To convert to mm Hg, multiply by 7.5. ²³⁵	408
22.6 A diagrammatic histological cross-section through a portion of lung tissue showing a normally inflated alveolus (at the end of a normal exhalation), and its walls containing the pulmonary capillaries (shown in cross-section). This illustrates how the pulmonary capillary blood is completely surrounded by alveolar air. In a normal human lung all the alveoli together contain about 3 liters of alveolar air. All the pulmonary capillaries contain about 100 ml blood. ²³⁶	409
22.7 The respiratory system of birds. ²³⁷ On inhalation, air travels to air sacs near the back of a bird. The air then passes through the lungs to air sacs near the front of the bird, from where the air is exhaled.	411
22.8 The axolotl (<i>Ambystoma mexicanum</i>) retains its larval form with gills into adulthood. ²³⁸	413
22.9 The book lungs of a spider. ²³⁹	414
 23.1 Upper and lower human gastrointestinal tract. ²⁴⁰	419
23.2 Location of the liver, gallbladder, spleen, stomach, colon, appendix, and small intestine in the human abdomen. ²⁴¹	419
23.3 The main salivary glands. ²⁴²	420
23.4 The human stomach. ²⁴³ 1. Body of stomach 2. Fundus 3. Anterior wall 4. Greater curvature 5. Lesser curvature 6. Cardia 9. Pyloric sphincter 10. Pyloric antrum 11. Pyloric canal 12. Angular incisure 13. Gastric canal 14. Rugae	423
23.5 Diagram of the microscopic structure of the liever. ²⁴⁴	424

²²⁹https://commons.wikimedia.org/wiki/File:T_cell_activation.svg

²³⁰https://commons.wikimedia.org/wiki/File:B_cell_activation.svg

²³¹https://commons.wikimedia.org/wiki/File:Antibody_chains.svg

²³²https://commons.wikimedia.org/wiki/File:Respiratory_system_complete_en.svg

²³³https://upload.wikimedia.org/wikipedia/commons/4/46/Alveolus_diagram.svg

²³⁴https://commons.wikimedia.org/wiki/File:Lungvolumes_Updated.png

²³⁵https://commons.wikimedia.org/wiki/File:Gas_exchange.jpg

²³⁶<https://commons.wikimedia.org/wiki/File:Alveolus.jpg>

²³⁷<https://commons.wikimedia.org/wiki/File:BirdRespiration.svg>

²³⁸https://commons.wikimedia.org/wiki/File:Axolotl_ganz.jpg

²³⁹https://upload.wikimedia.org/wikipedia/commons/2/22/Spider_internal_anatomy-en.svg

²⁴⁰https://en.wikipedia.org/wiki/File:Digestive_system_diagram_edit.svg

²⁴¹https://commons.wikimedia.org/wiki/File:Anatomy_Abdomen_Tiesworks.jpg

²⁴²https://en.wikipedia.org/wiki/Human_digestive_system#/media/File:Blausen_0780_SalivaryGlands.png

²⁴³https://commons.wikimedia.org/wiki/File:Illu_stomach.jpg

²⁴⁴https://commons.wikimedia.org/wiki/File:Hepatic_structure2.svg

23.6 The pancreas has a role in digestion, highlighted here. Ducts in the pancreas (green) conduct digestive enzymes into the duodenum. This image also shows a pancreatic islet, part of the endocrine pancreas, which contains cells responsible for secretion of insulin and glucagon. ²⁴⁵	425
24.1 Movement of water and ions in freshwater fish. ²⁴⁶	429
24.2 Movement of water and ions in saltwater fish. ²⁴⁷	429
24.3 The human urinary system. ²⁴⁸ 1. Human urinary system: 2. Kidney, 3. Renal pelvis, 4. Ureter, 5. Urinary bladder, 6. Urethra. (Left side with frontal section) 7. Adrenal gland Vessels: 8. Renal artery and vein, 9. Inferior vena cava, 10. Abdominal aorta, 11. Common iliac artery and vein Transparent: 12. Liver, 13. Large intestine, 14. Pelvis	431
24.4 The kidneys lie in the retroperitoneal space behind the abdomen, and act to filter blood to create urine. ²⁴⁹	432
24.5 The structure of the kidney. ²⁵⁰ 1. Renal pyramid • 2. Interlobular artery • 3. Renal artery • 4. Renal vein • 5. Renal hilum • 6. Renal pelvis • 7. Ureter • 8. Minor calyx • 9. Renal capsule • 10. Inferior renal capsule • 11. Superior renal capsule • 12. Interlobular vein • 13. Nephron • 14. Renal sinus • 15. Major calyx • 16. Renal papilla • 17. Renal column	432
24.6 The nephron, shown here, is the functional unit of the kidneys. Its parts are labelled except the (gray) connecting tubule located after the (dark red) distal convoluted tubule and before the large (gray) collecting duct (mislabelled collection duct). ²⁵¹	433
24.7 (ref:)	433
24.8 Secretion and reabsorption of water, ions and various substances throughout the nephron. ²⁵²	433
25.1 The human nervous system.	435
25.2 Transverse section of a nerve. a) a single nerve fibre (axon) surrounded by a thick layer of myelin. c) an interstitial cell. Histologie du système nerveux de l'homme & des vertébrés, Tome Premier ²⁵³ (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay.	436
25.3 A brainbow of mouse neurons. ²⁵⁴	436
25.4 A nerve cell from the cerebral cortex of a rabbit. Notice the extensive tree of dendrites at the top, the long axon at the bottom. Because of the pyramid-like shape of the cell body, this type of neuron is referred to as a pyramidal cell.	437
25.5 Comparison of nervous systems of invertebrates. Top left: A diffuse nerve net in <i>Actinia</i> (a genus of sea anemones in the family <i>Actiniidae</i> in the phylum <i>Cnidaria</i>); top right: The nervous system of <i>Anadonta anatina</i> , a freshwater mussel in the family <i>Unionidae</i> in the phylum <i>Mollusca</i> . c, foot; k, pedal ganglion; i, cerebro-pedal connective; g, cerebral ganglion; h, cerebral connective; a, anterior adductor muscle; r, q, anterior pallial nerves; d, liver; s, visceral nerve; l, cerebro-visceral connective; e, gill; f, edge of mantle; n, branchial nerves; m, visceral ganglion; o, posterior pallial nerves; b, posterior adductor muscle; p, lateral pallial nerves; bottom left: the nervous system of <i>Alitta virens</i> , a polychaete worm in the phylum <i>Annelida</i> . J, jaws; b, antennal nerves; c, palpal nerves; f, ganglia for the dorsal peristomial cirri; n ¹ , ganglion; n, nerves for the dissepimenta; m, parapodial nerves; i, parapodial branch; h, ventral chain of ganglia; C, cerebral ganglion; o, nerve passing through dissepiment to preceding segment; k, parapodial ganglion. Bottom right: the nervous system of an insect (<i>Arthropoda</i>). From Morphology of invertebrate types, by Alexander Petrunkevitch. New York, Macmillan company, 1916. ²⁵⁵	438

²⁴⁵https://upload.wikimedia.org/wikipedia/commons/e/e0/1820_The_Pancreas.jpg

²⁴⁶https://en.wikipedia.org/wiki/Osmoregulation#/media/File:Bachforelle_osmoregulatooin_bw_en2.png

²⁴⁷https://en.wikipedia.org/wiki/File:Osmoseragation_Carangoides_bartholomaei_bw_en2.png

²⁴⁸https://commons.wikimedia.org/wiki/File:Urinary_system.svg

²⁴⁹https://commons.wikimedia.org/wiki/File:Blausen_0592_KidneyAnatomy_01.png

²⁵⁰https://commons.wikimedia.org/wiki/File:KidneyStructures_PioM.svg

²⁵¹https://commons.wikimedia.org/wiki/File:Kidney_Nephron.png

²⁵²https://upload.wikimedia.org/wikipedia/commons/2/26/2618_Nephron_Secretion_Reabsorption.jpg

²⁵³<https://wellcomelibrary.org/item/b2129592x#?c=0&m=0&s=0&cv=14&z=0%2C-3.48%2C1%2C8.6591>

²⁵⁴[https://commons.wikimedia.org/wiki/File:Brainbow_\(Smith_2007\).jpg](https://commons.wikimedia.org/wiki/File:Brainbow_(Smith_2007).jpg)

²⁵⁵[https://www.biodiversitylibrary.org/ia/morphologyofinvertebrate0petr#/page/7\(mode/1up](https://www.biodiversitylibrary.org/ia/morphologyofinvertebrate0petr#/page/7(mode/1up)

25.6 Dorsal views of the central nervous systems of the teleosts (from left to right) <i>Trigla hirundo</i> (a) and <i>Mola mola</i> , the urodele <i>Ambystoma tigrinum</i> , the anuran <i>Xenopus laevis</i> , the tortoise <i>Testudo hermanni</i> , the tegu lizard <i>Tupinambis teguixin</i> , the pigeon, the cat and human. In a, c, d, e, f, g and j the full length of the spinalcord, including the filum terminale (where present) is shown; in <i>Mola mola</i> and the cat most of the filum terminale is cut. Vertical black bars correspond to 1 cm in length. Modified from Nieuwenhuys, R., ten Donkelaar, H. J., & Nicholson, C. (1998). The Meaning of It All. The Central Nervous System of Vertebrates, 2135–2195 ²⁵⁶	440
25.7 Variability of brain size and external topography. Photographs and weights of the brains of different species. Primates: human (<i>Homo sapiens</i> , 1.176 kg), chimpanzee (<i>Pan troglodytes</i> , 273 g), baboon (<i>Papio cynocephalus</i> , 151 g), mandrill (<i>Mandrillus sphinx</i> , 123 g), macaque (<i>Macaca tonkeana</i> , 110 g). Carnivores: bear (<i>Ursus arctos</i> , 289 g), lion (<i>Panthera leo</i> , 165 g), cheetah (<i>Acinonyx jubatus</i> , 119 g), dog (<i>Canis familiaris</i> , 95 g), cat (<i>Felis catus</i> , 32 g). Artiodactyls: giraffe (<i>Giraffa camelopardalis</i> , 700 g), kudu (<i>Tragelaphus strepsiceros</i> , 166 g), mouflon (<i>Ovis musimon</i> , 118 g), ibex (<i>Capra pyrenaica</i> , 115 g); peccary (<i>Tayassu pecari</i> , 41 g). Marsupials: wallaby (<i>Protemnodon rufogrisea</i> , 28 g). Lagomorphs: rabbit (<i>Oryctolagus cuniculus</i> , 5.2 g). Rodents: rat (<i>Rattus rattus</i> , 2.6 g), mouse (<i>Mus musculus</i> , 0.5 g). The chimpanzee brain was kindly supplied by Dr. Dean Falk. The rest of non-human brains were from material used in Ballesteros-Yáñez et al., 2005). Scale bar: 5 cm. From DeFelipe J (2011) The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. <i>Front. Neuroanat.</i> 5:29 ²⁵⁷	443
25.8 Principal lobes and fissures of the cerebrum viewed laterally. From Gray Henry, Anatomy of the Human Body. 20 th Edition, Lea & Febiger, Philadelphia & New York, 1918 ²⁵⁸	443
25.9 Diagram showing a lateral view of the ridges (gyri) and grooves (sulci) of the left hemisphere of the brain. From Gray Henry, Anatomy of the Human Body. 20 th Edition, Lea & Febiger, Philadelphia & New York, 1918 ²⁵⁹	443
25.10 Diagram showing a medial view of the ridges (gyri) and grooves (sulci) of the left hemisphere of the brain. From Gray Henry, Anatomy of the Human Body. 20 th Edition, Lea & Febiger, Philadelphia & New York, 1918 ²⁶⁰	444
25.11 Diagram showing a view from the bottom of the ridges (gyri) and grooves (sulci) of the left hemisphere of the brain.	444
25.12 Diagram of a myelinated vertebrate motor neuron. ²⁶¹	450
25.13 Morphologically distinct types of neurons after Cajal. A) Unipolar neurons; B) bipolar neurons; Golgi I neurons: C) a Purkinje cell; D) spinal motor neuron E) a pyramidal cell; F) Golgi II neuron. <i>Histologie du système nerveux de l'homme & des vertébrés</i> , Tome Premier ²⁶² (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay.	450
25.14 Astrocytes (A) and oligodendrocytes (B) are the major types of macroglia in the grey and white matter of the brain, respectively. <i>Histologie du système nerveux de l'homme & des vertébrés</i> , Tome Premier ²⁶³ (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay.	453
25.15 A cartoon representation of the molecular structure of the sodium – potassium pump based on atomic coordinates of PDB 2ZXE ²⁶⁴ , rendered with open source molecular visualization tool PyMol.	456
25.16 The structure of the potassium channel KcsA from <i>Streptomyces lividans</i> determined by X-ray crystallography. Top: side view; bottom: top view. KcsA shares sequence similarity with all known K ⁺ channels and was the first ion channel to have its structure solved at atomic resolution. It consists of four identical sub-units that together form a cone shaped structure (top) with a ion selectivity filter at its outer end. Three K ⁺ ions (red spheres) are present in the channel, two of which are held 7.5 angstroms apart by the selectivity filter, a third K ⁺ ion shown in the channel pore below. Data from PDB 1BL8 ²⁶⁵ , rendered with open source molecular visualization tool PyMol ²⁶⁶	458

²⁵⁶https://doi.org/10.1007/978-3-642-18262-4_24

²⁵⁷<https://www.frontiersin.org/article/10.3389/fnana.2011.00029>

²⁵⁸<https://archive.org/details/anatomyofhumanbo1918gray/page/n6/mode/2up>

²⁵⁹<https://archive.org/details/anatomyofhumanbo1918gray/page/n6/mode/2up>

²⁶⁰<https://archive.org/details/anatomyofhumanbo1918gray/page/n6/mode/2up>

²⁶¹https://commons.wikimedia.org/wiki/File:Complete_neuron_cell_diagram_en.svg

²⁶²<https://wellcomelibrary.org/item/b2129592x#?c=0&m=0&s=0&cv=14&z=0%2C-3.48%2C1%2C8.6591>

²⁶³<https://wellcomelibrary.org/item/b2129592x#?c=0&m=0&s=0&cv=14&z=0%2C-3.48%2C1%2C8.6591>

²⁶⁴<https://www.rcsb.org/structure/2ZXE>

²⁶⁵<https://www.rcsb.org/structure/1BL8>

²⁶⁶<https://pymol.org/2/>

²⁶⁷ <https://www.rcsb.org/structure/5EKO>

²⁶⁸ <https://www.rcsb.org/structure/2A79>

²⁶⁹ <https://www.rcsb.org/structure/3JBB>

²⁷⁰<https://www.rcsb.org/structure/1KPI>

²⁷¹<https://www.fcc.gov>

²⁷² <https://www.rcsb.org/structure/3PC9>

²⁷³<https://www.rcsb.org/structure/6H1Q>

²⁷⁴ <https://www.rcsb.org/structure/6H1Q>

²⁷⁵ <https://www.rcsb.org/structure/3IAD>

²⁷⁵ <https://www.rcsb.org/structure/3JAD>

<https://www.rcsb.org/structure/5IDE>

<https://www.rcsb.org/structure/4PEs>

[278 https://pymol.org/2/](https://pymol.org/2/)

https://commons.wikimedia.org/wiki/File:Membrane_Permeability_of_a_Neuron_During_an_Action_Potential.svg

<https://doi.org/10.1074/mcp.R115.051482>

²⁸¹ <https://www.rcsb.org/structure/4PE5>
²⁸² <https://www.ncbi.nlm.nih.gov/2/>

²⁸² <https://pymol.org/2/>

- 25.22 The tetrapartite synapse of principal neurons in the forebrain, consisting of the pre- and postsynaptic compartment, astrocytic endfeet, and the extracellular matrix has a tightly regulated protein composition. A microsecretory system is present in synapses and dendrites that allows for translation of mRNA, local synthesis of, processing and insertion of transmembrane proteins. Hence the turnover of the synaptic protein machinery is controlled by local and somatic de novo protein synthesis, protein degradation by the ubiquitin proteasome system, lysosomes and autophagosomes. In addition, the association of proteins with pre- and postsynaptic compartments is highly dynamic. Molecular machineries and organelles for proteostasis are shared between synapses in dendritic segments. Proteins are transported in and out of the synapse as well as by diffusion of transmembrane proteins. These processes govern the activity-dependent assembly of the pre- and postsynaptic scaffold and the synaptic surface expression of receptors, calcium channels and cell adhesion molecules. Abbreviations: CAM, cell adhesion molecules; CAZ, cytomatrix at the active zone; ECM, extracellular matrix; ER, endoplasmic reticulum; ERGIC, endoplasmic reticulum Golgi intermediate compartment; MT, microtubules; PSD, postsynaptic density; RE, recycling endosomes; Lys, lysomes; SV, synaptic vesicle. From Proteomics of the Synapse – A Quantitative Approach to Neuronal Plasticity Daniela C. Dieterich, Michael R. Kreutz Molecular & Cellular Proteomics February 1, 2016, First published on August 25, 2015, 15 (2) 368–381; DOI: 10.1074/mcp.R115.051482²⁸³ 468
- 25.23 A cartoon representation of the atomic structure of the the nicotinic acetylcholine receptor from the electric ray *Torpedo marmorata* at 4Å resolution. Data from PDB 2BG9²⁸⁴, rendered with open source molecular visualization tool PyMol²⁸⁵. 471
- 26.1 Horizontal section of the human eyeball. From Gray Henry, Anatomy of the Human Body. 20th Edition, Lea & Febiger, Philadelphia & New York, 1918²⁸⁶ 477
- 26.2 Vertical section of the adult human retina. Carmine and Nissl stain. A, Photoreceptor layer. B, Cell bodies of the photoreceptors. C, Outer plexiform layer. D, Internal granule layer. E, Internal plexiform layer. F, Ganglion cell layer. G, Ganglion cell axons. a, external limiting membrane. b, internal limiting membrane. c, Spherical endfeet of the rod photoreceptors. d, endfeet of the cones. e, a, cone. f, a rod g, horizontal cells. h, amacrine cells. Fig. 188 from *Histologie du système nerveux de l'homme & des vertébrés*²⁸⁷ (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay. 479
- 26.3 A semischematic diagram of the frog retina. a) green rods; b (left) red rods; c) cone; i) horizontal cell; h) bipolar cell; n,m,r,s,t) amacrine cells; o,p) ganglion cells; q) displaced amacrine cell. A) Pigment epithelial cell with extended process; B) Pigment epithelial cell with retracted process. 480
- 26.4 Rods and cones from the human retina. A) a rod from the peripheral retina; B) a cone from the peripheral retina; C) cones from the fovea. 480
- 26.5 The chemical reactions involved in the photoreceptor visual cycle.²⁸⁸ 481
- 26.6 Representation²⁸⁹ of molecular steps in photoactivation (modified from Leskov et al., 2000). Depicted is an outer membrane disk in a rod. Step 1: Incident photon (hv) is absorbed and activates a rhodopsin by conformational change in the disk membrane to R. Step 2: Next, R makes repeated contacts with transducin molecules, catalyzing its activation to G* by the release of bound GDP in exchange for cytoplasmic GTP, which expels its β and γ subunits. Step 3: G* binds inhibitory γ subunits of the phosphodiesterase (PDE) activating its α and β subunits. Step 4: Activated PDE hydrolyzes cGMP. Step 5: Guanylyl cyclase (GC) synthesizes cGMP, the second messenger in the phototransduction cascade. Reduced levels of cytosolic cGMP cause cyclic nucleotide gated channels to close preventing further influx of Na⁺ and Ca²⁺. 482
- 26.7 A simplified schema²⁹⁰ of the human visual pathway. 483
- 26.8 A visual field map²⁹¹ of the primary visual cortex and the numerous extrastriate areas. 485

²⁸³<https://doi.org/10.1074/mcp.R115.051482>

²⁸⁴<https://www.rcsb.org/structure/2BG9>

²⁸⁵<https://pymol.org/2/>

²⁸⁶<https://archive.org/details/anatomyofhumanbo1918gray/page/n6/mode/2up>

²⁸⁷<https://wellcomelibrary.org/item/b2129592x#?c=0&m=0&s=0&cv=0&z=-0.9137%2C-0.0887%2C2.8274%2C1.7747>

²⁸⁸https://commons.wikimedia.org/wiki/File:Visual_cycle.svg

²⁸⁹<https://commons.wikimedia.org/wiki/File:Phototransduction.png>

²⁹⁰https://commons.m.wikimedia.org/wiki/File:human_visual_pathway.svg

²⁹¹https://commons.wikimedia.org/wiki/File:Visual_field_maps.jpg

26.9 Front view of the right outer, middle and inner human ear. Gray, Henry, 1825–1861. Anatomy, descriptive and surgical; ed. by T. Pickering Pick and Robert Howden. A revised American, from the fifteenth English edition. Philadelphia, Lea, 1901 ²⁹²	487
26.10 A cross section of the cochlea illustrating the organ of Corti. ²⁹³	488
26.11 Anatomy of human skin. ²⁹⁴	492
26.12 A cartoon representation of the mechanosensitive Piezo1 channel in side and top view. PDB 5Z10 ²⁹⁵ , rendered with open source molecular visualization tool PyMol.	493
26.13 Tactile corpuscles (from left to right): Meissner, Krause, Pacini ²⁹⁶	494
26.14 A cartoon representation of the transient receptor potential cation channel subfamily V member 1 (TRPV1), also known as the capsaicin receptor and the vanilloid receptor 1 viewed from the side and the top. The function of TRPV1 is detection and regulation of body temperature. In addition, TRPV1 provides a sensation of scalding heat and pain (nociception). In primary afferent sensory neurons, it cooperates with TRPA1 (a chemical irritant receptor) to mediate the detection of noxious environmental stimuli. PDB 5IS0 ²⁹⁷ , rendered with open source molecular visualization tool PyMol.	495
26.15 Muscle spindle from the pectoral muscle of a frog. Bottom part of figure: B) extrafusal muscle fibers with efferent motor nerve; top part of figure: myelinated afferent nerve fiber. Histologie du système nerveux de l'homme & des vertébrés, Tome Premier ²⁹⁸ (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay.	496
26.16 Golgi tendon organ from an adult cat. a) Terminal arborisation and c) fine ultimate branches of the afferent nerve; b) myelinated afferent nerve fiber; e) muscle fibers. Histologie du système nerveux de l'homme & des vertébrés, Tome Premier ²⁹⁹ (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay.	497
26.17 View of the right side of the nasal septum showing the olfactory bulb and the filaments of the olfactory nerve. From Gray, Henry: Anatomy Descriptive And Surgical. 7 th Edition, Longmans, Green, And Co., London, 1875 ³⁰⁰ .	499
26.18 Diagram of the structure of the olfactory bulb and olfactory cortex. A) Olfactory mucosa; B) glomeruli in the olfactory bulb; C) mitral cells; D) granule cells; E) olfactory nerve; F) pyramidal cells in the olfactory cortex. Action potentials fired by the olfactory receptor cells and subsequently by the cells in the olfactory bulb, travel to the olfactory cortex (arrows). Histologie du système nerveux de l'homme & des vertébrés, Tome Premier ³⁰¹ (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay.	500
26.19 The human tongue and three kinds of papillae magnified. Bottom left to right: filiform, fungiform and circumvallate papillae. From Gray, Henry: Anatomy Descriptive And Surgical. 7 th Edition, Longmans, Green, And Co., London, 1875 ³⁰²	502
26.20 Section through a papilla vallata of the human tongue. 1) Papilla. 2) Vallum. 3) Taste buds. From Textbook of anatomy. Section 2. The muscular system: the nervous system: the organs of sense and integument edited by D. J. Cunningham ³⁰³	503
26.21 A. Three quarter surface view of a taste bud from the papilla foliata of a rabbit (highly magnified). B. Vertical section of taste bud from the papilla foliata of a rabbit (highly magnified). From Textbook of anatomy. Section 2. The muscular system: the nervous system: the organs of sense and integument edited by D. J. Cunningham ³⁰⁴	503

²⁹²<https://archive.org/details/anatomydescript00grayuoft/page/n6/mode/2up>

²⁹³<https://commons.wikimedia.org/wiki/File:Cochlea-crosssection.svg>

²⁹⁴<https://commons.wikimedia.org/wiki/File:Skin.jpg>

²⁹⁵<https://www.rcsb.org/structure/5Z10>

²⁹⁶<https://commons.wikimedia.org/wiki/File:Gray940.png>

²⁹⁷<https://www.rcsb.org/structure/5IS0>

²⁹⁸<https://wellcomelibrary.org/item/b2129592x#?c=0&m=0&s=0&cv=14&z=0%2C-3.48%2C1%2C8.6591>

²⁹⁹<https://wellcomelibrary.org/item/b2129592x#?c=0&m=0&s=0&cv=14&z=0%2C-3.48%2C1%2C8.6591>

³⁰⁰<https://wellcomelibrary.org/item/b21688692>

³⁰¹<https://wellcomelibrary.org/item/b2129592x#?c=0&m=0&s=0&cv=14&z=0%2C-3.48%2C1%2C8.6591>

³⁰²<https://wellcomelibrary.org/item/b21688692>

³⁰³<https://wellcomelibrary.org/item/b21271070>

³⁰⁴<https://wellcomelibrary.org/item/b21271070>

26.22 Isolated cells from taste bud of a rabbit. a, Supporting cells. b, Gustatory cells. From Textbook of anatomy. Section 2. The muscular system: the nervous system: the organs of sense and integument edited by D. J. Cunningham ³⁰⁵	503
27.1 Vertebrate skulls and skeletons ³⁰⁶ : 1. <i>Homo sapiens</i> ³⁰⁷ , 2. Human skull ³⁰⁸ , 3. <i>Australopithecus</i> ³⁰⁹ , 4. <i>Homo neanderthalensis</i> ³¹⁰ , 5. <i>Pan troglodytes</i> ³¹¹ , 6. <i>Papio hamadryas</i> ³¹² , 7. <i>Rhinopithecus roxellana</i> ³¹³ , 8. <i>Gorilla</i> ³¹⁴ , 9. <i>Sus scrofa</i> ³¹⁵ , 10. <i>Cattle</i> ³¹⁶ , 11. <i>Panthera leo</i> ³¹⁷ , 12. <i>Canis lupus</i> ³¹⁸ , 13. <i>Equus ca- ballus</i> ³¹⁹ , 14. <i>Elephantidae</i> ³²⁰ , 15. <i>Capra</i> ³²¹ , 16. <i>Hippopotamus</i> ³²² , 17. <i>Camelus</i> ³²³ , 18. <i>Macropus</i> ³²⁴ , 19. <i>Damaliscus korrigum</i> ³²⁵ , 20. <i>Odobenus rosmarus</i> ³²⁶ , 21. <i>Bat</i> ³²⁷ , 22. <i>Cetacea</i> ³²⁸ , 23. <i>Accipitridae</i> ³²⁹ , 24. <i>Psittacidae</i> ³³⁰ , 25. <i>Gallus gallus</i> ³³¹ , 26. <i>Roosters</i> ³³² , 27. <i>Ramphastos sulfuratus</i> ³³³ , 28. <i>Casuaridae</i> ³³⁴ , 29. <i>Spheniscidae</i> ³³⁵ , 30. <i>Gruidae</i> ³³⁶ , 32. <i>Sharovipteryx mirabilis</i> ³³⁷ , 32. <i>Natrix natrix</i> ³³⁸ , 33. <i>Crotalinae</i> ³³⁹ , 34. <i>Boa constrictor</i> ³⁴⁰ , 35. <i>Crocodile</i> ³⁴¹ , 36. <i>Lizard</i> ³⁴² , 37. <i>Testudines</i> ³⁴³ , 38. <i>Frog</i> ³⁴⁴ , 39. <i>Salamandra salamandra</i> ³⁴⁵ , 40. <i>Perca fluviatilis</i> ³⁴⁶ , 41. <i>Acipenser</i> ³⁴⁷ , 42. <i>Balistoides viridescens</i> ³⁴⁸ , 43. <i>Rajidae</i> ³⁴⁹ , 44. <i>Polypterus</i> ³⁵⁰	508
27.2 A synovial joint. ³⁵¹	510

³⁰⁵<https://wellcomelibrary.org/item/b21271070>

³⁰⁶<https://upload.wikimedia.org/wikipedia/commons/8/87/Skeletons.png>

³⁰⁷https://commons.wikimedia.org/wiki/File:S_kleleton_diagram.svg

³⁰⁸https://commons.wikimedia.org/wiki/File:Gray188_skull_left_lateral_index.png

³⁰⁹<https://commons.wikimedia.org/wiki/File:Australopithecus.jpg>

³¹⁰https://commons.wikimedia.org/wiki/File:Homo_sapiens_neanderthalensis.jpg

³¹¹https://commons.wikimedia.org/wiki/File:Pan_troglodytes01.jpg

³¹²https://commons.wikimedia.org/wiki/File:Huxley_-_Mans_Place_in_Nature.jpg

³¹³[https://commons.wikimedia.org/wiki/File:Goldstumpfnasen_\(Rhinopithecus_roxellana\).jpg](https://commons.wikimedia.org/wiki/File:Goldstumpfnasen_(Rhinopithecus_roxellana).jpg)

³¹⁴<https://commons.wikimedia.org/wiki/File:GorillaSkeleton.jpg>

³¹⁵<https://commons.wikimedia.org/wiki/File:PigSkelLyd2.png>

³¹⁶https://commons.wikimedia.org/wiki/File:Heubach_cattle_skull.jpg

³¹⁷https://commons.wikimedia.org/wiki/File:Lion_anatomy_lateral_skeleton_view.jpg

³¹⁸<https://commons.wikimedia.org/wiki/File:WolfsSkelLyd1.png>

³¹⁹https://commons.wikimedia.org/wiki/File:Horse_bones_uglan.jpg

³²⁰<https://commons.wikimedia.org/wiki/File:ElephantSkelLyd2.png>

³²¹https://commons.wikimedia.org/wiki/File:Goat_anatomy_lateral_skeleton_view.jpg

³²²<https://commons.wikimedia.org/wiki/File:HippoSkelLyd2.png>

³²³<https://commons.wikimedia.org/wiki/File:CamelSkelLyd2.png>

³²⁴https://commons.wikimedia.org/wiki/File:Macropodidae_skeleton.png

³²⁵<https://commons.wikimedia.org/wiki/File:ChausimgaSkullLyd2.png>

³²⁶<https://commons.wikimedia.org/wiki/File:WalrusLyd2.png>

³²⁷[https://commons.wikimedia.org/wiki/File:Fledermaus_\(Nycteris_fuliginosis\).png](https://commons.wikimedia.org/wiki/File:Fledermaus_(Nycteris_fuliginosis).png)

³²⁸<https://commons.wikimedia.org/wiki/File:GreenlandWhaleLyd3.jpg>

³²⁹<https://commons.wikimedia.org/wiki/File:Baldeagle-06jul16.jpg>

³³⁰<https://commons.wikimedia.org/wiki/File:ParrotSkelLyd.jpg>

³³¹<https://commons.wikimedia.org/wiki/File:Furcula.png>

³³²<https://commons.wikimedia.org/wiki/File:Furcula.png>

³³³https://commons.wikimedia.org/wiki/File:Toco_toucan_whipsnade.jpg

³³⁴<https://commons.wikimedia.org/wiki/File:CassowarySkullLyd4.png>

³³⁵https://commons.wikimedia.org/wiki/File:Skull_alca_torda.jpg

³³⁶https://commons.wikimedia.org/wiki/File:Grus_grus_skull.png

³³⁷https://commons.wikimedia.org/wiki/File:Sharovipteryx_BW.jpg

³³⁸https://commons.wikimedia.org/wiki/File:Python_gab_fbi.png

³³⁹

³⁴⁰https://commons.wikimedia.org/wiki/File:Acrantophis_dumerili.jpg

³⁴¹<https://commons.wikimedia.org/wiki/File:Crocodilelyd5.png>

³⁴²https://commons.wikimedia.org/wiki/File:Pleurodont_jaw_Mivart.png

³⁴³<https://commons.wikimedia.org/wiki/File:SmithTestudoSkeletonTagged.PNG>

³⁴⁴https://commons.wikimedia.org/wiki/File:Rana_skull.png

³⁴⁵https://commons.wikimedia.org/wiki/File:Albanerpeton_BW.jpg

³⁴⁶https://commons.wikimedia.org/wiki/File:Pesce_Persico.jpg

³⁴⁷<https://commons.wikimedia.org/wiki/File:Sturgeon2.jpg>

³⁴⁸https://commons.wikimedia.org/wiki/Category:Balistoides_viridescens

³⁴⁹https://commons.wikimedia.org/wiki/File:Smooth_skate.jpg

³⁵⁰<https://commons.wikimedia.org/wiki/Category:Polypterus>

³⁵¹<https://commons.wikimedia.org/wiki/File:Joint.svg>

27.3 Diagram of the right knee. Anterior cruciate ligament labeled at center left. ³⁵²	511
27.4 Major bones of the human skull viewed from the side. ³⁵³	511
27.5 The human skeleton viewed from the front. ³⁵⁴	512
27.6 The human skeleton viewed from the back. ³⁵⁵	512
27.7 The Axial skeleton ³⁵⁶ consists of the bones in the head and trunk of the human body. It is composed of five parts; the human skull, the ossicles of the inner ear, the hyoid bone of the throat, the chest, and the vertebral column. The axial skeleton and the appendicular skeleton together form the complete skeleton.	513
27.8 The appendicular skeleton (red) and the axial skeleton together form the complete skeleton. ³⁵⁷	513
27.9 Structure of a long bone. ³⁵⁸	514
27.10 Cross-section of human bone. ³⁵⁹	514
27.11 A cross section of a long bone, showing the internal structure. ³⁶⁰	514
27.12 Bone cells. ³⁶¹	515
27.13 The body contains three types of muscle tissue ³⁶² : (a) skeletal muscle, (b) smooth muscle, and (c) cardiac muscle.	517
27.14 Bundles of muscle fibers, called fascicles, are covered by the perimysium. Muscle fibers are covered by the endomysium. ³⁶³	517
27.15 A skeletal muscle fiber is surrounded by a plasma membrane called the sarcolemma, which contains sarcoplasm, the cytoplasm of muscle cells. A muscle fiber is composed of many fibrils, which give the cell its striated appearance. ³⁶⁴	517
27.16 On the anterior and posterior views of the muscular system above, superficial muscles (those at the surface) are shown on the right side of the body while deep muscles (those underneath the superficial muscles) are shown on the left half of the body. For the legs, superficial muscles are shown in the anterior view while the posterior view shows both superficial and deep muscles. ³⁶⁵	518
27.17 Structure of neuromuscular junction. ³⁶⁶	519
27.18 Crossbridge cycling. ³⁶⁷	521
28.1 Main glands of the human endocrine system. ³⁶⁸	525
28.2 Representative examples of the three chemical classes of hormones in the human body. ³⁶⁹	526
28.3 The left diagram shows a steroid (lipid) hormone (1) entering a cell and (2) binding to a receptor protein in the nucleus, causing (3) mRNA synthesis which is the first step of protein synthesis. The right side shows protein hormones (1) binding with receptors which (2) begins a transduction pathway. The transduction pathway ends (3) with transcription factors being activated in the nucleus, and protein synthesis beginning. In both diagrams, a is the hormone, b is the cell membrane, c is the cytoplasm, and d is the nucleus. ³⁷⁰	526
28.4 The Hypothalamus–Pituitary Complex. ³⁷¹	526
28.5 The hypothalamic-pituitary-thyroid axis (HPT axis for short, a.k.a. thyroid homeostasis or thyrotropic feedback control) is part of the neuroendocrine system responsible for the regulation of metabolism and also responds to stress. ³⁷²	531

³⁵²https://commons.wikimedia.org/wiki/File:Knee_diagram.svg

³⁵³[https://commons.wikimedia.org/wiki/File:Human_skull_side_simplified_\(bones\).svg](https://commons.wikimedia.org/wiki/File:Human_skull_side_simplified_(bones).svg)

³⁵⁴https://commons.wikimedia.org/wiki/File:Human_skeleton_front_en.svg

³⁵⁵https://commons.wikimedia.org/wiki/File:Human_skeleton_back_en.svg

³⁵⁶https://commons.wikimedia.org/wiki/File:Axial_skeleton_diagram.svg

³⁵⁷https://commons.wikimedia.org/wiki/File:Appendicular_skeleton_diagram.svg

³⁵⁸https://commons.wikimedia.org/wiki/File:603_Anatomy_of_Long_Bone.jpg

³⁵⁹https://commons.wikimedia.org/wiki/File:Human_skeleton_back_en.svg

³⁶⁰https://commons.wikimedia.org/wiki/File:Illu_compact_spongy_bone.jpg

³⁶¹https://commons.wikimedia.org/wiki/File:604_Bone_cells.jpg

³⁶²https://upload.wikimedia.org/wikipedia/commons/e/e5/414_Skeletal_Smooth_Cardiac.jpg

³⁶³https://upload.wikimedia.org/wikipedia/commons/d/dd/1007_Muscle_Fibes_%28large%29.jpg

³⁶⁴[https://commons.wikimedia.org/wiki/File:1022_Muscle_Fibers_\(small\).jpg](https://commons.wikimedia.org/wiki/File:1022_Muscle_Fibers_(small).jpg)

³⁶⁵https://commons.wikimedia.org/wiki/File:1105_Anterior_and_Posterior_VIEWS_of_Muscles.jpg

³⁶⁶https://upload.wikimedia.org/wikipedia/commons/5/57/1009_Motor_End_Plate_and_Innervation.jpg

³⁶⁷https://upload.wikimedia.org/wikipedia/commons/2/24/1008_Skeletal_Muscle_Contraction.jpg

³⁶⁸https://commons.wikimedia.org/wiki/File:Illu_endocrine_system_New.png

³⁶⁹https://commons.wikimedia.org/wiki/File:1802_Examples_of_Amine_Peptide_Protein_and_Steroid_Hormone_Structure.jpg

³⁷⁰https://commons.wikimedia.org/wiki/File:Steroid_and_Lipid_Hormones.svg

³⁷¹https://commons.wikimedia.org/wiki/File:1806_The_Hypothalamus–Pituitary_Complex.jpg

³⁷²https://commons.wikimedia.org/wiki/File:Thyroid_system.svg

28.6 The hypothalamic-pituitary-gonadal axis (HPG axis) refers to the hypothalamus, pituitary gland, and gonadal glands as if these individual endocrine glands were a single entity. Because these glands often act in concert, physiologists and endocrinologists find it convenient and descriptive to speak of them as a single system. ³⁷³	532
28.7 The hypothalamic-pituitary-adrenal axis (HPA axis or HTPA axis) is a complex set of direct influences and feedback interactions among three components: the hypothalamus, the pituitary gland (a pea-shaped structure located below the thalamus), and the adrenal (also called “suprarenal”) glands (small, conical organs on top of the kidneys). ³⁷⁴	533
28.8 Diagram of the renin-angiotensin system. ³⁷⁵	535
29.1 Diagram of the female reproductive system. ³⁷⁶	539
29.2 External and internal views of the vulva. ³⁷⁷	539
29.3 The progression of the menstrual cycle and the different hormones regulating it. ³⁷⁸	540
29.4 Diagram of the male reproductive system. ³⁷⁹	540
29.5 Diagram of inner structures of testes. ³⁸⁰	540
29.6 Differentiation of the male and female reproductive systems from a common origin. ³⁸¹	541
29.7 Normal spermatogenesis, testis biopsy. ³⁸²	541
29.8 The process of spermatogenesis as the cells progress from primary spermatocytes, to secondary spermatocytes, to spermatids, to Sperm. ³⁸³	541
29.9 Micrograph of the ovarian cortex from a rhesus monkey showing several round follicles embedded in a matrix of stromal cells. A secondary follicle sectioned through the nucleus of an oocyte is at the upper left, and earlier stage follicles are at the lower right. The tissue was stained with the dyes hematoxylin and eosin. ³⁸⁴	541
29.10 The process of ovulation and gamete production, oogenesis, in a human ovary. ³⁸⁵	541
30.1 Generalized scheme of the embryonic development of animals. ³⁸⁶	543
30.2 Stem cell differentiation into various tissue types. ³⁸⁷	544
30.3 Homologous hox genes in such different animals as insects and vertebrates control embryonic development and hence the form of adult bodies. These genes have been highly conserved through hundreds of millions of years of evolution. ³⁸⁸	546
30.4 Stages during pregnancy. Embryonic development is marked in green. Weeks and months are numbered by gestation. ³⁸⁹	547
30.5 Blastocyst with an inner cell mass and trophoblast. ³⁹⁰	547
30.6 The initial stages of human embryonic development. ³⁹¹	547
30.7 An embryo at the 8-cell stage, at 3 days. ³⁹²	547
30.8 Histogenesis of the three germ layers. ³⁹³	549

³⁷³https://commons.wikimedia.org/wiki/File:Hypothalamic-pituitary-gonadal_axis.svg

³⁷⁴[https://commons.wikimedia.org/wiki/File:HPA_Axis_Diagram_\(Brian_M_Sweis_2012\).svg](https://commons.wikimedia.org/wiki/File:HPA_Axis_Diagram_(Brian_M_Sweis_2012).svg)

³⁷⁵https://commons.wikimedia.org/wiki/File:Renin-angiotensin_system_in_man_shadow.svg

³⁷⁶https://commons.wikimedia.org/wiki/File:Female_anatomy_en.svg

³⁷⁷https://commons.wikimedia.org/wiki/File:Figure_28_02_02.jpg

³⁷⁸https://upload.wikimedia.org/wikipedia/commons/f/f3/Figure_28_02_07.jpg

³⁷⁹https://commons.wikimedia.org/wiki/File:Male_anatomy_en.svg

³⁸⁰https://commons.wikimedia.org/wiki/File:Figure_28_01_03.JPG

³⁸¹https://commons.wikimedia.org/wiki/File:2915_Sexual_Differentiation-02.jpg

³⁸²https://commons.wikimedia.org/wiki/File:Normal_spermatogenesis,_intermed._mag.jpg

³⁸³https://commons.wikimedia.org/wiki/File:Figure_28_01_04.jpg

³⁸⁴https://en.wikipedia.org/wiki/File:Ovarian_cortex_in_a_rhesus_monkey.jpg

³⁸⁵https://commons.wikimedia.org/wiki/File:Oogenesis_Labeled.svg

³⁸⁶https://commons.wikimedia.org/wiki/File:Slack_Essential_Dev_Biol_Fig_02-08.jpg

³⁸⁷[https://commons.wikimedia.org/wiki/File:Final_stem_cell_differentiation_\(1\).svg](https://commons.wikimedia.org/wiki/File:Final_stem_cell_differentiation_(1).svg)

³⁸⁸https://commons.wikimedia.org/wiki/File:Genes_hox.jpeg

³⁸⁹https://upload.wikimedia.org/wikipedia/commons/5/57/Prenatal_development_table.svg

³⁹⁰https://commons.wikimedia.org/wiki/File:Blastocyst_English.svg

³⁹¹<https://commons.wikimedia.org/wiki/File:HumanEmbryogenesis.svg>

³⁹²https://commons.wikimedia.org/wiki/File:Embryo,_8_cells.jpg

³⁹³https://commons.wikimedia.org/wiki/File:Germ_layers.jpg

30.9 Embryo attached to the placenta in the amniotic cavity. ³⁹⁴	549
30.10 Embryonic development of the heart. ³⁹⁵	550
30.11 A diagram of the stages of neural tube formation. ³⁹⁶	551
30.12 An 8 week old human embryo. ³⁹⁷	552

List of Tables

2.1	30
2.2 Average values of pH in common solutions	38
2.3 Values of pH in living systems	38
3.1 Some biologically important functional groups containing oxygen or nitrogen	45
3.2 Comparison of the main classes of biological macromolecules.	47
5.1 Major classes of membrane proteins.	85
6.1 The three essential polymeric macromolecules of life	98
8.1 Phases of the eukaryotic cell cycle	125
8.2 Human Genome Assembly GRCh38.p12 (nucleotides) and GRCh38.p13 (coding genes) ³⁹⁸ . Length of DNA sequence and number of coding genes of each human chromosome. Total lengths are calculated by summing the length of the sequenced bases and estimated gaps.	129
8.3 Chromosome numbers in some plants.	129
8.4 Chromosome numbers (2n) in some animals.	130
9.1 Terminology and examples of heteroploidy in humans.	145
10.1 Mendel's Laws of Inheritance.	151
11.1A list of major DNA replication enzymes that participate in the replisome	168
11.2 The genetic code: RNA codons.	185
13.1 Nutritional types in bacterial metabolism	209
13.2 A comparison of major characteristics of the domains of Bacteria, Archaea and Eukarya	216
13.3 Nutritional types in archaeal metabolism	221
14.1A recent (2012) classification of protists.	236
14.2 Nutritional types in protist metabolism	238
18.1 Four commonly used concepts of plants.	288
18.2 Distinctive features of angiosperms.	297
19.1 Estimated numbers of described extant species for the animal groups with the largest numbers of species, along with their principal habitats (terrestrial, fresh water, and marine), and free-living or parasitic ways of life.	319

³⁹⁴https://commons.wikimedia.org/wiki/File:2910_The_Placenta-02.jpg

³⁹⁵https://commons.wikimedia.org/wiki/File:2037_Embryonic_Development_of_Heart.jpg

³⁹⁶https://en.wikipedia.org/wiki/Neural_tube#/media/File:Neural_crest.svg

³⁹⁷https://commons.wikimedia.org/wiki/File:Human_embryo_8_weeks_4.JPG

³⁹⁸<https://www.ncbi.nlm.nih.gov/grc/human/data>

19.2Comparison of 4 classes of sponges.	325
19.3Comparison of 4 classes of cnidaria.	328
19.4Comparison of major characteristics of porifera, cnidaria, ctenophora and bilateria.	333
19.5The commonly recognized classes of living molluscs.	355
21.1Comparison of major characteristics of the innate and adaptive immune systems.	391
28.1Hormones of the anterior pituitary gland.	528
28.2Hormones of the posterior pituitary gland.	530

Acknowledgements

This text was assembled using various pages from Wikipedia³⁹⁹ as my source. I copied text passages from many Wikipedia pages, remixed, edited, adapted and supplemented it with my own words to assemble the chapters of this text. I make this text available as an open educational resource under Creative Commons Attribution-Share Alike 3.0 Unported⁴⁰⁰ United States License (same⁴⁰¹ as Wikipedia) for others to do as I did and freely copy, remix, edit and adapt as required. Illustrations with associated hyperlinks are from Wikimedia Commons, which should be consulted for copyright details. Pictures and illustrations that I generated myself are subject to the Creative Commons Attribution-Share Alike 3.0 Unported⁴⁰² United States License. I want to thank all the contributors and supporters of Wikipedia for making it possible and keeping it relevant. I am grateful for the support provided by the Roxbury Community College's OER initiative⁴⁰³ "RCC Go Open" and wish to thank the members of the OER Task Force and library director William Hoag for their unwavering commitment to OER. With your support and help this text can only get better over time. I urge you to email me with your criticisms and suggestions at nsucher@rcc.mass.edu.

³⁹⁹ <https://www.wikipedia.org>

⁴⁰⁰ <https://creativecommons.org/licenses/by-sa/3.0/deed.en>

⁴⁰¹ https://en.wikipedia.org/wiki/Wikipedia:Text_of_Creative_Commons_Attribution-ShareAlike_3.0_Unported_License

⁴⁰² <https://creativecommons.org/licenses/by-sa/3.0/deed.en>

⁴⁰³ <http://libguides.rcc.mass.edu/OER/proposal>

Chapter 1

Biology: The Science of Life



Figure 1.1: A Common Eastern Bumble Bee (*Bombus impatiens*) picking up nectar and pollen from a spear cistle (*Cirsium vulgare*) on Winter Island in Salem, Massachusetts.

Biology¹ is the natural science that studies life and living organisms, including their physical structure, chemical processes, molecular interactions, physiological mechanisms, development and evolution. Despite the complexity of the science, certain unifying concepts consolidate it into a single, coherent field. Biology recognizes the cell as the basic unit of life, genes as the basic unit of heredity, and evolution as the engine that propels the creation and extinction of species. Living organisms² (from Greek ὄργανισμός, *organismos*, from ὄργανον, *organon*, i.e. "instrument, implement, tool, organ of sense or apprehension") are open systems that survive by transforming energy and decreasing their local entropy to maintain a stable and vital condition defined as homeostasis.

Sub-disciplines of biology are defined by the research methods employed and the kind of system studied: theoretical biology uses mathematical methods to formulate

quantitative models while experimental biology performs empirical experiments to test the validity of proposed theories and understand the mechanisms underlying life and how it appeared and evolved from non-living matter about 4 billion years ago through a gradual increase in the complexity of the system.

Biology derives from the Ancient Greek words of βίος; romanized *bíos* meaning "life" and -λογία; romanized *logía* (-logy) meaning "branch of study" or "to speak". Those combined make the Greek word βιολογία; romanized *biología* meaning biology. Despite this, the term βιολογία as a whole didn't exist in Ancient Greek. The first to borrow it was the English and French (*biologie*). Since the advent of the scientific era, reanalyzable as a compound using the combining forms *bio* + *logy*.

The Latin-language form of the term first appeared in 1736 when Swedish scientist Carl Linnaeus³ (Carl von Linné) used *biologi* in his *Bibliotheca Botanica*. It was used again in 1766 in a work entitled *Philosophiae naturalis sive physicae: tomus III, continens geologian, biologian, phytologian generalis*, by Michael Christoph Hanov, a disciple of Christian Wolff. The first German use, *Biologie*, was in a 1771 translation of Linnaeus' work. In 1797, Theodor Georg August Roose used the term in the preface of a book, *Grundzüge der Lehre van der Lebenskraft*. Karl Friedrich Burdach used the term in 1800 in a more restricted sense of the study of human beings from a morphological, physiological and psychological perspective (*Propädeutik zum Studien der gesammten Heilkunst*). The term came into its modern usage with the six-volume treatise *Biologie, oder Philosophie der lebenden Natur* (1802–22) by Gottfried Reinhold Treviranus, who announced:

The objects of our research will be the different forms and manifestations of life, the conditions and laws under which these phenomena occur, and the causes through which they have been effected. The science that concerns

¹<https://en.wikipedia.org/wiki/Biology>

²<https://en.wikipedia.org/wiki/Organism>

³https://en.wikipedia.org/wiki/Carl_Linnaeus

itself with these objects we will indicate by the name biology [Biologie] or the doctrine of life [Lebenslehre].

Although modern biology is a relatively recent development, sciences related to and included within it have been studied since ancient times. Natural philosophy was studied as early as the ancient civilizations of Mesopotamia, Egypt, the Indian subcontinent, and China. However, the origins of modern biology and its approach to the study of nature are most often traced back to ancient Greece. While the formal study of medicine dates back to Pharaonic Egypt, it was Aristotle (384–322 BC) who contributed most extensively to the development of biology. Especially important are his *History of Animals* and other works where he showed naturalist leanings, and later more empirical works that focused on biological causation and the diversity of life. Aristotle's successor at the Lyceum, Theophrastus, wrote a series of books on botany that survived as the most important contribution of antiquity to the plant sciences, even into the Middle Ages.

Scholars of the medieval Islamic world who wrote on biology included al-Jahiz (781–869), Al-Dīnawarī (828–896), who wrote on botany, and Rhazes (865–925) who wrote on anatomy and physiology. Medicine was especially well studied by Islamic scholars working in Greek philosopher traditions, while natural history drew heavily on Aristotelian thought, especially in upholding a fixed hierarchy of life.

Biology began to quickly develop and grow with Anton van Leeuwenhoek's⁴ dramatic improvement of the microscope. It was then that scholars discovered spermatozoa, bacteria, infusoria and the diversity of microscopic life. Investigations by Jan Swammerdam led to new interest in entomology and helped to develop the basic techniques of microscopic dissection and staining.

Advances in microscopy also had a profound impact on biological thinking. In the early 19th century, a number of biologists pointed to the central importance of the cell. Then, in 1838, Schleiden and Schwann began promoting the now universal ideas that (1) the basic unit of organisms is the cell and (2) that individual cells have all the characteristics of life, although they opposed the idea that (3) all cells come from the division of other cells. Thanks to the work of Robert Remak and Rudolf Virchow, however, by the 1860s most biologists accepted all three tenets of what came to be known as cell theory.

Meanwhile, taxonomy and classification became the focus of natural historians. Carl Linnaeus published a basic taxonomy for the natural world in 1735 (variations of which have been in use ever since), and in the 1750s introduced

scientific names for all his species. Georges-Louis Leclerc⁵, Comte de Buffon, treated species as artificial categories and living forms as malleable—even suggesting the possibility of common descent. Although he was opposed to evolution, Buffon is a key figure in the history of evolutionary thought; his work influenced the evolutionary theories of both Lamarck and Darwin.

Serious evolutionary thinking originated with the works of Jean-Baptiste Lamarck⁶, who was the first to present a coherent theory of evolution. He posited that evolution was the result of environmental stress on properties of animals, meaning that the more frequently and rigorously an organ was used, the more complex and efficient it would become, thus adapting the animal to its environment. Lamarck believed that these acquired traits could then be passed on to the animal's offspring, who would further develop and perfect them. However, it was the British naturalist Charles Darwin⁷, combining the biogeographical approach of Humboldt, the uniformitarian geology of Lyell, Malthus's writings on population growth, and his own morphological expertise and extensive natural observations, who forged a more successful evolutionary theory based on natural selection; similar reasoning and evidence led Alfred Russel Wallace⁸ to independently reach the same conclusions. Although it was the subject of controversy (which continues to this day), Darwin's theory quickly spread through the scientific community and soon became a central axiom of the rapidly developing science of biology.

The discovery of the physical representation of heredity came along with evolutionary principles and population genetics. In the 1940s and early 1950s, experiments pointed to DNA as the component of chromosomes that held the trait-carrying units that had become known as genes. A focus on new kinds of model organisms such as viruses and bacteria, along with the discovery of the double-helical structure of DNA in 1953, marked the transition to the era of molecular genetics. From the 1950s to the present times, biology has been vastly extended in the molecular domain. The genetic code was cracked by Har Gobind Khorana, Robert W. Holley and Marshall Warren Nirenberg after DNA was understood to contain codons. Finally, the Human Genome Project was launched in 1990 with the goal of mapping the general human genome. This project was essentially completed in 2003, with further analysis still being published. The Human Genome Project was the first step in a globalized effort to incorporate accumulated knowledge of biology into a functional, molecular definition of the human body and the bodies of other organisms.

⁵https://en.wikipedia.org/wiki/Georges-Louis_Leclerc,_Comte_de_Buffon

⁶https://en.wikipedia.org/wiki/Jean-Baptiste_Lamarck

⁷https://en.wikipedia.org/wiki/Charles_Darwin

⁸https://en.wikipedia.org/wiki/Alfred_Russel_Wallace

1.1 Characteristics of Life

In the past, there have been many attempts to define what is meant by “life” through obsolete concepts such as odic force, hylomorphism, spontaneous generation and vitalism, that have now been disproved by biological discoveries. Aristotle is considered to be the first person to classify organisms. Later, Carl Linnaeus introduced his system of binomial nomenclature for the classification of species. Eventually new groups and categories of life were discovered, such as cells and microorganisms, forcing dramatic revisions of the structure of relationships between living organisms. Though currently only known on Earth, life need not be restricted to it, and many scientists speculate in the existence of extraterrestrial life. Artificial life is a computer simulation or human-made reconstruction of any aspect of life, which is often used to examine systems related to natural life.

Death is the permanent termination of all biological functions which sustain an organism, and as such, is the end of its life. Extinction is the term describing the dying out of a group or taxon, usually a species. Fossils are the preserved remains or traces of organisms.

The definition of life has long been a challenge for scientists and philosophers, with many varied definitions put forward. This is partially because life is a process, not a substance. This is complicated by a lack of knowledge of the characteristics of living entities, if any, that may have developed outside of Earth. Philosophical definitions of life have also been put forward, with similar difficulties on how to distinguish living things from the non-living. Legal definitions of life have also been described and debated, though these generally focus on the decision to declare a human dead, and the legal ramifications of this decision.

Since there is no unequivocal definition of life, most current definitions in biology are descriptive. Life is considered a characteristic of something that preserves, furthers or reinforces its existence in the given environment. This characteristic exhibits all or most of the following traits:

- Homeostasis: regulation of the internal environment to maintain a constant state; for example, sweating to reduce temperature
- Organization: being structurally composed of one or more cells – the basic units of life
- Metabolism: transformation of energy by converting chemicals and energy into cellular components (anabolism) and decomposing organic matter (catabolism). Living things require energy to maintain internal organization (homeostasis) and to produce the other phenomena associated with life.
- Growth: maintenance of a higher rate of anabolism than catabolism. A growing organism increases in

size in all of its parts, rather than simply accumulating matter.

- Adaptation: the ability to change over time in response to the environment. This ability is fundamental to the process of evolution and is determined by the organism's heredity, diet, and external factors.
- Response to stimuli: a response can take many forms, from the contraction of a unicellular organism to external chemicals, to complex reactions involving all the senses of multicellular organisms. A response is often expressed by motion; for example, the leaves of a plant turning toward the sun (phototropism), and chemotaxis.
- Reproduction: the ability to produce new individual organisms, either asexually from a single parent organism or sexually from two parent organisms.

These complex processes, called physiological functions, have underlying physical and chemical bases, as well as signaling and control mechanisms that are essential to maintaining life.

More than 99% of all species of life forms, amounting to over five billion species, that ever lived on Earth are estimated to be extinct.

Although the number of Earth's catalogued species of lifeforms is between 1.2 million and 2 million, the total number of species in the planet is uncertain. Estimates range from 8 million to 100 million, with a more narrow range between 10 and 14 million, but it may be as high as 1 trillion (with only one-thousandth of one percent of the species described) according to studies realized in May 2016. The total number of related DNA base pairs on Earth is estimated at 5.0×10^{37} and weighs 50 billion tonnes. In comparison, the total mass of the biosphere has been estimated to be as much as 4 TtC (trillion tons of carbon). In July 2016, scientists reported identifying a set of 355 genes from the Last Universal Common Ancestor (LUCA) of all organisms living on Earth.

1.2 Origin of Life

The Ancient Greeks believed that living things could spontaneously come into being from nonliving matter, and that the goddess Gaia could make life arise spontaneously from stones – a process known as Generatio spontanea. Aristotle⁹ disagreed, but he still believed that creatures could arise from dissimilar organisms or from soil. Variations of this concept of spontaneous generation still existed as late as the 17th century, but towards the end of the 17th century, a series of observations and arguments began that eventually discredited such ideas. This advance in sci-

⁹<https://en.wikipedia.org/wiki/Aristotle>

tific understanding was met with much opposition, with personal beliefs and individual prejudices often obscuring the facts.

William Harvey¹⁰ (1578–1657) was an early proponent of all life beginning from an egg, *omne vivum ex ovo*. Francesco Redi¹¹, an Italian physician, proved as early as 1668 that higher forms of life did not originate spontaneously by demonstrating that maggots come from eggs of flies. But proponents of spontaneous generation claimed that this did not apply to microbes and continued to hold that these could arise spontaneously. Attempts to disprove the spontaneous generation of life from non-life continued in the early 19th century with observations and experiments by Franz Schulze and Theodor Schwann¹². In 1745, John Needham¹³ added chicken broth to a flask and boiled it. He then let it cool and waited. Microbes grew, and he proposed it as an example of spontaneous generation. In 1768, Lazzaro Spallanzani¹⁴ repeated Needham's experiment but removed all the air from the flask. No growth occurred. In 1854, Heinrich G. F. Schröder (1810–1885) and Theodor von Dusch, and in 1859, Schröder alone, repeated the Helmholtz filtration experiment and showed that living particles can be removed from air by filtering it through cotton-wool.

In 1864, Louis Pasteur¹⁵ finally announced the results of his scientific experiments. In a series of experiments similar to those performed earlier by Needham and Spallanzani, Pasteur demonstrated that life does not arise in areas that have not been contaminated by existing life. Pasteur's empirical results were summarized in the phrase *Omne vivum ex vivo*, Latin for "all life [is] from life".

All known life forms share fundamental molecular mechanisms, reflecting their common descent; based on these observations, hypotheses on the origin of life attempt to find a mechanism explaining the formation of a universal common ancestor, from simple organic molecules via pre-cellular life to protocells and metabolism. Models have been divided into "genes-first" and "metabolism-first" categories, but a recent trend is the emergence of hybrid models that combine both categories.

Life on Earth is based on carbon and water. Carbon provides stable frameworks for complex chemicals and can be easily extracted from the environment, especially from carbon dioxide. There is no other chemical element whose properties are similar enough to carbon's to be called an analogue; silicon, the element directly below carbon on the

periodic table, does not form very many complex stable molecules, and because most of its compounds are water-insoluble and because silicon dioxide is a hard and abrasive solid in contrast to carbon dioxide at temperatures associated with living things, it would be more difficult for organisms to extract. The elements boron and phosphorus have more complex chemistries, but suffer from other limitations relative to carbon. Water is an excellent solvent and has two other useful properties: the fact that ice floats enables aquatic organisms to survive beneath it in winter; and its molecules have electrically negative and positive ends, which enables it to form a wider range of compounds than other solvents can. Other good solvents, such as ammonia, are liquid only at such low temperatures that chemical reactions may be too slow to sustain life, and lack water's other advantages. Organisms based on alternative biochemistry may, however, be possible on other planets.

Abiogenesis¹⁶, or informally the origin of life, is the natural process of life arising from non-living matter, such as simple organic compounds. The prevailing scientific hypothesis is that the transition from non-living to living entities was not a single event, but a gradual process of increasing complexity. Although the occurrence of abiogenesis is uncontroversial among scientists, its possible mechanisms are poorly understood. There are several principles and hypotheses for how abiogenesis could have occurred. Life on Earth first appeared as early as 4.28 billion years ago, soon after ocean formation 4.41 billion years ago, and not long after the formation of the Earth 4.54 billion years ago. The earliest known life forms are microfossils of bacteria.

There is no current scientific consensus as to how life originated. However, many accepted scientific models build on the Miller–Urey experiment and the work of Sidney Fox, which show that conditions on the primitive Earth favored chemical reactions that synthesize amino acids and other organic compounds from inorganic precursors, and phospholipids spontaneously form lipid bilayers, the basic structure of a cell membrane.

The classic 1952 Miller–Urey experiment (Figure 1.2) demonstrated that most amino acids, the chemical constituents of the proteins used in all living organisms, can be synthesized from inorganic compounds under conditions intended to replicate those of the early Earth. The experiment used water (H_2O), methane (CH_4), ammonia (NH_3), and hydrogen (H_2). The chemicals were all sealed inside a sterile 5-liter glass flask connected to a 500 ml flask half-full of water. The water in the smaller flask was heated to induce evaporation, and the water vapour was allowed to enter the larger flask. Continuous electrical sparks were fired between the electrodes to simulate lightning in the

¹⁰https://en.wikipedia.org/wiki/William_Harvey

¹¹https://en.wikipedia.org/wiki/Francesco_Redi

¹²https://en.wikipedia.org/wiki/Theodor_Schwann

¹³https://en.wikipedia.org/wiki/John_Needham

¹⁴https://en.wikipedia.org/wiki/Lazzaro_Spallanzani

¹⁵https://en.wikipedia.org/wiki/Louis_Pasteur

¹⁶<https://en.wikipedia.org/wiki/Abiogenesis>

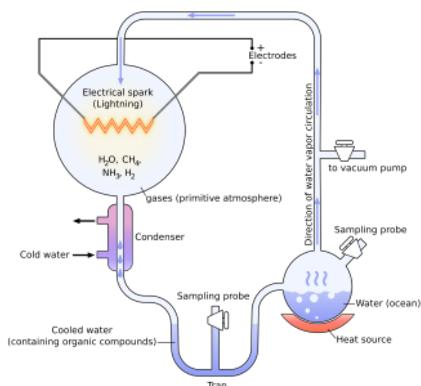


Figure 1.2: The Miller-Urey experiment¹⁷ was a chemical experiment that simulated the conditions thought at the time (1952) to be present on the early Earth and tested the chemical origin of life under those conditions. The experiment at the time supported Alexander Oparin's and J. B. S. Haldane's hypothesis that putative conditions on the primitive Earth favoured chemical reactions that synthesized more complex organic compounds from simpler inorganic precursors. Considered to be the classic experiment investigating abiogenesis, it was conducted in 1952 by Stanley Miller, with assistance from Harold Urey, at the University of Chicago and later the University of California, San Diego and published the following year.

water vapour and gaseous mixture, and then the simulated atmosphere was cooled again so that the water condensed and trickled into a U-shaped trap at the bottom of the apparatus.

After a day, the solution collected at the trap had turned pink in colour, and after a week of continuous operation the solution was deep red and turbid. The boiling flask was then removed, and mercuric chloride was added to prevent microbial contamination. The reaction was stopped by adding barium hydroxide and sulfuric acid, and evaporated to remove impurities. Using paper chromatography, Miller identified five amino acids present in the solution: glycine, α -alanine and β -alanine were positively identified, while aspartic acid and α -aminobutyric acid (AABA) were less certain, due to the spots being faint. Complex organic molecules occur in the Solar System and in interstellar space, and these molecules may have provided starting material for the development of life on Earth.

Living organisms synthesize proteins, which are polymers of amino acids using instructions encoded by deoxyribonucleic acid (DNA). Protein synthesis entails intermediary ribonucleic acid (RNA) polymers. One possibility for how life began is that genes originated first, followed by proteins; the alternative being that proteins came first

and then genes.

However, because genes and proteins are both required to produce the other, the problem of considering which came first is like that of the chicken or the egg. Most scientists have adopted the hypothesis that because of this, it is unlikely that genes and proteins arose independently.

Therefore, a possibility, first suggested by Francis Crick, is that the first life was based on RNA, which has the DNA-like properties of information storage and the catalytic properties of some proteins. This is called the RNA world hypothesis, and it is supported by the observation that many of the most critical components of cells (those that evolve the slowest) are composed mostly or entirely of RNA. Also, many critical cofactors (ATP, Acetyl-CoA, NADH, etc.) are either nucleotides or substances clearly related to them. The catalytic properties of RNA had not yet been demonstrated when the hypothesis was first proposed, but they were confirmed by Thomas Cech in 1986.

One issue with the RNA world hypothesis is that synthesis of RNA from simple inorganic precursors is more difficult than for other organic molecules. One reason for this is that RNA precursors are very stable and react with each other very slowly under ambient conditions, and it has also been proposed that living organisms consisted of other molecules before RNA. However, the successful synthesis of certain RNA molecules under the conditions that existed prior to life on Earth has been achieved by adding alternative precursors in a specified order with the precursor phosphate present throughout the reaction. This study makes the RNA world hypothesis more plausible.

Geological findings in 2013 showed that reactive phosphorus species (like phosphite) were in abundance in the ocean before 3.5 Ga, and that Schreibersite easily reacts with aqueous glycerol to generate phosphite and glycerol 3-phosphate. It is hypothesized that Schreibersite-containing meteorites from the Late Heavy Bombardment could have provided early reduced phosphorus, which could react with prebiotic organic molecules to form phosphorylated biomolecules, like RNA.

In 2009, experiments demonstrated Darwinian evolution of a two-component system of RNA enzymes (ribozymes) in vitro. The work was performed in the laboratory of Gerald Joyce, who stated "This is the first example, outside of biology, of evolutionary adaptation in a molecular genetic system."

Prebiotic compounds may have originated extraterrestrially. NASA findings in 2011, based on studies with meteorites found on Earth, suggest DNA and RNA components (adenine, guanine and related organic molecules) may be

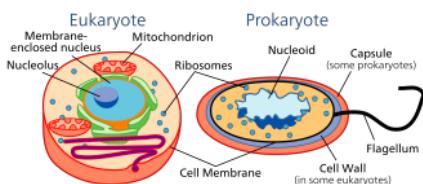


Figure 1.3: Cartoons of a eukaryotic and prokaryotic cell.¹⁸

formed in outer space.

In March 2015, NASA scientists reported that, for the first time, complex DNA and RNA organic compounds of life, including uracil, cytosine and thymine, have been formed in the laboratory under outer space conditions, using starting chemicals, such as pyrimidine, found in meteorites. Pyrimidine, like polycyclic aromatic hydrocarbons (PAHs), the most carbon-rich chemical found in the universe, may have been formed in red giants or in interstellar dust and gas clouds, according to the scientists.

According to the panspermia hypothesis, microscopic life—distributed by meteoroids, asteroids and other small Solar System bodies—may exist throughout the universe.

Since its primordial beginnings, life on Earth has changed its environment on a geologic time scale, but it has also adapted to survive in most ecosystems and conditions. Some microorganisms, called extremophiles, thrive in physically or geochemically extreme environments that are detrimental to most other life on Earth. The cell is considered the structural and functional unit of life. There are two kinds of cells, prokaryotic and eukaryotic, both of which consist of cytoplasm enclosed within a membrane and contain many biomolecules such as proteins and nucleic acids. Cells reproduce through a process of cell division, in which the parent cell divides into two or more daughter cells.

1.3 Foundations of Modern Biology

1.3.1 Cell theory

Cell theory¹⁹ states that the cell is the fundamental unit of life, that all living things are composed of one or more cells, and that all cells arise from pre-existing cells through cell division. In multicellular organisms, every cell in the organism's body derives ultimately from a single cell in a fertilized egg. The cell is also considered to be the basic unit in many pathological processes. In addition, the phenomenon of energy flow occurs in cells in processes that are part of the function known as metabolism. Finally, cells contain hereditary information (DNA), which is passed from

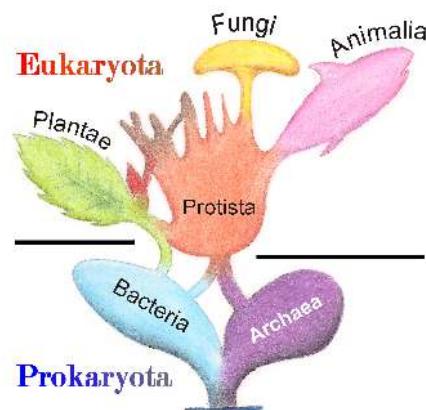


Figure 1.4: Tree diagram²⁴ illustrating the evolutionary relationship of living organisms]

cell to cell during cell division. Research into the origin of life, abiogenesis, amounts to an attempt to discover the origin of the first cells.

Cells are the basic unit of structure in every living thing, and all cells arise from pre-existing cells by division. Cell theory was formulated by Henri Dutrochet, Theodor Schwann, Rudolf Virchow and others during the early nineteenth century, and subsequently became widely accepted. The activity of an organism depends on the total activity of its cells, with energy flow occurring within and between them. Cells contain hereditary information that is carried forward as a genetic code during cell division.

There are two primary types of cells. Prokaryotes²⁰ lack a nucleus and other membrane-bound organelles, although they have circular DNA and ribosomes. Bacteria²¹ and Archaea²² are two domains of prokaryotes. The other primary type of cells are the eukaryotes²³, which have distinct nuclei bound by a nuclear membrane and membrane-bound organelles, including mitochondria, chloroplasts, lysosomes, rough and smooth endoplasmic reticulum, and vacuoles. In addition, they possess organized chromosomes that store genetic material. All species of large complex organisms are eukaryotes, including animals, plants and fungi, though most species of eukaryote are protist microorganisms. The conventional model is that eukaryotes evolved from prokaryotes, with the main organelles of the eukaryotes forming through endosymbiosis between bacteria and the progenitor eukaryotic cell.

²⁰<https://en.wikipedia.org/wiki/Prokaryote>

²¹<https://en.wikipedia.org/wiki/Bacteria>

²²<https://en.wikipedia.org/wiki/Archaea>

²³<https://en.wikipedia.org/wiki/Eukaryote>

¹⁹https://en.wikipedia.org/wiki/Cell_theory

A virus²⁵ is a submicroscopic infectious agent that replicates only inside the living cells of an organism. Scientific opinions differ on whether viruses are a form of life, or organic structures that interact with living organisms. They have been described as “organisms at the edge of life”, since they resemble organisms in that they possess genes, evolve by natural selection, and reproduce by creating multiple copies of themselves through self-assembly. Although they have genes, they do not have a cellular structure, which is seen as the basic unit of life. Viruses do not have their own metabolism, and require a host cell to make new products. They therefore cannot naturally reproduce outside a host cell—although bacterial species such as rickettsia and chlamydia are considered living organisms despite the same limitation. Accepted forms of life use cell division to reproduce, whereas viruses spontaneously assemble within cells. They differ from autonomous growth of crystals as they inherit genetic mutations while being subject to natural selection. Virus self-assembly within host cells has implications for the study of the origin of life, as it lends further credence to the hypothesis that life could have started as self-assembling organic molecules.

The molecular mechanisms of cell biology are based on proteins which are synthesized by the ribosomes through an enzyme-catalyzed process called protein biosynthesis. A sequence of amino acids is assembled and joined together based upon gene expression of the cell's nucleic acid. In eukaryotic cells, these proteins may then be transported and processed through the Golgi apparatus in preparation for dispatch to their destination.

Cells reproduce through a process of cell division in which the parent cell divides into two or more daughter cells. For prokaryotes, cell division occurs through a process of fission in which the DNA is replicated, then the two copies are attached to parts of the cell membrane. In eukaryotes, a more complex process of mitosis is followed. However, the end result is the same; the resulting cell copies are identical to each other and to the original cell (except for mutations), and both are capable of further division following an interphase period.

Multicellular organisms may have first evolved through the formation of colonies of identical cells. These cells can form group organisms through cell adhesion. The individual members of a colony are capable of surviving on their own, whereas the members of a true multi-cellular organism have developed specializations, making them dependent on the remainder of the organism for survival. Such organisms are formed clonally or from a single germ cell that is capable of forming the various specialized cells that form the adult organism. This specialization allows mul-

ticellular organisms to exploit resources more efficiently than single cells. In January 2016, scientists reported that, about 800 million years ago, a minor genetic change in a single molecule, called GK-PID, may have allowed organisms to go from a single cell organism to one of many cells.

Cells have evolved methods to perceive and respond to their microenvironment, thereby enhancing their adaptability. Cell signaling coordinates cellular activities, and hence governs the basic functions of multicellular organisms. Signaling between cells can occur through direct cell contact using juxtarine signalling, or indirectly through the exchange of agents as in the endocrine system. In more complex organisms, coordination of activities can occur through a dedicated nervous system.

1.3.2 The Theory of Evolution

A central organizing concept in biology is that life changes and develops through evolution²⁶, and that all life-forms known have a common origin. The theory of evolution postulates that all organisms on the Earth, both living and extinct, have descended from a common ancestor or an ancestral gene pool. This universal common ancestor of all organisms is believed to have appeared about 3.5 billion years ago. Biologists regard the ubiquity of the genetic code as definitive evidence in favor of the theory of universal common descent for all bacteria, archaea, and eukaryotes.

The term “evolution” was introduced into the scientific lexicon by Jean-Baptiste de Lamarck²⁷ in 1809, and fifty years later Charles Darwin²⁸ posited a scientific model of natural selection as evolution’s driving force. Alfred Russel Wallace²⁹ independently reached the same conclusions and is recognized as the co-discoverer of this concept. Evolution is now used to explain the great variations of life found on Earth.

Darwin theorized that species flourish or die when subjected to the processes of natural selection or selective breeding. Genetic drift was embraced as an additional mechanism of evolutionary development in the modern synthesis of the theory.

The evolutionary history of the species—which describes the characteristics of the various species from which it descended—together with its genealogical relationship to every other species is known as its phylogeny. Widely varied approaches to biology generate information about phylogeny. These include the comparisons of DNA sequences, a product of molecular biology (more

²⁶<https://en.wikipedia.org/wiki/Evolution>

²⁷https://en.wikipedia.org/wiki/Jean-Baptiste_Lamarck

²⁸https://en.wikipedia.org/wiki/Charles_Darwin

²⁹https://en.wikipedia.org/wiki/Alfred_Russel_Wallace

²⁵<https://en.wikipedia.org/wiki/Virus>

particularly genomics), and comparisons of fossils or other records of ancient organisms, a product of paleontology. Biologists organize and analyze evolutionary relationships through various methods, including phylogenetics, phenetics, and cladistics.

Evolution is relevant to the understanding of the natural history of life forms and to the understanding of the organization of current life forms. But, those organizations can only be understood in light of how they came to be by way of the process of evolution. Consequently, evolution is central to all fields of biology.

The evolutionary history of life on Earth traces the processes by which living and fossil organisms evolved, from the earliest emergence of life to the present. Earth formed about 4.5 billion years (Ga) ago and evidence suggests life emerged prior to 3.7 Ga. (Although there is some evidence of life as early as 4.1 to 4.28 Ga, it remains controversial due to the possible non-biological formation of the purported fossils.) The similarities among all known present-day species indicate that they have diverged through the process of evolution from a common ancestor. Approximately 1 trillion species currently live on Earth of which only 1.75–1.8 million have been named and 1.6 million documented in a central database. These currently living species represent less than one percent of all species that have ever lived on earth.

The earliest evidence of life comes from biogenic carbon signatures and stromatolite fossils discovered in 3.7 billion-year-old metasedimentary rocks from western Greenland. In 2015, possible “remains of biotic life” were found in 4.1 billion-year-old rocks in Western Australia. In March 2017, putative evidence of possibly the oldest forms of life on Earth was reported in the form of fossilized microorganisms discovered in hydrothermal vent precipitates in the Nuvvuagittuq Belt of Quebec, Canada, that may have lived as early as 4.28 billion years ago, not long after the oceans formed 4.4 billion years ago, and not long after the formation of the Earth 4.54 billion years ago.

Microbial mats of coexisting bacteria and archaea were the dominant form of life in the early Archean Epoch and many of the major steps in early evolution are thought to have taken place in this environment. The evolution of photosynthesis, around 3.5 Ga, eventually led to a buildup of its waste product, oxygen, in the atmosphere, leading to the great oxygenation event, beginning around 2.4 Ga. The earliest evidence of eukaryotes (complex cells with organelles) dates from 1.85 Ga, and while they may have been present earlier, their diversification accelerated when they started using oxygen in their metabolism. Later, around 1.7 Ga, multicellular organisms began to appear, with differentiated cells performing specialised functions.



Figure 1.5: Modern stromatolites in Shark Bay, Western Australia³⁰

Sexual reproduction, which involves the fusion of male and female reproductive cells (gametes) to create a zygote in a process called fertilization is, in contrast to asexual reproduction, the primary method of reproduction for the vast majority of macroscopic organisms, including almost all eukaryotes (which includes animals and plants). However the origin and evolution of sexual reproduction remain a puzzle for biologists though it did evolve from a common ancestor that was a single celled eukaryotic species. Bilateria, animals having a left and a right side that are mirror images of each other, appeared by 555 Ma (million years ago).

The earliest plants on land date back to around 850 million years ago (Ma), from carbon isotopes in Precambrian rocks, while algae-like multicellular land plants are dated back even to about 1 billion years ago, although evidence suggests that microorganisms formed the earliest terrestrial ecosystems, at least 2.7 billion years ago (Ga). Microorganisms are thought to have paved the way for the inception of land plants in the Ordovician. Land plants were so successful that they are thought to have contributed to the Late Devonian extinction event. (The long causal chain implied seems to involve the success of early tree *archaeopteris* (1) drew down CO₂ levels, leading to global cooling and lowered sea levels, (2) roots of *archeopteris* fostered soil development which increased rock weathering, and the subsequent nutrient run-off may have triggered algal blooms resulting in anoxic events which caused marine-life die-offs. Marine species were the primary victims of the Late Devonian extinction.)

Ediacara biota appear during the Ediacaran period, while vertebrates, along with most other modern phyla originated about 525 Ma during the Cambrian explosion. During the Permian period, synapsids, including the ancestors of mammals, dominated the land, but most of this group became extinct in the Permian-Triassic extinction event 252 Ma. During the recovery from this catastrophe, archosaurs became the most abundant

land vertebrates; one archosaur group, the dinosaurs, dominated the Jurassic and Cretaceous periods. After the Cretaceous–Paleogene extinction event 65 Ma killed off the non-avian dinosaurs, mammals increased rapidly in size and diversity. Such mass extinctions may have accelerated evolution by providing opportunities for new groups of organisms to diversify.

1.3.3 Genetics

Genes³¹ are the primary units of inheritance in all organisms. A gene is a unit of heredity and corresponds to a region of DNA that influences the form or function of an organism in specific ways. All organisms, from bacteria to animals, share the same basic machinery that copies and translates DNA into proteins. Cells transcribe a DNA gene into an RNA version of the gene, and a ribosome then translates the RNA into a sequence of amino acids known as a protein. The translation code from RNA codon to amino acid is the same for most organisms. For example, a sequence of DNA that codes for insulin in humans also codes for insulin when inserted into other organisms, such as plants.

DNA³² is found as linear chromosomes in eukaryotes, and circular chromosomes in prokaryotes. A chromosome is an organized structure consisting of DNA and histones. The set of chromosomes in a cell and any other hereditary information found in the mitochondria, chloroplasts, or other locations is collectively known as a cell's genome. In eukaryotes, genomic DNA is localized in the cell nucleus, or with small amounts in mitochondria and chloroplasts. In prokaryotes, the DNA is held within an irregularly shaped body in the cytoplasm called the nucleoid. The genetic information in a genome is held within genes, and the complete assemblage of this information in an organism is called its genotype.

1.3.4 Homeostasis

Homeostasis³³ is the ability of an open system to regulate its internal environment to maintain stable conditions by means of multiple dynamic equilibrium adjustments that are controlled by interrelated regulation mechanisms. All living organisms, whether unicellular or multicellular, exhibit homeostasis.

To maintain dynamic equilibrium and effectively carry out certain functions, a system must detect and respond to perturbations. After the detection of a perturbation, a biological system normally responds through negative feedback that stabilize conditions by reducing or increasing the

activity of an organ or system. One example is the release of glucagon when sugar levels are too low.

1.3.5 Energy

The survival of a living organism depends on the continuous input of energy³⁴. Chemical reactions that are responsible for its structure and function are tuned to extract energy from substances that act as its food and transform them to help form new cells and sustain them. In this process, molecules of chemical substances that constitute food play two roles; first, they contain energy that can be transformed and reused in that organism's biological, chemical reactions; second, food can be transformed into new molecular structures (biomolecules) that are of use to that organism.

The organisms responsible for the introduction of energy into an ecosystem are known as producers or autotrophs. Nearly all such organisms originally draw their energy from the sun. Plants and other phototrophs use solar energy via a process known as photosynthesis to convert raw materials into organic molecules, such as ATP, whose bonds can be broken to release energy. A few ecosystems, however, depend entirely on energy extracted by chemotrophs from methane, sulfides, or other non-luminal energy sources.

Some of the energy thus captured produces biomass and energy that is available for growth and development of other life forms. The majority of the rest of this biomass and energy are lost as waste molecules and heat. The most important processes for converting the energy trapped in chemical substances into energy useful to sustain life are metabolism and cellular respiration.

1.4 Areas And Levels of Study And Research in Biology

Molecular biology³⁵ is the study of biology at the molecular level. This field overlaps with other areas of biology, particularly those of genetics and biochemistry. Molecular biology is a study of the interactions of the various systems within a cell, including the interrelationships of DNA, RNA, and protein synthesis and how those interactions are regulated.

The next larger scale, cell biology³⁶, studies the structural and physiological properties of cells, including their internal behavior, interactions with other cells, and with their environment. This is done on both the microscopic

³¹<https://en.wikipedia.org/wiki/Gene>

³²<https://en.wikipedia.org/wiki/DNA>

³³<https://en.wikipedia.org/wiki/Homeostasis>

³⁴<https://en.wikipedia.org/wiki/Energy>

³⁵https://en.wikipedia.org/wiki/Molecular_biology

³⁶https://en.wikipedia.org/wiki/Cell_biology

and molecular levels, for unicellular organisms such as bacteria, as well as the specialized cells of multicellular organisms such as humans. Understanding the structure and function of cells is fundamental to all of the biological sciences. The similarities and differences between cell types are particularly relevant to molecular biology.

Anatomy³⁷ is a treatment of the macroscopic forms of such structures organs and organ systems.

Genetics³⁸ is the science of genes, heredity, and the variation of organisms. Genes encode the information needed by cells for the synthesis of proteins, which in turn play a central role in influencing the final phenotype of the organism. Genetics provides research tools used in the investigation of the function of a particular gene, or the analysis of genetic interactions. Within organisms, genetic information is physically represented as chromosomes, within which it is represented by a particular sequence of amino acids in particular DNA molecules.

Developmental biology³⁹ studies the process by which organisms grow and develop. Developmental biology, originated from embryology, studies the genetic control of cell growth, cellular differentiation, and “cellular morphogenesis,” which is the process that progressively gives rise to tissues, organs, and anatomy. Model organisms for developmental biology include the round worm *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster*, the zebrafish *Danio rerio*, the mouse *Mus musculus*, and the weed *Arabidopsis thaliana*. (A model organism is a species that is extensively studied to understand particular biological phenomena, with the expectation that discoveries made in that organism provide insight into the workings of other organisms.)

Physiology⁴⁰ is the study of the mechanical, physical, and biochemical processes of living organisms function as a whole. The theme of “structure to function” is central to biology. Physiological studies have traditionally been divided into plant physiology and animal physiology, but some principles of physiology are universal, no matter what particular organism is being studied. For example, what is learned about the physiology of yeast cells can also apply to human cells. The field of animal physiology extends the tools and methods of human physiology to non-human species. Plant physiology borrows techniques from both research fields.

Evolutionary biology⁴¹ is concerned with the origin and descent of species, and their change over time. It employs scientists from many taxonomically oriented disciplines,

for example, those with special training in particular organisms such as mammalogy, ornithology, botany, or herpetology, but are of use in answering more general questions about evolution.

Evolutionary biology is partly based on paleontology⁴², which uses the fossil record to answer questions about the mode and tempo of evolution, and partly on the developments in areas such as population genetics. In the 1980s, developmental biology re-entered evolutionary biology after its initial exclusion from the modern synthesis through the study of evolutionary developmental biology. Phylogenetics, systematics, and taxonomy are related fields often considered part of evolutionary biology.

Multiple speciation events create a tree structured system of relationships between species. The role of systematics⁴³ is to study these relationships and thus the differences and similarities between species and groups of species. However, systematics was an active field of research long before evolutionary thinking was common.

Traditionally, living things have been divided into five kingdoms: Monera; Protista; Fungi; Plantae; Animalia. However, many scientists now consider this five-kingdom system outdated. Modern alternative classification systems generally begin with the three-domain system: Archaea (originally Archaebacteria); Bacteria (originally Eubacteria) and Eukaryota (including protists, fungi, plants, and animals). These domains reflect whether the cells have nuclei or not, as well as differences in the chemical composition of key biomolecules such as ribosomes.

Further, each kingdom is broken down recursively until each species is separately classified. The order is: Domain; Kingdom; Phylum; Class; Order; Family; Genus; Species.

Outside of these categories, there are obligate intracellular parasites that are “on the edge of life” in terms of metabolic activity, meaning that many scientists do not actually classify such structures as alive, due to their lack of at least one or more of the fundamental functions or characteristics that define life. They are classified as viruses, viroids, prions, or satellites.

The scientific name of an organism is generated from its genus and species. For example, humans are listed as *Homo sapiens*. *Homo* is the genus, and *sapiens* the species. When writing the scientific name of an organism, it is proper to capitalize the first letter in the genus and put all of the species in lowercase. Additionally, the entire term may be italicized or underlined.

The dominant classification system is called the

³⁷<https://en.wikipedia.org/wiki/Anatomy>

³⁸<https://en.wikipedia.org/wiki/Genetics>

³⁹https://en.wikipedia.org/wiki/Developmental_biology

⁴⁰<https://en.wikipedia.org/wiki/Physiology>

⁴¹https://en.wikipedia.org/wiki/Evolutionary_biology

⁴²<https://en.wikipedia.org/wiki/Paleontology>

⁴³<https://en.wikipedia.org/wiki/Systematics>



Figure 1.6: Eukaryotes and some examples of their diversity⁴⁴ – clockwise from top left: Red mason bee, *Boletus edulis*, chimpanzee, *Isotricha intestinalis*, *Ranunculus asiaticus*, and *Volvox carteri*.

Linnaean taxonomy⁴⁶. It includes ranks and binomial nomenclature. How organisms are named is governed by international agreements such as the International Code of Nomenclature for algae, fungi, and plants (ICN), the International Code of Zoological Nomenclature (ICZN), and the International Code of Nomenclature of Bacteria (ICNB). The classification of viruses, viroids, prions, and all other sub-viral agents that demonstrate biological characteristics is conducted by the International Committee on Taxonomy of Viruses (ICTV) and is known as the International Code of Viral Classification and Nomenclature (ICVCN).

Ecology⁴⁷ is the study of the distribution and abundance of living organisms, the interaction between them and their environment. An organism shares an environment that includes other organisms and biotic factors as well as local abiotic factors (non-living) such as climate and ecology. One reason that biological systems can be difficult to study is that so many different interactions with other organisms and the environment are possible, even on small scales. A microscopic bacterium responding to a local sugar gradient is responding to its environment as much as a lion searching for food in the African savanna. For any species, behaviors can be co-operative, competitive, parasitic, or symbiotic. Matters become more complex when two or more species interact in an ecosystem.

Ecological systems are studied at several different levels, from the scale of the ecology of individual organisms, to those of populations, to the ecosystems and finally the

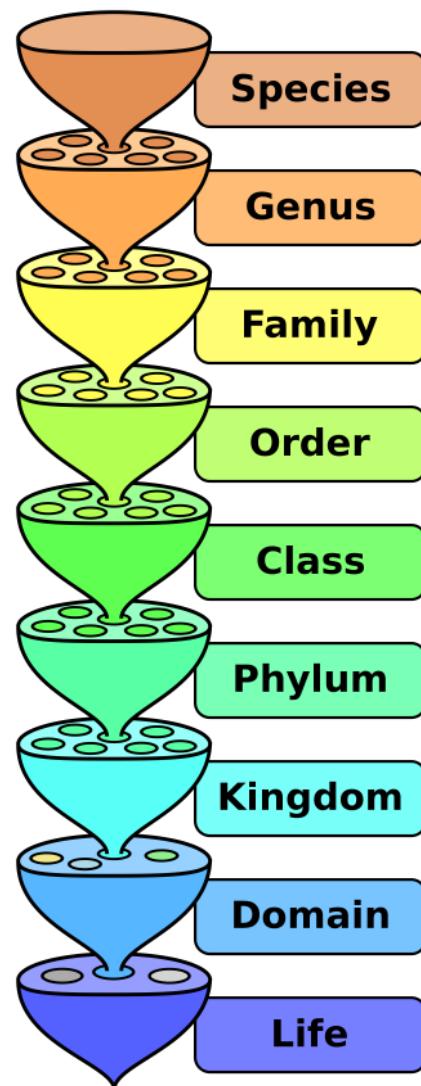


Figure 1.7: The hierarchy of biological classification's eight major taxonomic ranks from the most specific (top) to the most general (bottom). Intermediate minor rankings are not shown. This diagram uses a 3 Domains / 6 Kingdoms format⁴⁵

⁴⁶ https://en.wikipedia.org/wiki/Linnaean_taxonomy

⁴⁷ <https://en.wikipedia.org/wiki/Ecology>



Figure 1.8: A range of animal behaviours.⁴⁹

biosphere. The term population biology is often used interchangeably with population ecology, although population biology is more frequently used in the case of diseases, viruses, and microbes, while the term population ecology is more commonly applied to the study of plants and animals. Ecology draws on many subdisciplines.

Ethology⁴⁸ is the study of animal behavior (particularly that of social animals such as primates and canids), and is sometimes considered a branch of zoology. Ethologists have been particularly concerned with the evolution of behavior and the understanding of behavior in terms of the theory of natural selection. In one sense, the first modern ethologist was Charles Darwin, whose book, *The Expression of the Emotions in Man and Animals*, influenced many ethologists to come.

Biogeography studies the spatial distribution of organisms on the Earth, focusing on such topics as plate tectonics, climate change, dispersal and migration, and cladistics.

1.5 Science

Science⁵⁰ (from the Latin word *scientia*, meaning “knowledge”) is a systematic enterprise that builds and organizes knowledge in the form of testable explanations and predictions about the universe.

Scientists are individuals who conduct scientific

⁴⁸<https://en.wikipedia.org/wiki/Ethology>
⁵⁰<https://en.wikipedia.org/wiki/Science>



Figure 1.9: Two scientists working at the National Cancer Institute which is part of the National Institutes of Health of the United States of America.⁵¹

research to advance knowledge in an area of interest. In classical antiquity, there was no real ancient analog of a modern scientist. Instead, philosophers engaged in the philosophical study of nature called natural philosophy, a precursor of natural science. It was not until the 19th century that the term scientist came into regular use after it was coined by the theologian, philosopher, and historian of science William Whewell in 1833. In modern times, many professional scientists are trained in an academic setting and upon completion, attain an academic degree, with the highest degree being a doctorate such as a Doctor of Philosophy (PhD). Many scientists pursue careers in various sectors of the economy such as academia, industry, government, and nonprofit organizations.

The roles of “scientists”, and their predecessors before the emergence of modern scientific disciplines, have evolved considerably over time. Scientists of different eras (and before them, natural philosophers and others who contributed to the development of science) have had widely different places in society, and the social norms, ethical values, and epistemic virtues associated with scientists—and expected of them—have changed over time as well.

Some historians point to the Scientific Revolution that began in 16th century as the period when science in a recognizably modern form developed. It wasn’t until the 19th century that sufficient socioeconomic changes occurred for scientists to emerge as a major profession.

The presence of women in science spans the earliest times of the history of science wherein they have made significant contributions. Historians with an interest in gender and science have researched the scientific endeavors and accomplishments of women, the barriers they have faced, and the strategies implemented to have their work peer-reviewed and accepted in major scientific journals and other publications.

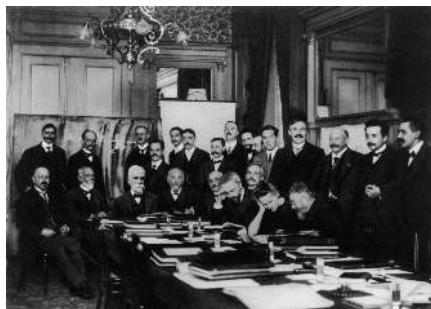


Figure 1.10: At First Solvay Conference (1911), Marie Curie (seated, second from right) confers with Henri Poincaré; standing, fourth from right, is Rutherford; second from right, Einstein; far right, Paul Langevin⁵³

The involvement of women in the field of medicine occurred in several early civilizations, and the study of natural philosophy in ancient Greece was open to women. Women contributed to the proto-science of alchemy in the first or second centuries AD. During the Middle Ages, religious convents were an important place of education for women, and some of these communities provided opportunities for women to contribute to scholarly research. The 11th century saw the emergence of the first universities; women were, for the most part, excluded from university education. Outside academia, botany was the science that benefitted most from contributions of women in early modern times. Gender roles were largely deterministic in the eighteenth century and women made substantial advances in science. During the nineteenth century, women were excluded from most formal scientific education, but they began to be admitted into learned societies during this period. In the later nineteenth century, the rise of the women's college provided jobs for women scientists and opportunities for education. Marie Curie⁵², a physicist and chemist who conducted pioneering research on radioactive decay, was the first woman to receive a Nobel Prize in Physics and became the first person to receive a second Nobel Prize in Chemistry. Forty women have been awarded the Nobel Prize between 1901 and 2010. Seventeen women have been awarded the Nobel Prize in physics, chemistry, physiology or medicine.

A list of African Americans inventors and scientists⁵⁴ documents many of the African-Americans who have invented a multitude of items or made discoveries in the course of their lives. These have ranged from practical everyday devices to applications and scientific discoveries in diverse fields, including physics, biology, math, plus the medical, nuclear and space science.

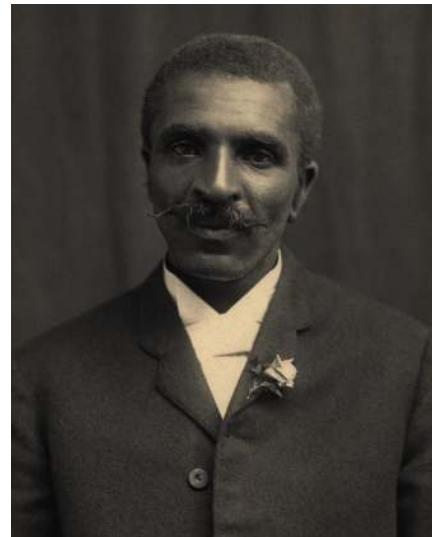


Figure 1.11: George Washington Carver⁵⁵ (1860s – January 5, 1943) was an American agricultural scientist and inventor who promoted alternative crops to cotton and methods to prevent soil depletion. He was the most prominent black scientist of the early 20th century. Apart from his work to improve the lives of farmers, Carver was also a leader in promoting environmentalism. He received numerous honors for his work, including the Spingarn Medal of the NAACP. In an era of high racial polarization, his fame reached beyond the black community. He was widely recognized and praised in the white community for his many achievements and talents. In 1941, Time magazine dubbed Carver a "Black Leonardo".

⁵²https://en.wikipedia.org/wiki/Marie_Curie

⁵⁴https://en.wikipedia.org/wiki/List_of_African-American_inventors_and_scientists

Scientists exhibit a strong curiosity about reality, with some scientists having a desire to apply scientific knowledge for the benefit of health, nations, environment, or industries. Other motivations include recognition by their peers and prestige. The Nobel Prize, a widely regarded prestigious award, is awarded annually to those who have achieved scientific advances in the fields of medicine, physics, chemistry, and economics.

The earliest roots of science can be traced to Ancient Egypt and Mesopotamia in around 3500 to 3000 BCE. Their contributions to mathematics, astronomy, and medicine entered and shaped Greek natural philosophy of classical antiquity, whereby formal attempts were made to provide explanations of events in the physical world based on natural causes. After the fall of the Western Roman Empire, knowledge of Greek conceptions of the world deteriorated in Western Europe during the early centuries (400 to 1000 CE) of the Middle Ages but was preserved in the Muslim world during the Islamic Golden Age. The recovery and assimilation of Greek works and Islamic inquiries into Western Europe from the 10th to 13th century revived "natural philosophy", which was later transformed by the Scientific Revolution that began in the 16th century as new ideas and discoveries departed from previous Greek conceptions and traditions. The scientific method soon played a greater role in knowledge creation and it was not until the 19th century that many of the institutional and professional features of science began to take shape; along with the changing of "natural philosophy" to "natural science."

Modern science is typically divided into three major branches that consist of the natural sciences (e.g., biology, chemistry, and physics), which study nature in the broadest sense; the social sciences (e.g., economics, psychology, and sociology), which study individuals and societies; and the formal sciences (e.g., logic, mathematics, and theoretical computer science), which study abstract concepts. There is disagreement, however, on whether the formal sciences actually constitute a science as they do not rely on empirical evidence. Disciplines that use existing scientific knowledge for practical purposes, such as engineering and medicine, are described as applied sciences.

Science is based on research, which is commonly conducted in academic and research institutions as well as in government agencies and companies. The practical impact of scientific research has led to the emergence of science policies that seek to influence the scientific enterprise by prioritizing the development of commercial products, armaments, health care, and environmental protection.

Natural science is concerned with the description, prediction, and understanding of natural phenomena based on empirical evidence from observation and experimentation.

1.5.1 Scientific Terminology

Scientific terminology⁵⁶ is the part of the language that is used by scientists in the context of their professional activities. While studying nature, scientists often encounter or create new material or immaterial objects and concepts and are compelled to name them.

In modern science and its applied fields such as technology and medicine, a knowledge of classical languages is not as rigid a prerequisite as it used to be. However, traces of their influence remain. Firstly, languages such as Greek, Latin and Arabic – either directly or via more recently derived languages such as French – have provided not only most of the technical terms used in Western science, but also a de facto vocabulary of roots, prefixes and suffixes for the construction of new terms as required. Echoes of the consequences sound in remarks such as "Television? The word is half Latin and half Greek."

Branches of science that are based, however tenuously, on fields of study known to the ancients, or that were established by more recent workers familiar with Greek and Latin, often use terminology that is fairly correct descriptive Latin, or occasionally Greek. Descriptive human anatomy or works on biological morphology often use such terms, for example, *musculus gluteus maximus* simply means the "largest rump muscle", where *musculus* was the Latin for "little mouse" and the name applied to muscles. During the last two centuries there has been an increasing tendency to modernise the terminology, though how beneficial that might be is subject to discussion. In other descriptive anatomical terms, whether in vertebrates or invertebrates, a *frenum* (a structure for keeping something in place) is simply the Latin for a bridle; and a *foramen* (a passage or perforation) also is the actual Latin word.

A special class of terminology that overwhelmingly is derived from classical sources, is biological classification, in which binomial nomenclature still is most often based on classical origins. Binomial nomenclature⁵⁷ ("two-term naming system") or binary nomenclature, is a formal system of naming species of living things by giving each a name composed of two parts, both of which use Latin grammatical forms, although they can be based on words from other languages. Such a name is called a binomial name (which may be shortened to just "binomial"), a binomen, binomial name or a scientific name; more informally it is also called a Latin name.

The first part of the name – the generic name – identifies the genus to which the species belongs, while the second part – the specific name or specific epithet – identifies

⁵⁶https://en.wikipedia.org/wiki/Scientific_terminology

⁵⁷https://en.wikipedia.org/wiki/Binomial_nomenclature

the species within the genus. For example, humans belong to the genus *Homo* and within this genus to the species *Homo sapiens*. *Tyrannosaurus rex* is probably the most widely known binomial. The formal introduction of this system of naming species is credited to Carl Linnaeus⁵⁸, effectively beginning with his work *Species Plantarum* in 1753. But Gaspard Bauhin⁵⁹, in as early as 1622, had introduced in his book *Pinax theatri botanici* (English, Illustrated exposition of plants) many names of genera that were later adopted by Linnaeus.

The application of binomial nomenclature is now governed by various internationally agreed codes of rules, of which the two most important are the International Code of Zoological Nomenclature (ICZN) for animals and the International Code of Nomenclature for algae, fungi, and plants (ICNfp).

In modern usage, the first letter of the first part of the name, the genus, is always capitalized in writing, while that of the second part is not, even when derived from a proper noun such as the name of a person or place. Similarly, both parts are italicized when a binomial name occurs in normal text (or underlined in handwriting). Thus the binomial name of the annual phlox (named after botanist Thomas Drummond) is now written as *Phlox drummondii*⁶⁰.

1.5.2 The Scientific Method

The scientific method⁶¹ is the process by which science is carried out. Scientific research involves using the scientific method, which seeks to objectively explain the events of nature in a reproducible way. An explanatory thought experiment or hypothesis is put forward as explanation using principles such as parsimony (also known as "Occam's Razor") and are generally expected to seek consilience – fitting well with other accepted facts related to the phenomena.

A hypothesis⁶² (plural hypotheses) is a proposed explanation for a phenomenon. For a hypothesis to be a scientific hypothesis, the scientific method requires that one can test it. An experiment is a procedure carried out to support, refute, or validate a hypothesis. Experiments provide insight into cause-and-effect by demonstrating what outcome occurs when a particular factor is manipulated. Experiments vary greatly in goal and scale, but always rely on repeatable procedure and logical analysis of the results. Scientists generally base scientific hypotheses on previous observations that cannot satisfactorily be explained with the available scientific theories. Even though the words

"hypothesis" and "theory" are often used synonymously, a scientific hypothesis is not the same as a scientific theory. A working hypothesis is a provisionally accepted hypothesis proposed for further research, in a process beginning with an educated guess or thought.

Occam's razor⁶³ or law of parsimony (Latin: *lex parsimoniae*) is the problem-solving principle that "entities should not be multiplied without necessity." The idea is attributed to English Franciscan friar William of Ockham (c. 1287–1347), a scholastic philosopher and theologian who used a preference for simplicity to defend the idea of divine miracles. It is variously paraphrased by statements like "the simplest explanation is most likely the right one". This philosophical razor advocates that when presented with competing hypotheses about the same prediction, one should select the solution with the fewest assumptions, and that this is not meant to be a way of choosing between hypotheses that make different predictions.

A hypothesis is used to make falsifiable predictions that are testable by experiment or observation. The predictions are to be posted before a confirming experiment or observation is sought, as proof that no tampering has occurred. Disproof of a prediction is evidence of progress. This is done partly through observation of natural phenomena, but also through experimentation that tries to simulate natural events under controlled conditions as appropriate to the discipline (in the observational sciences, such as astronomy or geology, a predicted observation might take the place of a controlled experiment). Experimentation is especially important in science to help establish causal relationships (to avoid the correlation fallacy).

Falsifiability⁶⁴ or refutability is the capacity for a statement, theory or hypothesis to be contradicted by evidence. For example, the statement "All swans are white" is falsifiable because one can observe that black swans exist

When a hypothesis proves unsatisfactory, it is either modified or discarded. If the hypothesis survived testing, it may become adopted into the framework of a scientific theory, a logically reasoned, self-consistent model or framework for describing the behavior of certain natural phenomena. A theory typically describes the behavior of much broader sets of phenomena than a hypothesis; commonly, a large number of hypotheses can be logically bound together by a single theory. Thus a theory is a hypothesis explaining various other hypotheses. In that vein, theories are formulated according to most of the same scientific principles as hypotheses. In addition to testing hypotheses, scientists may also generate a model, an attempt to describe or depict the phenomenon in terms of a logical, physical or mathematical representation and to generate

⁵⁸https://en.wikipedia.org/wiki/Carl_Linnaeus

⁵⁹https://en.wikipedia.org/wiki/Gaspard_Bauhin

⁶⁰https://en.wikipedia.org/wiki/Phlox_drummondii

⁶¹https://en.wikipedia.org/wiki/Scientific_method

⁶²<https://en.wikipedia.org/wiki/Hypothesis>

⁶³https://en.wikipedia.org/wiki/Occam%27s_razor

⁶⁴<https://en.wikipedia.org/wiki/Falsifiability>

new hypotheses that can be tested, based on observable phenomena.

While performing experiments to test hypotheses, scientists may have a preference for one outcome over another, and so it is important to ensure that science as a whole can eliminate this bias. This can be achieved by careful experimental design, transparency, and a thorough peer review process of the experimental results as well as any conclusions. After the results of an experiment are announced or published, it is normal practice for independent researchers to double-check how the research was performed, and to follow up by performing similar experiments to determine how dependable the results might be. Taken in its entirety, the scientific method allows for highly creative problem solving while minimizing any effects of subjective bias on the part of its users (especially the confirmation bias).

1.5.3 Dissemination of Scientific Knowledge

Scientific research is published in an enormous range of scientific literature. Scientific journals communicate and document the results of research carried out in universities and various other research institutions, serving as an archival record of science. The first scientific journals, *Journal des Sçavans* followed by the *Philosophical Transactions*, began publication in 1665. Since that time the total number of active periodicals has steadily increased. In 1981, one estimate for the number of scientific and technical journals in publication was 11,500. The United States National Library of Medicine currently indexes 5,516 journals that contain articles on topics related to the life sciences. Although the journals are in 39 languages, 91 percent of the indexed articles are published in English.

Most scientific journals cover a single scientific field and publish the research within that field; the research is normally expressed in the form of a scientific paper. Science has become so pervasive in modern societies that it is generally considered necessary to communicate the achievements, news, and ambitions of scientists to a wider populace.

An important aspect of the dissemination of scientific results is scholarly peer review⁶⁵ (also known as refereeing). This is the process of subjecting an author's scholarly work, research, or ideas to the scrutiny of others who are experts in the same field, before a paper describing this work is published in a journal, conference proceedings or as a book. The peer review helps the publisher (that is, the editor-in-chief, the editorial board or the program committee) decide whether the work should be accepted, considered acceptable with revisions, or rejected.

Researchers have peer reviewed manuscripts prior to publishing them in a variety of ways since the 18th century. The main goal of this practice is to improve the relevance and accuracy of scientific discussions. Even though experts often criticize peer review for a number of reasons, the process is still often considered the "gold standard" of science. Occasionally however, peer review approves studies that are later found to be wrong and rarely deceptive or fraudulent results are discovered prior to publication. Thus, there seems to be an element of discord between the ideology behind and the practice of peer review. By failing to effectively communicate that peer review is imperfect, the message conveyed to the wider public is that studies published in peer-reviewed journals are "true" and that peer review protects the literature from flawed science.

Open access (OA) is a set of principles and a range of practices through which research outputs are distributed online, free of cost or other access barriers. With open access strictly defined (according to the 2001 definition), or libre open access, barriers to copying or reuse are also reduced or removed by applying an open license for copyright.

The main focus of the open access movement is "peer reviewed research literature." Historically, this has centered mainly on print-based academic journals. Whereas conventional (non-open access) journals cover publishing costs through access tolls such as subscriptions, site licenses or pay-per-view charges, open-access journals are characterised by funding models which do not require the reader to pay to read the journal's contents but the authors may be charged for the cost of publishing .

1.6 Philosophy of Science

Philosophy of science⁶⁶ is a branch of philosophy concerned with the foundations, methods, and implications of science. The central questions of this study concern what qualifies as science, the reliability of scientific theories, and the ultimate purpose of science.

There is no consensus among philosophers about many of the central problems concerned with the philosophy of science, including whether science can reveal the truth about unobservable things and whether scientific reasoning can be justified at all.

Problems of philosophy of science include

- defining science
- scientific explanation
- justifying science
- separability of theory and observation

⁶⁵https://en.wikipedia.org/wiki/Peer_review

⁶⁶https://en.wikipedia.org/wiki/Philosophy_of_science

- purpose of science
- values and science

Scientists usually take for granted a set of basic assumptions that are needed to justify the scientific method: (1) that there is an objective reality shared by all rational observers; (2) that this objective reality is governed by natural laws; (3) that these laws can be discovered by means of systematic observation and experimentation. Philosophy of science seeks a deep understanding of what these underlying assumptions mean and whether they are valid.

The belief that scientific theories should and do represent metaphysical reality is known as realism. It can be contrasted with anti-realism, the view that the success of science does not depend on it being accurate about unobservable entities such as electrons. One form of anti-realism is idealism, the belief that the mind or consciousness is the most basic essence, and that each mind generates its own reality. In an idealistic world view, what is true for one mind need not be true for other minds.

There are different schools of thought in philosophy of science. The most popular position is empiricism, which holds that knowledge is created by a process involving observation and that scientific theories are the result of generalizations from such observations. Empiricism generally encompasses inductivism, a position that tries to explain the way general theories can be justified by the finite number of observations humans can make and hence the finite amount of empirical evidence available to confirm scientific theories. This is necessary because the number of predictions those theories make is infinite, which means that they cannot be known from the finite amount of evidence using deductive logic only. Many versions of empiricism exist, with the predominant ones being Bayesianism and the hypothetico-deductive method.

Logical positivism, later called logical empiricism, and both of which together are also known as neopositivism, is a branch in Western philosophy whose central thesis is the verification principle (also known as the verifiability criterion of meaning). This theory of knowledge asserts that only statements verifiable through direct observation or logical proof are meaningful. Starting in the late 1920s, groups of philosophers, scientists, and mathematicians formed the Berlin Circle and the Vienna Circle, which, in these two cities, would propound the ideas of logical positivism.

Flourishing in several European centres through the 1930s, the movement sought to prevent confusion rooted in unclear language and unverifiable claims by converting philosophy into “scientific philosophy”, which, according to the logical positivists, ought to share the bases and structures of empirical sciences’ best examples, such as

Albert Einstein’s general theory of relativity. Despite its ambition to overhaul philosophy by studying and mimicking the extant conduct of empirical science, logical positivism became erroneously stereotyped as a movement to regulate the scientific process and to place strict standards on it.

After World War II, the movement shifted to a milder variant, logical empiricism, led mainly by Carl Hempel⁶⁷, who, during the rise of Nazism, had immigrated to the United States. In the ensuing years, the movement’s central premises, still unresolved, were criticised by other philosophers, particularly Willard van Orman Quine⁶⁸ and by Austrian-British philosopher Karl Popper⁶⁹.

Empiricism has stood in contrast to rationalism, the position originally associated with Descartes, which holds that knowledge is created by the human intellect, not by observation. Critical rationalism is a contrasting 20th-century approach to science, first defined Karl Popper. Popper rejected the way that empiricism describes the connection between theory and observation. He claimed that theories are not generated by observation, but that observation is made in the light of theories and that the only way a theory can be affected by observation is when it comes in conflict with it. Popper proposed replacing verifiability with falsifiability as the landmark of scientific theories and replacing induction with falsification as the empirical method. Popper further claimed that there is actually only one universal method, not specific to science: the negative method of criticism, trial and error. It covers all products of the human mind, including science, mathematics, philosophy, and art.

Another approach, instrumentalism, colloquially termed “shut up and calculate⁷⁰,” emphasizes the utility of theories as instruments for explaining and predicting phenomena. It views scientific theories as black boxes with only their input (initial conditions) and output (predictions) being relevant. Consequences, theoretical entities, and logical structure are claimed to be something that should simply be ignored and that scientists should not make a fuss about (see interpretations of quantum mechanics). Close to instrumentalism is constructive empiricism, according to which the main criterion for the success of a scientific theory is whether what it says about observable entities is true.

In his book *The Structure of Scientific Revolutions*⁷¹ the American philosopher of science Thomas Samuel

⁶⁷ https://en.wikipedia.org/wiki/Carl_Gustav_Hempel

⁶⁸ https://en.wikipedia.org/wiki/Willard_Van_Orman_Quine

⁶⁹ https://en.wikipedia.org/wiki/Karl_Popper

⁷⁰ <https://physicstoday.scitation.org/doi/10.1063/1.1768652>

⁷¹ https://en.wikipedia.org/wiki/The_Structure_of_Scientific_Revolutions

Kuhn⁷² made several claims concerning the progress of scientific knowledge: that scientific fields undergo periodic “paradigm shifts” rather than solely progressing in a linear and continuous way, and that these paradigm shifts open up new approaches to understanding what scientists would never have considered valid before; and that the notion of scientific truth, at any given moment, cannot be established solely by objective criteria but is defined by a consensus of a scientific community. Competing paradigms are frequently incommensurable; that is, they are competing and irreconcilable accounts of reality. Thus, our comprehension of science can never rely wholly upon “objectivity” alone. Science must account for subjective perspectives as well, since all objective conclusions are ultimately founded upon the subjective conditioning/worldview of its researchers and participants.

A scientific theory is empirical and is always open to falsification if new evidence is presented. That is, no theory is ever considered strictly certain as science accepts the concept of fallibilism. The philosopher of science Karl Popper sharply distinguished truth from certainty. He wrote that scientific knowledge “consists in the search for truth,” but it “is not the search for certainty ... All human knowledge is fallible and therefore uncertain.”

New scientific knowledge rarely results in vast changes in our understanding. Knowledge in science is gained by a gradual synthesis of information from different experiments by various researchers across different branches of science; it is more like a climb than a leap. Theories vary in the extent to which they have been tested and verified, as well as their acceptance in the scientific community. For example, heliocentric theory, the theory of evolution, relativity theory, and germ theory still bear the name “theory” even though, in practice, they are considered factual. Philosopher Barry Stroud adds that, although the best definition for “knowledge” is contested, being skeptical and entertaining the possibility that one is incorrect is compatible with being correct. Therefore, scientists adhering to proper scientific approaches will doubt themselves even once they possess the truth.

The Polish and Israeli physician, biologist and philosopher of science Ludwik Fleck⁷³ wrote in his 1935 book “Entstehung und Entwicklung einer wissenschaftlichen Tatsache; Einführung in die Lehre vom Denkstil und Denkkollektiv” that the development of truth in scientific research was an unattainable ideal as different researchers were locked into thought collectives (or thought-styles). This means “that a pure and direct observation cannot exist: in the act of perceiving objects the observer, i.e. the

epistemological subject, is always influenced by the epoch and the environment to which he belongs, that is by what Fleck calls the thought style.” A “truth” was a relative value, expressed in the language or symbolism of the thought collective in which it belonged, and subject to the social and temporal structure of this collective. To state therefore that a specific truth is true or false is impossible. It is true in its own collective, but incomprehensible or unverifiable in most others. He felt that the development of scientific insights was not unidirectional and does not consist of just accumulating new pieces of information, but also in overthrowing the old ones. This overthrowing of old insights is difficult because a collective attains over time a specific way of investigating, bringing with it a blindness to alternative ways of observing and conceptualization. Change was especially possible when members of two thought collectives met and cooperated in observing, formulating hypothesis and ideas. He strongly advocated comparative epistemology. This approach anticipated later developments in social constructionism, and especially the development of critical science and technology studies.

In his book Against Method and Science in a Free Society the philosopher Paul Feyerabend⁷⁴ defended the idea that there are no methodological rules which are always used by scientists. He objected to any single prescriptive scientific method on the grounds that any such method would limit the activities of scientists, and hence restrict scientific progress. In his view, science would benefit most from a “dose” of theoretical anarchism. He also thought that theoretical anarchism was desirable because it was more humanitarian than other systems of organization, by not imposing rigid rules on scientists.

For is it not possible that science as we know it today, or a “search for the truth” in the style of traditional philosophy, will create a monster? Is it not possible that an objective approach that frowns upon personal connections between the entities examined will harm people, turn them into miserable, unfriendly, self-righteous mechanisms without charm or humour? “Is it not possible,” asks Kierkegaard, “that my activity as an objective [or critico-rational] observer of nature will weaken my strength as a human being?” I suspect the answer to many of these questions is affirmative and I believe that a reform of the sciences that makes them more anarchic and more subjective (in Kierkegaard’s sense) is urgently needed.

Against Method: Outline of an Anarchistic

⁷²https://en.wikipedia.org/wiki/Thomas_Kuhn

⁷³https://en.wikipedia.org/wiki/Ludwik_Fleck

⁷⁴https://en.wikipedia.org/wiki/Paul_Feyerabend

Theory of Knowledge (1975)

Feyerabend's position was seen as radical in the philosophy of science, because it implies that philosophy can neither succeed in providing a general description of science, nor in devising a method for differentiating products of science from non-scientific entities like myths. (Feyerabend's position also implies that philosophical guidelines should be ignored by scientists, if they are to aim for progress.)

To support his position that methodological rules generally do not contribute to scientific success, Feyerabend provides counterexamples to the claim that (good) science operates according to a certain fixed method. He took some examples of episodes in science that are generally regarded as indisputable instances of progress (e.g. the Copernican revolution), and argued that these episodes violated all common prescriptive rules of science. Moreover, he claimed that applying such rules in these historical situations would actually have prevented scientific revolution.

According to Feyerabend, new theories came to be accepted not because of their accord with scientific method, but because their supporters made use of any trick – rational, rhetorical or ribald – in order to advance their cause. Without a fixed ideology, or the introduction of religious tendencies, the only approach which does not inhibit progress (using whichever definition one sees fit) is “anything goes”: “‘anything goes’ is not a ‘principle’ I hold... but the terrified exclamation of a rationalist who takes a closer look at history.”

1.7 Science And Religion

Historians of science and of religion⁷⁵, philosophers, theologians, scientists, and others from various geographical regions and cultures have addressed numerous aspects of the relationship between religion and science. Critical questions in this debate include whether religion and science are compatible, whether religious beliefs can be conducive to science (or necessarily inhibit it), and what the nature of religious beliefs is. Another question is whether or not science has for some people become a new religion.

Even though the ancient and medieval worlds did not have conceptions resembling the modern understandings of “science” or of “religion”, certain elements of modern ideas on the subject recur throughout history. The pair-structured phrases “religion and science” and “science and religion” first emerged in the literature in the 19th century. This coincided with the refining of “science” (from the studies of “natural philosophy”) and of “religion” as distinct concepts in the preceding few centuries—partly due to profes-

sionalization of the sciences, the Protestant Reformation, colonization, and globalization. Since then the relationship between science and religion has been characterized in terms of ‘conflict’, ‘harmony’, ‘complexity’, and ‘mutual independence’, among others.

Both science and religion are complex social and cultural endeavors that vary across cultures and change over time. Most scientific (and technical) innovations prior to the scientific revolution were achieved by societies organized by religious traditions. Ancient pagan, Islamic, and Christian scholars pioneered individual elements of the scientific method. Roger Bacon, often credited with formalizing the scientific method, was a Franciscan friar. Confucian thought, whether religious or non-religious in nature, has held different views of science over time. Many 21st-century Buddhists view science as complementary to their beliefs. While the classification of the material world by the ancient Indians and Greeks into air, earth, fire and water was more metaphysical, and figures like Anaxagoras questioned certain popular views of Greek divinities, medieval Middle Eastern scholars empirically classified materials.

Events in Europe such as the Galileo affair of the early 17th century, associated with the scientific revolution and the Age of Enlightenment, led scholars such as John William Draper to postulate (c. 1874) a conflict thesis, suggesting that religion and science have been in conflict methodologically, factually and politically throughout history. Some contemporary scientists (such as Richard Dawkins, Lawrence Krauss, Peter Atkins, and Donald Prothero) subscribe to this thesis. However, the conflict thesis has lost favor among most contemporary historians of science.

Many scientists, philosophers, and theologians throughout history, such as Francisco Ayala, Kenneth R. Miller and Francis Collins, have seen compatibility or interdependence between religion and science. Biologist Stephen Jay Gould, other scientists, and some contemporary theologians regard religion and science as non-overlapping magisteria, addressing fundamentally separate forms of knowledge and aspects of life. Some theologians or historians of science, including John Lennox, Thomas Berry, Brian Swimme and Ken Wilber propose an interconnection between science and religion, while others such as Ian Barbour believe there are even parallels.

Public acceptance of scientific facts may sometimes be influenced by religious beliefs such as in the United States, where some reject the concept of evolution by natural selection, especially regarding human beings. Nevertheless, the American National Academy of Sciences has written that “the evidence for evolution can be fully compatible with religious faith”, a view endorsed by many religious de-

⁷⁵ https://en.wikipedia.org/wiki/Relationship_between_religion_and_science

nominations.

The concepts of "science" and "religion" are a recent invention: "religion" emerged in the 17th century in the midst of colonization and globalization and the Protestant Reformation, "science" emerged in the 19th century in the midst of attempts to narrowly define those who studied nature. Originally what is now known as "science" was pioneered as "natural philosophy".

It was in the 19th century that the terms "Buddhism", "Hinduism", "Taoism", "Confucianism" and "World Religions" first emerged. In the ancient and medieval world, the etymological Latin roots of both science (*scientia*) and religion (*religio*) were understood as inner qualities of the individual or virtues, never as doctrines, practices, or actual sources of knowledge.

It was in the 19th century that the concept of "science" received its modern shape with new titles emerging such as "biology" and "biologist", "physics", and "physicist", among other technical fields and titles; institutions and communities were founded, and unprecedented applications to and interactions with other aspects of society and culture occurred. Even in the 19th century, a treatise by Lord Kelvin and Peter Guthrie Tait's, which helped define much of modern physics, was titled *Treatise on Natural Philosophy* (1867).

It was in the 17th century that the concept of "religion" received its modern shape despite the fact that ancient texts like the Bible, the Quran, and other texts did not have a concept of religion in the original languages and neither did the people or the cultures in which these texts were written. In the 19th century, Max Müller noted that what is called ancient religion today, would have been called "law" in antiquity.

The development of sciences (especially natural philosophy) in Western Europe during the Middle Ages, has considerable foundation in the works of the Arabs who translated Greek and Latin compositions. The works of Aristotle played a major role in the institutionalization, systematization, and expansion of reason. Christianity accepted reason within the ambit of faith. In Christendom, reason was considered subordinate to revelation, which contained the ultimate truth and this truth could not be challenged. In medieval universities, the faculty for natural philosophy and theology were separate, and discussions pertaining to theological issues were often not allowed to be undertaken by the faculty of philosophy.

Natural philosophy, as taught in the arts faculties of the universities, was seen as an essential area of study in its own right and was considered necessary for almost every area of study. It was an independent field, separated from theology, and enjoyed a good deal of intellectual freedom

as long as it was restricted to the natural world. In general, there was religious support for natural science by the late Middle Ages and a recognition that it was an important element of learning.

The extent to which medieval science led directly to the new philosophy of the scientific revolution remains a subject for debate, but it certainly had a significant influence.

The Middle Ages laid ground for the developments that took place in science, during the Renaissance which immediately succeeded it. By 1630, ancient authority from classical literature and philosophy, as well as their necessity, started eroding, although scientists were still expected to be fluent in Latin, the international language of Europe's intellectuals. With the sheer success of science and the steady advance of rationalism, the individual scientist gained prestige. Along with the inventions of this period, especially the printing press by Johannes Gutenberg, allowed for the dissemination of the Bible in languages of the common people (languages other than Latin). This allowed more people to read and learn from the scripture, leading to the Evangelical movement. The people who spread this message, concentrated more on individual agency rather than the structures of the Church.

In the 17th century, founders of the Royal Society largely held conventional and orthodox religious views, and a number of them were prominent Churchmen. While theological issues that had the potential to be divisive were typically excluded from formal discussions of the early Society, many of its fellows nonetheless believed that their scientific activities provided support for traditional religious belief. Clerical involvement in the Royal Society remained high until the mid-nineteenth century, when science became more professionalised.

Albert Einstein supported the compatibility of some interpretations of religion with science. In "Science, Philosophy and Religion, A Symposium" published by the Conference on Science, Philosophy and Religion in Their Relation to the Democratic Way of Life, Inc., New York in 1941, Einstein stated:

Accordingly, a religious person is devout in the sense that he has no doubt of the significance and loftiness of those superpersonal objects and goals which neither require nor are capable of rational foundation. They exist with the same necessity and matter-of-factness as he himself. In this sense religion is the age-old endeavor of mankind to become clearly and completely conscious of these values and goals and constantly to strengthen and extend their effect. If one conceives of religion and science according

to these definitions then a conflict between them appears impossible. For science can only ascertain what is, but not what should be, and outside of its domain value judgments of all kinds remain necessary. Religion, on the other hand, deals only with evaluations of human thought and action: it cannot justifiably speak of facts and relationships between facts. According to this interpretation the well-known conflicts between religion and science in the past must all be ascribed to a misapprehension of the situation which has been described.

Prominent modern scientists who are atheists include evolutionary biologist Richard Dawkins and Nobel Prize-winning physicist Steven Weinberg. Prominent scientists advocating religious belief include Nobel Prize-winning physicist and United Church of Christ member Charles Townes, evangelical Christian and past head of the Human Genome Project Francis Collins, and climatologist John T. Houghton.

A modern view, described by Stephen Jay Gould as "non-overlapping magisteria" (NOMA), is that science and religion deal with fundamentally separate aspects of human experience and so, when each stays within its own domain, they co-exist peacefully. While Gould spoke of independence from the perspective of science, W. T. Stace viewed independence from the perspective of the philosophy of religion. Stace felt that science and religion, when each is viewed in its own domain, are both consistent and complete. They originate from different perceptions of reality, as Arnold O. Benz points out, but meet each other, for example, in the feeling of amazement and in ethics.

The USA's National Academy of Science supports the view that science and religion are independent.

Science and religion are based on different aspects of human experience. In science, explanations must be based on evidence drawn from examining the natural world. Scientifically based observations or experiments that conflict with an explanation eventually must lead to modification or even abandonment of that explanation. Religious faith, in contrast, does not depend on empirical evidence, is not necessarily modified in the face of conflicting evidence, and typically involves supernatural forces or entities. Because they are not a part of nature, supernatural entities cannot be investigated by science. In this sense, science and religion are separate and address aspects of human understanding in different

ways. Attempts to put science and religion against each other create controversy where none needs to exist.

Chapter 2

Basic General Chemistry For Biology

Chemistry is the scientific discipline involved with elements and compounds composed of atoms, molecules and ions: their composition, structure, properties, behavior and the changes they undergo during a reaction with other substances.

In the scope of its subject, chemistry occupies an intermediate position between physics and biology. For example, chemistry explains aspects of plant chemistry (botany), the formation of igneous rocks (geology), how atmospheric ozone is formed and how environmental pollutants are degraded (ecology), the properties of the soil on the moon (astrophysics), how medications work (pharmacology), and how to collect DNA evidence at a crime scene (forensics).

Chemistry addresses topics such as how atoms and molecules interact via chemical bonds to form new chemical compounds. There are four types of chemical bonds: covalent bonds, in which compounds share one or more electron(s); ionic bonds, in which a compound donates one or more electrons to another compound to produce ions (cations and anions); hydrogen bonds; and Van der Waals force bonds.

The current model of atomic structure is the quantum mechanical model. Traditional chemistry starts with the study of elementary particles, atoms, molecules, substances, metals, crystals and other aggregates of matter. Matter can be studied in solid, liquid, gas and plasma states, in isolation or in combination. The interactions, reactions and transformations that are studied in chemistry are usually the result of interactions between atoms, leading to rearrangements of the chemical bonds which hold atoms together.

A chemical reaction is a transformation of some substances into one or more different substances. The basis of such a chemical transformation is the rearrangement of electrons in the chemical bonds between atoms. It can be symbolically depicted through a chemical equation, which usually involves atoms as subjects. The number of atoms on the left and the right in the equation for a chemical transformation is equal. (When the number of atoms on

either side is unequal, the transformation is referred to as a nuclear reaction or radioactive decay.) The type of chemical reactions a substance may undergo and the energy changes that may accompany it are constrained by certain basic rules, known as chemical laws.

2.1 Matter

In classical physics and general chemistry, matter is any substance that has mass and takes up space by having volume. Matter should not be confused with mass, as the two are not the same in modern physics. Matter is a general term describing any ‘physical substance’. By contrast, mass is not a substance but rather a quantitative property of matter and other substances or systems. All everyday objects that can be touched are ultimately composed of atoms, which are made up of interacting subatomic particles, and in everyday as well as scientific usage, “matter” generally includes atoms and anything made up of them, and any particles (or combination of particles) that act as if they have both rest mass and volume. However it does not include massless particles such as photons, or other energy phenomena or waves such as light or sound. Matter exists in various states (also known as phases). These include classical everyday phases such as solid, liquid, and gas – for example water exists as ice, liquid water, and gaseous steam – but other states are possible, including plasma, Bose-Einstein condensates, fermionic condensates, and quark-gluon plasma.

Usually atoms can be imagined as a nucleus of protons and neutrons, and a surrounding “cloud” of orbiting electrons which “take up space”. However this is only somewhat correct, because subatomic particles and their properties are governed by their quantum nature, which means they do not act as everyday objects appear to act – they can act like waves as well as particles and they do not have well-defined sizes or positions. In the Standard Model of particle physics, matter is not a fundamental concept because the elementary constituents of atoms are quantum

entities which do not have an inherent “size” or “volume” in any everyday sense of the word. Due to the exclusion principle and other fundamental interactions, some “point particles” known as fermions (quarks, leptons), and many composites and atoms, are effectively forced to keep a distance from other particles under everyday conditions; this creates the property of matter which appears to us as matter taking up space.

2.1.1 The Atom

An atom is the smallest unit of ordinary matter that forms a chemical element. Every solid, liquid, gas, and plasma is composed of neutral or ionized atoms. Atoms are extremely small, typically around 100 picometers across. They are so small that accurately predicting their behavior using classical physics—as if they were tennis balls, for example—is not possible due to quantum effects. The word atom is derived from Ancient Greek ἄτομος (*átomos*, “indivisible”), from ἀ- (*a-*, “not”) + τέμνω (*témnō*, “I cut”), which means “uncuttable”.

Every atom is composed of a nucleus and one or more electrons surrounding the nucleus. The nucleus is made of one or more protons and a number of neutrons. Only the most common variety of hydrogen has no neutrons. More than 99.94% of an atom’s mass is in the nucleus. The protons have a positive electric charge, the electrons have a negative electric charge, and the neutrons have no electric charge. If the number of protons and electrons are equal, then the atom is electrically neutral. If an atom has more or fewer electrons than protons, then it has an overall negative or positive charge, respectively — such atoms are called ions.

The electrons of an atom are attracted to the protons in an atomic nucleus by the electromagnetic force. The protons and neutrons in the nucleus are attracted to each other by the nuclear force. This force is usually stronger than the electromagnetic force that repels the positively charged protons from one another. Under certain circumstances, the repelling electromagnetic force becomes stronger than the nuclear force. In this case, the nucleus splits and leaves behind different elements. This is a form of nuclear decay.

The number of protons in the nucleus is the atomic number and it defines to which chemical element the atom belongs. For example, any atom that contains 29 protons is copper. The number of neutrons defines the isotope of the element. Atoms can attach to one or more other atoms by chemical bonds to form chemical compounds such as molecules or crystals. The ability of atoms to associate and dissociate is responsible for most of the physical changes observed in nature. Chemistry is the discipline that studies these changes.

Atoms are extremely small, typically around 100 picometers (1 picometer = 10^{-12} m) across. They are so small that accurately predicting their behavior using classical physics—as if they were billiard balls, for example—is not possible due to quantum effects. Current atomic models use quantum mechanics¹ to better explain and predict this behavior. The word quantum derives from the Latin, meaning “how great” or “how much”. In quantum mechanics, it refers to a discrete unit assigned to certain physical quantities such as the energy of an atom at rest. The discovery that particles are discrete packets of energy with wave-like properties led to the branch of physics dealing with atomic and subatomic systems which is today called quantum mechanics. It underlies the mathematical framework of many fields of physics and chemistry.

Quantum mechanics is essential for understanding the behavior of systems at atomic length scales and smaller. If the physical nature of an atom were solely described by classical mechanics, electrons would not orbit the nucleus, since orbiting electrons emit radiation (due to circular motion) and so would quickly lose energy and collide with the nucleus. This framework was unable to explain the stability of atoms. Instead, electrons remain in an uncertain, non-deterministic, smeared, probabilistic wave-particle orbital about the nucleus, defying the traditional assumptions of classical mechanics and electromagnetism.

Quantum mechanics is also critically important for understanding how individual atoms are joined by covalent bonds to form molecules. The application of quantum mechanics to chemistry is known as quantum chemistry. Quantum mechanics can also provide quantitative insight into ionic and covalent bonding processes by explicitly showing which molecules are energetically favorable to which others and the magnitudes of the energies involved. Furthermore, most of the calculations performed in modern computational chemistry rely on quantum mechanics.

For much of the history of the natural sciences people have contemplated the exact nature of matter. The idea that matter was built of discrete building blocks, the so-called particulate theory of matter, independently appeared in ancient Greece and ancient India among Buddhists, Hindus and Jains in 1st-millennium BC. Ancient philosophers who proposed the particulate theory of matter include Kanada (c. 6th-century BC or after), Leucippus (~490 BC) and Democritus (~470–380 BC).

In the early 1800s, John Dalton³ compiled experimental data gathered by himself and other scientists and observed that chemical elements seemed to combine by weight in ratios of small whole numbers; he called this

¹https://en.wikipedia.org/wiki/Quantum_mechanics

³https://en.wikipedia.org/wiki/John_Dalton

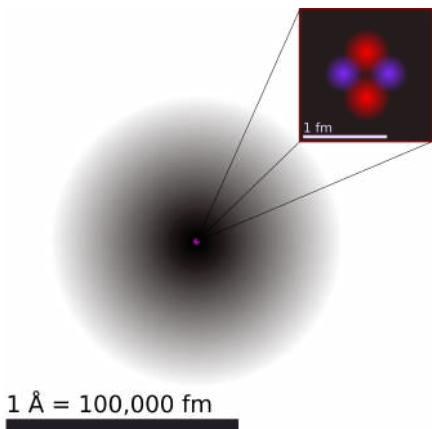


Figure 2.1: An illustration of the helium atom², depicting the nucleus (pink) and the electron cloud distribution (black). The nucleus (upper right) in helium-4 is in reality spherically symmetric and closely resembles the electron cloud, although for more complicated nuclei this is not always the case. The black bar is one angstrom (10^{-10} m or 100 pm).

pattern the “law of multiple proportions”. For instance, there are two types of tin oxide: one is 88.1% tin and 11.9% oxygen, and the other is 78.7% tin and 21.3% oxygen. Adjusting these figures, for every 100 g of tin there is either 13.5 g or 27 g of oxygen respectively. 13.5 and 27 form a ratio of 1:2, a ratio of small whole numbers. Similarly, there are two common types of iron oxide, in which for every 112 g of iron there is either 32 g or 48 g of oxygen respectively, which gives a ratio of 2:3. As a final example, there are three oxides of nitrogen in which for every 140 g of nitrogen, there is 80 g, 160 g, and 320 g of oxygen respectively, which gives a ratio of 1:2:4. This recurring pattern in the data suggested that elements always combine in multiples of discrete units, which Dalton concluded were atoms. In the case of the tin oxides, for every one tin atom, there are either one or two oxygen atoms (SnO and SnO_2). In the case of the iron oxides, for every two iron atoms, there are either two or three oxygen atoms (FeO and Fe_2O_3). In the case of the nitrogen oxides, their formulas are N_2O , NO , and NO_2 .

In the late 18th century, a number of scientists found that they could better explain the behavior of gases by describing them as collections of sub-microscopic particles and modelling their behavior using statistics and probability. Unlike Dalton’s atomic theory, the kinetic theory of gases describes not how gases react chemically with each other to form compounds, but how they behave physically: diffusion, viscosity, conductivity, pressure, etc.

In 1827, botanist Robert Brown⁴ used a microscope

⁴[https://en.wikipedia.org/wiki/Robert_Brown_\(botanist,_born_1773\)](https://en.wikipedia.org/wiki/Robert_Brown_(botanist,_born_1773))

to look at dust grains floating in water and discovered that they moved about erratically, a phenomenon that became known as “Brownian motion”. This was thought to be caused by water molecules knocking the grains about. In 1905, Albert Einstein⁵ proved the reality of these molecules and their motions by producing the first statistical physics analysis of Brownian motion. French physicist Jean Perrin⁶ used Einstein’s work to experimentally determine the mass and dimensions of atoms, thereby verifying Dalton’s atomic theory.

In 1897, J.J. Thomson⁷ discovered that cathode rays are not electromagnetic waves but made of particles that are 1,800 times lighter than hydrogen (the lightest atom). Therefore, they were not atoms, but a new particle, the first subatomic particle to be discovered. He called these new particles corpuscles but they were later renamed electrons. Thomson also showed that electrons were identical to particles given off by photoelectric and radioactive materials. It was quickly recognized that electrons are the particles that carry electric currents in metal wires, and carry the negative electric charge within atoms. Thus Thomson overturned the belief that atoms are the indivisible, fundamental particles of matter. The misnomer “atom” is still used, even though atoms are not literally “uncuttable”.

J. J. Thomson postulated that the negatively-charged electrons were distributed throughout the atom in a uniform sea of positive charge. This was known as the plum pudding model. In 1909, Hans Geiger⁸ and Ernest Marsden⁹, working under the direction of Ernest Rutherford¹⁰, bombarded metal foil with alpha particles to observe how they scattered. They expected all the charged particles to pass straight through with little deflection, because Thomson’s model said that the charges in the atom are so diffuse that their electric fields in the foil could not affect the alpha particles much. Yet Geiger and Marsden spotted alpha particles being deflected by angles greater than 90°, which was supposed to be impossible according to Thomson’s model. To explain this, Rutherford proposed that the positive charge of the atom is concentrated in a tiny nucleus at the center. Only such an intense concentration of charge could produce an electric field strong enough to deflect alpha particles that much.

While experimenting with the products of radioactive decay, in 1913 radiochemist Frederick Soddy discovered that there appeared to be more than one type of atom at each position on the periodic table. The term isotope was coined by Margaret Todd as a suitable name for different

⁵https://en.wikipedia.org/wiki/Albert_Einstein

⁶https://en.wikipedia.org/wiki/Jean_Baptiste_Perrin

⁷https://en.wikipedia.org/wiki/J._J._Thomson

⁸https://en.wikipedia.org/wiki/Hans_Geiger

⁹https://en.wikipedia.org/wiki/Ernest_Marsden

¹⁰https://en.wikipedia.org/wiki/Ernest_Rutherford

atoms that belong to the same element. J.J. Thomson created a technique for isotope separation through his work on ionized gases, which subsequently led to the discovery of stable isotopes.

The development of the mass spectrometer allowed the mass of atoms to be measured with increased accuracy. The device uses a magnet to bend the trajectory of a beam of ions, and the amount of deflection is determined by the ratio of an atom's mass to its charge. The chemist Francis William Aston¹¹ used this instrument to show that isotopes had different masses. The atomic mass of these isotopes varied by integer amounts, called the whole number rule. The explanation for these different isotopes awaited the discovery of the neutron, an uncharged particle with a mass similar to the proton, by the physicist James Chadwick¹² in 1932. Isotopes were then explained as elements with the same number of protons, but different numbers of neutrons within the nucleus.

In 1913 the physicist Niels Bohr¹³ proposed a model in which the electrons of an atom were assumed to orbit the nucleus but could only do so in a finite set of orbits, and could jump between these orbits only in discrete changes of energy corresponding to absorption or radiation of a photon. This quantization was used to explain why the electrons' orbits are stable (given that normally, charges in acceleration, including circular motion, lose kinetic energy which is emitted as electromagnetic radiation, see synchrotron radiation) and why elements absorb and emit electromagnetic radiation in discrete spectra.

Later in the same year Henry Moseley provided additional experimental evidence in favor of Niels Bohr's theory. These results refined Ernest Rutherford's and Antonius Van den Broek's model, which proposed that the atom contains in its nucleus a number of positive nuclear charges that is equal to its (atomic) number in the periodic table. Until these experiments, atomic number was not known to be a physical and experimental quantity. That it is equal to the atomic nuclear charge remains the accepted atomic model today.

Atomic dimensions are thousands of times smaller than the wavelengths of light (400–700 nm) so they cannot be viewed using an optical microscope, although individual atoms can be observed using a scanning tunneling microscope. To visualize the minuteness of the atom, consider that a typical human hair is about 1 million carbon atoms in width. A single drop of water contains about 2 sextillion (2×10^{21}) atoms of oxygen, and twice the number of hydrogen atoms. A single carat diamond with a mass of 2×10^{-4} kg contains about 10 sextillion

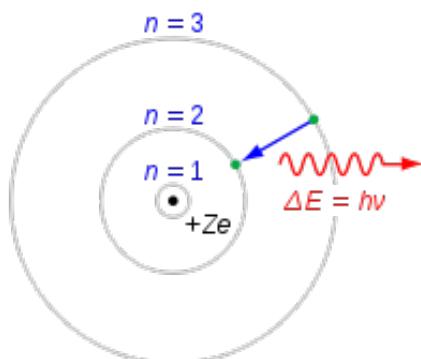


Figure 2.2: Niels Bohr's 1913 quantum model of the atom¹⁴, which incorporated an explanation of Johannes Rydberg's 1888 formula, Max Planck's 1900 quantum hypothesis, i.e. that atomic energy radiators have discrete energy values ($\epsilon = hv$), J. J. Thomson's 1904 plum pudding model, Albert Einstein's 1905 light quanta postulate, and Ernest Rutherford's 1907 discovery of the atomic nucleus. Note that the electron does not travel along the black line when emitting a photon. It "jumps", disappearing from the outer orbit and appearing in the inner one and cannot exist in the space between orbits 2 and 3. The Bohr model of the hydrogen atom ($Z = 1$) or a hydrogen-like ion ($Z > 1$), where the negatively charged electron confined to an atomic shell encircles a small, positively charged atomic nucleus and where an electron jumps between orbits, is accompanied by an emitted or absorbed amount of electromagnetic energy (hv). The orbits in which the electron may travel are shown as grey circles; their radius increases as n^2 , where n is the principal quantum number. The $3 \rightarrow 2$ transition depicted here produces the first line of the Balmer series, and for hydrogen ($Z = 1$) it results in a photon of wavelength 656 nm (red light).

¹¹https://en.wikipedia.org/wiki/Francis_William_Aston

¹²https://en.wikipedia.org/wiki/James_Chadwick

¹³https://en.wikipedia.org/wiki/Niels_Bohr

(10^{22}) atoms of carbon. If an apple were magnified to the size of the Earth, then an atom would be approximately the size of the original apple.

2.1.1.1 Structure of The Atom

Though the word atom originally denoted a particle that cannot be cut into smaller particles, in modern scientific usage the atom is composed of various subatomic particles. The constituent particles of an atom are the electron, the proton and the neutron.

The electron is by far the least massive of these particles at 9.11×10^{-31} kg, with a negative electrical charge and a size that is too small to be measured using available techniques. Under ordinary conditions, electrons are bound to the positively charged nucleus by the attraction created from opposite electric charges. If an atom has more or fewer electrons than its atomic number, then it becomes respectively negatively or positively charged as a whole; a charged atom is called an ion.

Protons have a positive charge and a mass 1,836 times that of the electron, at 1.6726×10^{-27} kg. The number of protons in an atom is called its atomic number.

Neutrons have no electrical charge and have a free mass of 1,839 times the mass of the electron, or 1.6749×10^{-27} kg. Neutrons are the heaviest of the three constituent particles, but their mass can be reduced by the nuclear binding energy. Neutrons and protons (collectively known as nucleons) have comparable dimensions—on the order of 2.5×10^{-15} m.

In the Standard Model of physics, electrons are truly elementary particles with no internal structure, whereas protons and neutrons are composite particles composed of elementary particles called quarks. There are two types of quarks in atoms, each having a fractional electric charge. Protons are composed of two up quarks (each with charge $+\frac{2}{3}$) and one down quark (with a charge of $\frac{1}{3}$). Neutrons consist of one up quark and two down quarks. This distinction accounts for the difference in mass and charge between the two particles.

The quarks are held together by the strong interaction (or strong force), which is mediated by gluons. The protons and neutrons, in turn, are held to each other in the nucleus by the nuclear force, which is a residuum of the strong force that has somewhat different range-properties (see the article on the nuclear force for more). The gluon is a member of the family of gauge bosons, which are elementary particles that mediate physical forces.

2.1.1.2 The Atomic Nucleus

The nucleus of an atom consists of neutrons and protons, which in turn are the manifestation of more elementary particles, called quarks, that are held together by the nuclear strong force. The nuclear strong force extends far enough so as to bind the neutrons and protons together against the repulsive electrical force between the positively charged protons. The nuclear strong force has a very short range, and essentially drops to zero just beyond the edge of the nucleus. The collective action of the positively charged nucleus is to hold the electrically negative charged electrons in their orbits about the nucleus. The collection of negatively charged electrons orbiting the nucleus display an affinity for certain configurations and numbers of electrons that make their orbits stable. Which chemical element an atom represents is determined by the number of protons in the nucleus; the neutral atom will have an equal number of electrons orbiting that nucleus. Individual chemical elements can create more stable electron configurations by combining to share their electrons. It is that sharing of electrons to create stable electronic orbits about the nucleus that appears to us as the chemistry of our world.

Protons define the entire charge of a nucleus, and hence its chemical identity. Neutrons are electrically neutral, but contribute to the mass of a nucleus to nearly the same extent as the protons. Neutrons can explain the phenomenon of isotopes (same atomic number with different atomic mass). The main role of neutrons is to reduce electrostatic repulsion inside the nucleus.

The electrons of an atom are attracted to the protons in an atomic nucleus by the electromagnetic force. The protons and neutrons in the nucleus are attracted to each other by the nuclear force. This force is usually stronger than the electromagnetic force that repels the positively charged protons from one another. Under certain circumstances, the repelling electromagnetic force becomes stronger than the nuclear force. In this case, the nucleus splits and leaves behind different elements. This is a form of nuclear decay.

The number of protons in the nucleus is the atomic number and it defines to which chemical element the atom belongs. For example, any atom that contains 6 protons is carbon. The number of neutrons defines the isotope of the element. For example, the most common isotope of carbon is ^{12}C with 6 neutrons. Another carbon isotope found in nature is ^{14}C , which has 8 neutrons. Atoms can attach to one or more other atoms by chemical bonds to form chemical compounds such as molecules or crystals. The ability of atoms to associate and dissociate is responsible for most of the physical changes observed in nature.

2.1.1.3 The Electron Cloud

The electron cloud is a region surrounding the nucleus. The electrons in an atom are attracted to the protons in the nucleus by the electromagnetic force. This force binds the electrons to the nucleus, which means that an external source of energy is needed for the electron to escape. The closer an electron is to the nucleus, the greater the attractive force. Hence the electrons that are closer to the nucleus require more energy to escape than those at greater separations.

Electrons, like other particles, have properties of both a particle and a wave. In atomic theory and quantum mechanics, an atomic orbital is a mathematical function describing the location and wave-like behavior of electrons surrounding an atom. The term "orbital" was coined by Robert Mulliken¹⁵ in 1932 as an abbreviation for one-electron orbital wave function. The term atomic orbital is also used to refer to the physical region or space where the electron can be calculated to be present using this function. Each atomic orbital corresponds to a particular energy level of the electron. The electron can change its state to a higher energy level by absorbing sufficient energy to boost it into the new quantum state. Likewise, through spontaneous emission, an electron in a higher energy state can drop to a lower energy state while radiating the excess energy.

The electrons in an atom are described by a unique set of values of four so-called quantum numbers n , ℓ , and m_ℓ , and s . As n increases, the electron is farther from the nucleus, on average, has a higher energy and is, therefore, less tightly bound to the nucleus. For each value of n , ℓ can take integer (whole number) values ranging from 0 to $n - 1$ and m_ℓ can take integer values ranging from $-\ell$ to $+\ell$, and two values of s ($+\frac{1}{2}$ or $-\frac{1}{2}$) hence higher- n electron states are more numerous. Thus, each n -shell can accommodate up to a maximum of $2n^2$ electrons. In chemistry, values $n = 1, 2, 3, 4, 5, 6, 7$ correspond to the so-called electron shells, which may be thought of as onion-like layers surrounding the nucleus that are filled with electrons. The shell terminology comes from Arnold Sommerfeld's¹⁶ modification of the Bohr model. Sommerfeld retained Bohr's planetary model (electrons orbiting the atomic nucleus like planets the sun), but added orbits that formed a thick "shell" instead of the infinitely thin circular orbit of Bohr's model. The closest shell to the nucleus is called the "1 shell" (also called the "K shell"), followed by the "2 shell" (or "L shell"), then the "3 shell" (or "M shell"), and so on farther and farther from the nucleus. The shells correspond to the principal quantum numbers ($n = 1, 2, 3, 4 \dots$) or are labeled alphabetically with the letters used

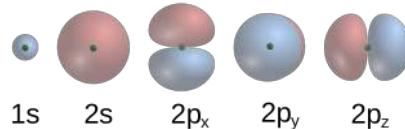


Figure 2.3: The five filled atomic orbitals of a neon atom¹⁸ separated and arranged in order of increasing energy from left to right, with the last three orbitals being equal in energy. Each orbital holds up to two electrons, which most probably exist in the zones represented by the colored bubbles. Each electron is equally present in both orbital zones, shown here by color only to highlight the different wave phase.

in X-ray notation¹⁷ (K, L, M, ...). Each shell is composed of one or more subshells (labeled with the letters s, p, d, and f, which are derived from the terms sharp, principal, diffuse, and fundamental historically used in the description of atomic spectral lines). For example, the first shell (with principal quantum number $n = 1$ also referred to as K shell) has one subshell, called 1s; the second (with principal quantum number $n = 2$ also referred to as L shell) shell has two subshells, called 2s ($\ell = 0$) and 2p ($\ell = 1$ and $m_\ell = -1, 0, 1$); the third shell (with principal quantum number $n = 3$ also referred to as M shell) has 3s, 3p, and 3d; the fourth shell (with principal quantum number $n = 4$ also referred to as N shell) has 4s, 4p, 4d and 4f; the fifth shell (with principal quantum number $n = 5$ also referred to as O shell) has 5s, 5p, 5d, and 5f and can theoretically hold more in the 5g subshell that is not occupied in the ground-state electron configuration of any known element.

Several rules govern the placement of electrons in orbitals in each atom referred to as the atom's electron configuration. The first dictates that no two electrons in an atom may have the same set of values of quantum numbers (this is the Pauli exclusion principle). These quantum numbers include the three that define orbitals, as well as s , or spin quantum number. Thus, two electrons may occupy a single orbital, so long as they have different values of s . However, only two electrons, because of their spin, can be associated with each orbital.

Additionally, an electron always tends to fall to the lowest possible energy state. It is possible for it to occupy any orbital so long as it does not violate the Pauli exclusion principle, but if lower-energy orbitals are available, this condition is unstable. The electron will eventually lose energy (by releasing a photon) and drop into the lower orbital. Thus, electrons fill orbitals in the order specified by the energy sequence given above. This behavior is responsible for the structure of the periodic table.

¹⁵ https://en.wikipedia.org/wiki/Robert_S._Mulliken

¹⁶ https://en.wikipedia.org/wiki/Arnold_Sommerfeld

¹⁷ https://en.wikipedia.org/wiki/x-ray_notation

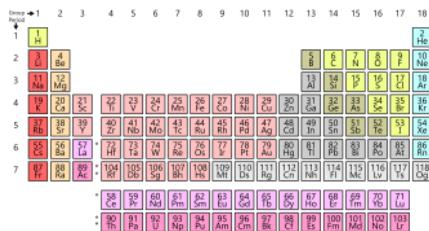


Figure 2.4: A simple depiction of the periodic table of the elements.²¹

2.1.2 The Periodic Table of The Elements

The periodic table¹⁹, also known as the periodic table of elements, is a tabular display of the chemical elements. In chemistry, an element is a pure substance which cannot be broken down by chemical means, consisting of atoms which have identical numbers of protons in their atomic nuclei. The number of protons in the nucleus is the defining property of an element, and is referred to as the atomic number (represented by the symbol Z). The mass number is the sum of the number of protons and neutrons in a nucleus. Although all the nuclei of all atoms belonging to one element will have the same atomic number, they may not necessarily have the same mass number; atoms of an element which have different mass numbers are known as isotopes. For example, all atoms with 6 protons in their nuclei are atoms of the chemical element carbon, but atoms of carbon may have mass numbers of 12 or 13.

In the periodic table, elements are arranged by atomic number, electron configuration, and recurring chemical properties. The organization of the periodic table can be used to derive relationships between the various element properties, and also to predict chemical properties and behaviours of undiscovered or newly synthesized elements. Russian chemist Dmitri Mendeleev²⁰ published the first recognizable periodic table in 1869, developed mainly to illustrate periodic trends of the then-known elements. He also predicted some properties of unidentified elements that were expected to fill gaps within the table. Most of his forecasts proved to be correct. Mendeleev's idea has been slowly expanded and refined with the discovery or synthesis of further new elements and the development of new theoretical models to explain chemical behaviour. In total, 118 elements have been identified. The first 94 occur naturally on Earth, and the remaining 24 are synthetic elements produced in nuclear reactions.

The structure of the table shows periodic trends. The seven rows of the table, called periods, generally have metals on the left and nonmetals on the right. The columns, called groups, contain elements with similar chemical be-

haviours. Six groups have accepted names as well as assigned numbers: for example, group 17 elements are the halogens; and group 18 are the noble gases. Also displayed are four simple rectangular areas or blocks associated with the filling of different atomic orbitals.

The periodic table may also be divided into several numbered rectangular 'blocks'. The elements belonging to a given block have this common feature: their highest-energy electrons all belong to the same ℓ -state (but the n associated with that ℓ -state depends upon the period). For instance, the leftmost two columns constitute the 's-block'. The outermost electrons of Li and Be respectively belong to the 2s subshell, and those of Na and Mg to the 3s subshell.

The following is the order for filling the "subshell" orbitals, which also gives the order of the "blocks" in the periodic table:

1s, 2s, 2p, 3s, 3p, 4s, 3d, 4p, 5s, 4d, 5p, 6s, 4f, 5d, 6p, 7s, 5f, 6d, 7p

The "periodic" nature of the filling of orbitals, as well as emergence of the s, p, d, and f "blocks", is more obvious if this order of filling is given in matrix form, with increasing principal quantum numbers starting the new rows ("periods") in the matrix. Then, each subshell (composed of the first two quantum numbers) is repeated as many times as required for each pair of electrons it may contain. The result is a compressed periodic table, with each entry representing two successive elements:

¹⁹ https://en.wikipedia.org/wiki/Periodic_table

²⁰ https://en.wikipedia.org/wiki/Dmitri_Mendeleev

Table 2.1

1s															2p	2p	2p
2s															2p	2p	2p
3s															3p	3p	3p
4s									3d	3d	3d	3d	3d	4p	4p	4p	4p
5s									4d	4d	4d	4d	4d	5p	5p	5p	5p
6s	4f	5d	5d	5d	5d	5d	6p	6p	6p	6p	6p						
7s	5f	6d	6d	6d	6d	6d	7p	7p	7p	7p	7p						

Although this is the general order of orbital filling, there are exceptions, and the actual electronic energies of each element are also dependent upon additional details of the atoms.

The number of electrons in an electrically neutral atom increases with the atomic number. The electrons in the outermost shell of a given atom, called valence electrons, tend to be responsible for an element's chemical behavior. Elements that contain the same number of valence electrons can be grouped together and display similar chemical properties. The elements at the far right of the table have their outer shell completely filled with electrons, which results in chemically inert elements known as the noble gases.

2.1.2.1 Chemical Bonds, Molecules, And Compounds

A chemical bond is an attraction between atoms. This attraction may be seen as the result of different behaviors of the outermost or valence electrons of atoms. Chemical bonds between atoms were explained by Gilbert Newton Lewis²² in 1916, as the interactions between their constituent electrons. As the chemical properties of the elements were known to largely repeat themselves according to the periodic law, in 1919 the American chemist Irving Langmuir²³ suggested that this could be explained if the electrons in an atom were connected or clustered in some manner. Groups of electrons were thought to occupy a set of electron shells about the nucleus.

A chemical bond can be a covalent bond²⁴, an ionic bond²⁵, a hydrogen bond²⁶ or just because of Van der Waals force. All bonds can be explained by quantum theory, but, in practice, simplification rules allow chemists to predict the strength, directionality, and polarity of bonds.

Covalent bonding is a common type of bonding in which two or more atoms share valence electrons more or less equally. The simplest and most common type is a single bond in which two atoms share two electrons. Other types include the double bond, the triple bond. In non-polar covalent bonds, the electrons are shared equally between the bonding atoms. Molecules that are formed primarily from non-polar covalent bonds are often immiscible in water or other polar solvents, but much more soluble in non-polar solvents such as hexane. A polar covalent bond is a covalent bond with a significant ionic character. This means that the two shared electrons are closer to one of the atoms than the other, creating an imbalance of charge.

Atoms will share valence electrons in such a way as to

create a noble gas electron configuration (eight electrons in their outermost shell) for each atom. Atoms that tend to combine in such a way that they each have eight electrons in their valence shell are said to follow the octet rule. However, some elements like hydrogen and lithium need only two electrons in their outermost shell to attain this stable configuration; these atoms are said to follow the duet rule, and in this way they are reaching the electron configuration of the noble gas helium, which has two electrons in its outer shell. In a polar covalent bond, one or more electrons are unequally shared between two nuclei.

An ionic bond is formed when a metal loses one or more of its electrons, becoming a positively charged cation, and the electrons are then gained by the non-metal atom, becoming a negatively charged anion. The two oppositely charged ions attract one another, and the ionic bond is the electrostatic force of attraction between them. For example, sodium (Na), a metal, loses one electron to become an Na^+ cation while chlorine (Cl), a non-metal, gains this electron to become Cl^- . The ions are held together due to electrostatic attraction, and that compound sodium chloride (NaCl), or common table salt, is formed.

A molecule (from French *molécule*, from New Latin *molecula* ("a molecule"), diminutive of Latin *moles* ("a mass"); see *mole* + *-cule*.) is a group of two or more atoms held together by chemical bonds. A molecule may be homonuclear, that is, it consists of atoms of one chemical element, as with two atoms in the oxygen molecule (O_2); or it may be heteronuclear, a chemical compound composed of more than one element, as with water (two hydrogen atoms and one oxygen atom; H_2O).

Molecules exist as electrically neutral units, unlike ions. When this rule is broken, giving the "molecule" a charge, the result is sometimes named a molecular ion or a polyatomic ion.

The "inert" or noble gas elements (helium, neon, argon, krypton, xenon and radon) are composed of lone atoms as their smallest discrete unit, but the other isolated chemical elements consist of either molecules or networks of atoms bonded to each other in some way. Identifiable molecules compose familiar substances such as water, air, and many organic compounds like alcohol, sugar, gasoline, and the various pharmaceuticals.

However, not all substances or chemical compounds consist of discrete molecules, and indeed most of the solid substances that make up the solid crust, mantle, and core of the Earth are chemical compounds without molecules. These other types of substances, such as ionic compounds and network solids, are organized in such a way as to lack the existence of identifiable molecules per se. Instead, these substances are discussed in terms of formula units

²²https://en.wikipedia.org/wiki/Gilbert_N._Lewis

²³https://en.wikipedia.org/wiki/Irving_Langmuir

²⁴https://en.wikipedia.org/wiki/Covalent_bond

²⁵https://en.wikipedia.org/wiki/Ionic_bonding

²⁶https://en.wikipedia.org/wiki/Hydrogen_bond

or unit cells as the smallest repeating structure within the substance. Examples of such substances are mineral salts (such as table salt), solids like carbon and diamond, metals, and familiar silica and silicate minerals such as quartz and granite.

One of the main characteristics of a molecule is its geometry often called its structure. While the structure of diatomic, triatomic or tetra-atomic molecules may be trivial, (linear, angular pyramidal etc.) the structure of polyatomic molecules, that are constituted of more than six atoms (of several elements) can be crucial for its chemical nature.

2.2 Chemical Bonds, Molecules And Compounds

A molecule is the smallest indivisible portion of a pure chemical substance that has its unique set of chemical properties, that is, its potential to undergo a certain set of chemical reactions with other substances. A compound is a pure chemical substance composed of more than one element. The properties of a compound bear little similarity to those of its elements. The standard nomenclature of compounds is set by the International Union of Pure and Applied Chemistry (IUPAC). Molecules are typically a set of atoms bound together by covalent bonds, such that the structure is electrically neutral and all valence electrons are paired with other electrons either in bonds or in lone pairs.

Thus, molecules exist as electrically neutral units, unlike ions. When this rule is broken, giving the “molecule” a charge, the result is sometimes named a molecular ion or a polyatomic ion. However, the discrete and separate nature of the molecular concept usually requires that molecular ions be present only in well-separated form, such as a directed beam in a vacuum in a mass spectrometer. Charged polyatomic collections residing in solids (for example, common sulfate or nitrate ions) are generally not considered “molecules” in chemistry. Some molecules contain one or more unpaired electrons, creating radicals. Most radicals are comparatively reactive, but some, such as nitric oxide (NO) can be stable.

The “inert” or noble gas elements (helium, neon, argon, krypton, xenon and radon) are composed of lone atoms as their smallest discrete unit, but the other isolated chemical elements consist of either molecules or networks of atoms bonded to each other in some way. Identifiable molecules compose familiar substances such as water, air, and many organic compounds like alcohol, sugar, gasoline, and the various pharmaceuticals.

However, not all substances or chemical compounds consist of discrete molecules, and indeed most of the solid

substances that make up the solid crust, mantle, and core of the Earth are chemical compounds without molecules. These other types of substances, such as ionic compounds and network solids, are organized in such a way as to lack the existence of identifiable molecules per se. Instead, these substances are discussed in terms of formula units or unit cells as the smallest repeating structure within the substance. Examples of such substances are mineral salts (such as table salt), solids like carbon and diamond, metals, and familiar silica and silicate minerals such as quartz and granite.

One of the main characteristics of a molecule is its geometry often called its structure. While the structure of diatomic, triatomic or tetra-atomic molecules may be trivial, (linear, angular pyramidal etc.) the structure of polyatomic molecules, that are constituted of more than six atoms (of several elements) can be crucial for its chemical nature.

Atoms sticking together in molecules or crystals are said to be bonded with one another. A chemical bond may be visualized as the multipole balance between the positive charges in the nuclei and the negative charges oscillating about them. More than simple attraction and repulsion, the energies and distributions characterize the availability of an electron to bond to another atom.

Chemical bonds between atoms were explained by Gilbert Newton Lewis²⁷ in 1916, as the interactions between their constituent electrons. As the chemical properties of the elements were known to largely repeat themselves according to the periodic law, in 1919 the American chemist Irving Langmuir²⁸ suggested that this could be explained if the electrons in an atom were connected or clustered in some manner. Groups of electrons were thought to occupy a set of electron shells about the nucleus.

A chemical bond can be a covalent bond³⁰, an ionic bond³¹, a hydrogen bond³² or just because of Van der Waals force. Each of these kinds of bonds is ascribed to some potential. These potentials create the interactions which hold atoms together in molecules or crystals. All bonds can be explained by quantum theory, but, in practice, simplification rules allow chemists to predict the strength, directionality, and polarity of bonds. In many simple compounds, valence bond theory, the Valence Shell Electron Pair Repulsion model (VSEPR), and the concept of oxidation number can be used to explain molecular structure and composition.

An ionic bond is formed when a metal loses one or

²⁷https://en.wikipedia.org/wiki/Gilbert_N._Lewis

²⁸https://en.wikipedia.org/wiki/Irving_Langmuir

³⁰https://en.wikipedia.org/wiki/Covalent_bond

³¹https://en.wikipedia.org/wiki/Ionic_bonding

³²https://en.wikipedia.org/wiki/Hydrogen_bond

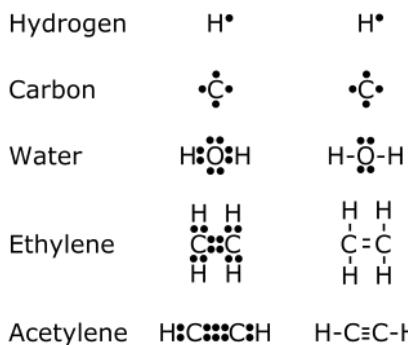


Figure 2.5: Examples of Lewis dot-style representations of chemical bonds²⁹ between carbon (C), hydrogen (H), and oxygen (O). Lewis dot diagrams were an early attempt to describe chemical bonding and are still widely used today.

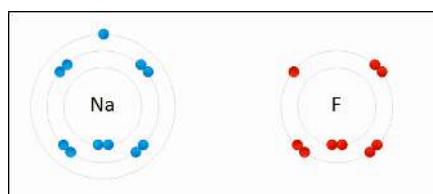


Figure 2.6: Formation of an ionic bond.³³ Sodium and fluorine atoms undergoing a redox reaction to form sodium fluoride. Sodium loses its outer electron to give it a stable electron configuration, and this electron enters the fluorine atom exothermically. The oppositely charged ions – typically a great many of them – are then attracted to each other to form a solid.

more of its electrons, becoming a positively charged cation, and the electrons are then gained by the non-metal atom, becoming a negatively charged anion. The two oppositely charged ions attract one another, and the ionic bond is the electrostatic force of attraction between them. For example, sodium (Na), a metal, loses one electron to become an Na^+ cation while chlorine (Cl), a non-metal, gains this electron to become Cl^- . The ions are held together due to electrostatic attraction, and that compound sodium chloride (NaCl), or common table salt, is formed.

In a covalent bond, one or more pairs of valence electrons are shared by two atoms: the resulting electrically neutral group of bonded atoms is termed a molecule. Atoms will share valence electrons in such a way as to create a noble gas electron configuration (eight electrons in their outermost shell) for each atom. Atoms that tend to combine in such a way that they each have eight electrons in their valence shell are said to follow the octet rule. However, some elements like hydrogen and lithium need only two electrons in their outermost shell to attain this stable configuration; these atoms are said to follow the

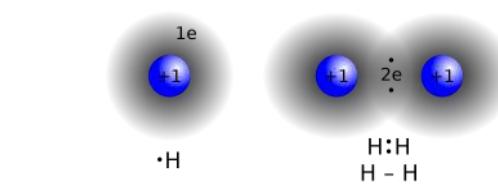


Figure 2.7: A covalent bond between two hydrogen atoms forming a hydrogen molecule.³⁴

duet rule, and in this way they are reaching the electron configuration of the noble gas helium, which has two electrons in its outer shell.

2.3 Energy

In physics, energy (from the Ancient Greek: ἐνέργεια, romanized: energeia, lit. ‘activity, operation’) is the quantitative property that must be transferred to an object in order to perform work on, or to heat, the object. Energy is a conserved quantity; the law of conservation of energy states that energy can be converted in form, but not created or destroyed. The SI unit of energy is the joule, which is the energy transferred to an object by the work of moving it a distance of 1 metre against a force of 1 newton.

The American physicist Richard Feynman³⁵ said during a 1961 lecture:

There is a fact, or if you wish, a law, governing all natural phenomena that are known to date. There is no known exception to this law – it is exact so far as we know. The law is called the conservation of energy. It states that there is a certain quantity, which we call energy, that does not change in manifold changes which nature undergoes. That is a most abstract idea, because it is a mathematical principle; it says that there is a numerical quantity which does not change when something happens. It is not a description of a mechanism, or anything concrete; it is just a strange fact that we can calculate some number and when we finish watching nature go through her tricks and calculate the number again, it is the same.

— The Feynman Lectures on Physics

Common forms of energy include the kinetic energy of a moving object, the potential energy stored by an object's position in a force field (gravitational, electric or magnetic), the elastic energy stored by stretching solid objects, the chemical energy released when a fuel burns, the radiant

³⁵ https://en.wikipedia.org/wiki/Richard_Feynman

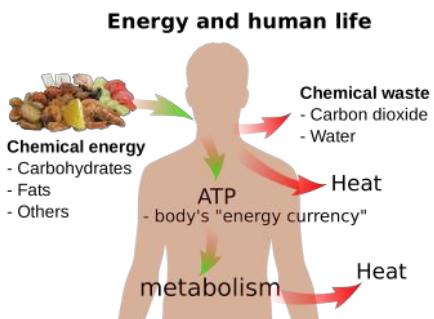


Figure 2.8: Basic overview of energy and human life.³⁶

energy carried by light, and the thermal energy due to an object's temperature.

Mass and energy are closely related. Due to mass-energy equivalence, any object that has mass when stationary (called rest mass) also has an equivalent amount of energy whose form is called rest energy, and any additional energy (of any form) acquired by the object above that rest energy will increase the object's total mass just as it increases its total energy. For example, after heating an object, its increase in energy could be measured as a small increase in mass, with a sensitive enough scale.

Living organisms require energy to stay alive, such as the energy humans get from food. Human civilization requires energy to function, which it gets from energy resources such as fossil fuels, nuclear fuel, or renewable energy. The processes of Earth's climate and ecosystem are driven by the radiant energy Earth receives from the sun and the geothermal energy contained within the earth.

In 1843, James Prescott Joule³⁷ discovered the link between mechanical work and the generation of heat in a series of experiments. The most famous of them used the "Joule apparatus": a descending weight, attached to a string, caused rotation of a paddle immersed in water, practically insulated from heat transfer. It showed that the gravitational potential energy lost by the weight in descending was equal to the internal energy gained by the water through friction with the paddle.

In the International System of Units (SI), the unit of energy is the joule, named after James Prescott Joule. It is a derived unit. It is equal to the energy expended (or work done) in applying a force of one newton through a distance of one metre. However energy is also expressed in many other units not part of the SI, such as ergs, calories, British Thermal Units, kilowatt-hours and kilocalories, which require a conversion factor when expressed in SI units.

The SI unit of energy rate (energy per unit time) is the

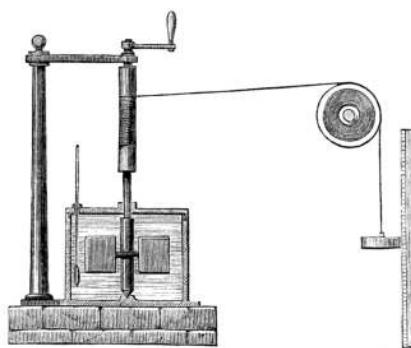


Figure 2.9: Joule's apparatus for measuring the mechanical equivalent of heat. A descending weight attached to a string causes a paddle immersed in water to rotate.³⁸

watt, which is a joule per second. Thus, one joule is one watt-second, and 3600 joules equal one watt-hour. The CGS energy unit is the erg and the imperial and US customary unit is the foot pound. Other energy units such as the electronvolt, food calorie or thermodynamic kcal (based on the temperature change of water in a heating process), and BTU are used in specific areas of science and commerce.

In the context of chemistry, energy is an attribute of a substance as a consequence of its atomic, molecular or aggregate structure. Since a chemical transformation is accompanied by a change in one or more of these kinds of structures, it is invariably accompanied by an increase or decrease of energy of the substances involved. Some energy is transferred between the surroundings and the reactants of the reaction in the form of heat or light; thus the products of a reaction may have more or less energy than the reactants.

A reaction is said to be exergonic if the final state is lower on the energy scale than the initial state; in the case of endergonic reactions the situation is the reverse. A reaction is said to be exothermic if the reaction releases heat to the surroundings; in the case of endothermic reactions, the reaction absorbs heat from the surroundings.

Chemical reactions are invariably not possible unless the reactants surmount an energy barrier known as the activation energy. The activation energy (E_a) of a reaction is measured in joules per mole (J/mol), kilojoules per mole (kJ/mol) or kilocalories per mole (kcal/mol). For a chemical reaction to proceed at a reasonable rate, the temperature of the system should be high enough such that there exists an appreciable number of molecules with translational energy equal to or greater than the activation energy. The term Activation Energy was introduced in 1889 by the Swedish scientist Svante Arrhenius³⁹. The Arrhenius

³⁷ https://en.wikipedia.org/wiki/James_Prescott_Joule

³⁹ https://en.wikipedia.org/wiki/Svante_Arrhenius

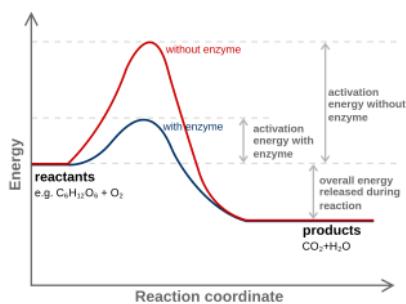


Figure 2.10: Example of an enzyme-catalysed exothermic reaction.⁴⁰

equation gives the quantitative basis of the relationship between the activation energy and the rate at which a reaction proceeds. From the equation, the activation energy can be found through the relation

$$k = Ae^{-E_a/(RT)}$$

where A is the pre-exponential factor for the reaction, R is the universal gas constant, T is the absolute temperature (usually in kelvins), and k is the reaction rate coefficient. Even without knowing A, E_a can be evaluated from the variation in reaction rate coefficients as a function of temperature (within the validity of the Arrhenius equation). The activation energy necessary for a chemical reaction to occur can be in the form of heat, light, electricity or mechanical force.

A substance that modifies the transition state to lower the activation energy is termed a catalyst; a catalyst composed only of protein and (if applicable) small molecule cofactors is termed an enzyme. A catalyst increases the rate of reaction without being consumed in the reaction. In addition, the catalyst lowers the activation energy, but it does not change the energies of the original reactants or products, and so does not change equilibrium. Rather, the reactant energy and the product energy remain the same and only the activation energy is altered (lowered).

A catalyst is able to reduce the activation energy by forming a transition state in a more favorable manner. Catalysts, by nature, create a more “comfortable” fit for the substrate of a reaction to progress to a transition state. This is possible due to a release of energy that occurs when the substrate binds to the active site of a catalyst. This energy is known as Binding Energy. Upon binding to a catalyst, substrates partake in numerous stabilizing forces while within the active site (i.e. Hydrogen bonding, van der Waals forces). Specific and favorable bonding occurs within the active site until the substrate forms to become the high-energy transition state. Forming the transition state is more favorable with the catalyst because the favor-

able stabilizing interactions within the active site release energy. A chemical reaction is able to manufacture a high-energy transition state molecule more readily when there is a stabilizing fit within the active site of a catalyst. The binding energy of a reaction is this energy released when favorable interactions between substrate and catalyst occur. The binding energy released assists in achieving the unstable transition state. Reactions otherwise without catalysts need a higher input of energy to achieve the transition state. Non-catalyzed reactions do not have free energy available from active site stabilizing interactions, such as catalytic enzyme reactions.

A related concept free energy, which also incorporates entropy considerations, is a very useful means for predicting the feasibility of a reaction and determining the state of equilibrium of a chemical reaction, in chemical thermodynamics. A reaction is feasible only if the total change in the Gibbs free energy is negative,

$$\Delta G \leq 0$$

if it is equal to zero the chemical reaction is said to be at equilibrium.

There exist only limited possible states of energy for electrons, atoms and molecules. These are determined by the rules of quantum mechanics, which require quantization of energy of a bound system. The atoms/molecules in a higher energy state are said to be excited. The molecules/atoms of substance in an excited energy state are often much more reactive; that is, more amenable to chemical reactions.

2.3.1 Chemical Reactions

When a chemical substance is transformed as a result of its interaction with another substance or with energy, a chemical reaction is said to have occurred. A chemical reaction is therefore a concept related to the “reaction” of a substance when it comes in close contact with another, whether as a mixture or a solution; exposure to some form of energy, or both. It results in some energy exchange between the constituents of the reaction as well as with the system environment, which may be designed vessels—often laboratory glassware.

Chemical reactions can result in the formation or dissociation of molecules, that is, molecules breaking apart to form two or more molecules or rearrangement of atoms within or across molecules. Chemical reactions usually involve the making or breaking of chemical bonds. Oxidation, reduction, dissociation, acid-base neutralization and molecular rearrangement are some of the commonly used kinds of chemical reactions.

A chemical reaction can be symbolically depicted through a chemical equation. In a chemical reaction the number and kind of atoms on both sides of the equation are always equal. This fact is referred to as the law of conservation of mass. The law implies that mass can neither be created nor destroyed, although it may be rearranged in space, or the entities associated with it may be changed in form. For example, in chemical reactions, the mass of the chemical components before the reaction is equal to the mass of the components after the reaction. Thus, during any chemical reaction and low-energy thermodynamic processes in an isolated system, the total mass of the reactants, or starting materials, must be equal to the mass of the products.

2.3.2 Radioactive Decay of Atoms

Radioactive decay (also known as nuclear decay, radioactivity, radioactive disintegration or nuclear disintegration) is the process by which an unstable atomic nucleus loses energy by radiation. Every element has one or more isotopes that have unstable nuclei that are subject to radioactive decay, causing the nucleus to emit particles or electromagnetic radiation. Radioactivity can occur when the radius of a nucleus is large compared with the radius of the strong force, which only acts over distances on the order of 1 fm.

The most common forms of radioactive decay are:

- Alpha decay: this process is caused when the nucleus emits an alpha particle, which is a helium nucleus consisting of two protons and two neutrons. The result of the emission is a new element with a lower atomic number.
- Beta decay (and electron capture): these processes are regulated by the weak force, and result from a transformation of a neutron into a proton, or a proton into a neutron. The neutron to proton transition is accompanied by the emission of an electron and an antineutrino, while proton to neutron transition (except in electron capture) causes the emission of a positron and a neutrino. The electron or positron emissions are called beta particles. Beta decay either increases or decreases the atomic number of the nucleus by one. Electron capture is more common than positron emission, because it requires less energy. In this type of decay, an electron is absorbed by the nucleus, rather than a positron emitted from the nucleus. A neutrino is still emitted in this process, and a proton changes to a neutron.
- Gamma decay: this process results from a change in the energy level of the nucleus to a lower state, resulting in the emission of electromagnetic radiation. The excited state of a nucleus which results in gamma emission usually occurs following the emis-

sion of an alpha or a beta particle. Thus, gamma decay usually follows alpha or beta decay.

Each radioactive isotope has a characteristic decay time period—the half-life—that is determined by the amount of time needed for half of a sample to decay. This is an exponential decay process that steadily decreases the proportion of the remaining isotope by 50% every half-life. Hence after two half-lives have passed only 25% of the isotope is present, and so forth.

Within living things, isotopic labels (both radioactive and nonradioactive) can be used to probe how the complex web of reactions which makes up the metabolism of an organism converts one substance to another. For instance a green plant uses light energy to convert water and carbon dioxide into glucose by photosynthesis. When plants performed photosynthesis using water in which the oxygen was the heavy oxygen (^{18}O) isotope, this oxygen isotope appeared in the oxygen gas formed by the plant and not in the glucose formed in the chloroplasts within the plant cells.

Radioactivity was discovered in 1896 by the French scientist Henri Becquerel⁴¹, while working with phosphorescent materials. These materials glow in the dark after exposure to light, and he suspected that the glow produced in cathode ray tubes by X-rays might be associated with phosphorescence. He wrapped a photographic plate in black paper and placed various phosphorescent salts on it. All results were negative until he used uranium salts. The uranium salts caused a blackening of the plate in spite of the plate being wrapped in black paper. These radiations were given the name “Becquerel Rays”.

It soon became clear that the blackening of the plate had nothing to do with phosphorescence, as the blackening was also produced by non-phosphorescent salts of uranium and by metallic uranium. It became clear from these experiments that there was a form of invisible radiation that could pass through paper and was causing the plate to react as if exposed to light.

At first, it seemed as though the new radiation was similar to the then recently discovered X-rays. Further research by Becquerel, Ernest Rutherford⁴², Paul Villard⁴³, Pierre Curie⁴⁴, Marie Curie⁴⁵, and others showed that this form of radioactivity was significantly more complicated. Rutherford was the first to realize that all such elements decay in accordance with the same mathematical exponential formula. Rutherford and his student Frederick Soddy were the first to realize that many decay processes resulted

⁴¹https://en.wikipedia.org/wiki/Henri_Becquerel

⁴²https://en.wikipedia.org/wiki/Ernest_Rutherford

⁴³https://en.wikipedia.org/wiki/Paul_Ulrich_Villard

⁴⁴https://en.wikipedia.org/wiki/Pierre_Curie

⁴⁵https://en.wikipedia.org/wiki/Marie_Curie

in the transmutation of one element to another. Subsequently, the radioactive displacement law of Fajans and Soddy was formulated to describe the products of alpha and beta decay.

The early researchers also discovered that many other chemical elements, besides uranium, have radioactive isotopes. A systematic search for the total radioactivity in uranium ores also guided Pierre and Marie Curie to isolate two new elements: polonium and radium. Except for the radioactivity of radium, the chemical similarity of radium to barium made these two elements difficult to distinguish.

Marie and Pierre Curie's study of radioactivity is an important factor in science and medicine. After their research on Becquerel's rays led them to the discovery of both radium and polonium, they coined the term "radioactivity". Their research on the penetrating rays in uranium and the discovery of radium launched an era of using radium for the treatment of cancer. Their exploration of radium could be seen as the first peaceful use of nuclear energy and the start of modern nuclear medicine.

2.4 Ions And Salts

An ion is a charged species, an atom or a molecule, that has lost or gained one or more electrons. When an atom loses an electron and thus has more protons than electrons, the atom is a positively charged ion or cation. When an atom gains an electron and thus has more electrons than protons, the atom is a negatively charged ion or anion. Cations and anions can form a crystalline lattice of neutral salts, such as the Na^+ and Cl^- ions forming sodium chloride, or NaCl . Examples of polyatomic ions that do not split up during acid-base reactions are hydroxide (OH^-) and phosphate (PO_4^{3-}).

Plasma is composed of gaseous matter that has been completely ionized, usually through high temperature.

2.5 Acids And Bases

A substance can often be classified as an acid or a base. There are several different theories which explain acid-base behavior. The simplest is Arrhenius theory, which states that acid is a substance that produces hydronium ions (H_3O^+) when it is dissolved in water, and a base is one that produces hydroxide ions (OH^-) when dissolved in water. According to Brønsted-Lowry acid-base theory, acids are substances that donate a positive hydrogen ion (H^+) to another substance in a chemical reaction; by extension, a base is the substance which receives that hydrogen ion.

A third common theory is Lewis acid-base theory,

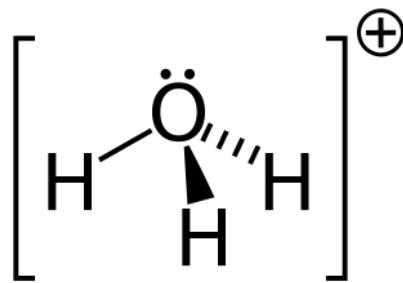


Figure 2.11: 3D diagram showing the pyramidal structure of the hydroxonium ion.⁴⁶

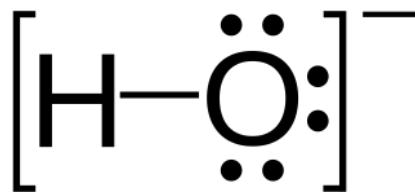


Figure 2.12: Lewis structure of the hydroxide ion showing three lone pairs on the oxygen atom⁴⁷

which is based on the formation of new chemical bonds. Lewis theory explains that an acid is a substance which is capable of accepting a pair of electrons from another substance during the process of bond formation, while a base is a substance which can provide a pair of electrons to form a new bond. According to this theory, the crucial things being exchanged are charges.

Acid strength is commonly measured by two methods. One measurement, based on the Arrhenius definition of acidity, is pH. The other measurement, based on the Brønsted-Lowry definition, is the acid dissociation constant (K_a), which measures the relative ability of a substance to act as an acid under the Brønsted-Lowry definition of an acid. That is, substances with a higher K_a are more likely to donate hydrogen ions in chemical reactions than those with lower K_a values.

In chemistry, pH ("potential of hydrogen" or "power of hydrogen) is a scale used to specify the acidity or basicity of an aqueous solution. Lower pH values correspond to solutions which are more acidic in nature, while higher values correspond to solutions which are more basic or alkaline. At room temperature (25 °C or 77 °F), pure water is neutral (neither acidic nor basic) and has a pH of 7.

The pH scale is logarithmic and inversely indicates the concentration of hydrogen ions in the solution (a lower pH indicates a higher concentration of hydrogen ions). This is because the formula used to calculate pH approximates the negative of the base 10 logarithm of the molar concen-

tration of hydrogen ions in the solution. More precisely, pH is the negative of the base 10 logarithm of the activity of the hydrogen ion:

$$pH = -\log_{10}[H_3O^+]$$

At 25 °C, solutions with a pH less than 7 are acidic, and solutions with a pH greater than 7 are basic. The neutral value of the pH depends on the temperature, being lower than 7 if the temperature increases. The pH value can be less than 0 for very strong acids, or greater than 14 for very strong bases.

Table 2.2: Average values of pH in common solutions

Substance	pH Range	Type
Battery acid	<1	Acid
Gastric acid	1.0 – 1.5	Acid
Vinegar	2.5	Acid
Orange juice	3.3 – 4.2	Acid
Black coffee	5 – 5.03	Acid
Milk	6.5 – 6.8	Acid
Pure water	7	Neutral
Sea water	7.5 – 8.4	Base
Ammonia	11.0 – 11.5	Base
Bleach	12.5	Base
Lye	13.0 – 13.	Base

The pH of different cellular compartments, body fluids, and organs is usually tightly regulated in a process called acid–base homeostasis. The most common disorder in acid–base homeostasis is acidosis, which means an acid overload in the body, generally defined by pH falling below 7.35. Alkalosis is the opposite condition, with blood pH being excessively high.

The pH of blood is usually slightly basic with a value of pH 7.365. This value is often referred to as physiological pH in biology and medicine. Plaque can create a local acidic environment that can result in tooth decay by demineralization. Enzymes and other proteins have an optimum pH range and can become inactivated or denatured outside this range.

Many biologically important molecules are acids. Nucleic acids, which contain acidic phosphate groups, include DNA and RNA. Nucleic acids contain the genetic code that determines many of an organism's characteristics, and is passed from parents to offspring. DNA contains the chemical blueprint for the synthesis of proteins which are made up of amino acid subunits. Cell membranes contain fatty acid esters such as phospholipids.

Table 2.3: Values of pH in living systems

Compartment	pH
Gastric acid	1.5–3.5
Lysosomes	4.5
Human skin	4.7
Granules of chromaffin cells	5.5
Urine	6
Cytosol	7.2
Blood (natural pH)	7.34–7.45
Cerebrospinal fluid (CSF)	7.5
Mitochondrial matrix	7.5
Pancreas secretions	8.1

An α -amino acid has a central carbon (the α or alpha carbon) which is covalently bonded to a carboxyl group (thus they are carboxylic acids), an amino group, a hydrogen atom and a variable group. The variable group, also called the R group or side chain, determines the identity and many of the properties of a specific amino acid. In glycine, the simplest amino acid, the R group is a hydrogen atom, but in all other amino acids it contains one or more carbon atoms bonded to hydrogens, and may contain other elements such as sulfur, oxygen or nitrogen. With the exception of glycine, naturally occurring amino acids are chiral and almost invariably occur in the L-configuration. Peptidoglycan, found in some bacterial cell walls contains some D-amino acids. At physiological pH, typically around 7, free amino acids exist in a charged form, where the acidic carboxyl group ($-COOH$) loses a proton ($-COO^-$) and the basic amine group ($-NH_2$) gains a proton ($-NH_3^+$). The entire molecule has a net neutral charge and is a zwitterion, with the exception of amino acids with basic or acidic side chains. Aspartic acid, for example, possesses one protonated amine and two deprotonated carboxyl groups, for a net charge of -1 at physiological pH.

Fatty acids and fatty acid derivatives are another group of carboxylic acids that play a significant role in biology. These contain long hydrocarbon chains and a carboxylic acid group on one end. The cell membrane of nearly all organisms is primarily made up of a phospholipid bilayer, a micelle of hydrophobic fatty acid esters with polar, hydrophilic phosphate "head" groups. Membranes contain additional components, some of which can participate in acid–base reactions.

In humans and many other animals, hydrochloric acid is a part of the gastric acid secreted within the stomach to help hydrolyze proteins and polysaccharides, as well as converting the inactive pro-enzyme, pepsinogen into the enzyme, pepsin. Some organisms produce acids for defense; for example, ants produce formic acid.

Acid–base equilibrium plays a critical role in regulating mammalian breathing. Oxygen gas (O_2) drives cellular respiration, the process by which animals release the chemical potential energy stored in food, producing carbon dioxide (CO_2) as a byproduct. Oxygen and carbon dioxide are exchanged in the lungs, and the body responds to changing energy demands by adjusting the rate of ventilation. For example, during periods of exertion the body rapidly breaks down stored carbohydrates and fat, releasing CO_2 into the blood stream. In aqueous solutions such as blood CO_2 exists in equilibrium with carbonic acid and bicarbonate ion.



It is the decrease in pH that signals the brain to breathe faster and deeper, expelling the excess CO_2 and resupplying the cells with O_2 .

Cell membranes are generally impermeable to charged or large, polar molecules because of the lipophilic fatty acyl chains comprising their interior. Many biologically important molecules, including a number of pharmaceutical agents, are organic weak acids which can cross the membrane in their protonated, uncharged form but not in their charged form (i.e. as the conjugate base). For this reason the activity of many drugs can be enhanced or inhibited by the use of antacids or acidic foods. The charged form, however, is often more soluble in blood and cytosol, both aqueous environments. When the extracellular environment is more acidic than the neutral pH within the cell, certain acids will exist in their neutral form and will be membrane soluble, allowing them to cross the phospholipid bilayer. Acids that lose a proton at the intracellular pH will exist in their soluble, charged form and are thus able to diffuse through the cytosol to their target. Ibuprofen, aspirin and penicillin are examples of drugs that are weak acids.

2.6 Redox

Redox (reduction–oxidation) reactions include all chemical reactions in which atoms have their oxidation state changed by either gaining electrons (reduction) or losing electrons (oxidation). Substances that have the ability to oxidize other substances are said to be oxidative and are known as oxidizing agents, oxidants or oxidizers. An oxidant removes electrons from another substance. Similarly, substances that have the ability to reduce other substances are said to be reductive and are known as reducing agents, reductants, or reducers. The chemical species from which the electron is removed is said to have been oxidized, while the chemical species to which the electron is added is said to have been reduced. In other words:

- Oxidation is the loss of electrons or an increase in the oxidation state of an atom, an ion, or of certain atoms in a molecule.
- Reduction is the gain of electrons or a decrease in the oxidation state of an atom, an ion, or of certain atoms in a molecule.

“Redox” is a portmanteau of the words “reduction” and “oxidation”. The word oxidation originally implied reaction with oxygen to form an oxide, since dioxygen ($O_2(g)$) was historically the first recognized oxidizing agent. Later, the term was expanded to encompass oxygen-like substances that accomplished parallel chemical reactions. Ultimately, the meaning was generalized to include all processes involving loss of electrons.

The word reduction originally referred to the loss in weight upon heating a metallic ore such as a metal oxide to extract the metal. In other words, ore was “reduced” to metal. Antoine Lavoisier showed that this loss of weight was due to the loss of oxygen as a gas. Later, scientists realized that the metal atom gains electrons in this process. The meaning of reduction then became generalized to include all processes involving a gain of electrons.

The term “hydrogenation” could often be used instead of reduction, since hydrogen is the reducing agent in a large number of reactions, especially in organic chemistry and biochemistry. But, unlike oxidation, which has been generalized beyond its root element, hydrogenation has maintained its specific connection to reactions that add hydrogen to another substance (e.g., the hydrogenation of unsaturated fats into saturated fats, $R-CH=CH-R + H_2 \rightarrow R-CH_2-CH_2-R$). The word “redox” was first used in 1928.

Many reactions in organic chemistry are redox reactions due to changes in oxidation states but without distinct electron transfer. For example, during the combustion of wood with molecular oxygen, the oxidation state of carbon atoms in the wood increases and that of oxygen atoms decreases as carbon dioxide and water are formed. The oxygen atoms undergo reduction, formally gaining electrons, while the carbon atoms undergo oxidation, losing electrons. Thus oxygen is the oxidizing agent and carbon is the reducing agent in this reaction.

Although oxidation reactions are commonly associated with the formation of oxides from oxygen molecules, oxygen is not necessarily included in such reactions, as other chemical species can serve the same function.

Redox reactions can occur relatively slowly, as in the formation of rust, or much more rapidly, as in the case of burning fuel. There are simple redox processes, such as the oxidation of carbon to yield carbon dioxide (CO_2) or the reduction of carbon by hydrogen to yield methane (CH_4), and more complex processes such as the oxidation

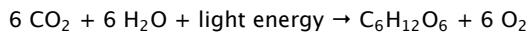
of glucose ($C_6H_{12}O_6$) in the human body. Analysis of bond energies and ionization energies in water allow calculation of the redox potentials.

A reductant transfers electrons to another substance and is thus oxidized itself. And because it "donates" electrons it is also called an electron donor. Oxidation and reduction properly refer to a change in oxidation number—the actual transfer of electrons may never occur. Thus, oxidation is better defined as an increase in oxidation number, and reduction as a decrease in oxidation number.

Many important biological processes involve redox reactions.

Cellular respiration, for instance, is the oxidation of glucose ($C_6H_{12}O_6$) to CO_2 and the reduction of oxygen to water. The summary equation for cell respiration is:

$C_6H_{12}O_6 + 6 O_2 \rightarrow 6 CO_2 + 6 H_2O$ The process of cell respiration also depends heavily on the reduction of NAD^+ to $NADH$ and the reverse reaction (the oxidation of $NADH$ to NAD^+). Photosynthesis and cellular respiration are complementary, but photosynthesis is not the reverse of the redox reaction in cell respiration:



Biological energy is frequently stored and released by means of redox reactions. Photosynthesis involves the reduction of carbon dioxide into sugars and the oxidation of water into molecular oxygen. The reverse reaction, respiration, oxidizes sugars to produce carbon dioxide and water. As intermediate steps, the reduced carbon compounds are used to reduce nicotinamide adenine dinucleotide (NAD^+) to $NADH$, which then contributes to the creation of a proton gradient, which drives the synthesis of adenosine triphosphate (ATP) and is maintained by the reduction of oxygen. In animal cells, mitochondria perform similar functions. See the Membrane potential article.

Free radical reactions are redox reactions that occur as a part of homeostasis and killing microorganisms, where an electron detaches from a molecule and then reattaches almost instantaneously. Free radicals are a part of redox molecules and can become harmful to the human body if they do not reattach to the redox molecule or an antioxidant. Unsatisfied free radicals can spur the mutation of cells they encounter and are, thus, causes of cancer.

The term redox state is often used to describe the balance of $GSH/GSSG$, $NAD^+/NADH$ and $NADP^+/NADPH$ in a biological system such as a cell or organ. The redox state is reflected in the balance of several sets of metabolites (e.g., lactate and pyruvate, beta-hydroxybutyrate, and acetoacetate), whose interconversion is dependent on these ratios. An abnormal redox state can develop in a variety of deleterious situations, such as hypoxia, shock, and sepsis. Re-

dox mechanism also control some cellular processes. Redox proteins and their genes must be co-located for redox regulation according to the CoRR hypothesis for the function of DNA in mitochondria and chloroplasts.

2.7 Water

Water is an inorganic, transparent, tasteless, odorless, and nearly colorless chemical substance, which is the main constituent of Earth's hydrosphere and the fluids of all known living organisms. It is vital for all known forms of life, even though it provides no calories or organic nutrients. Its chemical formula is H_2O , meaning that each of its molecules contains one oxygen and two hydrogen atoms, connected by covalent bonds.

From a biological standpoint, water has many distinct properties that are critical for the proliferation of life. It carries out this role by allowing organic compounds to react in ways that ultimately allow replication. All known forms of life depend on water. Water is vital both as a solvent in which many of the body's solutes dissolve and as an essential part of many metabolic processes within the body. Metabolism is the sum total of anabolism and catabolism. In anabolism, water is removed from molecules (through energy requiring enzymatic chemical reactions) in order to grow larger molecules (e.g., starches, triglycerides and proteins for storage of fuels and information). In catabolism, water is used to break bonds in order to generate smaller molecules (e.g., glucose, fatty acids and amino acids to be used for fuels for energy use or other purposes). Without water, these particular metabolic processes could not exist.

Water is fundamental to photosynthesis and respiration. Photosynthetic cells use the sun's energy to split off water's hydrogen from oxygen. Hydrogen is combined with CO_2 (absorbed from air or water) to form glucose and release oxygen. All living cells use such fuels and oxidize the hydrogen and carbon to capture the sun's energy and reform water and CO_2 in the process (cellular respiration).

Water is also central to acid-base neutrality and enzyme function. An acid, a hydrogen ion (H^+ , that is, a proton) donor, can be neutralized by a base, a proton acceptor such as a hydroxide ion (OH^-) to form water. Water is considered to be neutral, with a pH of 7. Acids have pH values less than 7 while bases have values greater than 7.

Earth surface waters are filled with life. The earliest life forms appeared in water; nearly all fish live exclusively in water, and there are many types of marine mammals, such as dolphins and whales. Some kinds of animals, such as amphibians, spend portions of their lives in water and portions on land. Plants such as kelp and algae grow in

the water and are the basis for some underwater ecosystems. Plankton is generally the foundation of the ocean food chain.

Aquatic vertebrates must obtain oxygen to survive, and they do so in various ways. Fish have gills instead of lungs, although some species of fish, such as the lungfish, have both. Marine mammals, such as dolphins, whales, and seals need to surface periodically to breathe air. Some amphibians are able to absorb oxygen through their skin. Invertebrates exhibit a wide range of modifications to survive in poorly oxygenated waters including breathing tubes (see insect and mollusc siphons) and gills (*Carcinus*). However as invertebrate life evolved in an aquatic habitat most have little or no specialization for respiration in water.

"Water" is the name of the liquid state of H₂O at standard ambient temperature and pressure. It forms precipitation in the form of rain and aerosols in the form of fog. Clouds are formed from suspended droplets of water and ice, its solid state. When finely divided, crystalline ice may precipitate in the form of snow. The gaseous state of water is steam or water vapor. Water moves continually through the water cycle of evaporation, transpiration (evapotranspiration), condensation, precipitation, and runoff, usually reaching the sea.

Water covers 71% of the Earth's surface, mostly in seas and oceans. Small portions of water occur as groundwater (1.7%), in the glaciers and the ice caps of Antarctica and Greenland (1.7%), and in the air as vapor, clouds (formed of ice and liquid water suspended in air), and precipitation (0.001%).

Water plays an important role in the world economy. Approximately 70% of the freshwater used by humans goes to agriculture. Fishing in salt and fresh water bodies is a major source of food for many parts of the world. Much of the long-distance trade of commodities (such as oil, natural gas, and manufactured products) is transported by boats through seas, rivers, lakes, and canals. Large quantities of water, ice, and steam are used for cooling and heating, in industry and homes. Water is an excellent solvent for a wide variety of substances both mineral and organic; as such it is widely used in industrial processes, and in cooking and washing.

Water (H₂O) is a polar inorganic compound that is at room temperature a tasteless and odorless liquid, nearly colorless with a hint of blue. This simplest hydrogen chalcogenide is by far the most studied chemical compound and is described as the "universal solvent" for its ability to dissolve many substances. This allows it to be the "solvent of life": indeed, water as found in nature almost always includes various dissolved substances, and special steps are required to obtain chemically pure water.

Water is the only common substance to exist as a solid, liquid, and gas in normal terrestrial conditions.

Water is one of the two official names for the chemical compound H₂O; it is also the liquid phase of H₂O. The other two common states of matter of water are the solid phase, ice, and the gaseous phase, water vapor or steam. The addition or removal of heat can cause phase transitions: freezing (water to ice), melting (ice to water), vaporization (water to vapor), condensation (vapor to water), sublimation (ice to vapor) and deposition (vapor to ice).

Water differs from most liquids in that it becomes less dense as it freezes. In 1 atm pressure, it reaches its maximum density of 1,000 kg/m³ (62.43 lb/cu ft) at 3.98 °C (39.16 °F). The density of ice is 917 kg/m³ (57.25 lb/cu ft), an expansion of 9%. This expansion can exert enormous pressure, bursting pipes and cracking rocks.

In a lake or ocean, water at 4 °C sinks to the bottom and ice forms on the surface, floating on the liquid water. This ice insulates the water below, preventing it from freezing solid. Without this protection, most aquatic organisms would perish during the winter.

At a pressure of one atmosphere (atm), ice melts or water freezes at 0 °C (32 °F) and water boils or vapor condenses at 100 °C (212 °F).

In a water molecule, the hydrogen atoms form a 104.5° angle with the oxygen atom. The hydrogen atoms are close to two corners of a tetrahedron centered on the oxygen. At the other two corners are lone pairs of valence electrons that do not participate in the bonding. In a perfect tetrahedron, the atoms would form a 109.5° angle, but the repulsion between the lone pairs is greater than the repulsion between the hydrogen atoms.

Other substances have a tetrahedral molecular structure, for example, methane (CH₄) and hydrogen sulfide (H₂S). However, oxygen is more electronegative (holds on to its electrons more tightly) than most other elements, so the oxygen atom retains a negative charge while the hydrogen atoms are positively charged. Along with the bent structure, this gives the molecule an electrical dipole moment and it is classified as a polar molecule.

Water is a good polar solvent, that dissolves many salts and hydrophilic organic molecules such as sugars and simple alcohols such as ethanol. Water also dissolves many gases, such as oxygen and carbon dioxide—the latter giving the fizz of carbonated beverages, sparkling wines and beers. In addition, many substances in living organisms, such as proteins, DNA and polysaccharides, are dissolved in water. The interactions between water and the sub-units of these biomacromolecules shape protein folding, DNA base pairing, and other phenomena crucial to life (hy-

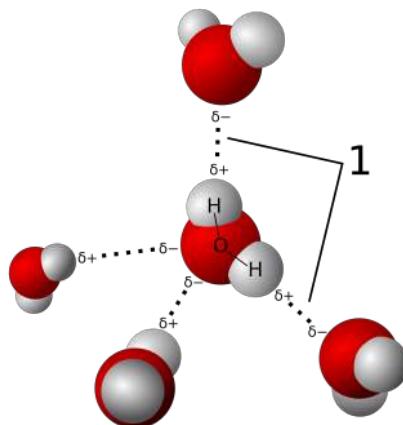


Figure 2.13: Model of hydrogen bonds (1) between molecules of water.⁴⁸

drophobic effect).

Many organic substances (such as fats and oils and alkanes) are hydrophobic, that is, insoluble in water. Many inorganic substances are insoluble too, including most metal oxides, sulfides, and silicates.

Because of its polarity, a molecule of water in the liquid or solid state can form up to four hydrogen bonds with neighboring molecules. Hydrogen bonds are about ten times as strong as the Van der Waals force that attracts molecules to each other in most liquids. This is the reason why the melting and boiling points of water are much higher than those of other analogous compounds like hydrogen sulfide. They also explain its exceptionally high specific heat capacity (about 4.2 J/g/K), heat of fusion (about 333 J/g), heat of vaporization (2257 J/g), and thermal conductivity (between 0.561 and 0.679 W/m/K). These properties make water more effective at moderating Earth's climate, by storing heat and transporting it between the oceans and the atmosphere. The hydrogen bonds of water are around 23 kJ/mol (compared to a covalent O-H bond at 492 kJ/mol). Of this, it is estimated that 90% is attributable to electrostatics, while the remaining 10% is partially covalent.

These bonds are the cause of water's high surface tension and capillary forces. The capillary action refers to the tendency of water to move up a narrow tube against the force of gravity. This property is relied upon by all vascular plants, such as trees.

Water is a weak solution of hydronium hydroxide – there is an equilibrium $2\text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+ + \text{OH}^-$ to form water. Water is considered to be neutral, with a pH of 7. Acids have pH values less than 7 while bases have values greater than 7.

Pure water has a low electrical conductivity, which in-

creases with the dissolution of a small amount of ionic material such as common salt.

Liquid water can be split into the elements hydrogen and oxygen by passing an electric current through it—a process called electrolysis. The decomposition requires more energy input than the heat released by the inverse process (285.8 kJ/mol, or 15.9 MJ/kg).

Chapter 3

Basic Organic Chemistry For Biology

Organic chemistry¹ is a branch of chemistry that studies the structure, properties and reactions of organic compounds, which contain carbon in covalent bonding. Study of structure determines their chemical composition and formula. Study of properties includes physical and chemical properties, and evaluation of chemical reactivity to understand their behavior. The study of organic reactions includes the chemical synthesis of natural products, drugs, and polymers, and study of individual organic molecules in the laboratory and via theoretical (in silico) study.

Organic compounds form the basis of all earthly life and constitute the majority of known chemicals. The bonding patterns of carbon, with its valence of four—formal single, double, and triple bonds, plus structures with delocalized electrons—make the array of organic compounds structurally diverse, and their range of applications enormous. They form the basis of, or are constituents of, many commercial products including pharmaceuticals; petrochemicals and agrichemicals, and products made from them including lubricants, solvents; plastics; fuels and explosives. The study of organic chemistry overlaps organometallic chemistry and biochemistry, but also with medicinal chemistry, polymer chemistry, and materials science.

The range of chemicals studied in organic chemistry includes hydrocarbons (compounds containing only carbon and hydrogen) as well as compounds based on carbon, but also containing other elements, especially oxygen, nitrogen, sulfur, phosphorus (included in many biochemicals) and the halogens.

Before the nineteenth century, chemists generally believed that compounds obtained from living organisms were endowed with a vital force that distinguished them from inorganic compounds. According to the concept of vitalism (vital force theory), organic matter was endowed with a “vital force”. During the first half of the nineteenth century, some of the first systematic studies of organic

compounds were reported. Around 1816 Michel Chevreul² started a study of soaps made from various fats and alkalies. He separated the acids that, in combination with the alkali, produced the soap. Since these were all individual compounds, he demonstrated that it was possible to make a chemical change in various fats (which traditionally come from organic sources), producing new compounds, without “vital force”. In 1828 Friedrich Wöhler³ produced the organic chemical urea (carbamide), a constituent of urine, from inorganic starting materials (the salts potassium cyanate and ammonium sulfate), in what is now called the Wöhler synthesis. Although Wöhler himself was cautious about claiming he had disproved vitalism, this was the first time a substance thought to be organic was synthesized in the laboratory without biological (organic) starting materials. The event is now generally accepted as indeed disproving the doctrine of vitalism.

A crucial breakthrough for organic chemistry was the concept of chemical structure, developed independently in 1858 by both Friedrich August Kekulé⁴ and Archibald Scott Couper⁵. Both researchers suggested that tetravalent carbon atoms could link to each other to form a carbon lattice, and that the detailed patterns of atomic bonding could be discerned by skillful interpretations of appropriate chemical reactions.

Organic molecules are described commonly by drawings or structural formulas, combinations of drawings and chemical symbols. The line-angle formula is simple and unambiguous. In this system, the endpoints and intersections of each line represent one carbon, and hydrogen atoms can either be notated explicitly or assumed to be present as implied by tetravalent carbon.

The era of the pharmaceutical industry began in the last decade of the 19th century when the manufacturing of acetylsalicylic acid—more commonly referred to as aspirin—in Germany was started by Bayer. By 1910

²https://en.wikipedia.org/wiki/Michel_Eugène_Chevreul

³https://en.wikipedia.org/wiki/Friedrich_Wöhler

⁴https://en.wikipedia.org/wiki/August_Kekulé

⁵https://en.wikipedia.org/wiki/Archibald_Scott_Couper

¹https://en.wikipedia.org/wiki/Organic_chemistry



Figure 3.1: This diagram⁶ shows 5 different structural representations of the organic compound butane. The left-most structure is a bond-line drawing where the hydrogen atoms are removed. The 2nd structure has the hydrogens added depicted—the dark wedged bonds indicate the hydrogen atoms are coming toward the reader, the hashed bonds indicate the atoms are oriented away from the reader, and the solid (plain) bonds indicate the bonds are in the plane of the screen/paper. The middle structure shows the four carbon atoms. The 4th structure is a representation just showing the atoms and bonds without 3-dimensions. The right-most structure is a condensed structure representation of butane.

Paul Ehrlich⁷ and his laboratory group began developing arsenic-based arsphenamine, (Salvarsan), as the first effective medicinal treatment of syphilis, and thereby initiated the medical practice of chemotherapy. Ehrlich popularized the concepts of “magic bullet” drugs and of systematically improving drug therapies. His laboratory made decisive contributions to developing antiserum for diphtheria and standardizing therapeutic serums.

In the early part of the 20th century, polymers and enzymes were shown to be large organic molecules, and petroleum was shown to be of biological origin.

The majority of chemical compounds occurring in biological organisms are carbon compounds, so the association between organic chemistry and biochemistry is so close that biochemistry might be regarded as in essence a branch of organic chemistry. Although the history of biochemistry might be taken to span some four centuries, fundamental understanding of the field only began to develop in the late 19th century and the actual term biochemistry was coined around the start of 20th century.

3.1 Functional groups

The concept of functional groups is central in organic chemistry, both as a means to classify structures and for predicting properties. A functional group is a molecular module, and the reactivity of that functional group is assumed, within limits, to be the same in a variety of molecules. Functional groups can have a decisive influence on the chemical and physical properties of organic compounds. Molecules are classified based on their functional groups. Alcohols, for example, all have the subunit C–O–H. All alcohols tend to be somewhat hydrophilic, usually form esters, and usually can be

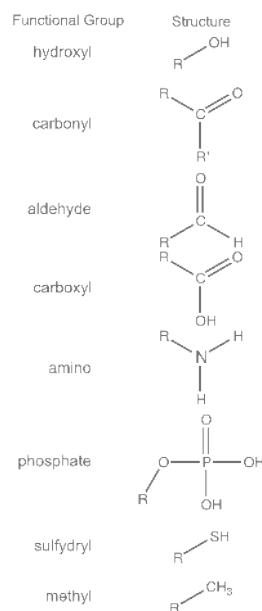


Figure 3.2: Biologically important functional groups.

converted to the corresponding halides. Most functional groups feature heteroatoms (atoms other than C and H). Organic compounds are classified according to functional groups, alcohols, carboxylic acids, amines, etc.

Combining the names of functional groups with the names of the parent alkanes generates what is termed a systematic nomenclature for naming organic compounds. In traditional nomenclature, the first carbon atom after the carbon that attaches to the functional group is called the alpha carbon; the second, beta carbon, the third, gamma carbon, etc. If there is another functional group at a carbon, it may be named with the Greek letter, e.g., the gamma-amino in gamma-aminobutyric acid is on the third carbon of the carbon chain attached to the carboxylic acid group. IUPAC conventions call for numeric labeling of the position, e.g. 4-aminobutanoic acid. In traditional names various qualifiers are used to label isomers, for example, isopropanol (IUPAC name: propan-2-ol) is an isomer of n-propanol (propan-1-ol). The term moiety has some overlap with the term “functional group”. However, a moiety is an entire “half” of a molecule, which can be not only a single functional group, but also a larger unit consisting of multiple functional groups. For example, an “aryl moiety” may be any group containing an aromatic ring, regardless of how many functional groups the said aryl has.

⁷https://en.wikipedia.org/wiki/Paul_Ehrlich

Table 3.1: Some biologically important functional groups containing oxygen or nitrogen

Chemical class	Group	Formula	Prefix	Example
Alcohol	Hydroxyl	ROH	hydroxy-	Methanol
Ketone	Carbonyl	RCOR'	-oyl- (-COR')oxo- (=O)	Butanone(Methyl ethyl ketone)
Aldehyde	Aldehyde	RCHO	formyl- (-COH)oxo- (=O)	Acetaldehyde(Ethanal)
Carboxylate	Carboxylate	RCOO-	carboxy-	Sodium acetate(Sodium ethanoate)
Carboxylic acid	Carboxyl	RCOOH	carboxy-	Acetic acid(Ethanoic acid)
Ester	Carboalkoxy	RCOOR'	alkanoyloxy- or alkoxycarbonyl	Ethyl butyrate(Ethyl butanoate)
	Carboxamide	RCONR'R"	carboxamido- or carbamoyl-	Acetamide(Ethanamide)
Amide	Primary amine	RNH2	amino-	Methylamine(Methanamine)
Amines	Secondary amine	R'R"NH	amino-	Dimethylamine
Amines	Tertiary amine	R3N	amino-	Trimethylamine
Amines	Quaternary ammonium ion	R4N+	ammonio-	Choline

3.2 Biomolecules

A biomolecule or biological molecule is a loosely used term for molecules present in organisms that are essential to one or more typically biological processes, such as cell division, morphogenesis, or development. Biomolecules include large macromolecules (or polyanions) such as proteins, carbohydrates, lipids, and nucleic acids, as well as small molecules such as primary metabolites, secondary metabolites and natural products. A more general name for this class of material is biological materials. Biomolecules are usually endogenous, produced within the organism but organisms usually need exogenous biomolecules, for example certain nutrients, to survive.

Biology and its subfields of biochemistry and molecular biology study biomolecules and their reactions. Most biomolecules are organic compounds, and just four elements—oxygen, carbon, hydrogen, and nitrogen—make up 96% of the human body's mass. But many other elements, such as the various biometals, are present in small amounts.

The uniformity of both specific types of molecules (the biomolecules) and of certain metabolic pathways are invariant features among the wide diversity of life forms; thus these biomolecules and metabolic pathways are referred to as “biochemical universals” or “theory of material unity of the living beings”, a unifying concept in biology, along with cell theory and evolution theory.

A macromolecule is a very large molecule, such as protein, commonly composed of the polymerization of smaller subunits called monomers. They are typically composed of thousands of atoms or more. A substance that is composed of macromolecules is called a polymer. The most common macromolecules in biochemistry are biopolymers (nucleic acids, proteins, and carbohydrates) and large non-polymeric molecules (such as lipids and macrocycles), synthetic fibers as well as experimental materials such as carbon nanotubes.

Macromolecules are large molecules composed of thousands of covalently connected atoms. Carbohydrates, lipids, proteins, and nucleic acids are all macromolecules. Macromolecules are formed by many monomers linking together, forming a polymer. Carbohydrates are composed of carbon, oxygen, and hydrogen. The monomer of carbohydrates are monosaccharides. There are three forms of carbohydrates: energy, storage, and structural molecules. A disaccharide is formed when a dehydration reaction joins two monosaccharides. Another type of macromolecules are lipids. Lipids are hydrocarbons that do not form polymers. Fats are constructed from glycerol and fatty acids. Phospholipids are commonly found in the phospholipid bilayer of membranes. They have hydrophilic

heads and hydrophobic tails. A protein is another type of macromolecules. Amino acids are the monomers of proteins. Proteins have many different functions. There are proteins that are used for structural support, storage, transport, cellular communication, movement, defense against foreign substances, and more. Nucleic acids transmit and help express hereditary information. They are made up of monomers called nucleotides. Two types of nucleic acids are DNA and RNA.

All living organisms are dependent on three essential biopolymers for their biological functions: DNA, RNA and proteins. Each of these molecules is required for life since each plays a distinct, indispensable role in the cell. The simple summary is that DNA makes RNA, and then RNA makes proteins.

DNA, RNA, and proteins all consist of a repeating structure of related building blocks (nucleotides in the case of DNA and RNA, amino acids in the case of proteins). In general, they are all unbranched polymers, and so can be represented in the form of a string. Indeed, they can be viewed as a string of beads, with each bead representing a single nucleotide or amino acid monomer linked together through covalent chemical bonds into a very long chain.

In most cases, the monomers within the chain have a strong propensity to interact with other amino acids or nucleotides. In DNA and RNA, this can take the form of Watson-Crick base pairs (G-C and A-T or A-U), although many more complicated interactions can and do occur.

Table 3.2: Comparison of the main classes of biological macromolecules.

Macromolecule (Polymer)	Building Block (Monomer)	Joining Bond
Proteins	Amino acids	Peptide
DNA	Nucleotides (a phosphate, ribose, and a base- adenine, guanine, thymine, or cytosine)	Phosphodiester
RNA	Nucleotides (a phosphate, ribose, and a base- adenine, guanine, Uracil, or cytosine)	Phosphodiester
Polysaccharides (carbohydrates)	Monosaccharides	Glycosidic
Lipids	unlike the other macromolecules, lipids are not defined by chemical Structure. Lipids are any organic nonpolar molecule.	Some lipids are held together by ester bonds; some are huge aggregates of small molecules held together by hydrophobic interactions.

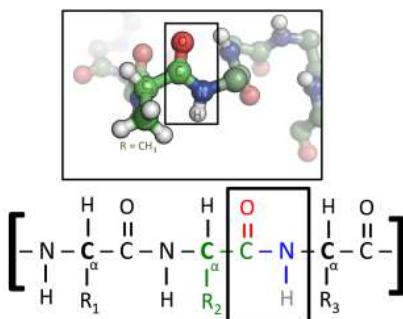


Figure 3.3: Chemical structure of the peptide bond (bottom) and the three-dimensional structure of a peptide bond between an alanine and an adjacent amino acid (top/inset). The bond itself is made of the CHON elements.⁹

3.3 Proteins

Proteins⁸ are large biomolecules, or macromolecules, consisting of one or more long chains of amino acid residues. Proteins perform a vast array of functions within organisms, including catalysing metabolic reactions, DNA replication, responding to stimuli, providing structure to cells, and organisms, and transporting molecules from one location to another. Proteins differ from one another primarily in their sequence of amino acids, which is dictated by the nucleotide sequence of their genes, and which usually results in protein folding into a specific 3D structure that determines its activity.

A linear chain of amino acid residues is called a polypeptide. A protein contains at least one long polypeptide. Short polypeptides, containing less than 20–30 residues, are rarely considered to be proteins and are commonly called peptides, or sometimes oligopeptides. The individual amino acid residues are bonded together by peptide bonds and adjacent amino acid residues. The sequence of amino acid residues in a protein is defined by the sequence of a gene, which is encoded in the genetic code. In general, the genetic code specifies 20 standard amino acids; but in certain organisms the genetic code can include selenocysteine and—in certain archaea—pyrrolysine. Shortly after or even during synthesis, the residues in a protein are often chemically modified by post-translational modification, which alters the physical and chemical properties, folding, stability, activity, and ultimately, the function of the proteins. Some proteins have non-peptide groups attached, which can be called prosthetic groups or cofactors. Proteins can also work together to achieve a particular function, and they often associate to form stable protein complexes.

⁸<https://en.wikipedia.org/wiki/Protein>

Once formed, proteins only exist for a certain period and are then degraded and recycled by the cell's machinery through the process of protein turnover. A protein's lifespan is measured in terms of its half-life and covers a wide range. They can exist for minutes or years with an average lifespan of 1–2 days in mammalian cells. Abnormal or misfolded proteins are degraded more rapidly either due to being targeted for destruction or due to being unstable.

Like other biological macromolecules such as polysaccharides and nucleic acids, proteins are essential parts of organisms and participate in virtually every process within cells. Many proteins are enzymes that catalyse biochemical reactions and are vital to metabolism. Proteins also have structural or mechanical functions, such as actin and myosin in muscle and the proteins in the cytoskeleton, which form a system of scaffolding that maintains cell shape. Other proteins are important in cell signalling, immune responses, cell adhesion, and the cell cycle. In animals, proteins are needed in the diet to provide the essential amino acids that cannot be synthesized. Digestion breaks the proteins down for use in the metabolism.

Proteins may be purified from other cellular components using a variety of techniques such as ultracentrifugation, precipitation, electrophoresis, and chromatography; the advent of genetic engineering has made possible a number of methods to facilitate purification. Methods commonly used to study protein structure and function include immunohistochemistry, site-directed mutagenesis, X-ray crystallography, nuclear magnetic resonance and mass spectrometry.

Proteins were recognized as a distinct class of biological molecules in the eighteenth century by Antoine Fourcroy and others, distinguished by the molecules' ability to coagulate or flocculate under treatments with heat or acid. Noted examples at the time included albumin from egg whites, blood serum albumin, fibrin, and wheat gluten.

Proteins were first described by the Dutch chemist Gerardus Johannes Mulder¹⁰ and named by the Swedish chemist Jöns Jacob Berzelius¹¹ in 1838. Mulder carried out elemental analysis of common proteins and found that nearly all proteins had the same empirical formula, $\text{C}_{400}\text{H}_{620}\text{N}_{100}\text{O}_{120}\text{P}_1\text{S}_1$. He came to the erroneous conclusion that they might be composed of a single type of (very large) molecule. The term "protein" to describe these molecules was proposed by Mulder's associate Berzelius; protein is derived from the Greek word πρώτεος (protoeios), meaning "primary", "in the lead", or "standing in front", + -in. Mulder went on to identify the products of protein degradation such as the amino acid leucine for

¹⁰https://en.wikipedia.org/wiki/Gerardus_Johannes_Mulder

¹¹https://en.wikipedia.org/wiki/J%C3%B6ns_Jacob_Berzelius

which he found a (nearly correct) molecular weight of 131 Da. Prior to “protein”, other names were used, like “albumins” or “albuminous materials” (*Eiweisskörper*, in German).

Early nutritional scientists such as the German Carl von Voit¹² believed that protein was the most important nutrient for maintaining the structure of the body, because it was generally believed that “flesh makes flesh.” Karl Heinrich Ritthausen¹³ extended known protein forms with the identification of glutamic acid. At the Connecticut Agricultural Experiment Station a detailed review of the vegetable proteins was compiled by Thomas Burr Osborne¹⁴. Working with Lafayette Mendel¹⁵ and applying Justus von Liebig’s¹⁶ law of the minimum in feeding laboratory rats, the nutritionally essential amino acids were established. The work was continued and communicated by William Cumming Rose¹⁷. The understanding of proteins as polypeptides came through the work of Franz Hofmeister¹⁸ and Hermann Emil Fischer¹⁹ in 1902. The central role of proteins as enzymes in living organisms was not fully appreciated until 1926, when James B. Sumner²⁰ showed that the enzyme urease was in fact a protein.

The difficulty in purifying proteins in large quantities made them very difficult for early protein biochemists to study. Hence, early studies focused on proteins that could be purified in large quantities, e.g., those of blood, egg white, various toxins, and digestive/metabolic enzymes obtained from slaughterhouses. In the 1950s, the Armour Hot Dog Co. purified 1 kg of pure bovine pancreatic ribonuclease A and made it freely available to scientists; this gesture helped ribonuclease A become a major target for biochemical study for the following decades.

Linus Pauling²¹ is credited with the successful prediction of regular protein secondary structures based on hydrogen bonding, an idea first put forth by William Astbury²² in 1933. Later work by Walter Kauzmann²³ on denaturation, based partly on previous studies by Kaj Linderstrøm-Lang²⁴, contributed an understanding of protein folding and structure mediated by hydrophobic interactions.

The first protein to be sequenced was insulin, by Fred-

erick Sanger²⁵, in 1949. Sanger correctly determined the amino acid sequence of insulin, thus conclusively demonstrating that proteins consisted of linear polymers of amino acids rather than branched chains, colloids, or cyclols. He won the Nobel Prize for this achievement in 1958.

The first protein structures to be solved were hemoglobin and myoglobin, by Max Perutz²⁶ and Sir John Cowdery Kendrew²⁷, respectively, in 1958. As of 2017, the Protein Data Bank has over 126,060 atomic-resolution structures of proteins. In more recent times, cryo-electron microscopy of large macromolecular assemblies and computational protein structure prediction of small protein domains are two methods approaching atomic resolution.

Most proteins consist of linear polymers built from series of up to 20 different L- α - amino acids. All proteinogenic amino acids possess common structural features, including an α -carbon to which an amino group, a carboxyl group, and a variable side chain are bonded. Only proline differs from this basic structure as it contains an unusual ring to the N-end amine group, which forces the CO-NH amide moiety into a fixed conformation. The side chains of the standard amino acids, detailed in the list of standard amino acids, have a great variety of chemical structures and properties; it is the combined effect of all of the amino acid side chains in a protein that ultimately determines its three-dimensional structure and its chemical reactivity. The amino acids in a polypeptide chain are linked by peptide bonds. Once linked in the protein chain, an individual amino acid is called a residue, and the linked series of carbon, nitrogen, and oxygen atoms are known as the main chain or protein backbone.

The peptide bond has two resonance forms that contribute some double-bond character and inhibit rotation around its axis, so that the alpha carbons are roughly coplanar. The other two dihedral angles in the peptide bond determine the local shape assumed by the protein backbone. The end with a free amino group is known as the N-terminus or amino terminus, whereas the end of the protein with a free carboxyl group is known as the C-terminus or carboxy terminus (the sequence of the protein is written from N-terminus to C-terminus, from left to right).

The words protein, polypeptide, and peptide are a little ambiguous and can overlap in meaning. Protein is generally used to refer to the complete biological molecule in a stable conformation, whereas peptide is generally reserved for a short amino acid oligomers often lacking a stable 3D structure. But the boundary between the two is not well defined and usually lies near 20–30 residues. Polypeptide

¹²https://en.wikipedia.org/wiki/Carl_von_Voit

¹³https://en.wikipedia.org/wiki/Karl_Heinrich_Ritthausen

¹⁴[https://en.wikipedia.org/wiki/Thomas_Burr_Osborne_\(chemist\)](https://en.wikipedia.org/wiki/Thomas_Burr_Osborne_(chemist))

¹⁵https://en.wikipedia.org/wiki/Lafayette_Mendel

¹⁶https://en.wikipedia.org/wiki/Justus_von_Liebig

¹⁷https://en.wikipedia.org/wiki/William_Cumming_Rose

¹⁸https://en.wikipedia.org/wiki/Franz_Hofmeister

¹⁹https://en.wikipedia.org/wiki/Emil_Fischer

²⁰https://en.wikipedia.org/wiki/James_B._Sumner

²¹https://en.wikipedia.org/wiki/Linus_Pauling

²²https://en.wikipedia.org/wiki/William_Astbury

²³https://en.wikipedia.org/wiki/Walter_Kauzmann

²⁴https://en.wikipedia.org/wiki/Kaj_Ulrik_Linderstrøm-Lang

²⁵https://en.wikipedia.org/wiki/Frederick_Sanger

²⁶https://en.wikipedia.org/wiki/Max_Perutz

²⁷https://en.wikipedia.org/wiki/John_Kendrew

can refer to any single linear chain of amino acids, usually regardless of length, but often implies an absence of a defined conformation.

Proteins can interact with many types of molecules, including with other proteins, with lipids, with carbohydrates, and with DNA.

It has been estimated that average-sized bacteria contain about 2 million proteins per cell (e.g. *E. coli* and *Staphylococcus aureus*). Smaller bacteria, such as Mycoplasma or spirochetes contain fewer molecules, on the order of 50,000 to 1 million. By contrast, eukaryotic cells are larger and thus contain much more protein. For instance, yeast cells have been estimated to contain about 50 million proteins and human cells on the order of 1 to 3 billion. The concentration of individual protein copies ranges from a few molecules per cell up to 20 million. Not all genes coding proteins are expressed in most cells and their number depends on, for example, cell type and external stimuli. For instance, of the 20,000 or so proteins encoded by the human genome, only 6,000 are detected in lymphoblastoid cells. Moreover, the number of proteins the genome encodes correlates well with the organism complexity. Eukaryotes have 15,000, bacteria have 3,200, archaea have 2,400, and viruses have 42 proteins on average coded in their respective genomes.

Proteins are assembled from amino acids using information encoded in genes. Each protein has its own unique amino acid sequence that is specified by the nucleotide sequence of the gene encoding this protein. The genetic code is a set of three-nucleotide sets called codons and each three-nucleotide combination designates an amino acid, for example AUG (adenine-uracil-guanine) is the code for methionine. Because DNA contains four nucleotides, the total number of possible codons is 64; hence, there is some redundancy in the genetic code, with some amino acids specified by more than one codon. Genes encoded in DNA are first transcribed into pre-messenger RNA (mRNA) by proteins such as RNA polymerase. Most organisms then process the pre-mRNA (also known as a primary transcript) using various forms of Post-transcriptional modification to form the mature mRNA, which is then used as a template for protein synthesis by the ribosome. In prokaryotes the mRNA may either be used as soon as it is produced, or be bound by a ribosome after having moved away from the nucleoid. In contrast, eukaryotes make mRNA in the cell nucleus and then translocate it across the nuclear membrane into the cytoplasm, where protein synthesis then takes place. The rate of protein synthesis is higher in prokaryotes than eukaryotes and can reach up to 20 amino acids per second.

The process of synthesizing a protein from an mRNA template is known as translation. The mRNA is loaded onto

the ribosome and is read three nucleotides at a time by matching each codon to its base pairing anticodon located on a transfer RNA molecule, which carries the amino acid corresponding to the codon it recognizes. The enzyme aminoacyl tRNA synthetase “charges” the tRNA molecules with the correct amino acids. The growing polypeptide is often termed the nascent chain. Proteins are always biosynthesized from N-terminus to C-terminus.

The size of a synthesized protein can be measured by the number of amino acids it contains and by its total molecular mass, which is normally reported in units of daltons (synonymous with atomic mass units), or the derivative unit kilodalton (kDa). The average size of a protein increases from Archaea to Bacteria to Eukaryote (283, 311, 438 residues and 31, 34, 49 kDa respectively) due to a bigger number of protein domains constituting proteins in higher organisms. For instance, yeast proteins are on average 466 amino acids long and 53 kDa in mass. The largest known proteins are the titins, a component of the muscle sarcomere, with a molecular mass of almost 3,000 kDa and a total length of almost 27,000 amino acids.

3.3.1 Structure

Most proteins fold into unique 3D structures. The shape into which a protein naturally folds is known as its native conformation. Although many proteins can fold unassisted, simply through the chemical properties of their amino acids, others require the aid of molecular chaperones to fold into their native states. Biochemists often refer to four distinct aspects of a protein's structure:

- Primary structure: the amino acid sequence. A protein is a polyamide.
- Secondary structure: regularly repeating local structures stabilized by hydrogen bonds. The most common examples are the α -helix, β -sheet and turns. Because secondary structures are local, many regions of different secondary structure can be present in the same protein molecule.
- Tertiary structure: the overall shape of a single protein molecule; the spatial relationship of the secondary structures to one another. Tertiary structure is generally stabilized by nonlocal interactions, most commonly the formation of a hydrophobic core, but also through salt bridges, hydrogen bonds, disulfide bonds, and even posttranslational modifications. The term “tertiary structure” is often used as synonymous with the term fold. The tertiary structure is what controls the basic function of the protein.
- Quaternary structure: the structure formed by several protein molecules (polypeptide chains), usually

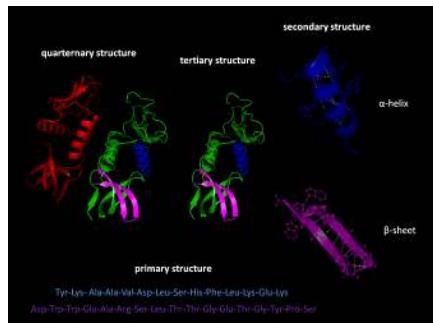


Figure 3.4: The amino-acid sequence, the primary structure of a protein, determines the secondary (α -helix and β -sheet), tertiary and quaternary protein structures²⁸

called protein subunits in this context, which function as a single protein complex.

- Quinary structure: the signatures of protein surface that organize the crowded cellular interior. Quinary structure is dependent on transient, yet essential, macromolecular interactions that occur inside living cells.

Proteins are not entirely rigid molecules. In addition to these levels of structure, proteins may shift between several related structures while they perform their functions. In the context of these functional rearrangements, these tertiary or quaternary structures are usually referred to as “conformations”, and transitions between them are called conformational changes. Such changes are often induced by the binding of a substrate molecule to an enzyme’s active site, or the physical region of the protein that participates in chemical catalysis. In solution proteins also undergo variation in structure through thermal vibration and the collision with other molecules.

Proteins can be informally divided into three main classes, which correlate with typical tertiary structures: globular proteins, fibrous proteins, and membrane proteins. Almost all globular proteins are soluble and many are enzymes. Fibrous proteins are often structural, such as collagen, the major component of connective tissue, or keratin, the protein component of hair and nails. Membrane proteins often serve as receptors or provide channels for polar or charged molecules to pass through the cell membrane.

A special case of intramolecular hydrogen bonds within proteins, poorly shielded from water attack and hence promoting their own dehydration, are called dehydrons.

Many proteins are composed of several protein domains, i.e. segments of a protein that fold into distinct structural units. Domains usually also have specific functions, such as enzymatic activities (e.g. kinase) or

they serve as binding modules (e.g. the SH3 domain binds to proline-rich sequences in other proteins).

Short amino acid sequences within proteins often act as recognition sites for other proteins. For instance, SH3 domains typically bind to short PxxP motifs (i.e. 2 prolines [P], separated by two unspecified amino acids [x], although the surrounding amino acids may determine the exact binding specificity). Many such motifs has been collected in the Eukaryotic Linear Motif (ELM) database.

3.3.2 Cellular Functions of Proteins

Proteins are the chief actors within the cell, said to be carrying out the duties specified by the information encoded in genes. With the exception of certain types of RNA, most other biological molecules are relatively inert elements upon which proteins act. Proteins make up half the dry weight of an *Escherichia coli* cell, whereas other macromolecules such as DNA and RNA make up only 3% and 20%, respectively. The set of proteins expressed in a particular cell or cell type is known as its proteome.

The chief characteristic of proteins that also allows their diverse set of functions is their ability to bind other molecules specifically and tightly. The region of the protein responsible for binding another molecule is known as the binding site and is often a depression or “pocket” on the molecular surface. This binding ability is mediated by the tertiary structure of the protein, which defines the binding site pocket, and by the chemical properties of the surrounding amino acids’ side chains. Protein binding can be extraordinarily tight and specific; for example, the ribonuclease inhibitor protein binds to human angiogenin with a sub-femtomolar dissociation constant (<10–15 M) but does not bind at all to its amphibian homolog onconase (>1 M). Extremely minor chemical changes such as the addition of a single methyl group to a binding partner can sometimes suffice to nearly eliminate binding; for example, the aminoacyl tRNA synthetase specific to the amino acid valine discriminates against the very similar side chain of the amino acid isoleucine.

Proteins can bind to other proteins as well as to small-molecule substrates. When proteins bind specifically to other copies of the same molecule, they can oligomerize to form fibrils; this process occurs often in structural proteins that consist of globular monomers that self-associate to form rigid fibers. Protein-protein interactions also regulate enzymatic activity, control progression through the cell cycle, and allow the assembly of large protein complexes that carry out many closely related reactions with a common biological function. Proteins can also bind to, or even be integrated into, cell membranes. The ability of binding partners to induce conformational changes in proteins allows the construction of enormously complex

signaling networks. As interactions between proteins are reversible, and depend heavily on the availability of different groups of partner proteins to form aggregates that are capable to carry out discrete sets of function, study of the interactions between specific proteins is a key to understand important aspects of cellular function, and ultimately the properties that distinguish particular cell types.

The best-known role of proteins in the cell is as enzymes, which catalyse chemical reactions. Enzymes are usually highly specific and accelerate only one or a few chemical reactions. Enzymes carry out most of the reactions involved in metabolism, as well as manipulating DNA in processes such as DNA replication, DNA repair, and transcription. Some enzymes act on other proteins to add or remove chemical groups in a process known as posttranslational modification. About 4,000 reactions are known to be catalysed by enzymes. The rate acceleration conferred by enzymatic catalysis is often enormous—as much as 10¹⁷-fold increase in rate over the uncatalysed reaction in the case of orotate decarboxylase (78 million years without the enzyme, 18 milliseconds with the enzyme).

The molecules bound and acted upon by enzymes are called substrates. Although enzymes can consist of hundreds of amino acids, it is usually only a small fraction of the residues that come in contact with the substrate, and an even smaller fraction—three to four residues on average—that are directly involved in catalysis. The region of the enzyme that binds the substrate and contains the catalytic residues is known as the active site.

Dirigent proteins are members of a class of proteins that dictate the stereochemistry of a compound synthesized by other enzymes.

3.3.3 Cell Signaling And Ligand Binding

Many proteins are involved in the process of cell signaling and signal transduction. Some proteins, such as insulin, are extracellular proteins that transmit a signal from the cell in which they were synthesized to other cells in distant tissues. Others are membrane proteins that act as receptors whose main function is to bind a signaling molecule and induce a biochemical response in the cell. Many receptors have a binding site exposed on the cell surface and an effector domain within the cell, which may have enzymatic activity or may undergo a conformational change detected by other proteins within the cell.

Antibodies are protein components of an adaptive immune system whose main function is to bind antigens, or foreign substances in the body, and target them for destruction. Antibodies can be secreted into the extracellular environment or anchored in the membranes of specialized B cells known as plasma cells. Whereas enzymes are

limited in their binding affinity for their substrates by the necessity of conducting their reaction, antibodies have no such constraints. An antibody's binding affinity to its target is extraordinarily high.

Many ligand transport proteins bind particular small biomolecules and transport them to other locations in the body of a multicellular organism. These proteins must have a high binding affinity when their ligand is present in high concentrations, but must also release the ligand when it is present at low concentrations in the target tissues. The canonical example of a ligand-binding protein is haemoglobin, which transports oxygen from the lungs to other organs and tissues in all vertebrates and has close homologs in every biological kingdom. Lectins are sugar-binding proteins which are highly specific for their sugar moieties. Lectins typically play a role in biological recognition phenomena involving cells and proteins. Receptors and hormones are highly specific binding proteins.

Transmembrane proteins can also serve as ligand transport proteins that alter the permeability of the cell membrane to small molecules and ions. The membrane alone has a hydrophobic core through which polar or charged molecules cannot diffuse. Membrane proteins contain internal channels that allow such molecules to enter and exit the cell. Many ion channel proteins are specialized to select for only a particular ion; for example, potassium and sodium channels often discriminate for only one of the two ions.

3.3.4 Structural Proteins

Structural proteins confer stiffness and rigidity to otherwise-fluid biological components. Most structural proteins are fibrous proteins; for example, collagen and elastin are critical components of connective tissue such as cartilage, and keratin is found in hard or filamentous structures such as hair, nails, feathers, hooves, and some animal shells. Some globular proteins can also play structural functions, for example, actin and tubulin are globular and soluble as monomers, but polymerize to form long, stiff fibers that make up the cytoskeleton, which allows the cell to maintain its shape and size.

Other proteins that serve structural functions are motor proteins such as myosin, kinesin, and dynein, which are capable of generating mechanical forces. These proteins are crucial for cellular motility of single celled organisms and the sperm of many multicellular organisms which reproduce sexually. They also generate the forces exerted by contracting muscles and play essential roles in intracellular transport.

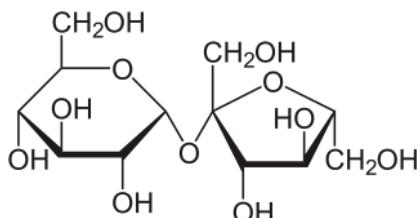


Figure 3.5: The disaccharide sucrose³⁰

3.4 Carbohydrates

A carbohydrate²⁹ is a biomolecule consisting of carbon (C), hydrogen (H) and oxygen (O) atoms, usually with a hydrogen–oxygen atom ratio of 2:1 (as in water) and thus with the empirical formula $C_m(H_2O)_n$ (where m may be different from n). This formula holds true for monosaccharides. Some exceptions exist; for example, deoxyribose, a sugar component of DNA, has the empirical formula $C_5H_{10}O_4$. The carbohydrates are technically hydrates of carbon; structurally it is more accurate to view them as aldoses and ketoses.

The term is most common in biochemistry, where it is a synonym of saccharide, a group that includes sugars, starch, and cellulose. The saccharides are divided into four chemical groups: monosaccharides, disaccharides, oligosaccharides, and polysaccharides. Monosaccharides and disaccharides, the smallest (lower molecular weight) carbohydrates, are commonly referred to as sugars. The word saccharide comes from the Greek word σάκχαρον (sákkharon), meaning “sugar”. While the scientific nomenclature of carbohydrates is complex, the names of the monosaccharides and disaccharides very often end in the suffix -ose, as in the monosaccharides fructose (fruit sugar) and glucose (starch sugar) and the disaccharides sucrose (cane or beet sugar) and lactose (milk sugar).

Carbohydrates perform numerous roles in living organisms. Polysaccharides serve for the storage of energy (e.g. starch and glycogen) and as structural components (e.g. cellulose in plants and chitin in arthropods). The 5-carbon monosaccharide ribose is an important component of coenzymes (e.g. ATP, FAD and NAD) and the backbone of the genetic molecule known as RNA. The related deoxyribose is a component of DNA. Saccharides and their derivatives include many other important biomolecules that play key roles in the immune system, fertilization, preventing pathogenesis, blood clotting, and development.

They are found in a wide variety of natural and processed foods. Starch is a polysaccharide. It is abundant in cereals (wheat, maize, rice), potatoes, and processed food

based on cereal flour, such as bread, pizza or pasta. Sugars appear in human diet mainly as table sugar (sucrose, extracted from sugarcane or sugar beets), lactose (abundant in milk), glucose and fructose, both of which occur naturally in honey, many fruits, and some vegetables. Table sugar, milk, or honey are often added to drinks and many prepared foods such as jam, biscuits and cakes.

Cellulose, a polysaccharide found in the cell walls of all plants, is one of the main components of insoluble dietary fiber. Although it is not digestible, insoluble dietary fiber helps to maintain a healthy digestive system by easing defecation. Other polysaccharides contained in dietary fiber include resistant starch and inulin, which feed some bacteria in the microbiota of the large intestine, and are metabolized by these bacteria to yield short-chain fatty acids.

In scientific literature, the term “carbohydrate” has many synonyms, like “sugar” (in the broad sense), “saccharide”, “ose”, “glucide”, “hydrate of carbon” or “polyhydroxy compounds with aldehyde or ketone”. Some of these terms, specially “carbohydrate” and “sugar”, are also used with other meanings.

Formerly the name “carbohydrate” was used in chemistry for any compound with the formula $C_m(H_2O)_n$. Following this definition, some chemists considered formaldehyde (CH_2O) to be the simplest carbohydrate, while others claimed that title for glycolaldehyde. Today, the term is generally understood in the biochemistry sense, which excludes compounds with only one or two carbons and includes many biological carbohydrates which deviate from this formula. For example, while the above representative formulas would seem to capture the commonly known carbohydrates, ubiquitous and abundant carbohydrates often deviate from this. For example, carbohydrates often display chemical groups such as: N-acetyl (e.g. chitin), sulphate (e.g. glycosaminoglycans), carboxylic acid (e.g. sialic acid) and deoxy modifications (e.g. fucose and sialic acid).

Natural saccharides are generally built of simple carbohydrates called monosaccharides with general formula $(CH_2O)_n$ where n is three or more. A typical monosaccharide has the structure $H-(CHOH)_x(C=O)-(CHOH)_y-H$, that is, an aldehyde or ketone with many hydroxyl groups added, usually one on each carbon atom that is not part of the aldehyde or ketone functional group. Examples of monosaccharides are glucose, fructose, and glyceraldehydes. However, some biological substances commonly called “monosaccharides” do not conform to this formula (e.g. uronic acids and deoxy-sugars such as fucose) and there are many chemicals that do conform to this formula but are not considered to be monosaccharides (e.g. formaldehyde CH_2O and inositol $(CH_2O)_6$).

²⁹<https://en.wikipedia.org/wiki/Carbohydrate>

The open-chain form of a monosaccharide often co-exists with a closed ring form where the aldehyde/ketone carbonyl group carbon ($C=O$) and hydroxyl group ($-OH$) react forming a hemiacetal with a new $C-O-C$ bridge.

Monosaccharides can be linked together into what are called polysaccharides (or oligosaccharides) in a large variety of ways. Many carbohydrates contain one or more modified monosaccharide units that have had one or more groups replaced or removed. For example, deoxyribose, a component of DNA, is a modified version of ribose; chitin is composed of repeating units of N-acetyl glucosamine, a nitrogen-containing form of glucose.

Monosaccharides are the simplest carbohydrates in that they cannot be hydrolyzed to smaller carbohydrates. They are aldehydes or ketones with two or more hydroxyl groups. The general chemical formula of an unmodified monosaccharide is $(CH_2O)_n$, literally a “carbon hydrate”. Monosaccharides are important fuel molecules as well as building blocks for nucleic acids. The smallest monosaccharides, for which $n=3$, are dihydroxyacetone and D- and L-glyceraldehydes.

Monosaccharides are the major fuel source for metabolism, being used both as an energy source (glucose being the most important in nature) and in biosynthesis. When monosaccharides are not immediately needed by many cells, they are often converted to more space-efficient forms, often polysaccharides. In many animals, including humans, this storage form is glycogen, especially in liver and muscle cells. In plants, starch is used for the same purpose. The most abundant carbohydrate, cellulose, is a structural component of the cell wall of plants and many forms of algae. Ribose is a component of RNA. Deoxyribose is a component of DNA. Lyxose is a component of lyxoflavin found in the human heart. Ribulose and xylulose occur in the pentose phosphate pathway. Galactose, a component of milk sugar lactose, is found in galactolipids in plant cell membranes and in glycoproteins in many tissues. Mannose occurs in human metabolism, especially in the glycosylation of certain proteins. Fructose, or fruit sugar, is found in many plants and humans, it is metabolized in the liver, absorbed directly into the intestines during digestion, and found in semen. Trehalose, a major sugar of insects, is rapidly hydrolyzed into two glucose molecules to support continuous flight.

Two joined monosaccharides are called a disaccharide and these are the simplest polysaccharides. Examples include sucrose and lactose. They are composed of two monosaccharide units bound together by a covalent bond known as a glycosidic linkage formed via a dehydration reaction, resulting in the loss of a hydrogen atom from one monosaccharide and a hydroxyl group from the other.

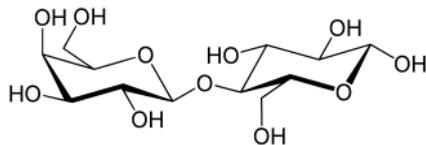


Figure 3.6: The disaccharide lactose³¹

The formula of unmodified disaccharides is $C_{12}H_{22}O_{11}$. Although there are numerous kinds of disaccharides, a handful of disaccharides are particularly notable.

- Its monosaccharides: glucose and fructose
- Their ring types: glucose is a pyranose and fructose is a furanose
- How they are linked together: the oxygen on carbon number 1 ($C1$) of α -D-glucose is linked to the $C2$ of D-fructose.
- The -oside suffix indicates that the anomeric carbon of both monosaccharides participates in the glycosidic bond.

Lactose, a disaccharide composed of one D-galactose molecule and one D-glucose molecule, occurs naturally in mammalian milk. The systematic name for lactose is $O-\beta-D\text{-galactopyranosyl}\text{-(1}\rightarrow 4\text{)}\text{-D\text{-glucopyranose}}$. Other notable disaccharides include maltose (two D-glucoses linked α -1,4) and cellobiose (two D-glucoses linked β -1,4). Disaccharides can be classified into two types: reducing and non-reducing disaccharides. If the functional group is present in bonding with another sugar unit, it is called a reducing disaccharide or biose.

3.5 Lipids

In biology and biochemistry, a lipid³² is a macro-biomolecule that is soluble in nonpolar solvents. Non-polar solvents are typically hydrocarbons used to dissolve other naturally occurring hydrocarbon lipid molecules that do not (or do not easily) dissolve in water, including fatty acids, waxes, sterols, fat-soluble vitamins (such as vitamins A, D, E, and K), monoglycerides, diglycerides, triglycerides, and phospholipids.

The functions of lipids include storing energy, signaling, and acting as structural components of cell membranes. Lipids have applications in the cosmetic and food industries as well as in nanotechnology.

Scientists sometimes define lipids as hydrophobic or amphiphilic small molecules; the amphiphilic nature of some lipids allows them to form structures such as vesicles, multilamellar/unilamellar liposomes, or membranes in an aqueous environment. Biological lipids originate

³²<https://en.wikipedia.org/wiki/Lipid>

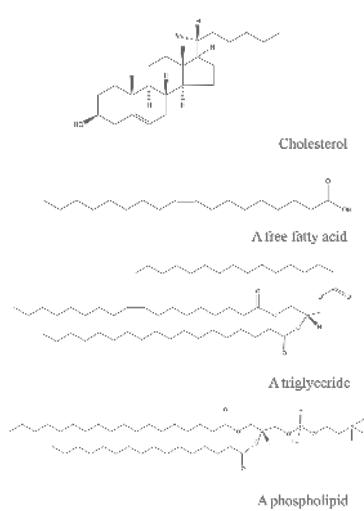


Figure 3.7: Structures of some common lipids.³³ At the top are cholesterol and oleic acid. The middle structure is a triglyceride composed of oleoyl, stearoyl, and palmitoyl chains attached to a glycerol backbone. At the bottom is the common phospholipid phosphatidylcholine.

entirely or in part from two distinct types of biochemical subunits or “building-blocks”: ketoacyl and isoprene groups. Using this approach, lipids may be divided into eight categories: fatty acids, glycerolipids, glycerophospholipids, sphingolipids, saccharolipids, and polyketides (derived from condensation of ketoacyl subunits); and sterol lipids and prenol lipids (derived from condensation of isoprene subunits).

Although the term “lipid” is sometimes used as a synonym for fats, fats are a subgroup of lipids called triglycerides. Lipids also encompass molecules such as fatty acids and their derivatives (including tri-, di-, monoglycerides, and phospholipids), as well as other sterol-containing metabolites such as cholesterol. Although humans and other mammals use various biosynthetic pathways both to break down and to synthesize lipids, some essential lipids can’t be made this way and must be obtained from the diet.

Lipid may be regarded as organic substances relatively insoluble in water, soluble in organic solvents(alcohol, ether etc.) actually or potentially related to fatty acid and utilized by the living cells.

In 1815, Henri Braconnot classified lipids (*graisses*) in two categories, suifs (solid greases or tallow) and huiles (fluid oils). In 1823, Michel Eugène Chevreul developed a more detailed classification, including oils, greases, tallow, waxes, resins, balsams and volatile oils (or essential oils).

In 1827, William Prout recognized fat (“oily” alimentary matters), along with protein (“albuminous”) and carbohydrate (“saccharine”), as an important nutrient for humans and animals.

The word “lipide”, which stems etymologically from the Greek *lipos* (fat), was introduced in 1923 by the french pharmacologist Gabriel Bertrand³⁴. Bertrands included in the concept not only the traditional fats (glycerides), but also the “lipoids”, with a complex constitution. Despite the word “lipide” was unanimously approved by the international commission of Société de Chimie Biologique during the plenary session on the 3rd of July 1923. The word “lipide” has been later anglicized as “lipid” because of its pronunciation (‘lɪpɪd). In the french language, the suffixe “-ide”, from the ancient greek “-ίδης” (meaning ‘son of’ or ‘descendant of’), is always pronounced (ɪd).

In 1947, T. P. Hilditch divided lipids into “simple lipids”, with greases and waxes (true waxes, sterols, alcohols).

Fatty acids, or fatty acid residues when they are part of a lipid, are a diverse group of molecules synthesized by chain-elongation of an acetyl-CoA primer with malonyl-CoA or methylmalonyl-CoA groups in a process called fatty acid synthesis. They are made of a hydrocarbon chain that terminates with a carboxylic acid group; this arrangement confers the molecule with a polar, hydrophilic end, and a nonpolar, hydrophobic end that is insoluble in water. The fatty acid structure is one of the most fundamental categories of biological lipids and is commonly used as a building-block of more structurally complex lipids. The carbon chain, typically between four and 24 carbons long, may be saturated or unsaturated, and may be attached to functionalgroups containing oxygen, halogens, nitrogen, and sulfur. If a fatty acid contains a double bond, there is the possibility of either a cis or trans geometric isomerism, which significantly affects the molecule’s configuration. Cis-double bonds cause the fatty acid chain to bend, an effect that is compounded with more double bonds in the chain. Three double bonds in 18-carbon linolenic acid, the most abundant fatty-acyl chains of plant thylakoid membranes, render these membranes highly fluid despite environmental low-temperatures, and also makes linolenic acid give dominating sharp peaks in high resolution 13-C NMR spectra of chloroplasts. This in turn plays an important role in the structure and function of cell membranes. Most naturally occurring fatty acids are of the cis configuration, although the trans form does exist in some natural and partially hydrogenated fats and oils.

Examples of biologically important fatty acids include the eicosanoids, derived primarily from arachidonic acid

³⁴https://en.wikipedia.org/wiki/Gabriel_Bertrand

and eicosapentaenoic acid, that include prostaglandins, leukotrienes, and thromboxanes. Docosahexaenoic acid is also important in biological systems, particularly with respect to sight. Other major lipid classes in the fatty acid category are the fatty esters and fatty amides. Fatty esters include important biochemical intermediates such as wax esters, fatty acid thioester coenzyme A derivatives, fatty acid thioester ACP derivatives and fatty acid carnitines. The fatty amides include N-acyl ethanolamines, such as the cannabinoid neurotransmitter anandamide.

Sterols, such as cholesterol and its derivatives, are an important component of membrane lipids, along with the glycerophospholipids and sphingomyelins. Other examples of sterols are the bile acids and their conjugates, which in mammals are oxidized derivatives of cholesterol and are synthesized in the liver. The plant equivalents are the phytosterols, such as β -sitosterol, stigmasterol, and brassicasterol; the latter compound is also used as a biomarker for algal growth. The predominant sterol in fungal cell membranes is ergosterol.

Sterols are steroids in which one of the hydrogen atoms is substituted with a hydroxyl group, at position 3 in the carbon chain. They have in common with steroids the same fused four-ring core structure. Steroids have different biological roles as hormones and signaling molecules. The eighteen-carbon (C_{18}) steroids include the estrogen family whereas the C_{19} steroids comprise the androgens such as testosterone and androsterone. The C_{21} subclass includes the progestogens as well as the glucocorticoids and mineralocorticoids. The secosteroids, comprising various forms of vitamin D, are characterized by cleavage of the B ring of the core structure.

Eukaryotic cells feature the compartmentalized membrane-bound organelles that carry out different biological functions. The glycerophospholipids are the main structural component of biological membranes, as the cellular plasma membrane and the intracellular membranes of organelles; in animal cells, the plasma membrane physically separates the intracellular components from the extracellular environment. The glycerophospholipids are amphipathic molecules (containing both hydrophobic and hydrophilic regions) that contain a glycerol core linked to two fatty acid-derived "tails" by ester linkages and to one "head" group by a phosphate ester linkage. While glycerophospholipids are the major component of biological membranes, other non-glyceride lipid components such as sphingomyelin and sterols (mainly cholesterol in animal cell membranes) are also found in biological membranes. In plants and algae, the galactosyldiacylglycerols, and sulfoquinovosyldiacylglycerol, which lack a phosphate group, are important components of membranes of chloroplasts and related organelles and are the most

abundant lipids in photosynthetic tissues, including those of higher plants, algae and certain bacteria.

Plant thylakoid membranes have the largest lipid component of a non-bilayer forming monogalactosyl diglyceride (MGDG), and little phospholipids; despite this unique lipid composition, chloroplast thylakoid membranes have been shown to contain a dynamic lipid-bilayer matrix as revealed by magnetic resonance and electron microscope studies.

A biological membrane is a form of lamellar phase lipid bilayer. The formation of lipid bilayers is an energetically preferred process when the glycerophospholipids described above are in an aqueous environment. This is known as the hydrophobic effect. In an aqueous system, the polar heads of lipids align towards the polar, aqueous environment, while the hydrophobic tails minimize their contact with water and tend to cluster together, forming a vesicle; depending on the concentration of the lipid, this biophysical interaction may result in the formation of micelles, liposomes, or lipid bilayers. Other aggregations are also observed and form part of the polymorphism of amphiphile (lipid) behavior. Micelles and bilayers form in the polar medium by a process known as the hydrophobic effect. When dissolving a lipophilic or amphiphilic substance in a polar environment, the polar molecules (i.e., water in an aqueous solution) become more ordered around the dissolved lipophilic substance, since the polar molecules cannot form hydrogen bonds to the lipophilic areas of the amphiphile. So in an aqueous environment, the water molecules form an ordered "clathrate" cage around the dissolved lipophilic molecule.

The formation of lipids into protocell membranes represents a key step in models of abiogenesis, the origin of life.

Triglycerides, stored in adipose tissue, are a major form of energy storage both in animals and plants. They are a major source of energy because carbohydrates are fully reduced structures. In comparison to glycogen which would contribute only half of the energy per its pure mass, triglyceride carbons are all bonded to hydrogens, unlike in carbohydrates. The adipocyte, or fat cell, is designed for continuous synthesis and breakdown of triglycerides in animals, with breakdown controlled mainly by the activation of hormone-sensitive enzyme lipase. The complete oxidation of fatty acids provides high caloric content, about 38 kJ/g (9 kcal/g), compared with 17 kJ/g (4 kcal/g) for the breakdown of carbohydrates and proteins. Migratory birds that must fly long distances without eating use stored energy of triglycerides to fuel their flights.

In recent years, evidence has emerged showing that

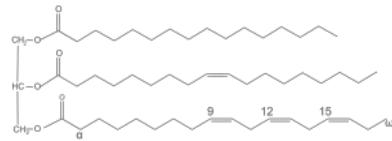


Figure 3.8: Example of an unsaturated fat triglyceride³⁵ ($C_{55}H_{98}O_6$). Left part: glycerol; right part, from top to bottom: palmitic acid, oleic acid, alpha-linolenic acid.

lipid signaling is a vital part of the cell signaling. Lipid signaling may occur via activation of G protein-coupled or nuclear receptors, and members of several different lipid categories have been identified as signaling molecules and cellular messengers. These include sphingosine-1-phosphate, a sphingolipid derived from ceramide that is a potent messenger molecule involved in regulating calcium mobilization, cell growth, and apoptosis; diacylglycerol (DAG) and the phosphatidylinositol phosphates (PIPs), involved in calcium-mediated activation of protein kinase C; the prostaglandins, which are one type of fatty-acid derived eicosanoid involved in inflammation and immunity; the steroid hormones such as estrogen, testosterone and cortisol, which modulate a host of functions such as reproduction, metabolism and blood pressure; and the oxysterols such as 25-hydroxy-cholesterol that are liver X receptor agonists. Phosphatidylserine lipids are known to be involved in signaling for the phagocytosis of apoptotic cells or pieces of cells. They accomplish this by being exposed to the extracellular face of the cell membrane after the inactivation of flippases which place them exclusively on the cytosolic side and the activation of scramblases, which scramble the orientation of the phospholipids. After this occurs, other cells recognize the phosphatidylserines and phagocytose the cells or cell fragments exposing them.

The “fat-soluble” vitamins (A, D, E and K) – which are isoprene-based lipids – are essential nutrients stored in the liver and fatty tissues, with a diverse range of functions. Acyl-carnitines are involved in the transport and metabolism of fatty acids in and out of mitochondria, where they undergo beta oxidation. Polyprenols and their phosphorylated derivatives also play important transport roles, in this case the transport of oligosaccharides across membranes. Polyprenol phosphate sugars and polyprenol diphosphate sugars function in extra-cytoplasmic glycosylation reactions, in extracellular polysaccharide biosynthesis (for instance, peptidoglycan polymerization in bacteria), and in eukaryotic protein N-glycosylation. Cardiolipins are a subclass of glycerophospholipids containing four acyl chains and three glycerol groups that are particularly abundant in the inner mitochondrial membrane. They are believed to activate enzymes involved

with oxidative phosphorylation. Lipids also form the basis of steroid hormones.

3.6 Nucleic Acids

Nucleic acids are the biopolymers, or large biomolecules, essential to all known forms of life. The term nucleic acid is the overall name for DNA and RNA. They are composed of nucleotides, which are the monomers made of three components: a 5-carbon sugar, a phosphate group and a nitrogenous base. If the sugar is a compound ribose, the polymer is RNA (ribonucleic acid); if the sugar is derived from ribose as deoxyribose, the polymer is DNA (deoxyribonucleic acid).

Nucleic acids are the most important of all biomolecules. These are found in abundance in all living things, where they function to create and encode and then store information of every living cell of every life-form organism on Earth. In turn, they function to transmit and express that information inside and outside the cell nucleus—to the interior operations of the cell and ultimately to the next generation of each living organism. The encoded information is contained and conveyed via the nucleic acid sequence, which provides the ‘ladder-step’ ordering of nucleotides within the molecules of RNA and DNA.

Strings of nucleotides are bonded to form helical backbones—typically, one for RNA, two for DNA—and assembled into chains of base-pairs selected from the five primary, or canonical, nucleobases, which are: adenine, cytosine, guanine, thymine, and uracil. Thymine occurs only in DNA and uracil only in RNA. Using amino acids and the process known as protein synthesis, the specific sequencing in DNA of these nucleobase-pairs enables storing and transmitting coded instructions as genes. In RNA, base-pair sequencing provides for manufacturing new proteins that determine the frames and parts and most chemical processes of all life forms

The term nucleic acid is the overall name for DNA and RNA, members of a family of biopolymers, and is synonymous with polynucleotide. Nucleic acids were named for their initial discovery within the nucleus, and for the presence of phosphate groups (related to phosphoric acid). Although first discovered within the nucleus of eukaryotic cells, nucleic acids are now known to be found in all life forms including within bacteria, archaea, mitochondria, chloroplasts, and viruses. All living cells contain both DNA and RNA (except some cells such as mature red blood cells), while viruses contain either DNA or RNA, but usually not both. The basic component of biological nucleic acids is the nucleotide, each of which contains a pentose sugar (ribose or deoxyribose), a phosphate group,

and a nucleobase. Nucleic acids are also generated within the laboratory, through the use of enzymes (DNA and RNA polymerases) and by solid-phase chemical synthesis. The chemical methods also enable the generation of altered nucleic acids that are not found in nature, for example peptide nucleic acids.

Nucleic acids are generally very large molecules. Indeed, DNA molecules are probably the largest individual molecules known. Well-studied biological nucleic acid molecules range in size from 21 nucleotides (small interfering RNA) to large chromosomes (human chromosome 1 is a single molecule that contains 247 million base pairs).

In most cases, naturally occurring DNA molecules are double-stranded and RNA molecules are single-stranded. There are numerous exceptions, however—some viruses have genomes made of double-stranded RNA and other viruses have single-stranded DNA genomes, and, in some circumstances, nucleic acid structures with three or four strands can form.

Nucleic acids are linear polymers (chains) of nucleotides. Each nucleotide consists of three components: a purine or pyrimidine nucleobase (sometimes termed nitrogenous base or simply base), a pentose sugar, and a phosphate group. The substructure consisting of a nucleobase plus sugar is termed a nucleoside. Nucleic acid types differ in the structure of the sugar in their nucleotides—DNA contains 2'-deoxyribose while RNA contains ribose (where the only difference is the presence of a hydroxyl group). Also, the nucleobases found in the two nucleic acid types are different: adenine, cytosine, and guanine are found in both RNA and DNA, while thymine occurs in DNA and uracil occurs in RNA.

The sugars and phosphates in nucleic acids are connected to each other in an alternating chain (sugar-phosphate backbone) through phosphodiester linkages. In conventional nomenclature, the carbons to which the phosphate groups attach are the 3'-end and the 5'-end carbons of the sugar. This gives nucleic acids directionality, and the ends of nucleic acid molecules are referred to as 5'-end and 3'-end. The nucleobases are joined to the sugars via an N-glycosidic linkage involving a nucleobase ring nitrogen (N-1 for pyrimidines and N-9 for purines) and the 1' carbon of the pentose sugar ring.

3.6.1 Deoxyribonucleic Acid (DNA)

Deoxyribonucleic acid (DNA) is a nucleic acid containing the genetic instructions used in the development and functioning of all known living organisms. The DNA segments carrying this genetic information are called genes. Likewise, other DNA sequences have structural purposes or are involved in regulating the use of this

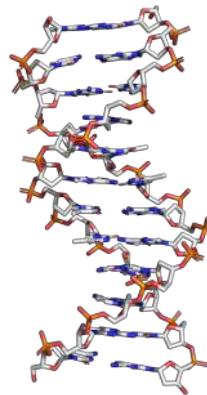


Figure 3.9: The structure of the DNA double helix. A section of DNA. The bases lie horizontally between the two spiraling strands. The atoms in the structure are colour-coded by element (based on atomic coordinates of PDB 1bna³⁶ rendered with open source molecular visualization tool PyMol.)

genetic information. Along with RNA and proteins, DNA is one of the three major macromolecules that are essential for all known forms of life. DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds. These two strands run in opposite directions to each other and are, therefore, anti-parallel. Attached to each sugar is one of four types of molecules called nucleobases (informally, bases). It is the sequence of these four nucleobases along the backbone that encodes information. This information is read using the genetic code, which specifies the sequence of the amino acids within proteins. The code is read by copying stretches of DNA into the related nucleic acid RNA in a process called transcription. Within cells, DNA is organized into long structures called chromosomes. During cell division these chromosomes are duplicated in the process of DNA replication, providing each cell its own complete set of chromosomes. Eukaryotic organisms (animals, plants, fungi, and protists) store most of their DNA inside the cell nucleus and some of their DNA in organelles, such as mitochondria or chloroplasts. In contrast, prokaryotes (bacteria and archaea) store their DNA only in the cytoplasm. Within the chromosomes, chromatin proteins such as histones compact and organize DNA. These compact structures guide the interactions between DNA and other proteins, helping control which parts of the DNA are transcribed.

3.6.2 Ribonucleic Acid (RNA)

Ribonucleic acid (RNA) functions in converting genetic information from genes into the amino acid sequences of proteins. The three universal types of RNA include transfer RNA (tRNA), messenger RNA (mRNA), and ribosomal RNA (rRNA). Messenger RNA acts to carry genetic sequence information between DNA and ribosomes, directing protein synthesis. Ribosomal RNA is a major component of the ribosome, and catalyzes peptide bond formation. Transfer RNA serves as the carrier molecule for amino acids to be used in protein synthesis, and is responsible for decoding the mRNA. In addition, many other classes of RNA are now known.

Chapter 4

The Cell

The cell¹ (from Latin *cella*, meaning “small room”) is the basic structural, functional, and biological unit of all known organisms. A cell is the smallest unit of life. Cells are often called the “building blocks of life”. The study of cells is called cell biology, cellular biology, or cytology.

Cells consist of cytoplasm enclosed within a membrane, which contains many biomolecules such as proteins and nucleic acids. Most plant and animal cells are only visible under a microscope, with dimensions between 1 and 100 micrometres. Organisms can be classified as unicellular (consisting of a single cell such as bacteria) or multicellular (including plants and animals). Most unicellular organisms are classed as microorganisms.

The number of cell in plants and animals varies from species to species; it has been estimated that humans contain somewhere around 40 trillion (4×10^{13}) cells. The human brain accounts for around 80 billion of these cells.

In biology, cell theory is the historic scientific theory, now universally accepted, that living organisms are made up of cells, that they are the basic structural/organizational unit of all organisms, and that all cells come from pre-existing cells. Cells are the basic unit of structure in all organisms and also the basic unit of reproduction.

Credit for developing cell theory is usually given to two scientists: Theodor Schwann² and Matthias Jakob Schleiden³. While Rudolf Virchow⁴ contributed to the theory, he is not as credited for his attributions toward it. In 1839, Schleiden suggested that every structural part of a plant was made up of cells or the result of cells. He also suggested that cells were made by a crystallization process either within other cells or from the outside. However, this was not an original idea of Schleiden. He claimed this theory as his own, though Barthélemy Dumortier⁵ had stated it years before him. This crystallization process is no

longer accepted with modern cell theory. In 1839, Theodor Schwann states that along with plants, animals are composed of cells or the product of cells in their structures. This was a major advancement in the field of biology since little was known about animal structure up to this point compared to plants. From these conclusions about plants and animals, two of the three tenets of cell theory were postulated.

1. All living organisms are composed of one or more cells
2. The cell is the most basic unit of life

Schleiden's theory of free cell formation through crystallization was refuted in the 1850s by Robert Remak⁶, Rudolf Virchow⁷, and Albert von Kölliker⁸. In 1855, Rudolf Virchow added the third tenet to cell theory. In Latin, this tenet states *Omnis cellula e cellula*. This translated to:

3. All cells arise only from pre-existing cells

However, the idea that all cells come from pre-existing cells had in fact already been proposed by Robert Remak; it has been suggested that Virchow plagiarized Remak and did not give him credit. Remak published observations in 1852 on cell division, claiming Schleiden and Schwann were incorrect about generation schemes. He instead said that binary fission, which was first introduced by Dumortier, was how reproduction of new animal cells were made. Once this tenet was added, the classical cell theory was complete.

The generally accepted parts of modern cell theory include:

1. All living cells arise from pre-existing cells by division.
2. The cell is the fundamental unit of structure and function in all living organisms.
3. The activity of an organism depends on the total activity of independent cells.

¹[https://en.wikipedia.org/wiki/Cell_\(biology\)](https://en.wikipedia.org/wiki/Cell_(biology))

²https://en.wikipedia.org/wiki/Theodor_Schwann

³https://en.wikipedia.org/wiki/Matthias_Jakob_Schleiden

⁴https://en.wikipedia.org/wiki/Rudolf_Virchow

⁵https://en.wikipedia.org/wiki/Barthélémy_Charles_Joseph_Dumortier

⁶https://en.wikipedia.org/wiki/Robert_Remak

⁷https://en.wikipedia.org/wiki/Rudolf_Virchow

⁸https://en.wikipedia.org/wiki/Albert_von_Kölliker

4. Energy flow (metabolism and biochemistry) occurs within cells.
5. Cells contain DNA which is found specifically in the chromosome and RNA found in the cell nucleus and cytoplasm.
6. All cells are basically the same in chemical composition in organisms of similar species.

The discovery of the cell was made possible through the invention of the microscope. In the first century BC, Romans were able to make glass, discovering that objects appeared to be larger under the glass. In Italy during the 12th century, Salvino D'Armata made a piece of glass fit over one eye, allowing for a magnification effect to that eye. The expanded use of lenses in eyeglasses in the 13th century probably led to wider spread use of simple microscopes (magnifying glasses) with limited magnification. Compound microscopes, which combine an objective lens with an eyepiece to view a real image achieving much higher magnification, first appeared in Europe around 1620. In 1665, Robert Hooke⁹ used a microscope about six inches long with two convex lenses inside and examined specimens under reflected light for the observations in his book Micrographia¹⁰. Hooke also used a simpler microscope with a single lens for examining specimens with directly transmitted light, because this allowed for a clearer image.

Extensive microscopic study was done by Anton van Leeuwenhoek¹⁴, a draper who took the interest in microscopes after seeing one while on an apprenticeship in Amsterdam in 1648. At some point in his life before 1668, he was able to learn how to grind lenses. This eventually led to Leeuwenhoek making his own unique microscope. His was a single lens simple microscope, rather than a compound microscope. This was because he was able to use a single lens that was a small glass sphere but allowed for a magnification of 270x. This was a large progression since the magnification before was only a maximum of 50x. After Leeuwenhoek, there was not much progress in microscope technology until the 1850s, two hundred years later. Carl Zeiss, a German engineer who manufactured microscopes, began to make changes to the lenses used. But the optical quality did not improve until the 1880s when he hired Otto Schott and eventually Ernst Abbe.

Optical microscopes can focus on objects the size of a wavelength or larger, giving restrictions still to advancement in discoveries with objects smaller than the wavelengths of visible light. The development of the electron microscope in the 1920s made it possible to view objects that are smaller than optical wavelengths, once again open-

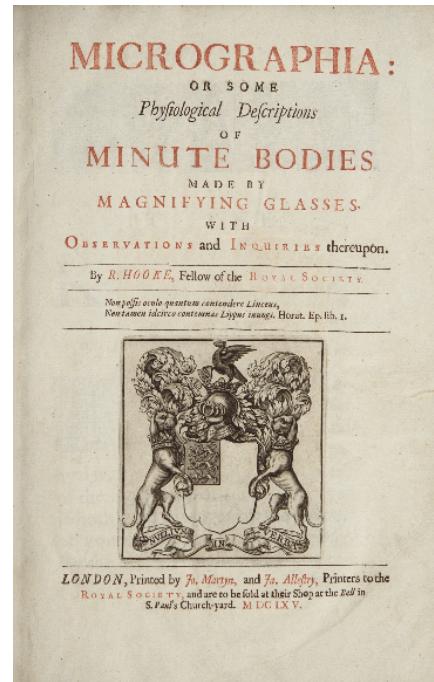


Figure 4.1: Title page¹¹ of “MICROGRAPHIA or some physiological descriptions of minute bodies made by magnifying glasses with observations and inquiries thereupon”.

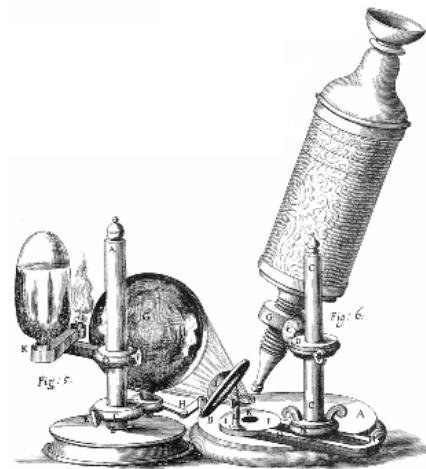


Figure 4.2: Hooke's microscope, from an engraving in Micrographia¹²

⁹https://en.wikipedia.org/wiki/Robert_Hooke

¹⁰<https://en.wikipedia.org/wiki/Micrographia>

¹⁴https://en.wikipedia.org/wiki/Antonie_van_Leeuwenhoek

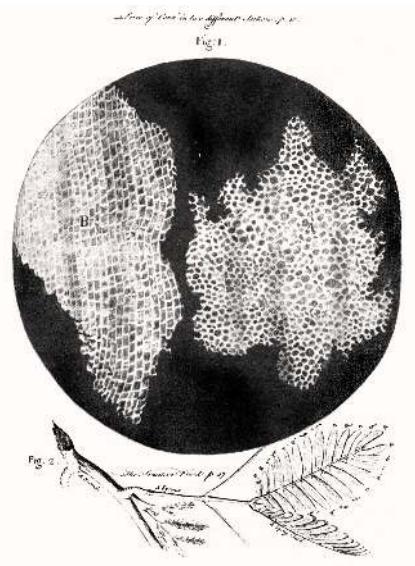


Figure 4.3: Hooke was the first to apply the word “cell” to biological objects. Cell structure of cork by Hooke¹³.

ing up new possibilities in science.

The cell was first discovered by Robert Hooke in 1665, which can be found to be described in his book *Micrographia*. In this book, he gave 60 ‘observations’ in detail of various objects under a coarse, compound microscope. One observation was from very thin slices of bottle cork. Hooke discovered a multitude of tiny pores that he named “cells”. This came from the Latin word *Cella*, meaning ‘a small room’ like monks lived in and also *Cellulae*, which meant the six sided cell of a honeycomb. However, Hooke did not know their real structure or function. What Hooke had thought were cells, were actually empty cell walls of plant tissues. With microscopes during this time having a low magnification, Hooke was unable to see that there were other internal components to the cells he was observing. Therefore, he did not think the “cellulae” were alive. His cell observations gave no indication of the nucleus and other organelles found in most living cells. In *Micrographia*, Hooke also observed mould, bluish in color, found on leather. After studying it under his microscope, he was unable to observe “seeds” that would have indicated how the mould was multiplying in quantity. This led to Hooke suggesting that spontaneous generation, from either natural or artificial heat, was the cause. Since this was an old Aristotelian theory still accepted at the time, others did not reject it and was not disproved until Leeuwenhoek later discovered that generation was achieved otherwise.

Anton van Leeuwenhoek¹⁵ is another scientist who saw these cells soon after Hooke did. He made use of a micro-

scope containing improved lenses that could magnify objects almost 300-fold, or 270x. Under these microscopes, Leeuwenhoek found motile objects. In a letter to The Royal Society on October 9, 1676, he states that motility is a quality of life therefore these were living organisms. Over time, he wrote many more papers in which described many specific forms of microorganisms. Leeuwenhoek named these “animalcules,” which included protozoa and other unicellular organisms, like bacteria. Though he did not have much formal education, he was able to identify the first accurate description of red blood cells and discovered bacteria after gaining interest in the sense of taste that resulted in Leeuwenhoek to observe the tongue of an ox, then leading him to study “pepper water” in 1676. He also found for the first time the sperm cells of animals and humans. Once discovering these types of cells, Leeuwenhoek saw that the fertilization process requires the sperm cell to enter the egg cell. This put an end to the previous theory of spontaneous generation. After reading letters by Leeuwenhoek, Hooke was the first to confirm his observations that were thought to be unlikely by other contemporaries.

The cells in animal tissues were observed after plants were because the tissues were so fragile and susceptible to tearing, it was difficult for such thin slices to be prepared for studying. Biologists believed that there was a fundamental unit to life, but were unsure what this was. It would not be until over a hundred years later that this fundamental unit was connected to cellular structure and existence of cells in animals or plants. This conclusion was not made until Henri Dutrochet. Besides stating “the cell is the fundamental element of organization”, Dutrochet also claimed that cells were not just a structural unit, but also a physiological unit.

In 1804, Karl Rudolphi¹⁶ and Johann Heinrich Friedrich Link¹⁷ were awarded the prize for “solving the problem of the nature of cells”, meaning they were the first to prove that cells had independent cell walls by the Königliche Societät der Wissenschaft (Royal Society of Science), Göttingen. Before, it had been thought that cells shared walls and the fluid passed between them this way.

Cells can be subdivided into the following subcategories:

1. Prokaryotes: Prokaryotes are relatively small cells surrounded by the plasma membrane, with a characteristic cell wall that may differ in composition depending on the particular organism. Prokaryotes lack a nucleus (although they do have circular or linear DNA) and other membrane-bound organelles (though they do contain ribosomes). The protoplasm of a prokaryote contains the chromosomal

¹⁶https://en.wikipedia.org/wiki/Karl_Rudolphi

¹⁷https://en.wikipedia.org/wiki/Johann_Heinrich_Friedrich_Link

region that appears as fibrous deposits under the microscope, and the cytoplasm. Bacteria and Archaea are the two domains of prokaryotes.

- Eukaryotes: Eukaryotic cells are also surrounded by the plasma membrane, but on the other hand, they have distinct nuclei bound by a nuclear membrane or envelope. Eukaryotic cells also contain membrane-bound organelles, such as (mitochondria, chloroplasts, lysosomes, rough and smooth endoplasmic reticulum, vacuoles). In addition, they possess organized chromosomes which store genetic material. Animals have evolved a greater diversity of cell types in a multicellular body (100–150 different cell types), compared with 10–20 in plants, fungi, and protocista.

The distinction between prokaryotes and eukaryotes was firmly established by the microbiologists Roger Stanier¹⁸ and C. B. van Niel¹⁹ in their 1962 paper The concept of a bacterium (though spelled procaryote and eucaryote there). That paper cites Édouard Chatton's 1937 book *Titres et Travaux Scientifiques* for using those terms and recognizing the distinction. One reason for this classification was so that what was then often called blue-green algae (now called cyanobacteria) would not be classified as plants but grouped with bacteria.

4.1 Prokaryotic Cells

A prokaryote is a cellular organism that lacks an envelope-enclosed nucleus. The word prokaryote comes from the Greek πρό (pro, 'before') and κάρυον (karyon, 'nut' or 'kernel'). In the two-empire system arising from the work of Édouard Chatton, prokaryotes were classified within the empire Prokaryota. But in the three-domain system, based upon molecular analysis, prokaryotes are divided into two domains: Bacteria (formerly Eubacteria) and Archaea (formerly Archaebacteria). Organisms with nuclei are placed in a third domain, Eukaryota. In the study of the origins of life, prokaryotes are thought to have arisen before eukaryotes.

Prokaryotes lack mitochondria, or any other eukaryotic membrane-bound organelles; and it was once thought that prokaryotes lacked cellular compartments, and therefore all cellular components within the cytoplasm were unenclosed, except for an outer cell membrane. But bacterial microcompartments, which are thought to be primitive organelles enclosed in protein shells, have been discovered; and there is also evidence of prokaryotic membrane-bound organelles. While typically being unicellular, some prokaryotes, such as cyanobacteria, may

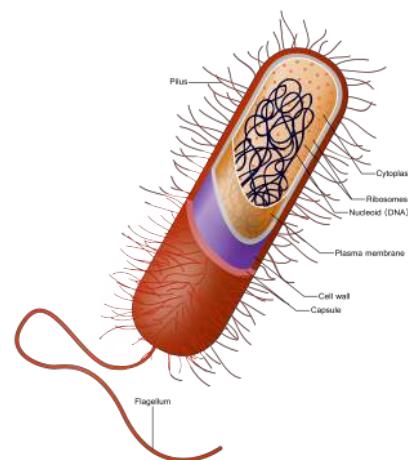


Figure 4.4: Structure of a typical prokaryotic cell²⁰

form large colonies. Others, such as myxobacteria, have multicellular stages in their life cycles. Prokaryotes are asexual, reproducing without fusion of gametes, although horizontal gene transfer also takes place.

Molecular studies have provided insight into the evolution and interrelationships of the three domains of life. The division between prokaryotes and eukaryotes reflects the existence of two very different levels of cellular organization; only eukaryotic cells have an enveloped nucleus that contains its chromosomal DNA, and other characteristic membrane-bound organelles including mitochondria. Distinctive types of prokaryotes include extremophiles and methanogens; these are common in some extreme environments.

Prokaryotes include bacteria and archaea, two of the three domains of life. Prokaryotic cells were the first form of life on Earth, characterized by having vital biological processes including cell signaling. They are simpler and smaller than eukaryotic cells, and lack a nucleus, and other membrane-bound organelles. The DNA of a prokaryotic cell consists of a single circular chromosome that is in direct contact with the cytoplasm. The nuclear region in the cytoplasm is called the nucleoid. Most prokaryotes are the smallest of all organisms ranging from 0.5 to 2.0 μm in diameter.

A prokaryotic cell has three regions:

- Enclosing the cell is the cell envelope – generally consisting of a plasma membrane covered by a cell wall which, for some bacteria, may be further covered by a third layer called a capsule. Though most prokaryotes have both a cell membrane and a cell wall, there are exceptions such as Mycoplasma (bacteria) and Thermoplasma (archaea) which only possess the cell

¹⁸https://en.wikipedia.org/wiki/Roger_Stanier

¹⁹https://en.wikipedia.org/wiki/C._B._van_Niel

membrane layer. The envelope gives rigidity to the cell and separates the interior of the cell from its environment, serving as a protective filter. The cell wall consists of peptidoglycan in bacteria, and acts as an additional barrier against exterior forces. It also prevents the cell from expanding and bursting (cytolysis) from osmotic pressure due to a hypotonic environment. Some eukaryotic cells (plant cells and fungal cells) also have a cell wall.

- Inside the cell is the cytoplasmic region that contains the genome (DNA), ribosomes and various sorts of inclusions. The genetic material is freely found in the cytoplasm. Prokaryotes can carry extrachromosomal DNA elements called plasmids, which are usually circular. Linear bacterial plasmids have been identified in several species of spirochete bacteria, including members of the genus *Borrelia* notably *Borrelia burgdorferi*, which causes Lyme disease. Though not forming a nucleus, the DNA is condensed in a nucleoid. Plasmids encode additional genes, such as antibiotic resistance genes.
- On the outside, flagella and pili project from the cell's surface. These are structures (not present in all prokaryotes) made of proteins that facilitate movement and communication between cells.

4.2 Eukaryotic Cells

Eukaryotes²¹ are organisms whose cells have a nucleus enclosed within a nuclear envelope. Eukaryotes belong to the domain Eukaryota or Eukarya; their name comes from the Greek εὖ (eu, "well" or "good") and κάρυον (karyon, "nut" or "kernel"). The domain Eukaryota makes up one of the domains of life in the three-domain system; the two other domains are Bacteria and Archaea (together known as prokaryotes). Eukaryotes represent a tiny minority of the number of living organisms; however, due to their generally much larger size, their collective worldwide biomass is estimated to be about equal to that of prokaryotes. Eukaryotes evolved approximately 1.6–2.1 billion years ago, during the Proterozoic eon.

Eukaryotic cells typically contain membrane-bound organelles such as mitochondria and Golgi apparatus, and chloroplasts can be found in plants and algae; these organelles are unique to eukaryotes, although primitive organelles can be found in prokaryotes. As well as being unicellular, eukaryotes may also be multicellular and include many cell types forming different kinds of tissue; in comparison, prokaryotes are typically unicellular. Animals, plants, and fungi are the most familiar eukaryotes; other

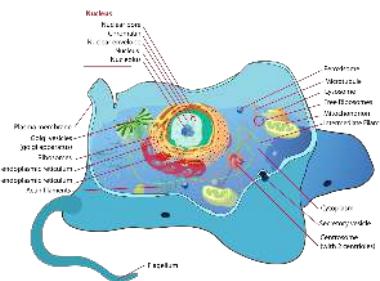


Figure 4.5: Cartoon of the structure of a typical animal cell²²

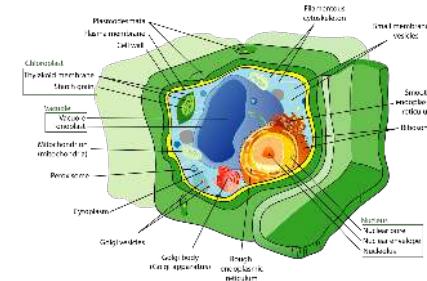


Figure 4.6: Cartoon of the structure of a typical plant cell²³

eukaryotes are sometimes called protists.

Eukaryotes can reproduce both asexually through mitosis and sexually through meiosis and gamete fusion. In mitosis, one cell divides to produce two genetically identical cells. In meiosis, DNA replication is followed by two rounds of cell division to produce four haploid daughter cells. These act as sex cells (gametes). Each gamete has just one set of chromosomes, each a unique mix of the corresponding pair of parental chromosomes resulting from genetic recombination during meiosis.

Plants, animals, fungi, and protists are all eukaryotic. These cells are about fifteen times wider than a typical prokaryote and can be as much as a thousand times greater in volume. The main distinguishing feature of eukaryotes as compared to prokaryotes is compartmentalization: the presence of membrane-bound organelles (compartments) in which specific activities take place. Most important among these is a cell nucleus, an organelle that houses the cell's DNA. This nucleus gives the eukaryote its name, which means "true kernel (nucleus)". Other differences include:

- The eukaryotic DNA is organized in one or more linear molecules, called chromosomes, which are associated with histone proteins. All chromosomal DNA is stored in the cell nucleus, separated from the cytoplasm by a membrane. Some eukaryotic organelles

²¹<https://en.wikipedia.org/wiki/Eukaryote>

- such as mitochondria also contain some DNA.
- Many eukaryotic cells are ciliated with primary cilia. Primary cilia play important roles in chemosensation, mechanosensation, and thermosensation. Each cilium may thus be “viewed as a sensory cellular antennae that coordinates a large number of cellular signaling pathways, sometimes coupling the signaling to ciliary motility or alternatively to cell division and differentiation.”
 - Motile eukaryotes can move using motile cilia or flagella. Motile cells are absent in conifers and flowering plants. Eukaryotic flagella are more complex than those of prokaryotes.

Table 4.1: Comparison of the main features of prokaryotes and eukaryotes.

	Prokaryotes	Eukaryotes
Typical organisms	bacteria, archaea	protists; fungi; plants; animals
Typical size	~1–5 µm	~10–100 µm
Type of nucleus	nucleoid region; no true nucleus	true nucleus with double membrane
DNA	circular (usually)	linear molecules (chromosomes) with histone proteins
RNA/protein synthesis	coupled in the cytoplasm	RNA synthesis in the nucleus; protein synthesis in the cytoplasm
Ribosomes	50S and 30S	60S and 40S
Cytoplasmic structure	very few structures	highly structured by endomembranes and a cytoskeleton
Cell movement	flagella made of flagellin	flagella and cilia containing microtubules; lamellipodia and filopodia containing actin
Mitochondria	none	one to several thousand
Chloroplasts	none	in algae and plants
Organization	usually single cells	single cells, colonies, higher multicellular organisms with specialized cells
Cell division	binary fission (simple division)	mitosis (fission or budding); meiosis
Chromosomes	single chromosome	more than one chromosome
Membranes	cell membrane	Cell membrane and membrane-bound organelles

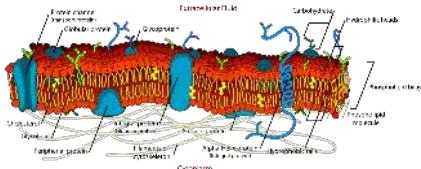


Figure 4.7: Detailed diagram of lipid bilayer cell membrane²⁴

All cells, whether prokaryotic or eukaryotic, have a membrane that envelops the cell, regulates what moves in and out (selectively permeable), and maintains the electric potential of the cell. Inside the membrane, the cytoplasm takes up most of the cell's volume. All cells (except red blood cells which lack a cell nucleus and most organelles to accommodate maximum space for hemoglobin) possess DNA, the hereditary material of genes, and RNA, containing the information necessary to build various proteins such as enzymes, the cell's primary machinery. There are also other kinds of biomolecules in cells. This article lists these primary cellular components, then briefly describes their function.

4.3 The Cell Membrane

The cell membrane, or plasma membrane, is a biological membrane that surrounds the cytoplasm of a cell. In animals, the plasma membrane is the outer boundary of the cell, while in plants and prokaryotes it is usually covered by a cell wall. This membrane serves to separate and protect a cell from its surrounding environment and is made mostly from a double layer of phospholipids, which are amphiphilic (partly hydrophobic and partly hydrophilic). Hence, the layer is called a phospholipid bilayer, or sometimes a fluid mosaic membrane. Embedded within this membrane is a macromolecular structure called the porosome the universal secretory portal in cells and a variety of protein molecules that act as channels and pumps that move different molecules into and out of the cell. The membrane is semi-permeable, and selectively permeable, in that it can either let a substance (molecule or ion) pass through freely, pass through to a limited extent or not pass through at all. Cell surface membranes also contain receptor proteins that allow cells to detect external signaling molecules such as hormones.

4.4 The Cytoskeleton

The cytoskeleton is a complex, dynamic network of inter-linking protein filaments present in the cytoplasm of all cells, including bacteria and archaea. It extends from the cell nucleus to the cell membrane and is composed of sim-

ilar proteins in the various organisms. In eukaryotes, it is composed of three main components, microfilaments, intermediate filaments and microtubules, and these are all capable of rapid growth or disassembly dependent on the cell's requirements.

A multitude of functions can be performed by the cytoskeleton. Its primary function is to give the cell its shape and mechanical resistance to deformation, and through association with extracellular connective tissue and other cells it stabilizes entire tissues. The cytoskeleton can also contract, thereby deforming the cell and the cell's environment and allowing cells to migrate. Moreover, it is involved in many cell signaling pathways and in the uptake of extracellular material (endocytosis), the segregation of chromosomes during cellular division, the cytokinesis stage of cell division, as scaffolding to organize the contents of the cell in space and in intracellular transport (for example, the movement of vesicles and organelles within the cell) and can be a template for the construction of a cell wall. Furthermore, it can form specialized structures, such as flagella, cilia, lamellipodia and podosomes. The structure, function and dynamic behavior of the cytoskeleton can be very different, depending on organism and cell type. Even within one cell, the cytoskeleton can change through association with other proteins and the previous history of the network.

A large-scale example of an action performed by the cytoskeleton is muscle contraction. This is carried out by groups of highly specialized cells working together. A main component in the cytoskeleton that helps show the true function of this muscle contraction is the microfilament. Microfilaments are composed of the most abundant cellular protein known as actin. During contraction of a muscle, within each muscle cell, myosin molecular motors collectively exert forces on parallel actin filaments. Muscle contraction starts from nerve impulses which then causes increased amounts of calcium to be released from the sarcoplasmic reticulum. Increases in calcium in the cytosol allows muscle contraction to begin with the help of two proteins, tropomyosin and troponin. Tropomyosin inhibits the interaction between actin and myosin, while troponin senses the increase in calcium and releases the inhibition. This action contracts the muscle cell, and through the synchronous process in many muscle cells, the entire muscle.

Eukaryotic cells contain three main kinds of cytoskeletal filaments: microfilaments, microtubules, and intermediate filaments. Each type is formed by the polymerization of a distinct type of protein subunit and has its own characteristic shape and intracellular distribution. Microfilaments are polymers of the protein actin and are 7 nm in diameter. Microtubules are composed of tubulin and are 25 nm

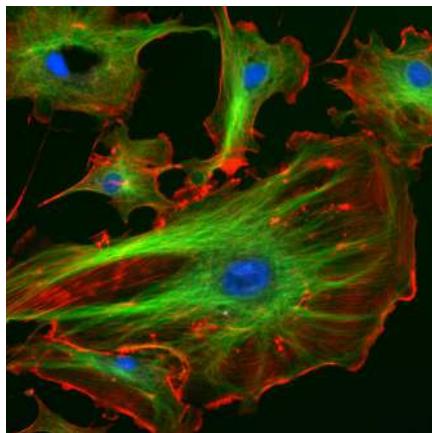


Figure 4.8: The eukaryotic cytoskeleton.²⁵ Actin filaments are shown in red, and microtubules composed of beta tubulin are in green.

in diameter. Intermediate filaments are composed of various proteins, depending on the type of cell in which they are found; they are normally 8–12 nm in diameter. The cytoskeleton provides the cell with structure and shape, and by excluding macromolecules from some of the cytosol, it adds to the level of macromolecular crowding in this compartment. Cytoskeletal elements interact extensively and intimately with cellular membranes.

Microfilaments, also known as actin filaments, are composed of linear polymers of G-actin proteins, and generate force when the growing (plus) end of the filament pushes against a barrier, such as the cell membrane. They also act as tracks for the movement of myosin molecules that affix to the microfilament and “walk” along them. In general, the major component or protein of microfilaments are actin. The G-actin monomer combines to form a polymer which continues to form the microfilament (actin filament). These subunits then assemble into two chains that intertwine into what are called F-actin chains. Myosin motoring along F-actin filaments generates contractile forces in so-called actomyosin fibers, both in muscle as well as most non-muscle cell types.

Intermediate filaments are a part of the cytoskeleton of many eukaryotic cells. These filaments, averaging 10 nanometers in diameter, are more stable (strongly bound) than microfilaments, and heterogeneous constituents of the cytoskeleton. Like actin filaments, they function in the maintenance of cell-shape by bearing tension (microtubules, by contrast, resist compression but can also bear tension during mitosis and during the positioning of the centrosome). Intermediate filaments organize the internal tridimensional structure of the cell, anchoring organelles and serving as structural components of the nuclear lamina. They also participate in some cell-cell and cell-matrix

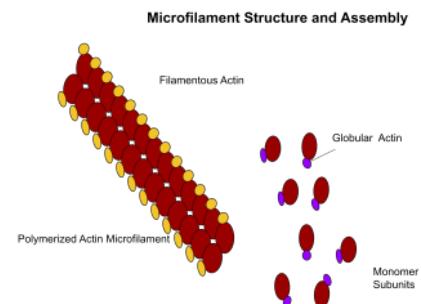


Figure 4.9: Cartoon of the structure of a microfilament, which creates the structure in the cytoskeleton of an eukaryotic cell.²⁶

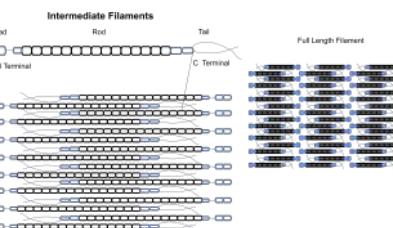


Figure 4.10: Cartoon of the structure of an intermediate filament.²⁷

junctions. Nuclear lamina exist in all animals and all tissues. Some animals like the fruit fly do not have any cytoplasmic intermediate filaments. In those animals that express cytoplasmic intermediate filaments, these are tissue specific. Keratin intermediate filaments in epithelial cells provide protection for different mechanical stresses the skin may endure. They also provide protection for organs against metabolic, oxidative, and chemical stresses. Strengthening of epithelial cells with these intermediate filaments may prevent onset of apoptosis, or cell death, by reducing the probability of stress.

Microtubules are hollow cylinders about 23 nm in diameter (lumen diameter of approximately 15 nm), most commonly comprising 13 protofilaments that, in turn, are polymers of alpha and beta tubulin. They have a very dynamic behavior, binding GTP for polymerization. They are commonly organized by the centrosome.

In nine triplet sets (star-shaped), they form the centrioles, and in nine doublets oriented about two additional microtubules (wheel-shaped), they form cilia and flagella. The latter formation is commonly referred to as a “9+2” arrangement, wherein each doublet is connected to another by the protein dynein. As both flagella and cilia are structural components of the cell, and are maintained by microtubules, they can be considered part of the cytoskeleton. There are two types of cilia: motile and non-motile

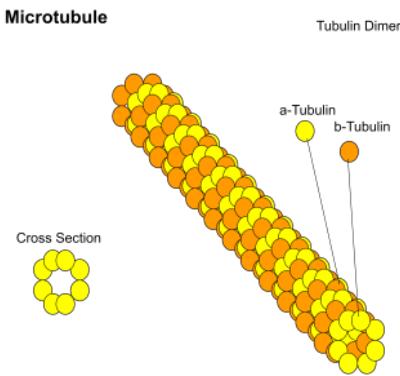


Figure 4.11: Cartoon of the the structure of a microtubule.²⁸

cilia. Cilia are short and more numerous than flagella. The motile cilia have a rhythmic waving or beating motion compared to the non-motile cilia which receive sensory information for the cell; processing signals from the other cells or the fluids surrounding it. Additionally, the microtubules control the beating (movement) of the cilia and flagella. Also, the dynein arms attached to the microtubules function as the molecular motors. The motion of the cilia and flagella is created by the microtubules sliding past one another, which requires ATP. Cytoplasmic streaming, also known as cyclosis, is the active movement of a cell's contents along the components of the cytoskeleton. While mainly seen in plants, all cell types use this process for transportation of waste, nutrients, and organelles to other parts of the cell. Plant and algae cells are generally larger than many other cells; so cytoplasmic streaming is important in these types of cells. This is because the cell's extra volume requires cytoplasmic streaming in order to move organelles throughout the entire cell. Organelles move along microfilaments in the cytoskeleton driven by myosin motors binding and pushing along actin filament bundles.

4.5 The Genetic Material

Two different kinds of genetic material exist: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Cells use DNA for their long-term information storage. The biological information contained in an organism is encoded in its DNA sequence. RNA is used for information transport (e.g., mRNA) and enzymatic functions (e.g., ribosomal RNA). Transfer RNA (tRNA) molecules are used to add amino acids during protein translation.

Prokaryotic genetic material is organized in a simple circular bacterial chromosome in the nucleoid region of the cytoplasm. Eukaryotic genetic material is divided into different, linear molecules called chromosomes inside a dis-

crete nucleus, usually with additional genetic material in some organelles like mitochondria and chloroplasts (see endosymbiotic theory).

A human cell has genetic material contained in the cell nucleus (the nuclear genome) and in the mitochondria (the mitochondrial genome). In humans the nuclear genome is divided into 46 linear DNA molecules called chromosomes, including 22 homologous chromosome pairs and a pair of sex chromosomes. The mitochondrial genome is a circular DNA molecule distinct from the nuclear DNA. Although the mitochondrial DNA is very small compared to nuclear chromosomes, it codes for 13 proteins involved in mitochondrial energy production and specific tRNAs.

Foreign genetic material (most commonly DNA) can also be artificially introduced into the cell by a process called transfection. This can be transient, if the DNA is not inserted into the cell's genome, or stable, if it is. Certain viruses also insert their genetic material into the genome.

4.6 Organelles of Eukaryotic Cells

Organelles are parts of the cell which are adapted and/or specialized for carrying out one or more vital functions, analogous to the organs of the human body (such as the heart, lung, and kidney, with each organ performing a different function). Both eukaryotic and prokaryotic cells have organelles, but prokaryotic organelles are generally simpler and are not membrane-bound.

There are several types of organelles in a cell. Some (such as the nucleus and golgi apparatus) are typically solitary, while others (such as mitochondria, chloroplasts, peroxisomes and lysosomes) can be numerous (hundreds to thousands). The cytosol is the gelatinous fluid that fills the cell and surrounds the organelles.

- Cell nucleus: A cell's information center, the cell nucleus is the most conspicuous organelle found in a eukaryotic cell. It houses the cell's chromosomes, and is the place where almost all DNA replication and RNA synthesis (transcription) occur. The nucleus is spherical and separated from the cytoplasm by a double membrane called the nuclear envelope. The nuclear envelope isolates and protects a cell's DNA from various molecules that could accidentally damage its structure or interfere with its processing. During processing, DNA is transcribed, or copied into a special RNA, called messenger RNA (mRNA). This mRNA is then transported out of the nucleus, where it is translated into a specific protein molecule. The nucleolus is a specialized region within the nucleus where ribosome subunits are assembled. In prokaryotes, DNA processing takes place in the cy-

- toplasm.
- Mitochondria and Chloroplasts: generate energy for the cell. Mitochondria are self-replicating organelles that occur in various numbers, shapes, and sizes in the cytoplasm of all eukaryotic cells. Respiration occurs in the cell mitochondria, which generate the cell's energy by oxidative phosphorylation, using oxygen to release energy stored in cellular nutrients (typically pertaining to glucose) to generate ATP. Mitochondria multiply by binary fission, like prokaryotes. Chloroplasts can only be found in plants and algae, and they capture the sun's energy to make carbohydrates through photosynthesis.
 - Endoplasmic reticulum: The endoplasmic reticulum (ER) is a transport network for molecules targeted for certain modifications and specific destinations, as compared to molecules that float freely in the cytoplasm. The ER has two forms: the rough ER, which has ribosomes on its surface that secrete proteins into the ER, and the smooth ER, which lacks ribosomes. The smooth ER plays a role in calcium sequestration and release.
 - Golgi apparatus: The primary function of the Golgi apparatus is to process and package the macromolecules such as proteins and lipids that are synthesized by the cell.
 - Lysosomes and Peroxisomes: Lysosomes contain digestive enzymes (acid hydrolases). They digest excess or worn-out organelles, food particles, and engulfed viruses or bacteria. Peroxisomes have enzymes that rid the cell of toxic peroxides. The cell could not house these destructive enzymes if they were not contained in a membrane-bound system.
 - Centrosome: the cytoskeleton organiser: The centrosome produces the microtubules of a cell – a key component of the cytoskeleton. It directs the transport through the ER and the Golgi apparatus. Centrosomes are composed of two centrioles, which separate during cell division and help in the formation of the mitotic spindle. A single centrosome is present in the animal cells. They are also found in some fungi and algae cells.
 - Vacuoles: Vacuoles sequester waste products and in plant cells store water. They are often described as liquid filled space and are surrounded by a membrane. Some cells, most notably Amoeba, have contractile vacuoles, which can pump water out of the cell if there is too much water. The vacuoles of plant cells and fungal cells are usually larger than those of animal cells.

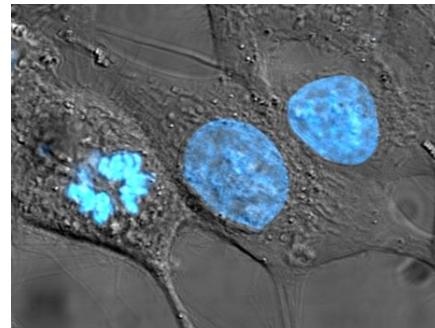


Figure 4.12: HeLa cells stained for nuclear DNA with the blue fluorescent Hoechst dye.²⁹ The central and rightmost cell are in interphase, thus their entire nuclei are labeled. On the left, a cell is going through mitosis and its DNA has condensed. HeLa is an immortal cell line used in scientific research. It is the oldest and most commonly used human cell line. The line is derived from cervical cancer cells taken on February 8, 1951, from Henrietta Lacks³⁰, a 31-year-old African-American mother of five, who died of cancer on October 4, 1951. The cell line was found to be remarkably durable and prolific, which allows it to be used extensively in scientific study. The cells from Lacks' cancerous cervical tumor were taken without her knowledge or consent, which was common practice at the time. Cell biologist George Otto Gey found that they could be kept alive, and developed a cell line. Previously, cells cultured from other human cells would only survive for a few days. Cells from Lacks' tumor behaved differently. As was custom for Gey's lab assistant, she labeled the culture 'HeLa', the first two letters of the patient's first and last name; this became the name of the cell line. HeLa was the subject of a 2010 book by Rebecca Skloot, *The Immortal Life of Henrietta Lacks*³¹, investigating the historical context of the cell line and how the Lacks family was involved its use.

4.6.1 The Cell Nucleus

In cell biology, the nucleus (pl. nuclei; from Latin *nucleus* or *nuculeus*, meaning kernel or seed) is a membrane-bound organelle found in eukaryotic cells. Eukaryotes usually have a single nucleus, but a few cell types, such as mammalian red blood cells, have no nuclei, and a few others including osteoclasts have many. Inside its fully enclosed nuclear membrane, it contains the majority of the cell's genetic material. This material is organized as DNA molecules, along with a variety of proteins, to form chromosomes.

The nucleus was the first organelle to be discovered. What is most likely the oldest preserved drawing dates back to the early microscopist Antonie van Leeuwenhoek (1632–1723). He observed a "lumen", the nucleus, in the red blood cells of salmon (Figure ??). Unlike mammalian red



Figure 4.13: Oldest known depiction of cells and their nuclei by Antonie van Leeuwenhoek, 1719³²

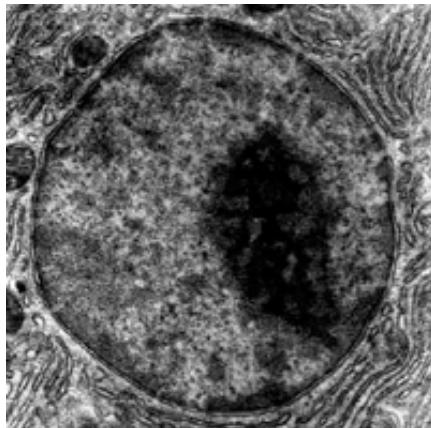


Figure 4.14: An electron micrograph of a cell nucleus, showing the darkly stained nucleolus³³

blood cells, those of other vertebrates still contain nuclei.

The cell nucleus contains all of the cell's genome, except for a small fraction of mitochondrial DNA, organized as multiple long linear DNA molecules in a complex with a large variety of proteins, such as histones, to form chromosomes. The genes within these chromosomes are structured in such a way to promote cell function. The nucleus maintains the integrity of genes and controls the activities of the cell by regulating gene expression—the nucleus is, therefore, the control center of the cell. The main structures making up the nucleus are the nuclear envelope, a double membrane that encloses the entire organelle and isolates its contents from the cellular cytoplasm, and the nuclear matrix (which includes the nuclear lamina), a network within the nucleus that adds mechanical support, much like the cytoskeleton, which supports the cell as a whole.

Because the nuclear envelope is impermeable to large molecules, nuclear pores are required to regulate nuclear transport of molecules across the envelope. The pores cross both nuclear membranes, providing a channel through which larger molecules must be actively transported by carrier proteins while allowing free movement of small molecules and ions. Movement of large molecules such as proteins and RNA through the pores is required for both gene expression and the maintenance of chromosomes. Although the interior of the nucleus does

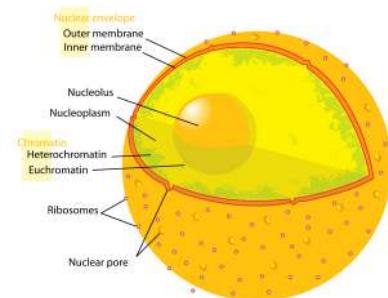


Figure 4.15: The eukaryotic cell nucleus.³⁴ Visible in this diagram are the ribosome-studded double membranes of the nuclear envelope, the DNA (complexed as chromatin), and the nucleolus. Within the cell nucleus is a viscous liquid called nucleoplasm, similar to the cytoplasm found outside the nucleus.

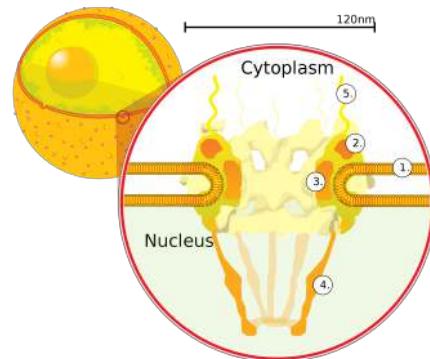


Figure 4.16: A cross section of a nuclear pore on the surface of the nuclear envelope (1).³⁵ Other diagram labels show (2) the outer ring, (3) spokes, (4) basket, and (5) filaments.

not contain any membrane-bound subcompartments, its contents are not uniform, and a number of nuclear bodies exist, made up of unique proteins, RNA molecules, and particular parts of the chromosomes. The best-known of these is the nucleolus, which is mainly involved in the assembly of ribosomes. After being produced in the nucleolus, ribosomes are exported to the cytoplasm where they translate mRNA.

The nuclear envelope, otherwise known as nuclear membrane, consists of two cellular membranes, an inner and an outer membrane, arranged parallel to one another and separated by 10 to 50 nanometres (nm). The nuclear envelope completely encloses the nucleus and separates the cell's genetic material from the surrounding cytoplasm, serving as a barrier to prevent macromolecules from diffusing freely between the nucleoplasm and the cytoplasm. The outer nuclear membrane is continuous

with the membrane of the rough endoplasmic reticulum (RER), and is similarly studded with ribosomes. The space between the membranes is called the perinuclear space and is continuous with the RER lumen.

Nuclear pores, which provide aqueous channels through the envelope, are composed of multiple proteins, collectively referred to as nucleoporins. The pores are about 125 million daltons in molecular weight and consist of around 50 (in yeast) to several hundred proteins (in vertebrates). The pores are 100 nm in total diameter; however, the gap through which molecules freely diffuse is only about 9 nm wide, due to the presence of regulatory systems within the center of the pore. This size selectively allows the passage of small water-soluble molecules while preventing larger molecules, such as nucleic acids and larger proteins, from inappropriately entering or exiting the nucleus. These large molecules must be actively transported into the nucleus instead. The nucleus of a typical mammalian cell will have about 3000 to 4000 pores throughout its envelope, each of which contains an eightfold-symmetric ring-shaped structure at a position where the inner and outer membranes fuse. Attached to the ring is a structure called the nuclear basket that extends into the nucleoplasm, and a series of filamentous extensions that reach into the cytoplasm. Both structures serve to mediate binding to nuclear transport proteins.

Most proteins, ribosomal subunits, and some DNAs are transported through the pore complexes in a process mediated by a family of transport factors known as karyopherins. Those karyopherins that mediate movement into the nucleus are also called importins, whereas those that mediate movement out of the nucleus are called exportins. Most karyopherins interact directly with their cargo, although some use adaptor proteins. Steroid hormones such as cortisol and aldosterone, as well as other small lipid-soluble molecules involved in intercellular signaling, can diffuse through the cell membrane and into the cytoplasm, where they bind nuclear receptor proteins that are trafficked into the nucleus. There they serve as transcription factors when bound to their ligand; in the absence of a ligand, many such receptors function as histone deacetylases that repress gene expression.

The cell nucleus contains the majority of the cell's genetic material in the form of multiple linear DNA molecules organized into structures called chromosomes. Each human cell contains roughly two meters of DNA. During most of the cell cycle these are organized in a DNA-protein complex known as chromatin, and during cell division the chromatin can be seen to form the well-defined chromosomes familiar from a karyotype. A small fraction of the cell's genes are located instead in the mitochondria.

There are two types of chromatin. Euchromatin is the

less compact DNA form, and contains genes that are frequently expressed by the cell. The other type, heterochromatin, is the more compact form, and contains DNA that is infrequently transcribed. This structure is further categorized into facultative heterochromatin, consisting of genes that are organized as heterochromatin only in certain cell types or at certain stages of development, and constitutive heterochromatin that consists of chromosome structural components such as telomeres and centromeres. During interphase the chromatin organizes itself into discrete individual patches, called chromosome territories. Active genes, which are generally found in the euchromatic region of the chromosome, tend to be located towards the chromosome's territory boundary.

The nucleolus is the largest of the discrete densely stained, membraneless structures known as nuclear bodies found in the nucleus. It forms around tandem repeats of rDNA, DNA coding for ribosomal RNA (rRNA). These regions are called nucleolar organizer regions (NOR). The main roles of the nucleolus are to synthesize rRNA and assemble ribosomes. The structural cohesion of the nucleolus depends on its activity, as ribosomal assembly in the nucleolus results in the transient association of nucleolar components, facilitating further ribosomal assembly, and hence further association. This model is supported by observations that inactivation of rDNA results in intermingling of nucleolar structures.

In the first step of ribosome assembly, a protein called RNA polymerase I transcribes rDNA, which forms a large pre-rRNA precursor. This is cleaved into the subunits 5.8S, 18S, and 28S rRNA. The transcription, post-transcriptional processing, and assembly of rRNA occurs in the nucleolus, aided by small nucleolar RNA (snoRNA) molecules, some of which are derived from spliced introns from messenger RNAs encoding genes related to ribosomal function. The assembled ribosomal subunits are the largest structures passed through the nuclear pores.

When observed under the electron microscope, the nucleolus can be seen to consist of three distinguishable regions: the innermost fibrillar centers (FCs), surrounded by the dense fibrillar component (DFC) (that contains fibrillarin and nucleolin), which in turn is bordered by the granular component (GC) (that contains the protein nucleophosmin). Transcription of the rDNA occurs either in the FC or at the FC-DFC boundary, and, therefore, when rDNA transcription in the cell is increased, more FCs are detected. Most of the cleavage and modification of rRNAs occurs in the DFC, while the latter steps involving protein assembly onto the ribosomal subunits occur in the GC.



Figure 4.17: Two mitochondria from mammalian lung tissue displaying their matrix and membranes as shown by electron microscopy.³⁷

4.6.2 The Mitochondria

The mitochondrion³⁶ (plural mitochondria) is a semi-autonomous double-membrane-bound organelle found in most eukaryotic organisms. Some cells in some multicellular organisms may, however, lack mitochondria (for example, mature mammalian red blood cells).

The word mitochondrion comes from the Greek μίτος, *mitos*, “thread”, and χονδρίον, *chondrion*, “granule” or “grain-like”. Mitochondria generate most of the cell’s supply of adenosine triphosphate (ATP), used as a source of chemical energy. A mitochondrion is thus termed the powerhouse of the cell.

In addition to supplying cellular energy, mitochondria are involved in other tasks, such as signaling, cellular differentiation, and cell death, as well as maintaining control of the cell cycle and cell growth.

The organelle is composed of compartments that carry out specialized functions. These compartments or regions include the outer membrane, the intermembrane space, the inner membrane, and the cristae and matrix.

Although most of a cell’s DNA is contained in the cell nucleus, the mitochondrion has its own independent genome (“mitogenome”) that shows substantial similarity to bacterial genomes. Mitochondrial proteins (proteins transcribed from mitochondrial DNA) vary depending on the tissue and the species.

The endosymbiotic hypothesis suggests that mitochondria were originally prokaryotic cells, capable of implementing oxidative mechanisms that were not possible for eukaryotic cells; they became endosymbionts living inside the eukaryote. The endosymbiotic hypothesis suggests that mitochondria descended from bacteria that somehow survived endocytosis by another cell, and became incorporated into the cytoplasm. The ability of

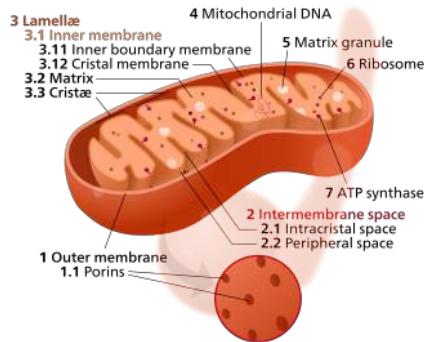


Figure 4.18: Cartoon of the structure of a mitochondrion.³⁸ Components of a typical mitochondrion 1 Outer membrane 1.1 Porin 2 Intermembrane space 2.1 Intracristal space 2.2 Peripheral space 3 Lamella 3.1 Inner membrane 3.11 Inner boundary membrane 3.12 Cristal membrane 3.2 Matrix 3.3 Cristæ 4 Mitochondrial DNA 5 Matrix granule 6 Ribosome 7 ATP synthase

these bacteria to conduct respiration in host cells that had relied on glycolysis and fermentation would have provided a considerable evolutionary advantage. This symbiotic relationship probably developed 1.7 to 2 billion years ago.

A mitochondrion contains outer and inner membranes composed of phospholipid bilayers and proteins. The two membranes have different properties. Because of this double-membraned organization, there are five distinct parts to a mitochondrion. They are:

- the outer mitochondrial membrane,
- the intermembrane space (the space between the outer and inner membranes),
- the inner mitochondrial membrane,
- the cristae space (formed by infoldings of the inner membrane), and
- the matrix (space within the inner membrane).

The inner mitochondrial membrane contains proteins with three types of functions:

1. Those that perform the electron transport chain redox reactions
2. ATP synthase, which generates ATP in the matrix
3. Specific transport proteins that regulate metabolite passage into and out of the mitochondrial matrix

The most prominent roles of mitochondria are to produce the energy currency of the cell, ATP (i.e., phosphorylation of ADP), through respiration, and to regulate cellular metabolism. The central set of reactions involved in ATP production are collectively known as the citric acid cycle, or the Krebs cycle. However, the mitochondrion has many other functions in addition to the production of ATP.

³⁶<https://en.wikipedia.org/wiki/Mitochondrion>

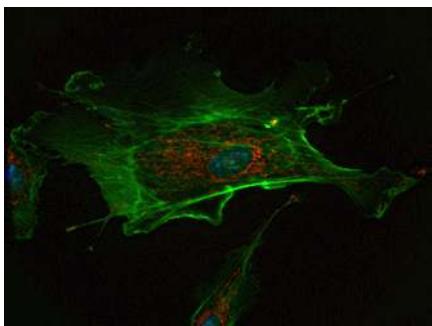


Figure 4.19: A fluorescent image of an endothelial cell. Nuclei are stained blue, mitochondria are stained red, and microfilaments are stained green.³⁹

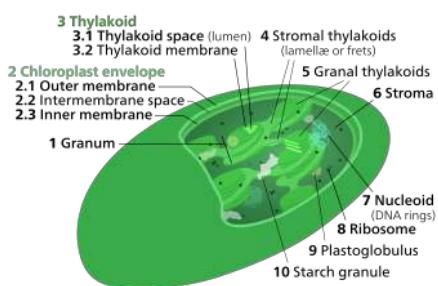


Figure 4.20: Cartoon of the structure of a chloroplast.⁴¹ Components of a typical chloroplast 1 Granum 2 Chloroplast envelope 2.1 Outer membrane 2.2 Intermembrane space 2.3 Inner membrane 3 Thylakoid □ You are here 3.1 Thylakoid space (lumen) 3.2 Thylakoid membrane 4 Stromal thylakoids 5 Granal thylakoids 6 Stroma 7 Nucleoid (DNA ring) 7 Ribosome 8 Plastoglobulus 9 Starch granule

4.6.3 The Chloroplasts

Chloroplasts⁴⁰ are organelles that conduct photosynthesis, where the photosynthetic pigment chlorophyll captures the energy from sunlight, converts it, and stores it in the energy-storage molecules ATP and NADPH while freeing oxygen from water in plant and algal cells. They then use the ATP and NADPH to make organic molecules from carbon dioxide in a process known as the Calvin cycle. Chloroplasts carry out a number of other functions, including fatty acid synthesis, much amino acid synthesis, and the immune response in plants. The number of chloroplasts per cell varies from one, in unicellular algae, up to 100 in plants like *Arabidopsis* and wheat.

A chloroplast is a type of organelle known as a plastid, characterized by its two membranes and a high concentration of chlorophyll. Other plastid types, such as the leucoplast and the chromoplast, contain little chlorophyll and do not carry out photosynthesis.

Chloroplasts are highly dynamic—they circulate and are moved around within plant cells, and occasionally pinch in two to reproduce. Their behavior is strongly influenced by environmental factors like light color and intensity. Chloroplasts, like mitochondria, contain their own DNA, which is thought to be inherited from their ancestor—a photosynthetic cyanobacterium that was engulfed by an early eukaryotic cell. Chloroplasts cannot be made by the plant cell and must be inherited by each daughter cell during cell division.

Chloroplasts are considered endosymbiotic Cyanobacteria. Cyanobacteria are sometimes called blue-green algae even though they are prokaryotes. Chloroplasts can probably be traced back to a single endosymbiotic event, when a cyanobacterium was engulfed by the eukaryote. Despite this, chloroplasts can be found in an extremely wide set of organisms, some not even directly related to each other—a consequence of many secondary and even tertiary endosymbiotic events.

The word chloroplast is derived from the Greek words *chloros* (χλωρός), which means green, and *plastes* (πλάστης), which means “the one who forms”.

In land plants, chloroplasts are generally lens-shaped, 3–10 µm in diameter and 1–3 µm thick.

All chloroplasts have at least three membrane systems—the outer chloroplast membrane, the inner chloroplast membrane, and the thylakoid system. Inside the outer and inner chloroplast membranes is the chloroplast stroma, a semi-gel-like fluid that makes up much of a chloroplast's volume, and in which the thylakoid system floats.

The chloroplast stroma contains many proteins, though the most common and important is RuBisCO, which is probably also the most abundant protein on the planet. RuBisCO is the enzyme that fixes CO₂ into sugar molecules. In C3 plants, RuBisCO is abundant in all chloroplasts, though in C4 plants, it is confined to the bundle sheath chloroplasts, where the Calvin cycle is carried out in C4 plants.

Suspended within the chloroplast stroma is the thylakoid system, a highly dynamic collection of membranous sacks called thylakoids where chlorophyll is found and the light reactions of photosynthesis happen. In most vascular plant chloroplasts, the thylakoids are arranged in stacks called grana, though in certain C4 plant chloroplasts and some algal chloroplasts, the thylakoids are free floating.

Thylakoids (sometimes spelled thylakoïds), are small interconnected sacks which contain the membranes that the light reactions of photosynthesis take place on. The word thylakoid comes from the Greek word *thylakos* which

⁴⁰<https://en.wikipedia.org/wiki/Chloroplast>

means “sack”.

Embedded in the thylakoid membranes are important protein complexes which carry out the light reactions of photosynthesis. Photosystem II and photosystem I contain light-harvesting complexes with chlorophyll and carotenoids that absorb light energy and use it to energize electrons. Molecules in the thylakoid membrane use the energized electrons to pump hydrogen ions into the thylakoid space, decreasing the pH and turning it acidic. ATP synthase is a large protein complex that harnesses the concentration gradient of the hydrogen ions in the thylakoid space to generate ATP energy as the hydrogen ions flow back out into the stroma—much like a dam turbine.

There are two types of thylakoids—granal thylakoids, which are arranged in grana, and stromal thylakoids, which are in contact with the stroma. Granal thylakoids are pancake-shaped circular disks about 300–600 nanometers in diameter. Stromal thylakoids are helicoid sheets that spiral around grana. The flat tops and bottoms of granal thylakoids contain only the relatively flat photosystem II protein complex. This allows them to stack tightly, forming grana with many layers of tightly appressed membrane, called granal membrane, increasing stability and surface area for light capture.

In contrast, photosystem I and ATP synthase are large protein complexes which jut out into the stroma. They can't fit in the appressed granal membranes, and so are found in the stromal thylakoid membrane—the edges of the granal thylakoid disks and the stromal thylakoids. These large protein complexes may act as spacers between the sheets of stromal thylakoids.

The number of thylakoids and the total thylakoid area of a chloroplast is influenced by light exposure. Shaded chloroplasts contain larger and more grana with more thylakoid membrane area than chloroplasts exposed to bright light, which have smaller and fewer grana and less thylakoid area. Thylakoid extent can change within minutes of light exposure or removal.

Inside the photosystems embedded in chloroplast thylakoid membranes are various photosynthetic pigments, which absorb and transfer light energy. The types of pigments found are different in various groups of chloroplasts, and are responsible for a wide variety of chloroplast colorations.

Chlorophyll a is found in all chloroplasts, as well as their cyanobacterial ancestors. Chlorophyll a is a blue-green pigment partially responsible for giving most cyanobacteria and chloroplasts their color.

Chlorophyll b is an olive green pigment found only

in the chloroplasts of plants, green algae, any secondary chloroplasts obtained through the secondary endosymbiosis of a green alga, and a few cyanobacteria. It is the chlorophylls a and b together that make most plant and green algal chloroplasts green.

In addition to chlorophylls, another group of yellow-orange pigments called carotenoids are also found in the photosystems. There are about thirty photosynthetic carotenoids. They help transfer and dissipate excess energy, and their bright colors sometimes override the chlorophyll green, like during the fall, when the leaves of some land plants change color. β-carotene is a bright red-orange carotenoid found in nearly all chloroplasts, like chlorophyll a. Xanthophylls, especially the orange-red zeaxanthin, are also common. Many other forms of carotenoids exist that are only found in certain groups of chloroplasts.

The chloroplasts of plant and algal cells can orient themselves to best suit the available light. In low-light conditions, they will spread out in a sheet—maximizing the surface area to absorb light. Under intense light, they will seek shelter by aligning in vertical columns along the plant cell's cell wall or turning sideways so that light strikes them edge-on. This reduces exposure and protects them from photooxidative damage. This ability to distribute chloroplasts so that they can take shelter behind each other or spread out may be the reason why land plants evolved to have many small chloroplasts instead of a few big ones. Chloroplast movement is considered one of the most closely regulated stimulus-response systems that can be found in plants.

One of the main functions of the chloroplast is its role in photosynthesis, the process by which light is transformed into chemical energy, to subsequently produce food in the form of sugars. Water (H_2O) and carbon dioxide (CO_2) are used in photosynthesis, and sugar and oxygen (O_2) is made, using light energy. Photosynthesis is divided into two stages—the light reactions, where water is split to produce oxygen, and the dark reactions, or Calvin cycle, which builds sugar molecules from carbon dioxide. The two phases are linked by the energy carriers adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide phosphate ($NADP^+$).

4.6.4 The Endoplasmic Reticulum

The endoplasmic reticulum⁴² (ER) is a type of organelle made up of two subunits – rough endoplasmic reticulum (RER), and smooth endoplasmic reticulum (SER). The endoplasmic reticulum is found in most eukaryotic cells and forms an interconnected network of flattened, membrane-

⁴²https://en.wikipedia.org/wiki/Endoplasmic_reticulum

enclosed sacs known as cisternae (in the RER), and tubular structures in the SER. The membranes of the ER are continuous with the outer nuclear membrane. The endoplasmic reticulum is not found in red blood cells, or spermatozoa.

The two types of ER share many of the same proteins and engage in certain common activities such as the synthesis of certain lipids and cholesterol. Different types of cells contain different ratios of the two types of ER depending on the activities of the cell.

The outer (cytosolic) face of the rough endoplasmic reticulum is studded with ribosomes that are the sites of protein synthesis. The rough endoplasmic reticulum is especially prominent in cells such as hepatocytes. The smooth endoplasmic reticulum lacks ribosomes and functions in lipid synthesis but not metabolism, the production of steroid hormones, and detoxification. The smooth endoplasmic reticulum is especially abundant in mammalian liver and gonad cells.

The general structure of the endoplasmic reticulum is a network of membranes called cisternae. These sac-like structures are held together by the cytoskeleton. The phospholipid membrane encloses the cisternal space (or lumen), which is continuous with the perinuclear space but separate from the cytosol. The functions of the endoplasmic reticulum can be summarized as the synthesis and export of proteins and membrane lipids, but varies between ER and cell type and cell function. The quantity of both rough and smooth endoplasmic reticulum in a cell can slowly interchange from one type to the other, depending on the changing metabolic activities of the cell. Transformation can include embedding of new proteins in membrane as well as structural changes. Changes in protein content may occur without noticeable structural changes.

The surface of the rough endoplasmic reticulum (often abbreviated RER or rough ER; also called granular endoplasmic reticulum) is studded with protein-manufacturing ribosomes giving it a "rough" appearance (hence its name). The binding site of the ribosome on the rough endoplasmic reticulum is the translocon. However, the ribosomes are not a stable part of this organelle's structure as they are constantly being bound and released from the membrane. A ribosome only binds to the RER once a specific protein-nucleic acid complex forms in the cytosol. This special complex forms when a free ribosome begins translating the mRNA of a protein destined for the secretory pathway. The first 5-30 amino acids polymerized encode a signal peptide, a molecular message that is recognized and bound by a signal recognition particle (SRP). Translation pauses and the ribosome complex binds to the RER translocon where translation continues with the nascent (new) protein forming into the RER lumen and/or membrane. The protein is processed in the ER lumen by an en-

zyme (a signal peptidase), which removes the signal peptide. Ribosomes at this point may be released back into the cytosol; however, non-translating ribosomes are also known to stay associated with translocons.

In most cells the smooth endoplasmic reticulum (abbreviated SER) is scarce. Instead there are areas where the ER is partly smooth and partly rough, this area is called the transitional ER. The transitional ER gets its name because it contains ER exit sites. These are areas where the transport vesicles that contain lipids and proteins made in the ER, detach from the ER and start moving to the Golgi apparatus. Specialized cells can have a lot of smooth endoplasmic reticulum and in these cells the smooth ER has many functions. It synthesizes lipids, phospholipids, and steroids. Cells which secrete these products, such as those in the testes, ovaries, and sebaceous glands have an abundance of smooth endoplasmic reticulum. It also carries out the metabolism of carbohydrates, detoxification of natural metabolism products and of alcohol and drugs, attachment of receptors on cell membrane proteins, and steroid metabolism. In muscle cells, it regulates calcium ion concentration.

4.6.5 The Golgi Apparatus

The Golgi apparatus⁴³, also known as the Golgi complex, Golgi body, or simply the Golgi, is an organelle found in most eukaryotic cells. Part of the endomembrane system in the cytoplasm, it packages proteins into membrane-bound vesicles inside the cell before the vesicles are sent to their destination. It resides at the intersection of the secretory, lysosomal, and endocytic pathways. It is of particular importance in processing proteins for secretion, containing a set of glycosylation enzymes that attach various sugar monomers to proteins as the proteins move through the apparatus.

It was identified in 1897 by the Italian scientist Camillo Golgi⁴⁴ and was named after him in 1898.

In most eukaryotes, the Golgi apparatus is made up of a series of compartments and is a collection of fused, flattened membrane-enclosed disks known as cisternae (singular: cisterna, also called "dictyosomes"), originating from vesicular clusters that bud off the endoplasmic reticulum. A mammalian cell typically contains 40 to 100 stacks of cisternae. Between four and eight cisternae are usually present in a stack; however, in some protists as many as sixty cisternae have been observed. This collection of cisternae is broken down into cis, medial, and trans compartments, making up two main networks: the cis Golgi network (CGN) and the trans Golgi network

⁴³https://en.wikipedia.org/wiki/Golgi_apparatus

⁴⁴https://en.wikipedia.org/wiki/Camillo_Golgi

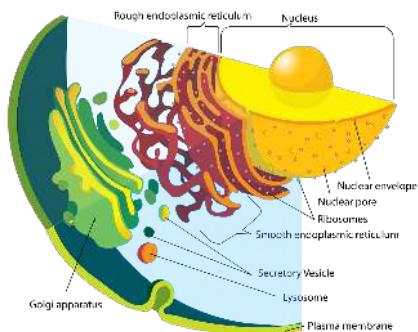


Figure 4.21: Diagram of the endomembrane system⁴⁵

(TGN). The CGN is the first cisternal structure, and the TGN is the final, from which proteins are packaged into vesicles destined to lysosomes, secretory vesicles, or the cell surface. The TGN is usually positioned adjacent to the stack, but can also be separate from it. The TGN may act as an early endosome in yeast and plants.

The Golgi apparatus is a major collection and dispatch station of protein products received from the endoplasmic reticulum (ER). Proteins synthesized in the ER are packaged into vesicles, which then fuse with the Golgi apparatus. These cargo proteins are modified and destined for secretion via exocytosis or for use in the cell. In this respect, the Golgi can be thought of as similar to a post office: it packages and labels items which it then sends to different parts of the cell or to the extracellular space. The Golgi apparatus is also involved in lipid transport and lysosome formation.

4.6.6 The Ribosomes

The ribosome is a large complex of RNA and protein molecules. They each consist of two subunits, and act as an assembly line where RNA from the nucleus is used to synthesise proteins from amino acids. Ribosomes can be found either floating freely or bound to a membrane (the rough endoplasmatic reticulum in eukaryotes, or the cell membrane in prokaryotes).

4.7 Structures Outside The Cell Membrane

Many cells also have structures which exist wholly or partially outside the cell membrane. These structures are notable because they are not protected from the external environment by the semipermeable cell membrane. In order to assemble these structures, their components must be carried across the cell membrane by export processes.

Many types of prokaryotic and eukaryotic cells have a

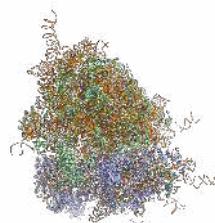


Figure 4.22: Crystal structure of the human 80S ribosome (based on atomic coordinates of PDB 4V6X⁴⁶ rendered with open source molecular visualization tool PyMol). The 40S (small) ribosomal subunit proteins are shown in lightblue, the 60S (large) subunit proteins in palegreen, the ribosomal RNA in orange.

cell wall. The cell wall acts to protect the cell mechanically and chemically from its environment, and is an additional layer of protection to the cell membrane. Different types of cell have cell walls made up of different materials; plant cell walls are primarily made up of cellulose, fungi cell walls are made up of chitin and bacteria cell walls are made up of peptidoglycan.

A gelatinous capsule is present in some bacteria outside the cell membrane and cell wall. The capsule may be polysaccharide as in pneumococci, meningococci or polypeptide as *Bacillus anthracis* or hyaluronic acid as in streptococci. Capsules are not marked by normal staining protocols and can be detected by India ink or methyl blue; which allows for higher contrast between the cells for observation.^{:87}

Flagella are organelles for cellular mobility. The bacterial flagellum stretches from cytoplasm through the cell membrane(s) and extrudes through the cell wall. They are long and thick thread-like appendages, protein in nature. A different type of flagellum is found in archaea and a different type is found in eukaryotes.

A fimbria (plural fimbriae also known as a pilus, plural pili) is a short, thin, hair-like filament found on the surface of bacteria. Fimbriae are formed of a protein called pilin (antigenic) and are responsible for the attachment of bacteria to specific receptors on human cells (cell adhesion). There are special types of pili involved in bacterial conjugation.

4.8 Cellular Replication

Cell division involves a single cell (called a mother cell) dividing into two daughter cells. This leads to growth in multicellular organisms (the growth of tissue) and to procreation (vegetative reproduction) in unicellular organisms. Prokaryotic cells divide by binary fission, while eukaryotic

cells usually undergo a process of nuclear division, called mitosis, followed by division of the cell, called cytokinesis. A diploid cell may also undergo meiosis to produce haploid cells, usually four. Haploid cells serve as gametes in multicellular organisms, fusing to form new diploid cells.

DNA replication, or the process of duplicating a cell's genome, always happens when a cell divides through mitosis or binary fission. This occurs during the S phase of the cell cycle.

In meiosis, the DNA is replicated only once, while the cell divides twice. DNA replication only occurs before meiosis I. DNA replication does not occur when the cells divide the second time, in meiosis II. Replication, like all cellular activities, requires specialized proteins for carrying out the job.

4.9 DNA Repair

In general, cells of all organisms contain enzyme systems that scan their DNA for damages and carry out repair processes when damages are detected. Diverse repair processes have evolved in organisms ranging from bacteria to humans. The widespread prevalence of these repair processes indicates the importance of maintaining cellular DNA in an undamaged state in order to avoid cell death or errors of replication due to damages that could lead to mutation. *E. coli* bacteria are a well-studied example of a cellular organism with diverse well-defined DNA repair processes. These include: (1) nucleotide excision repair, (2) DNA mismatch repair, (3) non-homologous end joining of double-strand breaks, (4) recombinational repair and (5) light-dependent repair (photoreactivation).

4.10 Cellular Growth And Metabolism

Between successive cell divisions, cells grow through the functioning of cellular metabolism. Cell metabolism is the process by which individual cells process nutrient molecules. Metabolism has two distinct divisions: catabolism, in which the cell breaks down complex molecules to produce energy and reducing power, and anabolism, in which the cell uses energy and reducing power to construct complex molecules and perform other biological functions. Complex sugars consumed by the organism can be broken down into simpler sugar molecules called monosaccharides such as glucose. Once inside the cell, glucose is broken down to make adenosine triphosphate (ATP), a molecule that possesses readily available energy, through two different pathways.

4.11 Protein Synthesis

Cells are capable of synthesizing new proteins, which are essential for the modulation and maintenance of cellular activities. This process involves the formation of new protein molecules from amino acid building blocks based on information encoded in DNA/RNA. Protein synthesis generally consists of two major steps: transcription and translation.

Transcription is the process where genetic information in DNA is used to produce a complementary RNA strand. This RNA strand is then processed to give messenger RNA (mRNA), which is free to migrate through the cell. mRNA molecules bind to protein-RNA complexes called ribosomes located in the cytosol, where they are translated into polypeptide sequences. The ribosome mediates the formation of a polypeptide sequence based on the mRNA sequence. The mRNA sequence directly relates to the polypeptide sequence by binding to transfer RNA (tRNA) adapter molecules in binding pockets within the ribosome. The new polypeptide then folds into a functional three-dimensional protein molecule.

4.12 Cell Motility

Unicellular organisms can move in order to find food or escape predators. Common mechanisms of motion include flagella and cilia.

In multicellular organisms, cells can move during processes such as wound healing, the immune response and cancer metastasis. For example, in wound healing in animals, white blood cells move to the wound site to kill the microorganisms that cause infection. Cell motility involves many receptors, crosslinking, bundling, binding, adhesion, motor and other proteins. The process is divided into three steps – protrusion of the leading edge of the cell, adhesion of the leading edge and de-adhesion at the cell body and rear, and cytoskeletal contraction to pull the cell forward. Each step is driven by physical forces generated by unique segments of the cytoskeleton.

Multicellular organisms are organisms that consist of more than one cell, in contrast to single-celled organisms.

In complex multicellular organisms, cells specialize into different cell types that are adapted to particular functions. In mammals, major cell types include skin cells, muscle cells, neurons, blood cells, fibroblasts, stem cells, and others. Cell types differ both in appearance and function, yet are genetically identical. Cells are able to be of the same genotype but of different cell type due to the differential expression of the genes they contain.

Most distinct cell types arise from a single totipotent

cell, called a zygote, that differentiates into hundreds of different cell types during the course of development. Differentiation of cells is driven by different environmental cues (such as cell-cell interaction) and intrinsic differences (such as those caused by the uneven distribution of molecules during division).

4.13 Origin of The First Cell

There are several theories about the origin of small molecules that led to life on the early Earth. They may have been carried to Earth on meteorites (see Murchison meteorite), created at deep-sea vents, or synthesized by lightning in a reducing atmosphere (see Miller-Urey experiment). There is little experimental data defining what the first self-replicating forms were. RNA is thought to be the earliest self-replicating molecule, as it is capable of both storing genetic information and catalyzing chemical reactions (see RNA world hypothesis), but some other entity with the potential to self-replicate could have preceded RNA, such as clay or peptide nucleic acid.

Cells emerged at least 3.5 billion years ago. The current belief is that these cells were heterotrophs. The early cell membranes were probably more simple and permeable than modern ones, with only a single fatty acid chain per lipid. Lipids are known to spontaneously form bilayered vesicles in water, and could have preceded RNA, but the first cell membranes could also have been produced by catalytic RNA, or even have required structural proteins before they could form.

4.14 Origin of Eukaryotic Cells

The eukaryotic cell seems to have evolved from a symbiotic community of prokaryotic cells. DNA-bearing organelles like the mitochondria and the chloroplasts are descended from ancient symbiotic oxygen-breathing proteobacteria and cyanobacteria, respectively, which were endosymbiosed by an ancestral archaean prokaryote.

Chapter 5

The Cell Membrane

The cell membrane¹ (also known as the plasma membrane (PM) or cytoplasmic membrane, and historically referred to as the plasmalemma) is a biological membrane that separates the interior of all cells from the outside environment (the extracellular space) which protects the cell from its environment. The cell membrane consists of a lipid bilayer, including sterols (e. g. cholesterol) that sit between phospholipids to maintain their fluidity at various temperatures. The membrane also contains membrane proteins, including integral proteins that go across the membrane serving as membrane transporters, and peripheral proteins that loosely attach to the outer (peripheral) side of the cell membrane, acting as enzymes shaping the cell. The cell membrane controls the movement of substances in and out of cells and organelles. In this way, it is selectively permeable to ions and organic molecules. In addition, cell membranes are involved in a variety of cellular processes such as cell adhesion, ion conductivity and cell signalling and serve as the attachment surface for several extracellular structures, including the cell wall, the carbohydrate layer called the glycocalyx, and the intracellular network of protein fibers called the cytoskeleton. In the field of synthetic biology, cell membranes can be artificially reassembled.

While Robert Hooke's⁷ discovery of cells in 1665 led to the proposal of the Cell Theory, Hooke misled the cell membrane theory that all cells contained a hard cell wall since only plant cells could be observed at the time. Microscopists focused on the cell wall for well over 150 years until advances in microscopy were made. In the early 19th century, cells were recognized as being separate entities, unconnected, and bound by individual cell walls after it was found that plant cells could be separated. This theory extended to include animal cells to suggest a universal mechanism for cell protection and development. By the second half of the 19th century, microscopy was still not advanced enough to make a distinction between cell membranes and cell walls. However, some microscopists correctly identified

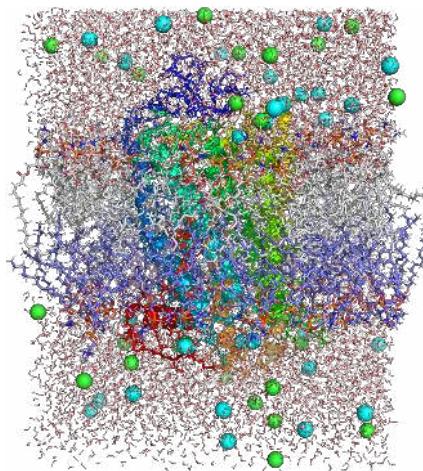


Figure 5.1: Picture of a molecular dynamics simulation² of a cell membrane/protein complex consisting of bovine rhodopsin³ incorporated of a phosphatidylcholine (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, POPC) lipid bilayer. POPC and water molecules are depicted as sticks. The lipid layers facing the extracellular and cytoplasmic spaces are shown in white and blue, respectively. Both the extra- and intracellular interfaces are covered with layers of water. The secondary structure of rhodopsin is depicted in rainbow colored cartoon representation. Potassium and chloride ions are shown as spheres (colored in cyan and green, respectively). Image generated from PDB file⁴ obtained from the CHARMM-GUI Archive – Protein/Membrane Complex Library⁵ using the open source molecular visualization tool PyMol⁶.

¹https://en.wikipedia.org/wiki/Cell_membrane

⁷https://en.wikipedia.org/wiki/Robert_Hooke

at this time that while invisible, it could be inferred that cell membranes existed in animal cells due to intracellular movement of components internally but not externally and that membranes were not the equivalent of a cell wall to plant cell. It was also inferred that cell membranes were not vital components to all cells. Many refuted the existence of a cell membrane still towards the end of the 19th century. In 1890, an update to the Cell Theory stated that cell membranes existed, but were merely secondary structures. It was not until later studies with osmosis and permeability that cell membranes gained more recognition. In 1895, Ernest Overton proposed that cell membranes were made of lipids.

The lipid bilayer hypothesis, proposed in 1925 by Gorter and Grendel, created speculation to the description of the cell membrane bilayer structure based on crystallographic studies and soap bubble observations. In an attempt to accept or reject the hypothesis, researchers measured membrane thickness. In 1925 it was determined by Fricke that the thickness of erythrocyte and yeast cell membranes ranged between 3.3 and 4 nm, a thickness compatible with a lipid monolayer. The choice of the dielectric constant used in these studies was called into question but future tests could not disprove the results of the initial experiment. Independently, the leptoscope was invented in order to measure very thin membranes by comparing the intensity of light reflected from a sample to the intensity of a membrane standard of known thickness. The instrument could resolve thicknesses that depended on pH measurements and the presence of membrane proteins that ranged from 8.6 to 23.2 nm, with the lower measurements supporting the lipid bilayer hypothesis. Later in the 1930s, the membrane structure model developed in general agreement to be the paucimolecular model of Davson and Danielli (1935). This model was based on studies of surface tension between oils and echinoderm eggs. Since the surface tension values appeared to be much lower than would be expected for an oil-water interface, it was assumed that some substance was responsible for lowering the interfacial tensions in the surface of cells. It was suggested that a lipid bilayer was in between two thin protein layers. The paucimolecular model immediately became popular and it dominated cell membrane studies for the following 30 years, until it became rivaled by the fluid mosaic model of Singer and Nicolson (1972).

Despite the numerous models of the cell membrane proposed prior to the fluid mosaic model, it remains the primary archetype for the cell membrane long after its inception in the 1970s. Although the fluid mosaic model has been modernized to detail contemporary discoveries, the basics have remained constant: the membrane is a lipid bilayer composed of hydrophilic exterior heads

and a hydrophobic interior where proteins can interact with hydrophilic heads through polar interactions, but proteins that span the bilayer fully or partially have hydrophobic amino acids that interact with the non-polar lipid interior. The fluid mosaic model not only provided an accurate representation of membrane mechanics, it enhanced the study of hydrophobic forces, which would later develop into an essential descriptive limitation to describe biological macromolecules.

For many centuries, the scientists cited disagreed with the significance of the structure they were seeing as the cell membrane. For almost two centuries, the membranes were seen but mostly disregarded this as an important structure with cellular function. It was not until the 20th century that the significance of the cell membrane as it was acknowledged. Finally, two scientists Gorter and Grendel (1925) made the discovery that the membrane is "lipid-based". From this, they furthered the idea that this structure would have to be in a formation that mimicked layers. Once studied further, it was found by comparing the sum of the cell surfaces and the surfaces of the lipids, a 2:1 ratio was estimated; thus, providing the first basis of the bilayer structure known today. This discovery initiated many new studies that arose globally within various fields of scientific studies, confirming that the structure and functions of the cell membrane are widely accepted.

The structure has been variously referred to by different writers as the ectoplast (de Vries, 1885), Plasmahaut (plasma skin, Pfeffer, 1877, 1891), Hautschicht (skin layer, Pfeffer, 1886; used with a different meaning by Hofmeister, 1867), plasmatic membrane (Pfeffer, 1900), plasma membrane, cytoplasmic membrane, cell envelope and cell membrane. Some authors who did not believe that there was a functional permeable boundary at the surface of the cell preferred to use the term plasmalemma (coined by Mast, 1924) for the external region of the cell.

Cell membranes contain a variety of biological molecules, notably lipids and proteins. Composition is not set, but constantly changing for fluidity and changes in the environment, even fluctuating during different stages of cell development. Specifically, the amount of cholesterol in human primary neuron cell membrane changes, and this change in composition affects fluidity throughout development stages.

Material is incorporated into the membrane, or deleted from it, by a variety of mechanisms:

Fusion of intracellular vesicles with the membrane (exocytosis) not only excretes the contents of the vesicle but also incorporates the vesicle membrane's components into the cell membrane. The membrane may form blebs around extracellular material that pinch off to become vesicles (en-

ocytosis). If a membrane is continuous with a tubular structure made of membrane material, then material from the tube can be drawn into the membrane continuously. Although the concentration of membrane components in the aqueous phase is low (stable membrane components have low solubility in water), there is an exchange of molecules between the lipid and aqueous phases.

The cell membrane consists of three classes of amphipathic lipids: phospholipids, glycolipids, and sterols. The amount of each depends upon the type of cell, but in the majority of cases phospholipids are the most abundant, often contributing for over 50% of all lipids in plasma membranes. Glycolipids only account for a minute amount of about 2% and sterols make up the rest. In RBC studies, 30% of the plasma membrane is lipid. However, for the majority of eukaryotic cells, the composition of plasma membranes is about half lipids and half proteins by weight.

The fatty chains in phospholipids and glycolipids usually contain an even number of carbon atoms, typically between 16 and 20. The 16- and 18-carbon fatty acids are the most common. Fatty acids may be saturated or unsaturated, with the configuration of the double bonds nearly always "cis". The length and the degree of unsaturation of fatty acid chains have a profound effect on membrane fluidity as unsaturated lipids create a kink, preventing the fatty acids from packing together as tightly, thus decreasing the melting temperature (increasing the fluidity) of the membrane. The ability of some organisms to regulate the fluidity of their cell membranes by altering lipid composition is called homeoviscous adaptation.

The entire membrane is held together via non-covalent interaction of hydrophobic tails, however the structure is quite fluid and not fixed rigidly in place. Under physiological conditions phospholipid molecules in the cell membrane are in the liquid crystalline state. It means the lipid molecules are free to diffuse and exhibit rapid lateral diffusion along the layer in which they are present. However, the exchange of phospholipid molecules between intracellular and extracellular leaflets of the bilayer is a very slow process. Lipid rafts and caveolae are examples of cholesterol-enriched microdomains in the cell membrane. Also, a fraction of the lipid in direct contact with integral membrane proteins, which is tightly bound to the protein surface is called annular lipid shell; it behaves as a part of protein complex.

In animal cells cholesterol is normally found dispersed in varying degrees throughout cell membranes, in the irregular spaces between the hydrophobic tails of the membrane lipids, where it confers a stiffening and strengthening effect on the membrane. Additionally, the amount of cholesterol in biological membranes varies between organisms, cell types, and even in individual cells.

Cholesterol, a major component of animal plasma membranes, regulates the fluidity of the overall membrane, meaning that cholesterol controls the amount of movement of the various cell membrane components based on its concentrations. In high temperatures, cholesterol inhibits the movement of phospholipid fatty acid chains, causing a reduced permeability to small molecules and reduced membrane fluidity. The opposite is true for the role of cholesterol in cooler temperatures. Cholesterol production, and thus concentration, is up-regulated (increased) in response to cold temperature. At cold temperatures, cholesterol interferes with fatty acid chain interactions. Acting as antifreeze, cholesterol maintains the fluidity of the membrane. Cholesterol is more abundant in cold-weather animals than warm-weather animals. In plants, which lack cholesterol, related compounds called sterols perform the same function as cholesterol.

Plasma membranes also contain carbohydrates, predominantly glycoproteins, but with some glycolipids (cerebrosides and gangliosides). Carbohydrates are important in the role of cell-cell recognition in eukaryotes; they are located on the surface of the cell where they recognize host cells and share information, viruses that bind to cells using these receptors cause an infection. For the most part, no glycosylation occurs on membranes within the cell; rather generally glycosylation occurs on the extracellular surface of the plasma membrane. The glycocalyx is an important feature in all cells, especially epithelia with microvilli. Recent data suggest the glycocalyx participates in cell adhesion, lymphocyte homing, and many others. The penultimate sugar is galactose and the terminal sugar is sialic acid, as the sugar backbone is modified in the Golgi apparatus. Sialic acid carries a negative charge, providing an external barrier to charged particles.

The cell membrane has large content of proteins, typically around 50% of membrane volume. These proteins are important for the cell because they are responsible for various biological activities. Approximately a third of the genes in yeast code specifically for them, and this number is even higher in multicellular organisms. Membrane proteins consist of three main types: integral proteins, peripheral proteins, and lipid-anchored proteins.

As shown in the adjacent table, integral proteins are amphipathic transmembrane proteins. Examples of integral proteins include ion channels, proton pumps, and g-protein coupled receptors. Ion channels allow inorganic ions such as sodium, potassium, calcium, or chlorine to diffuse down their electrochemical gradient across the lipid bilayer through hydrophilic pores across the membrane. The electrical behavior of cells (i.e. nerve cells) are controlled by ion channels. Proton pumps are protein pumps that are embedded in the lipid bilayer that allow protons

to travel through the membrane by transferring from one amino acid side chain to another. Processes such as electron transport and generating ATP use proton pumps. A G-protein coupled receptor is a single polypeptide chain that crosses the lipid bilayer seven times responding to signal molecules (i.e. hormones and neurotransmitters). G-protein coupled receptors are used in processes such as cell to cell signaling, the regulation of the production of cAMP, and the regulation of ion channels.

The cell membrane, being exposed to the outside environment, is an important site of cell-cell communication. As such, a large variety of protein receptors and identification proteins, such as antigens, are present on the surface of the membrane. Functions of membrane proteins can also include cell-cell contact, surface recognition, cytoskeleton contact, signaling, enzymatic activity, or transporting substances across the membrane.

Most membrane proteins must be inserted in some way into the membrane. For this to occur, an N-terminus "signal sequence" of amino acids directs proteins to the endoplasmic reticulum, which inserts the proteins into a lipid bilayer. Once inserted, the proteins are then transported to their final destination in vesicles, where the vesicle fuses with the target membrane.

Table 5.1: Major classes of membrane proteins.

Type	Description	Examples
Integral proteins or transmembrane proteins	Span the membrane and have a hydrophilic cytosolic domain, which interacts with internal molecules, a hydrophobic membrane-spanning domain that anchors it within the cell membrane, and a hydrophilic extracellular domain that interacts with external molecules. The hydrophobic domain consists of one, multiple, or a combination of α -helices and β sheet protein motifs.	Ion channels, proton pumps, G protein-coupled receptor
Lipid anchored proteins	Covalently bound to single or multiple lipid molecules; hydrophobically insert into the cell membrane and anchor the protein. The protein itself is not in contact with the membrane.	G proteins
Peripheral proteins	Attached to integral membrane proteins, or associated with peripheral regions of the lipid bilayer. These proteins tend to have only temporary interactions with biological membranes, and once reacted, the molecule dissociates to carry on its work in the cytoplasm.	Some enzymes, some hormones

5.1 Structure And Function of The Cell Membrane

The cell membrane surrounds the cytoplasm of living cells, physically separating the intracellular components from the extracellular environment. The cell membrane also plays a role in anchoring the cytoskeleton to provide shape to the cell, and in attaching to the extracellular matrix and other cells to hold them together to form tissues. Fungi, bacteria, most archaea, and plants also have a cell wall, which provides a mechanical support to the cell and precludes the passage of larger molecules.

The cell membrane is selectively permeable and able to regulate what enters and exits the cell, thus facilitating the transport of materials needed for survival. The movement of substances across the membrane can be either “passive”, occurring without the input of cellular energy, or “active”, requiring the cell to expend energy in transporting it. The membrane also maintains the cell potential. The cell membrane thus works as a selective filter that allows only certain things to come inside or go outside the cell. The cell employs a number of transport mechanisms that involve biological membranes:

1. Passive osmosis and diffusion: Some substances (small molecules, ions) such as carbon dioxide (CO_2) and oxygen (O_2), can move across the plasma membrane by diffusion, which is a passive transport process. Because the membrane acts as a barrier for certain molecules and ions, they can occur in different concentrations on the two sides of the membrane. Diffusion occurs when small molecules and ions move freely from high concentration to low concentration in order to equilibrate the membrane. It is considered a passive transport process because it does not require energy and is propelled by the concentration gradient created by each side of the membrane. Such a concentration gradient across a semipermeable membrane sets up an osmotic flow for the water. Osmosis, in biological systems involves a solvent, moving through a semipermeable membrane similarly to passive diffusion as the solvent still moves with the concentration gradient and requires no energy. While water is the most common solvent in cell, it can also be other liquids as well as supercritical liquids and gases.
2. Transmembrane protein channels and transporters: Transmembrane proteins extend through the lipid bilayer of the membranes; they function on both sides of the membrane to transport molecules across it. Nutrients, such as sugars or amino acids, must enter the cell, and certain products of metabolism must leave the cell. Such molecules

can diffuse passively through protein channels such as aquaporins in facilitated diffusion or are pumped across the membrane by transmembrane transporters. Protein channel proteins, also called permeases, are usually quite specific, and they only recognize and transport a limited variety of chemical substances, often limited to a single substance. Another example of a transmembrane protein is a cell-surface receptor, which allow cell signaling molecules to communicate between cells.

3. Endocytosis: Endocytosis is the process in which cells absorb molecules by engulfing them. The plasma membrane creates a small deformation inward, called an invagination, in which the substance to be transported is captured. This invagination is caused by proteins on the outside on the cell membrane, acting as receptors and clustering into depressions that eventually promote accumulation of more proteins and lipids on the cytosolic side of the membrane. The deformation then pinches off from the membrane on the inside of the cell, creating a vesicle containing the captured substance. Endocytosis is a pathway for internalizing solid particles (“cell eating” or phagocytosis), small molecules and ions (“cell drinking” or pinocytosis), and macromolecules. Endocytosis requires energy and is thus a form of active transport.
4. Exocytosis: Just as material can be brought into the cell by invagination and formation of a vesicle, the membrane of a vesicle can be fused with the plasma membrane, extruding its contents to the surrounding medium. This is the process of exocytosis. Exocytosis occurs in various cells to remove undigested residues of substances brought in by endocytosis, to secrete substances such as hormones and enzymes, and to transport a substance completely across a cellular barrier. In the process of exocytosis, the undigested waste-containing food vacuole or the secretory vesicle budded from Golgi apparatus, is first moved by cytoskeleton from the interior of the cell to the surface. The vesicle membrane comes in contact with the plasma membrane. The lipid molecules of the two bilayers rearrange themselves and the two membranes are, thus, fused. A passage is formed in the fused membrane and the vesicles discharges its contents outside the cell.

5.1.1 The Fluid Mosaic Model of The Cell Membrane

According to the fluid mosaic model of S. J. Singer⁸ and G. L. Nicolson⁹ (1972), which replaced the earlier model of Davson and Danielli, biological membranes can be considered as a two-dimensional liquid in which lipid and protein molecules diffuse more or less easily. Although the lipid bilayers that form the basis of the membranes do indeed form two-dimensional liquids by themselves, the plasma membrane also contains a large quantity of proteins, which provide more structure. Examples of such structures are protein-protein complexes, pickets and fences formed by the actin-based cytoskeleton, and potentially lipid rafts.

5.1.2 Lipid bilayer

Lipid bilayers form through the process of self-assembly. The cell membrane consists primarily of a thin layer of amphipathic phospholipids that spontaneously arrange so that the hydrophobic "tail" regions are isolated from the surrounding water while the hydrophilic "head" regions interact with the intracellular (cytosolic) and extracellular faces of the resulting bilayer. This forms a continuous, spherical lipid bilayer. Hydrophobic interactions (also known as the hydrophobic effect) are the major driving forces in the formation of lipid bilayers. An increase in interactions between hydrophobic molecules (causing clustering of hydrophobic regions) allows water molecules to bond more freely with each other, increasing the entropy of the system. This complex interaction can include noncovalent interactions such as van der Waals, electrostatic and hydrogen bonds.

Lipid bilayers are generally impermeable to ions and polar molecules. The arrangement of hydrophilic heads and hydrophobic tails of the lipid bilayer prevent polar solutes (ex. amino acids, nucleic acids, carbohydrates, proteins, and ions) from diffusing across the membrane, but generally allows for the passive diffusion of hydrophobic molecules. This affords the cell the ability to control the movement of these substances via transmembrane protein complexes such as pores, channels and gates. Flippases and scramblases concentrate phosphatidyl serine, which carries a negative charge, on the inner membrane. Along with NANA, this creates an extra barrier to charged moieties moving through the membrane.

Membranes serve diverse functions in eukaryotic and prokaryotic cells. One important role is to regulate the movement of materials into and out of cells. The phospholipid bilayer structure (fluid mosaic model) with specific membrane proteins accounts for the selective permeability

of the membrane and passive and active transport mechanisms. In addition, membranes in prokaryotes and in the mitochondria and chloroplasts of eukaryotes facilitate the synthesis of ATP through chemiosmosis.

5.1.3 Membrane Polarity

The apical membrane of a polarized cell is the surface of the plasma membrane that faces inward to the lumen. This is particularly evident in epithelial and endothelial cells, but also describes other polarized cells, such as neurons. The basolateral membrane of a polarized cell is the surface of the plasma membrane that forms its basal and lateral surfaces. It faces outwards, towards the interstitium, and away from the lumen. Basolateral membrane is a compound phrase referring to the terms "basal (base) membrane" and "lateral (side) membrane", which, especially in epithelial cells, are identical in composition and activity. Proteins (such as ion channels and pumps) are free to move from the basal to the lateral surface of the cell or vice versa in accordance with the fluid mosaic model. Tight junctions join epithelial cells near their apical surface to prevent the migration of proteins from the basolateral membrane to the apical membrane. The basal and lateral surfaces thus remain roughly equivalent[clarification needed] to one another, yet distinct from the apical surface.

Cell membrane can form different types of "supramembrane" structures such as caveola, postsynaptic density, podosome, invadopodium, focal adhesion, and different types of cell junctions. These structures are usually responsible for cell adhesion, communication, endocytosis and exocytosis. They can be visualized by electron microscopy or fluorescence microscopy. They are composed of specific proteins, such as integrins and cadherins.

5.1.4 The Cytoskeleton

The cytoskeleton is found underlying the cell membrane in the cytoplasm and provides a scaffolding for membrane proteins to anchor to, as well as forming organelles that extend from the cell. Indeed, cytoskeletal elements interact extensively and intimately with the cell membrane. Anchoring proteins restricts them to a particular cell surface — for example, the apical surface of epithelial cells that line the vertebrate gut — and limits how far they may diffuse within the bilayer. The cytoskeleton is able to form appendage-like organelles, such as cilia, which are microtubule-based extensions covered by the cell membrane, and filopodia, which are actin-based extensions. These extensions are ensheathed in membrane and project from the surface of the cell in order to sense the external environment and/or make contact with the substrate or other cells. The apical surfaces of epithelial cells are dense with actin-based finger-like projections

⁸https://en.wikipedia.org/wiki/Seymour_Jonathan_Singer

⁹https://en.wikipedia.org/wiki/Garth_L._Nicolson

known as microvilli, which increase cell surface area and thereby increase the absorption rate of nutrients. Localized decoupling of the cytoskeleton and cell membrane results in formation of a bleb.

5.1.5 Intracellular Membranes In Eukaryotic Cells

The content of the cell, inside the cell membrane, is composed of numerous membrane-bound organelles, which contribute to the overall function of the cell. The origin, structure, and function of each organelle leads to a large variation in the cell composition due to the individual uniqueness associated with each organelle.

- Mitochondria and chloroplasts are considered to have evolved from bacteria, known as the endosymbiotic theory. This theory arose from the idea that Paracoccus and Rhodopseudomonas, types of bacteria, share similar functions to mitochondria and blue-green algae, or cyanobacteria, share similar functions to chloroplasts. The endosymbiotic theory proposes that through the course of evolution, a eukaryotic cell engulfed these 2 types of bacteria, leading to the formation of mitochondria and chloroplasts inside eukaryotic cells. This engulfment lead to the 2 membranes systems of these organelles in which the outer membrane originated from the host's plasma membrane and the inner membrane was the endosymbiont's plasma membrane. Considering that mitochondria and chloroplasts both contain their own DNA is further support that both of these organelles evolved from engulfed bacteria that thrived inside a eukaryotic cell.
- In eukaryotic cells, the nuclear membrane separates the contents of the nucleus from the cytoplasm of the cell. The nuclear membrane is formed by an inner and outer membrane, providing the strict regulation of materials in to and out of the nucleus. Materials move between the cytosol and the nucleus through nuclear pores in the nuclear membrane. If a cell's nucleus is more active in transcription, its membrane will have more pores. The protein composition of the nucleus can vary greatly from the cytosol as many proteins are unable to cross through pores via diffusion. Within the nuclear membrane, the inner and outer membranes vary in protein composition, and only the outer membrane is continuous with the endoplasmic reticulum (ER) membrane. Like the ER, the outer membrane also possesses ribosomes responsible for producing and transporting proteins into the space between the two membranes. The nuclear membrane disassembles during

the early stages of mitosis and reassembles in later stages of mitosis.

- The ER, which is part of the endomembrane system, which makes up a very large portion of the cell's total membrane content. The ER is an enclosed network of tubules and sacs, and its main functions include protein synthesis, and lipid metabolism. There are 2 types of ER, smooth and rough. The rough ER has ribosomes attached to it used for protein synthesis, while the smooth ER is used more for the processing of toxins and calcium regulation in the cell.
- The Golgi apparatus has two interconnected round Golgi cisternae. Compartments of the apparatus forms multiple tubular-reticular networks responsible for organization, stack connection and cargo transport that display a continuous grape-like stringed vesicles ranging from 50–60 nm. The apparatus consists of three main compartments, a flat disc-shaped cisterna with tubular-reticular networks and vesicles.

5.2 Membrane Permeability

The permeability of a membrane is the rate of passive diffusion of molecules through the membrane. These molecules are known as permeant molecules. Permeability depends mainly on the electric charge and polarity of the molecule and to a lesser extent the molar mass of the molecule. Due to the cell membrane's hydrophobic nature, small electrically neutral molecules pass through the membrane more easily than charged, large ones. The inability of charged molecules to pass through the cell membrane results in pH partition of substances throughout the fluid compartments of the body. Because of these properties, the cell membrane is referred to as a selectively (or semi-) permeable membrane.

5.2.1 Diffusion

Diffusion¹⁰ is the net movement of material from an area of high concentration to an area with lower concentration. The difference of concentration between the two areas is often termed as the concentration gradient, and diffusion will continue until this gradient has been eliminated. Since diffusion moves materials from an area of higher concentration to an area of lower concentration, it is described as moving solutes "down the concentration gradient" (compared with active transport, which often moves material from area of low concentration to area of higher concentration, and therefore referred to as moving the material "against the concentration gradient"). However, in many cases (e.g. passive drug transport) the driving force of pas-

¹⁰<https://en.wikipedia.org/wiki/Diffusion>

sive transport can not be simplified to the concentration gradient. If there are different solutions at the two sides of the membrane with different equilibrium solubility of the drug, the difference in degree of saturation is the driving force of passive membrane transport. It is also true for supersaturated solutions which are more and more important owing to the spreading of the application of amorphous solid dispersions for drug bioavailability enhancement.

Simple diffusion and osmosis are in some ways similar. Simple diffusion is the passive movement of solute from a high concentration to a lower concentration until the concentration of the solute is uniform throughout and reaches equilibrium. Osmosis is much like simple diffusion but it specifically describes the movement of water (not the solute) across a selectively permeable membrane until there is an equal concentration of water and solute on both sides of the membrane. Simple diffusion and osmosis are both forms of passive transport and require none of the cell's ATP energy.

5.2.2 Facilitated Diffusion

Facilitated diffusion, also called carrier-mediated osmosis, is the movement of molecules across the cell membrane via special transport proteins that are embedded in the plasma membrane by actively taking up or excluding ions. Active transport of protons by H⁺ ATPases alters membrane potential allowing for facilitated passive transport of particular ions such as potassium down their charge gradient through high affinity transporters and channels.

5.2.3 Osmosis

Osmosis¹¹ is the movement of water molecules across a selectively permeable membrane. The net movement of water molecules through a partially permeable membrane from a solution of high water potential to an area of low water potential. A cell with a less negative water potential will draw in water but this depends on other factors as well such as solute potential (pressure in the cell e.g. solute molecules) and pressure potential (external pressure e.g. cell wall). There are three types of Osmosis solutions: the isotonic solution, hypotonic solution, and hypertonic solution. Isotonic solution is when the extracellular solute concentration is balanced with the concentration inside the cell. In the Isotonic solution, the water molecules still move between the solutions, but the rates are the same from both directions, thus the water movement is balanced between the inside of the cell as well as the outside of the cell. A hypotonic solution is when the solute concentration outside the cell is lower than the concentration inside the cell. In hypotonic solutions, the water moves

into the cell, down its concentration gradient (from higher to lower water concentrations). That can cause the cell to swell. Cells that don't have a cell wall, such as animal cells, could burst in this solution. A hypertonic solution is when the solute concentration is higher (think of hyper – as high) than the concentration inside the cell. In hypertonic solution, the water will move out, causing the cell to shrink.

5.2.4 Active Transport

Unlike passive transport, which uses the kinetic energy and natural entropy of molecules moving down a gradient, active transport uses cellular energy to move them against a gradient, polar repulsion, or other resistance. Active transport is usually associated with accumulating high concentrations of molecules that the cell needs, such as ions, glucose and amino acids. Examples of active transport include the uptake of glucose in the intestines in humans and the uptake of mineral ions into root hair cells of plants.

There are two types of active transport: primary active transport that uses adenosine triphosphate (ATP), and secondary active transport that uses an electrochemical gradient. An example of active transport in human physiology is the uptake of glucose in the intestines.

5.2.5 Bulk transport

Endocytosis and exocytosis are both forms of bulk transport that move materials into and out of cells, respectively, via vesicles. In the case of endocytosis, the cellular membrane folds around the desired materials outside the cell. The ingested particle becomes trapped within a pouch, known as a vesicle, inside the cytoplasm. Often enzymes from lysosomes are then used to digest the molecules absorbed by this process. Substances that enter the cell via signal mediated electrolysis include proteins, hormones and growth and stabilization factors. Viruses enter cells through a form of endocytosis that involves their outer membrane fusing with the membrane of the cell. This forces the viral DNA into the host cell.

Biologists distinguish two main types of endocytosis: pinocytosis and phagocytosis.

- In pinocytosis, cells engulf liquid particles (in humans this process occurs in the small intestine, where cells engulf fat droplets).
- In phagocytosis, cells engulf solid particles. Exocytosis involves the removal of substances through the fusion of the outer cell membrane and a vesicle membrane. An example of exocytosis would be the transmission of neurotransmitters across a synapse between brain cells.

¹¹<https://en.wikipedia.org/wiki/Osmosis>

5.3 The Extracellular Matrix

In biology, the extracellular matrix (ECM) is a three-dimensional network of extracellular macromolecules, such as collagen, enzymes, and glycoproteins, that provide structural and biochemical support to surrounding cells. Because multicellularity evolved independently in different multicellular lineages, the composition of ECM varies between multicellular structures; however, cell adhesion, cell-to-cell communication and differentiation are common functions of the ECM.

The animal extracellular matrix includes the interstitial matrix and the basement membrane. Interstitial matrix is present between various animal cells (i.e., in the intercellular spaces). Gels of polysaccharides and fibrous proteins fill the interstitial space and act as a compression buffer against the stress placed on the ECM. Basement membranes are sheet-like depositions of ECM on which various epithelial cells rest. Each type of connective tissue in animals has a type of ECM: collagen fibers and bone mineral comprise the ECM of bone tissue; reticular fibers and ground substance comprise the ECM of loose connective tissue; and blood plasma is the ECM of blood.

The plant ECM includes cell wall components, like cellulose, in addition to more complex signaling molecules. Some single-celled organisms adopt multicellular biofilms in which the cells are embedded in an ECM composed primarily of extracellular polymeric substances (EPS).

Due to its diverse nature and composition, the ECM can serve many functions, such as providing support, segregating tissues from one another, and regulating intercellular communication. The extracellular matrix regulates a cell's dynamic behavior. In addition, it sequesters a wide range of cellular growth factors and acts as a local store for them. Changes in physiological conditions can trigger protease activities that cause local release of such stores. This allows the rapid and local growth factor-mediated activation of cellular functions without de novo synthesis.

Formation of the extracellular matrix is essential for processes like growth, wound healing, and fibrosis. An understanding of ECM structure and composition also helps in comprehending the complex dynamics of tumor invasion and metastasis in cancer biology as metastasis often involves the destruction of extracellular matrix by enzymes such as serine proteases, threonine proteases, and matrix metalloproteinases.

The stiffness and elasticity of the ECM has important implications in cell migration, gene expression, and differentiation. Cells actively sense ECM rigidity and migrate preferentially towards stiffer surfaces in a phenomenon called durotaxis. They also detect elasticity and adjust their gene expression accordingly which has increasingly

become a subject of research because of its impact on differentiation and cancer progression.

Many cells bind to components of the extracellular matrix. Cell adhesion can occur in two ways; by focal adhesions, connecting the ECM to actin filaments of the cell, and hemidesmosomes, connecting the ECM to intermediate filaments such as keratin. This cell-to-ECM adhesion is regulated by specific cell-surface cellular adhesion molecules (CAM) known as integrins. Integrins are cell-surface proteins that bind cells to ECM structures, such as fibronectin and laminin, and also to integrin proteins on the surface of other cells.

Components of the ECM are produced intracellularly by resident cells and secreted into the ECM via exocytosis. Once secreted, they then aggregate with the existing matrix. The ECM is composed of an interlocking mesh of fibrous proteins and glycosaminoglycans (GAGs).

Glycosaminoglycans (GAGs) are carbohydrate polymers and mostly attached to extracellular matrix proteins to form proteoglycans (hyaluronic acid is a notable exception; see below). Proteoglycans have a net negative charge that attracts positively charged sodium ions (Na^+), which attracts water molecules via osmosis, keeping the ECM and resident cells hydrated. Proteoglycans may also help to trap and store growth factors within the ECM.

Collagens are the most abundant protein in the ECM. In fact, collagen is the most abundant protein in the human body and accounts for 90% of bone matrix protein content. Collagens are present in the ECM as fibrillar proteins and give structural support to resident cells. Collagen is exocytosed in precursor form (procollagen), which is then cleaved by procollagen proteases to allow extracellular assembly. Disorders such as Ehlers Danlos Syndrome, osteogenesis imperfecta, and epidermolysis bullosa are linked with genetic defects in collagen-encoding genes. The collagen can be divided into several families according to the types of structure they form:

Elastins, in contrast to collagens, give elasticity to tissues, allowing them to stretch when needed and then return to their original state. This is useful in blood vessels, the lungs, in skin, and the ligamentum nuchae, and these tissues contain high amounts of elastins. Elastins are synthesized by fibroblasts and smooth muscle cells. Elastins are highly insoluble, and tropoelastins are secreted inside a chaperone molecule, which releases the precursor molecule upon contact with a fiber of mature elastin. Tropoelastins are then deaminated to become incorporated into the elastin strand. Disorders such as cutis laxa and Williams syndrome are associated with deficient or absent elastin fibers in the ECM.

Fibronectins are glycoproteins that connect cells with

collagen fibers in the ECM, allowing cells to move through the ECM. Fibronectins bind collagen and cell-surface integrins, causing a reorganization of the cell's cytoskeleton to facilitate cell movement. Fibronectins are secreted by cells in an unfolded, inactive form. Binding to integrins unfolds fibronectin molecules, allowing them to form dimers so that they can function properly. Fibronectins also help at the site of tissue injury by binding to platelets during blood clotting and facilitating cell movement to the affected area during wound healing.

Laminins are proteins found in the basal laminae of virtually all animals. Rather than forming collagen-like fibers, laminins form networks of web-like structures that resist tensile forces in the basal lamina. They also assist in cell adhesion. Laminins bind other ECM components such as collagens and nidogens.

5.4 Cell Junctions

Cell junctions (or intercellular bridges) are a class of cellular structures consisting of multiprotein complexes that provide contact or adhesion between neighboring cells or between a cell and the extracellular matrix in animals. They also maintain the paracellular barrier of epithelia and control paracellular transport. Cell junctions are especially abundant in epithelial tissues. Combined with cell adhesion molecules and extracellular matrix, cell junctions help hold animal cells together.

Cell junctions are also especially important in enabling communication between neighboring cells via specialized protein complexes called communicating (gap) junctions. Cell junctions are also important in reducing stress placed upon cells.

In plants, similar communication channels are known as plasmodesmata, and in fungi they are called septal pores.

In vertebrates, there are three major types of cell junctions:

- Adherens junctions, desmosomes and hemidesmosomes (anchoring junctions)
- Gap junctions (communicating junction)
- Tight junctions (occluding junctions)

Invertebrates have several other types of specific junctions.

In multicellular plants, the structural functions of cell junctions are instead provided for by cell walls. The analogues of communicative cell junctions in plants are called plasmodesmata.

5.4.1 Anchoring Junctions

Cells within tissues and organs must be anchored to one another and attached to components of the extracellular matrix. Cells have developed several types of junctional complexes to serve these functions, and in each case, anchoring proteins extend through the plasma membrane to link cytoskeletal proteins in one cell to cytoskeletal proteins in neighboring cells as well as to proteins in the extracellular matrix.

Anchoring-type junctions not only hold cells together but provide tissues with structural cohesion. These junctions are most abundant in tissues that are subject to constant mechanical stress such as skin and heart.

5.4.2 Desmosomes

Desmosomes, also termed as maculae adherentes, can be visualized as rivets through the plasma membrane of adjacent cells. Intermediate filaments composed of keratin or desmin are attached to membrane-associated attachment proteins that form a dense plaque on the cytoplasmic face of the membrane. Cadherin molecules form the actual anchor by attaching to the cytoplasmic plaque, extending through the membrane and binding strongly to cadherins coming through the membrane of the adjacent cell.

5.4.3 Hemidesmosomes

Hemidesmosomes form rivet-like links between cytoskeleton and extracellular matrix components such as the basal laminae that underlie epithelia. Like desmosomes, they tie to intermediate filaments in the cytoplasm, but in contrast to desmosomes, their transmembrane anchors are integrins rather than cadherins.

5.4.4 Adherens Junctions

Adherens junctions share the characteristic of anchoring cells through their cytoplasmic actin filaments. Similarly to desmosomes and hemidesmosomes, their transmembrane anchors are composed of cadherins in those that anchor to other cells and integrins in those that anchor to extracellular matrix. There is considerable morphologic diversity among adherens junctions. Those that tie cells to one another are seen as isolated streaks or spots, or as bands that completely encircle the cell. The band-type of adherens junctions is associated with bundles of actin filaments that also encircle the cell just below the plasma membrane. Spot-like adherens junctions help cells adhere to extracellular matrix both *in vivo* and *in vitro* where they are called focal adhesions. The cytoskeletal actin filaments that tie into adherens junctions are contractile proteins and in addition to providing an anchoring function, adherens

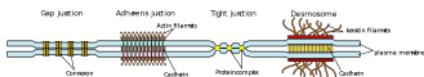


Figure 5.2: Simplified diagram of two types of cell junctions.¹²

junctions are thought to participate in folding and bending of epithelial cell sheets. Thinking of the bands of actin filaments as being similar to ‘drawstrings’ allows one to envision how contraction of the bands within a group of cells would distort the sheet into interesting patterns

5.4.5 Communicating (gap) Junctions

Communicating junctions, or gap junctions allow for direct chemical communication between adjacent cellular cytoplasm through diffusion without contact with the extracellular fluid. This is possible due to six connexin proteins interacting to form a cylinder with a pore in the centre called a connexon. The connexon complexes stretches across the cell membrane and when two adjacent cell connexons interact, they form a complete gap junction channel. Connexon pores vary in size, polarity and therefore can be specific depending on the connexin proteins that constitute each individual connexon. Whilst variation in gap junction channels do occur, their structure remains relatively standard, and this interaction ensures efficient communication without the escape of molecules or ions to the extracellular fluid.

Gap junctions play vital roles in the human body, including their role in the uniform contractile of the heart muscle. They are also relevant in signal transfers in the brain, and their absence shows a decreased cell density in the brain. Retinal and skin cells are also dependent on gap junctions in cell differentiation and proliferation.

5.4.6 Tight Junctions

Found in vertebrate epithelia, tight junctions act as barriers that regulate the movement of water and solutes between epithelial layers. Tight junctions are classified as a paracellular barrier which is defined as not having directional discrimination; however, movement of the solute is largely dependent upon size and charge. There is evidence to suggest that the structures in which solutes pass through are somewhat like pores.

Physiological pH plays a part in the selectivity of solutes passing through tight junctions with most tight junctions being slightly selective for cations. Tight junctions present in different types of epithelia are selective for solutes of differing size, charge, and polarity.

5.4.7 Cell Junction Molecules

The molecules responsible for creating cell junctions include various cell adhesion molecules. There are four main types: selectins, cadherins, integrins, and the immunoglobulin superfamily.

Selectins are cell adhesion molecules that play an important role in the initiation of inflammatory processes. The functional capacity of selectin is limited to leukocyte collaborations with vascular endothelium. There are three types of selectins found in humans; L-selectin, P-selectin and E-selectin. L-selectin deals with lymphocytes, monocytes and neutrophils, P-selectin deals with platelets and endothelium and E-selectin deals only with endothelium. They have extracellular regions made up of an amino-terminal lectin domain, attached to a carbohydrate ligand, growth factor-like domain, and short repeat units (numbered circles) that match the complementary binding protein domains.

Cadherins are calcium-dependent adhesion molecules. Cadherins are extremely important in the process of morphogenesis – fetal development. Together with an alpha-beta catenin complex, the cadherin can bind to the microfilaments of the cytoskeleton of the cell. This allows for homophilic cell-cell adhesion. The β -catenin- α -catenin linked complex at the adherens junctions allows for the formation of a dynamic link to the actin cytoskeleton.

Integrins act as adhesion receptors, transporting signals across the plasma membrane in multiple directions. These molecules are an invaluable part of cellular communication, as a single ligand can be used for many integrins. Unfortunately these molecules still have a long way to go in the ways of research.

Immunoglobulin superfamily are a group of calcium independent proteins capable of homophilic and heterophilic adhesion. Homophilic adhesion involves the immunoglobulin-like domains on the cell surface binding to the immunoglobulin-like domains on an opposing cell's surface while heterophilic adhesion refers to the binding of the immunoglobulin-like domains to integrins and carbohydrates instead.

Cell adhesion is a vital component of the body. Loss of this adhesion effects cell structure, cellular functioning and communication with other cells and the extracellular matrix and can lead to severe health issues and diseases.

Chapter 6

Bioenergetics

Bioenergetics¹ is a field in biochemistry and cell biology that concerns energy flow through living systems. This is an active area of biological research that includes the study of the transformation of energy in living organisms and the study of thousands of different cellular processes such as cellular respiration and the many other metabolic and enzymatic processes that lead to production and utilization of energy in forms such as adenosine triphosphate (ATP) molecules. That is, the goal of bioenergetics is to describe how living organisms acquire and transform energy in order to perform biological work. The study of metabolism and metabolic pathways is thus essential to bioenergetics.

The ability to harness energy from a variety of metabolic pathways is a property of all living organisms that contains earth science. Growth, development, anabolism and catabolism are some of the central processes in the study of biological organisms, because the role of energy is fundamental to such biological processes. Life is dependent on energy transformations; living organisms survive because of exchange of energy between living tissues/ cells and the outside environment. Some organisms, such as autotrophs, can acquire energy from sunlight (through photosynthesis) without needing to consume nutrients and break them down. Other organisms, like heterotrophs, must intake nutrients from food to be able to sustain energy by breaking down chemical bonds in nutrients during metabolic processes such as glycolysis and the citric acid cycle. Importantly, as a direct consequence of the first law of thermodynamics, autotrophs and heterotrophs participate in a universal metabolic network—by eating autotrophs (plants), heterotrophs harness energy that was initially transformed by the plants during photosynthesis.

In a living organism, chemical bonds are broken and made as part of the exchange and transformation of energy. Energy is available for work (such as mechanical work) or for other processes (such as chemical synthesis and anabolic processes in growth), when weak bonds are

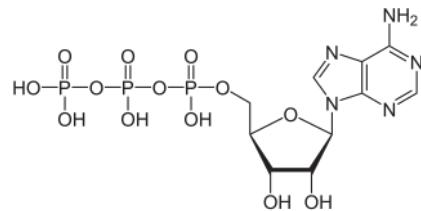


Figure 6.1: Structure of adenosine triphosphate (ATP), protonated²

broken and stronger bonds are made. The production of stronger bonds allows release of usable energy.

Adenosine triphosphate (ATP) is the main “energy currency” for organisms; the goal of metabolic and catabolic processes are to synthesize ATP from available starting materials (from the environment), and to break- down ATP (into adenosine diphosphate (ADP) and inorganic phosphate) by utilizing it in biological processes. In a cell, the ratio of ATP to ADP concentrations is known as the “energy charge” of the cell. A cell can use this energy charge to relay information about cellular needs; if there is more ATP than ADP available, the cell can use ATP to do work, but if there is more ADP than ATP available, the cell must synthesize ATP via oxidative phosphorylation.

Living organisms produce ATP from energy sources, such as sunlight and organic compounds including carbohydrates, lipids and proteins, mainly via oxidative phosphorylation. The terminal phosphate bonds of ATP are relatively weak compared with the stronger bonds formed when ATP is hydrolyzed to adenosine diphosphate and inorganic phosphate. Here it is the thermodynamically favorable free energy of hydrolysis that results in energy release; the phosphoanhydride bond between the terminal phosphate group and the rest of the ATP molecule does not itself contain this energy. An organism’s stockpile of ATP is used as a battery to store energy in cells. Utilization of chemical energy from such molecular bond rearrangement powers biological processes in every biological organism.

¹<https://en.wikipedia.org/wiki/Bioenergetics>

Living organisms obtain energy from organic and inorganic materials; i.e. ATP can be synthesized from a variety of biochemical precursors. For example, lithotrophs can oxidize minerals such as nitrites or forms of sulfur, such as elemental sulfur, sulfites, and hydrogen sulfide to produce ATP. In photosynthesis, autotrophs produce ATP using light energy, whereas heterotrophs must consume organic compounds, mostly including carbohydrates, fats, and proteins. The amount of energy actually obtained by the organism is lower than the amount released in combustion of the food; there are losses in digestion, metabolism, and thermogenesis.

6.1 Thermodynamics of Living Organisms

Thermodynamics is a branch of physics that deals with heat, work, and temperature, and their relation to energy, radiation, and properties of matter. The behavior of these quantities is governed by the four laws of thermodynamics which convey a quantitative description using measurable macroscopic physical quantities, but may be explained in terms of microscopic constituents by statistical mechanics.

Historically, thermodynamics developed out of a desire to increase the efficiency of early steam engines, particularly through the work of French physicist Nicolas Léonard Sadi Carnot³ (1824) who believed that engine efficiency was the key that could help France win the Napoleonic Wars. Scots-Irish physicist William Thomson⁴ (Lord Kelvin) was the first to formulate a concise definition of thermodynamics in 1854 which stated, "Thermo-dynamics is the subject of the relation of heat to forces acting between contiguous parts of bodies, and the relation of heat to electrical agency."

The initial application of thermodynamics to mechanical heat engines was quickly extended to the study of chemical compounds and chemical reactions. Biological thermodynamics is the quantitative study of the energy transductions that occur in or between living organisms, structures, and cells and of the nature and function of the chemical processes underlying these transductions. Biological thermodynamics may address the question of whether the benefit associated with any particular phenotypic trait is worth the energy investment it requires.

Living organisms must obey the laws of thermodynamics, which describe the transfer of heat and work. The First Law of Thermodynamics is a statement of the conservation of energy; though it can be changed from one form

to another, energy can be neither created nor destroyed. The Second Law of Thermodynamics is concerned primarily with whether or not a given process is possible. The Second Law states that no natural process can occur unless it is accompanied by an increase in the entropy of the universe. Stated differently, an isolated system will always tend to disorder. Although living organisms' amazing complexity appears to contradict this law, life is possible as all organisms are open systems that exchange matter and energy with their surroundings. Thus living systems are not in equilibrium, but instead are dissipative systems that maintain their state of high complexity by causing a larger increase in the entropy of their environments. The metabolism of a cell achieves this by coupling the spontaneous processes of catabolism to the non-spontaneous processes of anabolism. In thermodynamic terms, metabolism maintains order by creating disorder.

As the environments of most organisms are constantly changing, the reactions of metabolism must be finely regulated to maintain a constant set of conditions within cells, a condition called homeostasis. Metabolic regulation also allows organisms to respond to signals and interact actively with their environments. Two closely linked concepts are important for understanding how metabolic pathways are controlled. Firstly, the regulation of an enzyme in a pathway is how its activity is increased and decreased in response to signals. Secondly, the control exerted by this enzyme is the effect that these changes in its activity have on the overall rate of the pathway (the flux through the pathway). For example, an enzyme may show large changes in activity (i.e. it is highly regulated) but if these changes have little effect on the flux of a metabolic pathway, then this enzyme is not involved in the control of the pathway.

There are multiple levels of metabolic regulation. In intrinsic regulation, the metabolic pathway self-regulates to respond to changes in the levels of substrates or products; for example, a decrease in the amount of product can increase the flux through the pathway to compensate. This type of regulation often involves allosteric regulation of the activities of multiple enzymes in the pathway. Extrinsic control involves a cell in a multicellular organism changing its metabolism in response to signals from other cells. These signals are usually in the form of water soluble messengers such as hormones and growth factors and are detected by specific receptors on the cell surface. These signals are then transmitted inside the cell by second messenger systems that often involve the phosphorylation of proteins.

A very well understood example of extrinsic control is the regulation of glucose metabolism by the hormone insulin. Insulin is produced in response to rises in blood glucose levels. Binding of the hormone to insulin recep-

³https://en.wikipedia.org/wiki/Nicolas_Léonard_Sadi_Carnot

⁴https://en.wikipedia.org/wiki/William_Thomson,_1st_Baron_Kelvin

tors on cells then activates a cascade of protein kinases that cause the cells to take up glucose and convert it into storage molecules such as fatty acids and glycogen. The metabolism of glycogen is controlled by activity of phosphorylase, the enzyme that breaks down glycogen, and glycogen synthase, the enzyme that makes it. These enzymes are regulated in a reciprocal fashion, with phosphorylation inhibiting glycogen synthase, but activating phosphorylase. Insulin causes glycogen synthesis by activating protein phosphatases and producing a decrease in the phosphorylation of these enzymes.

The central pathways of metabolism described below, such as glycolysis and the citric acid cycle, are present in all three domains of living things and were present in the last universal common ancestor. This universal ancestral cell was prokaryotic and probably a methanogen that had extensive amino acid, nucleotide, carbohydrate and lipid metabolism. The retention of these ancient pathways during later evolution may be the result of these reactions having been an optimal solution to their particular metabolic problems, with pathways such as glycolysis and the citric acid cycle producing their end products highly efficiently and in a minimal number of steps. The first pathways of enzyme-based metabolism may have been parts of purine nucleotide metabolism, while previous metabolic pathways were a part of the ancient RNA world.

Many models have been proposed to describe the mechanisms by which novel metabolic pathways evolve. These include the sequential addition of novel enzymes to a short ancestral pathway, the duplication and then divergence of entire pathways as well as the recruitment of pre-existing enzymes and their assembly into a novel reaction pathway. The relative importance of these mechanisms is unclear, but genomic studies have shown that enzymes in a pathway are likely to have a shared ancestry, suggesting that many pathways have evolved in a step-by-step fashion with novel functions created from pre-existing steps in the pathway. An alternative model comes from studies that trace the evolution of proteins' structures in metabolic networks, this has suggested that enzymes are pervasively recruited, borrowing enzymes to perform similar functions in different metabolic pathways. These recruitment processes result in an evolutionary enzymatic mosaic. A third possibility is that some parts of metabolism might exist as "modules" that can be reused in different pathways and perform similar functions on different molecules.

As well as the evolution of new metabolic pathways, evolution can also cause the loss of metabolic functions. For example, in some parasites metabolic processes that are not essential for survival are lost and preformed amino acids, nucleotides and carbohydrates may instead be scav-

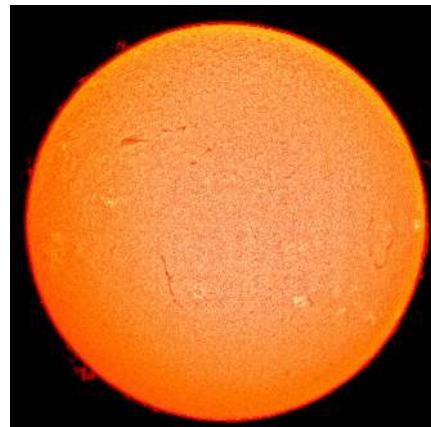


Figure 6.2: The Sun⁵ is the source of energy for most of life on Earth. It derives its energy mainly from nuclear fusion in its core, converting mass to energy as protons are combined to form helium. This energy is transported to the sun's surface then released into space mainly in the form of radiant (light) energy.

enged from the host. Similar reduced metabolic capabilities are seen in endosymbiotic organisms.

The sun is the primary source of energy for living organisms on earth. The relationship between the energy of the incoming sunlight and its wavelength λ or frequency ν is given by

$$E = \frac{hc}{\lambda} = h\nu$$

where h is the Planck constant (6.63×10^{-34} Js) and c is the speed of light (2.998×10^8 m/s). Plants trap this energy from the sunlight and perform photosynthesis, effectively converting solar energy into chemical energy. To transfer the energy once again, animals will feed on plants (or other animals) and use the energy of digested plant (or animal) materials to create biological macromolecules.

Energy flow, also called the calorific flow, refers to the flow of energy through a food chain.

A general energy flow scenario follows:

- Solar energy is fixed by the photoautotrophs, called primary producers, like green plants. Primary consumers absorb most of the stored energy in the plant through digestion, and transform it into the form of energy they need, such as adenosine triphosphate (ATP), through respiration. A part of the energy received by primary consumers, herbivores, is converted to body heat (an effect of respiration), which is radiated away and lost from the system. The loss of energy through body heat

is far greater in warm-blooded animals, which must eat much more frequently than those that are cold-blooded. Energy loss also occurs in the expulsion of undigested food (egesta) by excretion or regurgitation.

- Secondary consumers, carnivores, then consume the primary consumers, although omnivores also consume primary producers. Energy that had been used by the primary consumers for growth and storage is thus absorbed into the secondary consumers through the process of digestion. As with primary consumers, secondary consumers convert this energy into a more suitable form (ATP) during respiration. Again, some energy is lost from the system, since energy which the primary consumers had used for respiration and regulation of body temperature cannot be utilized by the secondary consumers.
- Tertiary consumers, which may or may not be apex predators, then consume the secondary consumers, with some energy passed on and some lost, as with the lower levels of the food chain.
- A final link in the food chain are decomposers which break down the organic matter of the tertiary consumers (or whichever consumer is at the top of the chain) and release nutrients into the soil. They also break down plants, herbivores and carnivores that were not eaten by organisms higher on the food chain, as well as the undigested food that is excreted by herbivores and carnivores. Saprotrophic bacteria and fungi are decomposers, and play a pivotal role in the nitrogen and carbon cycles.

The energy is passed on from trophic level to trophic level and each time about 90% of the energy is lost, with some being lost as heat into the environment (an effect of respiration) and some being lost as incompletely digested food (egesta). Therefore, primary consumers get about 10% of the energy produced by autotrophs, while secondary consumers get 1% and tertiary consumers get 0.1%. This means the top consumer of a food chain receives the least energy, as much of the food chain's energy has been lost between trophic levels. This loss of energy at each level limits typical food chains to only four to six links.

6.2 Metabolism

Metabolism (from Greek: μεταβολή metabolē, "change") is the set of life-sustaining chemical reactions in organisms. The three main purposes of metabolism are: the conversion of food to energy to run cellular processes; the conversion of food/fuel to building blocks for proteins, lipids, nucleic acids, and some carbohydrates; and the elimination

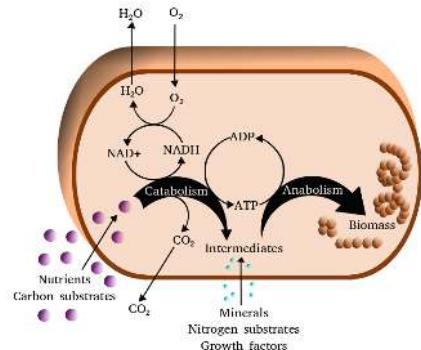


Figure 6.3: Simplified view of the cellular metabolism⁶

of nitrogenous wastes. These enzyme-catalyzed reactions allow organisms to grow and reproduce, maintain their structures, and respond to their environments. (The word metabolism can also refer to the sum of all chemical reactions that occur in living organisms, including digestion and the transport of substances into and between different cells, in which case the above described set of reactions within the cells is called intermediary metabolism or intermediate metabolism).

Metabolic reactions may be categorized as catabolic – the breaking down of compounds (for example, the breaking down of glucose to pyruvate by cellular respiration); or anabolic – the building up (synthesis) of compounds (such as proteins, carbohydrates, lipids, and nucleic acids). Usually, catabolism releases energy, and anabolism consumes energy.

The chemical reactions of metabolism are organized into metabolic pathways, in which one chemical is transformed through a series of steps into another chemical, each step being facilitated by a specific enzyme. Enzymes are crucial to metabolism because they allow organisms to drive desirable reactions that require energy that will not occur by themselves, by coupling them to spontaneous reactions that release energy. Enzymes act as catalysts – they allow a reaction to proceed more rapidly – and they also allow the regulation of the rate of a metabolic reaction, for example in response to changes in the cell's environment or to signals from other cells.

The metabolic system of a particular organism determines which substances it will find nutritious and which poisonous. For example, some prokaryotes use hydrogen sulfide as a nutrient, yet this gas is poisonous to animals. The basal metabolic rate of an organism is the measure of the amount of energy consumed by all of these chemical reactions.

A striking feature of metabolism is the similarity of the basic metabolic pathways among vastly different

species. For example, the set of carboxylic acids that are best known as the intermediates in the citric acid cycle are present in all known organisms, being found in species as diverse as the unicellular bacterium *Escherichia coli* and huge multicellular organisms like elephants. These similarities in metabolic pathways are likely due to their early appearance in evolutionary history, and their retention because of their efficacy. The metabolism of cancer cells is also different from the metabolism of normal cells and these differences can be used to find targets for therapeutic intervention in cancer.

Most of the structures that make up animals, plants and microbes are made from four basic classes of molecule: amino acids, carbohydrates , nucleic acid and lipids (often called fats). As these molecules are vital for life, metabolic reactions either focus on making these molecules during the construction of cells and tissues, or by breaking them down and using them as a source of energy, by their digestion. These biochemicals can be joined together to make polymers such as DNA and proteins, essential macromolecules of life.

Table 6.1: The three essential polymeric macromolecules of life

Type of molecule	Name of monomer forms	Name of polymer forms	Examples of polymer forms
Amino acids	Amino acids	Proteins (made of polypeptides)	Fibrous proteins and globular proteins
Carbohydrates	Monosaccharides	Polysaccharides	Starch, glycogen and cellulose
Nucleic acids	Nucleotides	Polynucleotides	DNA and RNA

The history of the scientific study of metabolism spans several centuries and has moved from examining whole animals in early studies, to examining individual metabolic reactions in modern biochemistry. The first controlled experiments in human metabolism were published by Santorio Santorio in 1614 in his book *Ars de statica medicina*. He described how he weighed himself before and after eating, sleep, working, sex, fasting, drinking, and excreting. He found that most of the food he took in was lost through what he called “insensible perspiration”.

In these early studies, the mechanisms of these metabolic processes had not been identified and a vital force was thought to animate living tissue. In the 19th century, when studying the fermentation of sugar to alcohol by yeast, Louis Pasteur⁷ concluded that fermentation was catalyzed by substances within the yeast cells he called “ferments”. He wrote that “alcoholic fermentation is an act correlated with the life and organization of the yeast cells, not with the death or putrefaction of the cells.” This discovery, along with the publication by Friedrich Wöhler in 1828 of a paper on the chemical synthesis of urea, and is notable for being the first organic compound prepared from wholly inorganic precursors. This proved that the organic compounds and chemical reactions found in cells were no different in principle than any other part of chemistry.

It was the discovery of enzymes at the beginning of the 20th century by Eduard Buchner⁸ that separated the study of the chemical reactions of metabolism from the biological study of cells, and marked the beginnings of biochemistry. The mass of biochemical knowledge grew rapidly throughout the early 20th century. One of the most prolific of these modern biochemists was Hans Krebs who made huge contributions to the study of metabolism. He discovered the urea cycle and later, working with Hans Kornberg, the citric acid cycle and the glyoxylate cycle. Modern biochemical research has been greatly aided by the development of new techniques such as chromatography, X-ray diffraction, NMR spectroscopy, radioisotopic labelling, electron microscopy and molecular dynamics simulations. These techniques have allowed the discovery and detailed analysis of the many molecules and metabolic pathways in cells.

6.2.1 Catabolism

Catabolism is the set of metabolic processes that break down large molecules. These include breaking down and oxidizing food molecules. The purpose of the catabolic reactions is to provide the energy and components needed by anabolic reactions which build molecules. The exact nature of these catabolic reactions differ from organism to

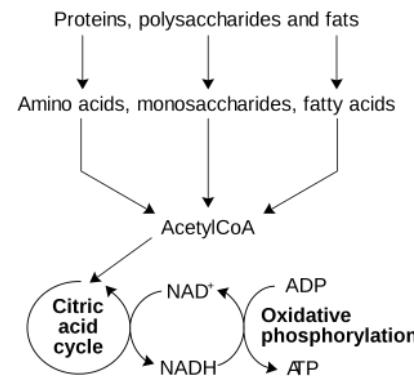


Figure 6.4: Simplified diagram of catabolism of proteins, carbohydrates and fats.⁹

organism, and organisms can be classified based on their sources of energy and carbon (their primary nutritional groups), as shown in the table below. Organic molecules are used as a source of energy by organotrophs, while lithotrophs use inorganic substrates, and phototrophs capture sunlight as chemical energy. However, all these different forms of metabolism depend on redox reactions that involve the transfer of electrons from reduced donor molecules such as organic molecules, water, ammonia, hydrogen sulfide or ferrous ions to acceptor molecules such as oxygen, nitrate or sulfate. In animals, these reactions involve complex organic molecules that are broken down to simpler molecules, such as carbon dioxide and water. In photosynthetic organisms, such as plants and cyanobacteria, these electron-transfer reactions do not release energy but are used as a way of storing energy absorbed from sunlight.

The most common set of catabolic reactions in animals can be separated into three main stages. In the first stage, large organic molecules, such as proteins, polysaccharides or lipids, are digested into their smaller components outside cells. Next, these smaller molecules are taken up by cells and converted to smaller molecules, usually acetyl coenzyme A (acetyl-CoA), which releases some energy. Finally, the acetyl group on the CoA is oxidised to water and carbon dioxide in the citric acid cycle and electron transport chain, releasing the energy that is stored by reducing the coenzyme nicotinamide adenine dinucleotide (NAD⁺) into NADH.

6.2.2 Digestion

Macromolecules cannot be directly processed by cells. Macromolecules must be broken into smaller units before they can be used in cell metabolism. Different classes of enzymes were being used to digest these polymers. These digestive enzymes include proteases that digest

⁷https://en.wikipedia.org/wiki/Louis_Pasteur

⁸https://en.wikipedia.org/wiki/Eduard_Buchner

proteins into amino acids, as well as glycoside hydrolases that digest polysaccharides into simple sugars known as monosaccharides

Microbes simply secrete digestive enzymes into their surroundings, while animals only secrete these enzymes from specialized cells in their guts, including the stomach and pancreas, and salivary glands. The amino acids or sugars released by these extracellular enzymes are then pumped into cells by active transport proteins.

6.2.3 Energy From Organic Compounds

Carbohydrate catabolism is the breakdown of carbohydrates into smaller units. Carbohydrates are usually taken into cells once they have been digested into monosaccharides. Once inside, the major route of breakdown is glycolysis, where sugars such as glucose and fructose are converted into pyruvate and some ATP is generated. Pyruvate is an intermediate in several metabolic pathways, but the majority is converted to acetyl-CoA through aerobic (with oxygen) glycolysis and fed into the citric acid cycle. Although some more ATP is generated in the citric acid cycle, the most important product is NADH, which is made from NAD^+ as the acetyl-CoA is oxidized. This oxidation releases carbon dioxide as a waste product. In anaerobic conditions, glycolysis produces lactate, through the enzyme lactate dehydrogenase re-oxidizing NADH to NAD^+ for re-use in glycolysis. An alternative route for glucose breakdown is the pentose phosphate pathway, which reduces the coenzyme NADPH and produces pentose sugars such as ribose, the sugar component of nucleic acids.

Fats are catabolised by hydrolysis to free fatty acids and glycerol. The glycerol enters glycolysis and the fatty acids are broken down by beta oxidation to release acetyl-CoA, which then is fed into the citric acid cycle. Fatty acids release more energy upon oxidation than carbohydrates because carbohydrates contain more oxygen in their structures. Steroids are also broken down by some bacteria in a process similar to beta oxidation, and this breakdown process involves the release of significant amounts of acetyl-CoA, propionyl-CoA, and pyruvate, which can all be used by the cell for energy. *M. tuberculosis* can also grow on the lipid cholesterol as a sole source of carbon, and genes involved in the cholesterol use pathway(s) have been validated as important during various stages of the infection lifecycle of *M. tuberculosis*.

Amino acids are either used to synthesize proteins and other biomolecules, or oxidized to urea and carbon dioxide as a source of energy. The oxidation pathway starts with the removal of the amino group by a transaminase. The amino group is fed into the urea cycle, leaving a deaminated carbon skeleton in the form of a keto acid. Sev-

eral of these keto acids are intermediates in the citric acid cycle, for example the deamination of glutamate forms α -ketoglutarate. The glucogenic amino acids can also be converted into glucose, through gluconeogenesis (discussed below).

6.3 Energy Transformations

6.3.1 Oxidative Phosphorylation

In oxidative phosphorylation, the electrons removed from organic molecules are transferred to oxygen and the energy released is used to make ATP. This is done in eukaryotes by a series of proteins in the membranes of mitochondria called the electron transport chain. In prokaryotes, these proteins are found in the cell's inner membrane. These proteins use the energy released from passing electrons from reduced molecules like NADH onto oxygen to pump protons across a membrane.

Pumping protons out of the mitochondria creates a proton concentration difference across the membrane and generates an electrochemical gradient. This force drives protons back into the mitochondrion through the base of an enzyme called ATP synthase. The flow of protons makes the stalk subunit rotate, causing the active site of the synthase domain to change shape and phosphorylate adenosine diphosphate – turning it into ATP.

6.3.2 Energy From Inorganic Compounds

Chemolithotrophy is a type of metabolism found in prokaryotes where energy is obtained from the oxidation of inorganic compounds. These organisms can use hydrogen, reduced sulfur compounds (such as sulfide, hydrogen sulfide and thiosulfate), ferrous iron (Fe²⁺) or ammonia as sources of reducing power and they gain energy from the oxidation of these compounds with electron acceptors such as oxygen or nitrite. These microbial processes are important in global biogeochemical cycles such as acetogenesis, nitrification and denitrification and are critical for soil fertility.

6.3.3 Energy From Light

The energy in sunlight is captured by plants, cyanobacteria, purple bacteria, green sulfur bacteria and some protists. This process is often coupled to the conversion of carbon dioxide into organic compounds, as part of photosynthesis, which is discussed below. The energy capture and carbon fixation systems can however operate separately in prokaryotes, as purple bacteria and green sulfur bacteria can use sunlight as a source of energy, while switching between carbon fixation and the fermentation of organic compounds.

In many organisms, the capture of solar energy is similar in principle to oxidative phosphorylation, as it involves the storage of energy as a proton concentration gradient. This proton motive force then drives ATP synthesis. The electrons needed to drive this electron transport chain come from light-gathering proteins called photosynthetic reaction centres. Reaction centers are classed into two types depending on the nature of photosynthetic pigment present, with most photosynthetic bacteria only having one type, while plants and cyanobacteria have two.

In plants, algae, and cyanobacteria, photosystem II uses light energy to remove electrons from water, releasing oxygen as a waste product. The electrons then flow to the cytochrome b6f complex, which uses their energy to pump protons across the thylakoid membrane in the chloroplast. These protons move back through the membrane as they drive the ATP synthase, as before. The electrons then flow through photosystem I and can then either be used to reduce the coenzyme NADP⁺. These coenzyme can be used in the Calvin cycle, which is discussed below, or recycled for further ATP generation.

6.3.4 Anabolism

Anabolism is the set of constructive metabolic processes where the energy released by catabolism is used to synthesize complex molecules. In general, the complex molecules that make up cellular structures are constructed step-by-step from small and simple precursors. Anabolism involves three basic stages. First, the production of precursors such as amino acids, monosaccharides, isoprenoids and nucleotides, secondly, their activation into reactive forms using energy from ATP, and thirdly, the assembly of these precursors into complex molecules such as proteins, polysaccharides, lipids and nucleic acids.

Anabolism in organisms can be different according to the source of constructed molecules in their cells. Autotrophs such as plants can construct the complex organic molecules in cells such as polysaccharides and proteins from simple molecules like carbon dioxide and water. Heterotrophs, on the other hand, require a source of more complex substances, such as monosaccharides and amino acids, to produce these complex molecules. Organisms can be further classified by ultimate source of their energy: photoautotrophs and photoheterotrophs obtain energy from light, whereas chemoautotrophs and chemoheterotrophs obtain energy from inorganic oxidation reactions.

6.3.5 Carbon Fixation

Photosynthesis is the synthesis of carbohydrates from sunlight and carbon dioxide (CO₂). In plants, cyanobacteria

and algae, oxygenic photosynthesis splits water, with oxygen produced as a waste product. This process uses the ATP and NADPH produced by the photosynthetic reaction centres, as described above, to convert CO₂ into glyceralate 3-phosphate, which can then be converted into glucose. This carbon-fixation reaction is carried out by the enzyme RuBisCO as part of the Calvin cycle. Three types of photosynthesis occur in plants, C3 carbon fixation, C4 carbon fixation and CAM photosynthesis. These differ by the route that carbon dioxide takes to the Calvin cycle, with C3 plants fixing CO₂ directly, while C4 and CAM photosynthesis incorporate the CO₂ into other compounds first, as adaptations to deal with intense sunlight and dry conditions.

In photosynthetic prokaryotes the mechanisms of carbon fixation are more diverse. Here, carbon dioxide can be fixed by the Calvin cycle, a reversed citric acid cycle, or the carboxylation of acetyl-CoA. Prokaryotic chemoautotrophs also fix CO₂ through the Calvin cycle, but use energy from inorganic compounds to drive the reaction.

6.3.6 Carbohydrates And Glycans

In carbohydrate anabolism, simple organic acids can be converted into monosaccharides such as glucose and then used to assemble polysaccharides such as starch. The generation of glucose from compounds like pyruvate, lactate, glycerol, glyceralate 3-phosphate and amino acids is called gluconeogenesis. Gluconeogenesis converts pyruvate to glucose-6-phosphate through a series of intermediates, many of which are shared with glycolysis. However, this pathway is not simply glycolysis run in reverse, as several steps are catalyzed by non-glycolytic enzymes. This is important as it allows the formation and breakdown of glucose to be regulated separately, and prevents both pathways from running simultaneously in a futile cycle.

Although fat is a common way of storing energy, in vertebrates such as humans the fatty acids in these stores cannot be converted to glucose through gluconeogenesis as these organisms cannot convert acetyl-CoA into pyruvate; plants do, but animals do not, have the necessary enzymatic machinery. As a result, after long-term starvation, vertebrates need to produce ketone bodies from fatty acids to replace glucose in tissues such as the brain that cannot metabolize fatty acids. In other organisms such as plants and bacteria, this metabolic problem is solved using the glyoxylate cycle, which bypasses the decarboxylation step in the citric acid cycle and allows the transformation of acetyl-CoA to oxaloacetate, where it can be used for the production of glucose. Other than fat, glucose is stored in most tissues, as an energy resource available within the tissue through glycogenesis which was usually being used to maintained glucose level in blood.

Polysaccharides and glycans are made by the sequen-

tial addition of monosaccharides by glycosyltransferase from a reactive sugar-phosphate donor such as uridine diphosphate glucose (UDP-Glc) to an acceptor hydroxyl group on the growing polysaccharide. As any of the hydroxyl groups on the ring of the substrate can be acceptors, the polysaccharides produced can have straight or branched structures. The polysaccharides produced can have structural or metabolic functions themselves, or be transferred to lipids and proteins by enzymes called oligosaccharyltransferases.

6.3.7 Fatty Acids, Isoprenoids and Sterol

Simplified version of the steroid synthesis pathway with the intermediates isopentenyl pyrophosphate (IPP), dimethylallyl pyrophosphate, geranyl pyrophosphate (GPP) and squalene shown. Some intermediates are omitted for clarity. Fatty acids are made by fatty acid synthases that polymerize and then reduce acetyl-CoA units. The acyl chains in the fatty acids are extended by a cycle of reactions that add the acyl group, reduce it to an alcohol, dehydrate it to an alkene group and then reduce it again to an alkane group. The enzymes of fatty acid biosynthesis are divided into two groups: in animals and fungi, all these fatty acid synthase reactions are carried out by a single multifunctional type I protein, while in plant plastids and bacteria separate type II enzymes perform each step in the pathway.

Terpenes and isoprenoids are a large class of lipids that include the carotenoids and form the largest class of plant natural products. These compounds are made by the assembly and modification of isoprene units donated from the reactive precursors isopentenyl pyrophosphate and dimethylallyl pyrophosphate. These precursors can be made in different ways. In animals and archaea, the mevalonate pathway produces these compounds from acetyl-CoA, while in plants and bacteria the non-mevalonate pathway uses pyruvate and glyceraldehyde 3-phosphate as substrates. One important reaction that uses these activated isoprene donors is sterol biosynthesis. Here, the isoprene units are joined together to make squalene and then folded up and formed into a set of rings to make lanosterol. Lanosterol can then be converted into other sterol such as cholesterol and ergosterol.

6.3.8 Proteins

Organisms vary in their ability to synthesize the 20 common amino acids. Most bacteria and plants can synthesize all twenty, but mammals can only synthesize eleven nonessential amino acids, so nine essential amino acids must be obtained from food. Some simple parasites, such as the bacteria *Mycoplasma pneumoniae*, lack all amino acid synthesis and take their amino acids directly from their hosts. All amino acids are synthesized from intermediates

in glycolysis, the citric acid cycle, or the pentose phosphate pathway. Nitrogen is provided by glutamate and glutamine. Nonessential amino acid synthesis depends on the formation of the appropriate alpha-keto acid, which is then transaminated to form an amino acid.

Amino acids are made into proteins by being joined together in a chain of peptide bonds. Each different protein has a unique sequence of amino acid residues: this is its primary structure. Just as the letters of the alphabet can be combined to form an almost endless variety of words, amino acids can be linked in varying sequences to form a huge variety of proteins. Proteins are made from amino acids that have been activated by attachment to a transfer RNA molecule through an ester bond. This aminoacyl-tRNA precursor is produced in an ATP-dependent reaction carried out by an aminoacyl tRNA synthetase. This aminoacyl-tRNA is then a substrate for the ribosome, which joins the amino acid onto the elongating protein chain, using the sequence information in a messenger RNA.

6.3.9 Nucleotide Synthesis And Salvage

Nucleotides are made from amino acids, carbon dioxide and formic acid in pathways that require large amounts of metabolic energy. Consequently, most organisms have efficient systems to salvage preformed nucleotides. Purines are synthesized as nucleosides (bases attached to ribose). Both adenine and guanine are made from the precursor nucleoside inosine monophosphate, which is synthesized using atoms from the amino acids glycine, glutamine, and aspartic acid, as well as formate transferred from the coenzyme tetrahydrofolate. Pyrimidines, on the other hand, are synthesized from the base orotate, which is formed from glutamine and aspartate.

6.4 Enzymes

Enzymes are proteins that act as biological catalysts (biocatalysts). Catalysts accelerate chemical reactions. A catalyst increases the rate of reaction without being consumed in the reaction. In addition, the catalyst lowers the activation energy, but it does not change the energies of the original reactants or products, and so does not change equilibrium. Rather, the reactant energy and the product energy remain the same and only the activation energy is altered (lowered). A catalyst is able to reduce the activation energy by forming a transition state in a more favorable manner. Catalysts, by nature, create a more "comfortable" fit for the substrate of a reaction to progress to a transition state. This is possible due to a release of energy that occurs when the substrate binds to the active site of a catalyst. This energy is known as Binding Energy. Upon binding to a

catalyst, substrates partake in numerous stabilizing forces while within the active site (i.e. Hydrogen bonding, van der Waals forces). Specific and favorable bonding occurs within the active site until the substrate forms to become the high-energy transition state. Forming the transition state is more favorable with the catalyst because the favorable stabilizing interactions within the active site release energy. A chemical reaction is able to manufacture a high-energy transition state molecule more readily when there is a stabilizing fit within the active site of a catalyst. The binding energy of a reaction is this energy released when favorable interactions between substrate and catalyst occur. The binding energy released assists in achieving the unstable transition state. Reactions otherwise without catalysts need a higher input of energy to achieve the transition state. Non-catalyzed reactions do not have free energy available from active site stabilizing interactions, such as catalytic enzyme reactions.

Enzymes are known to catalyze more than 5,000 biochemical reaction types. Other biocatalysts are catalytic RNA molecules, called ribozymes. Enzymes' specificity comes from their unique three-dimensional structures.

The molecules upon which enzymes may act are called substrates, and the enzyme converts the substrates into different molecules known as products. Almost all metabolic processes in the cell need enzyme catalysis in order to occur at rates fast enough to sustain life. Metabolic pathways depend upon enzymes to catalyze individual steps.

Like all catalysts, enzymes increase the reaction rate by lowering its activation energy. Some enzymes can make their conversion of substrate to product occur many millions of times faster. An extreme example is orotidine 5'-phosphate decarboxylase, which allows a reaction that would otherwise take millions of years to occur in milliseconds. Chemically, enzymes are like any catalyst and are not consumed in chemical reactions, nor do they alter the equilibrium of a reaction. Enzymes differ from most other catalysts by being much more specific. Enzyme activity can be affected by other molecules: inhibitors are molecules that decrease enzyme activity, and activators are molecules that increase activity. Many therapeutic drugs and poisons are enzyme inhibitors. An enzyme's activity decreases markedly outside its optimal temperature and pH, and many enzymes are (permanently) denatured when exposed to excessive heat, losing their structure and catalytic properties.

Some enzymes are used commercially, for example, in the synthesis of antibiotics. Some household products use enzymes to speed up chemical reactions: enzymes in biological washing powders break down protein, starch or fat stains on clothes, and enzymes in meat tenderizer break

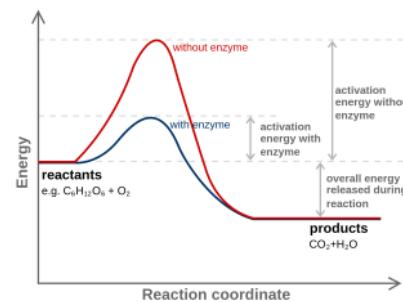


Figure 6.5: Example of an enzyme-catalysed exothermic reaction¹⁰

down proteins into smaller molecules, making the meat easier to chew.

By the late 17th and early 18th centuries, the digestion of meat by stomach secretions and the conversion of starch to sugars by plant extracts and saliva were known but the mechanisms by which these occurred had not been identified.

French chemist Anselme Payen¹¹ was the first to discover an enzyme, diastase, in 1833. A few decades later, when studying the fermentation of sugar to alcohol by yeast, Louis Pasteur concluded that this fermentation was caused by a vital force contained within the yeast cells called "ferments", which were thought to function only within living organisms. He wrote that "alcoholic fermentation is an act correlated with the life and organization of the yeast cells, not with the death or putrefaction of the cells."

Eduard Buchner submitted his first paper on the study of yeast extracts in 1897. In a series of experiments at the University of Berlin, he found that sugar was fermented by yeast extracts even when there were no living yeast cells in the mixture. He named the enzyme that brought about the fermentation of sucrose "zymase". In 1907, he received the Nobel Prize in Chemistry for "his discovery of cell-free fermentation". Following Buchner's example, enzymes are usually named according to the reaction they carry out: the suffix -ase is combined with the name of the substrate (e.g., lactase is the enzyme that cleaves lactose) or to the type of reaction (e.g., DNA polymerase forms DNA polymers).

The biochemical identity of enzymes was still unknown in the early 1900s. Many scientists observed that enzymatic activity was associated with proteins, but others (such as Nobel laureate Richard Willstätter¹²) argued that proteins were merely carriers for the true enzymes and

¹¹https://en.wikipedia.org/wiki/Anselme_Payen

¹²https://en.wikipedia.org/wiki/Richard_Willst%C3%A4tter

that proteins per se were incapable of catalysis. In 1926, James B. Sumner¹³ showed that the enzyme urease was a pure protein and crystallized it; he did likewise for the enzyme catalase in 1937. The conclusion that pure proteins can be enzymes was definitively demonstrated by John Howard Northrop¹⁴ and Wendell Meredith Stanley¹⁵, who worked on the digestive enzymes pepsin (1930), trypsin and chymotrypsin. These three scientists were awarded the 1946 Nobel Prize in Chemistry.

The discovery that enzymes could be crystallized eventually allowed their structures to be solved by x-ray crystallography. This was first done for lysozyme, an enzyme found in tears, saliva and egg whites that digests the coating of some bacteria; the structure was solved by a group led by David Chilton Phillips¹⁶ and published in 1965. This high-resolution structure of lysozyme marked the beginning of the field of structural biology and the effort to understand how enzymes work at an atomic level of detail.

An enzyme's name is often derived from its substrate or the chemical reaction it catalyzes, with the word ending in -ase. Examples are lactase, alcohol dehydrogenase and DNA polymerase. Different enzymes that catalyze the same chemical reaction are called isozymes.

Enzymes are generally globular proteins, acting alone or in larger complexes. The sequence of the amino acids specifies the structure which in turn determines the catalytic activity of the enzyme. Although structure determines function, a novel enzymatic activity cannot yet be predicted from structure alone. Enzyme structures unfold (denature) when heated or exposed to chemical denaturants and this disruption to the structure typically causes a loss of activity. Enzyme denaturation is normally linked to temperatures above a species' normal level; as a result, enzymes from bacteria living in volcanic environments such as hot springs are prized by industrial users for their ability to function at high temperatures, allowing enzyme-catalysed reactions to be operated at a very high rate.

Enzymes are usually much larger than their substrates. Sizes range from just 62 amino acid residues, for the monomer of 4-oxalocrotonate tautomerase, to over 2,500 residues in the animal fatty acid synthase. Only a small portion of their structure (around 2–4 amino acids) is directly involved in catalysis: the catalytic site. This catalytic site is located next to one or more binding sites where residues orient the substrates. The catalytic site and binding site together compose the enzyme's active site. The remaining majority of the enzyme structure serves to maintain the precise orientation and dynamics

of the active site.

In some enzymes, no amino acids are directly involved in catalysis; instead, the enzyme contains sites to bind and orient catalytic cofactors. Enzyme structures may also contain allosteric sites where the binding of a small molecule causes a conformational change that increases or decreases activity.

A small number of RNA-based biological catalysts called ribozymes exist, which again can act alone or in complex with proteins. The most common of these is the ribosome which is a complex of protein and catalytic RNA components.

Enzymes must bind their substrates before they can catalyse any chemical reaction. Enzymes are usually very specific as to what substrates they bind and then the chemical reaction catalysed. Specificity is achieved by binding pockets with complementary shape, charge and hydrophilic/hydrophobic characteristics to the substrates. Enzymes can therefore distinguish between very similar substrate molecules to be chemoselective, regioselective and stereospecific.

Some of the enzymes showing the highest specificity and accuracy are involved in the copying and expression of the genome. Some of these enzymes have "proof-reading" mechanisms. Here, an enzyme such as DNA polymerase catalyzes a reaction in a first step and then checks that the product is correct in a second step. This two-step process results in average error rates of less than 1 error in 100 million reactions in high-fidelity mammalian polymerases. Similar proofreading mechanisms are also found in RNA polymerase, aminoacyl tRNA synthetases and ribosomes.

Conversely, some enzymes display enzyme promiscuity, having broad specificity and acting on a range of different physiologically relevant substrates. Many enzymes possess small side activities which arose fortuitously (i.e. neutrally), which may be the starting point for the evolutionary selection of a new function.

To explain the observed specificity of enzymes, in 1894 Emil Fischer¹⁷ proposed that both the enzyme and the substrate possess specific complementary geometric shapes that fit exactly into one another. This is often referred to as "the lock and key" model. This early model explains enzyme specificity, but fails to explain the stabilization of the transition state that enzymes achieve.

In 1958, Daniel Koshland¹⁸ suggested a modification to the lock and key model: since enzymes are rather flexible structures, the active site is continuously reshaped by interactions with the substrate as the substrate interacts

¹³https://en.wikipedia.org/wiki/James_B._Sumner

¹⁴https://en.wikipedia.org/wiki/John_Howard_Northrop

¹⁵https://en.wikipedia.org/wiki/Wendell_Meredith_Stanley

¹⁶https://en.wikipedia.org/wiki/David_Chilton_Phillips

¹⁷https://en.wikipedia.org/wiki/Emil_Fischer

¹⁸https://en.wikipedia.org/wiki/Daniel_E._Koshland_Jr.

with the enzyme. As a result, the substrate does not simply bind to a rigid active site; the amino acid side-chains that make up the active site are molded into the precise positions that enable the enzyme to perform its catalytic function. In some cases, such as glycosidases, the substrate molecule also changes shape slightly as it enters the active site. The active site continues to change until the substrate is completely bound, at which point the final shape and charge distribution is determined. Induced fit may enhance the fidelity of molecular recognition in the presence of competition and noise via the conformational proofreading mechanism.

Enzymes can accelerate reactions in several ways, all of which lower the activation energy (ΔG , Gibbs free energy)

- By stabilizing the transition state:
- Creating an environment with a charge distribution complementary to that of the transition state to lower its energy
- By providing an alternative reaction pathway:
- Temporarily reacting with the substrate, forming a covalent intermediate to provide a lower energy transition state
- By destabilising the substrate ground state:
- Distorting bound substrate(s) into their transition state form to reduce the energy required to reach the transition state
- By orienting the substrates into a productive arrangement to reduce the reaction entropy change (the contribution of this mechanism to catalysis is relatively small)

Enzymes may use several of these mechanisms simultaneously. For example, proteases such as trypsin perform covalent catalysis using a catalytic triad, stabilise charge build-up on the transition states using an oxyanion hole, complete hydrolysis using an oriented water substrate.

Enzymes are not rigid, static structures; instead they have complex internal dynamic motions – that is, movements of parts of the enzyme's structure such as individual amino acid residues, groups of residues forming a protein loop or unit of secondary structure, or even an entire protein domain. These motions give rise to a conformational ensemble of slightly different structures that interconvert with one another at equilibrium. Different states within this ensemble may be associated with different aspects of an enzyme's function. For example, different conformations of the enzyme dihydrofolate reductase are associated with the substrate binding, catalysis, cofactor release, and product release steps of the catalytic cycle, consistent with catalytic resonance theory.

Substrate presentation is a process where the enzyme is sequestered away from its substrate. Enzymes can be

sequestered to the plasma membrane away from a substrate in the nucleus or cytosol. Or within the membrane, an enzyme can be sequestered into lipid rafts away from its substrate in the disordered region. When the enzyme is released it mixes with its substrate. Alternatively, the enzyme can be sequestered near its substrate to activate the enzyme. For example, the enzyme can be soluble and upon activation bind to a lipid in the plasma membrane and then act upon molecules in the plasma membrane.

6.4.1 Allosteric Modulation

Allosteric sites are pockets on the enzyme, distinct from the active site, that bind to molecules in the cellular environment. These molecules then cause a change in the conformation or dynamics of the enzyme that is transduced to the active site and thus affects the reaction rate of the enzyme. In this way, allosteric interactions can either inhibit or activate enzymes. Allosteric interactions with metabolites upstream or downstream in an enzyme's metabolic pathway cause feedback regulation, altering the activity of the enzyme according to the flux through the rest of the pathway.

6.4.2 Cofactors of Enzymes and Coenzymes

Some enzymes do not need additional components to show full activity. Others require non-protein molecules called cofactors to be bound for activity. Cofactors can be either inorganic (e.g., metal ions and iron-sulfur clusters) or organic compounds (e.g., flavin and heme). These cofactors serve many purposes; for instance, metal ions can help in stabilizing nucleophilic species within the active site. Organic cofactors can be either coenzymes, which are released from the enzyme's active site during the reaction, or prosthetic groups, which are tightly bound to an enzyme. Organic prosthetic groups can be covalently bound (e.g., biotin in enzymes such as pyruvate carboxylase).

An example of an enzyme that contains a cofactor is carbonic anhydrase, which uses a zinc cofactor bound as part of its active site. These tightly bound ions or molecules are usually found in the active site and are involved in catalysis. For example, flavin and heme cofactors are often involved in redox reactions.

Enzymes that require a cofactor but do not have one bound are called apoenzymes or apoproteins. An enzyme together with the cofactor(s) required for activity is called a holoenzyme (or haloenzyme). The term holoenzyme can also be applied to enzymes that contain multiple protein subunits, such as the DNA polymerases; here the holoenzyme is the complete complex containing all the subunits needed for activity.

Coenzymes are small organic molecules that can be loosely or tightly bound to an enzyme. Coenzymes transport chemical groups from one enzyme to another. Examples include NADH, NADPH and adenosine triphosphate (ATP). Some coenzymes, such as flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD), thiamine pyrophosphate (TPP), and tetrahydrofolate (THF), are derived from vitamins. These coenzymes cannot be synthesized by the body de novo and closely related compounds (vitamins) must be acquired from the diet. The chemical groups carried include:

- the hydride ion (H^-), carried by NAD or NAD^+
- the phosphate group, carried by adenosine triphosphate
- the acetyl group, carried by coenzyme A
- formyl, methenyl or methyl groups, carried by folic acid and
- the methyl group, carried by S-adenosylmethionine

Since coenzymes are chemically changed as a consequence of enzyme action, it is useful to consider coenzymes to be a special class of substrates, or second substrates, which are common to many different enzymes. For example, about 1000 enzymes are known to use the coenzyme NADH.

Coenzymes are usually continuously regenerated and their concentrations maintained at a steady level inside the cell. For example, NADPH is regenerated through the pentose phosphate pathway and S-adenosylmethionine by methionine adenosyltransferase. This continuous regeneration means that small amounts of coenzymes can be used very intensively. For example, the human body turns over its own weight in ATP each day.

6.4.3 Inhibition of Enzymes

An enzyme binding site that would normally bind substrate can alternatively bind a competitive inhibitor, preventing substrate access. Dihydrofolate reductase is inhibited by methotrexate which prevents binding of its substrate, folic acid. Binding site in blue, inhibitor in green, and substrate in black. (PDB: 4QI9) Two dimensional representations of the chemical structure of folic acid and methotrexate highlighting the differences between these two substances (amidation of pyrimidine and methylation of secondary amine). The coenzyme folic acid (left) and the anti-cancer drug methotrexate (right) are very similar in structure (differences shown in green). As a result, methotrexate is a competitive inhibitor of many enzymes that use folates. Main article: Enzyme inhibitor Enzyme reaction rates can be decreased by various types of enzyme inhibitors.

A competitive inhibitor and substrate cannot bind to the enzyme at the same time. Often competitive inhibitors

strongly resemble the real substrate of the enzyme. For example, the drug methotrexate is a competitive inhibitor of the enzyme dihydrofolate reductase, which catalyzes the reduction of dihydrofolate to tetrahydrofolate. The similarity between the structures of dihydrofolate and this drug are shown in the accompanying figure. This type of inhibition can be overcome with high substrate concentration. In some cases, the inhibitor can bind to a site other than the binding-site of the usual substrate and exert an allosteric effect to change the shape of the usual binding-site.

A non-competitive inhibitor binds to a site other than where the substrate binds. The substrate still binds with its usual affinity and hence K_m remains the same. However the inhibitor reduces the catalytic efficiency of the enzyme so that V_{max} is reduced. In contrast to competitive inhibition, non-competitive inhibition cannot be overcome with high substrate concentration.

An uncompetitive inhibitor cannot bind to the free enzyme, only to the enzyme-substrate complex; hence, these types of inhibitors are most effective at high substrate concentration. In the presence of the inhibitor, the enzyme-substrate complex is inactive. This type of inhibition is rare.

A mixed inhibitor binds to an allosteric site and the binding of the substrate and the inhibitor affect each other. The enzyme's function is reduced but not eliminated when bound to the inhibitor. This type of inhibitor does not follow the Michaelis-Menten equation.

An irreversible inhibitor permanently inactivates the enzyme, usually by forming a covalent bond to the protein. Penicillin and aspirin are common drugs that act in this manner.

In many organisms, inhibitors may act as part of a feedback mechanism. If an enzyme produces too much of one substance in the organism, that substance may act as an inhibitor for the enzyme at the beginning of the pathway that produces it, causing production of the substance to slow down or stop when there is sufficient amount. This is a form of negative feedback. Major metabolic pathways such as the citric acid cycle make use of this mechanism.

Since inhibitors modulate the function of enzymes they are often used as drugs. Many such drugs are reversible competitive inhibitors that resemble the enzyme's native substrate, similar to methotrexate above; other well-known examples include statins used to treat high cholesterol, and protease inhibitors used to treat retroviral infections such as HIV. A common example of an irreversible inhibitor that is used as a drug is aspirin, which inhibits the COX-1 and COX-2 enzymes that produce the inflammation messenger prostaglandin. Other enzyme inhibitors are poisons. For example, the poison

cyanide is an irreversible enzyme inhibitor that combines with the copper and iron in the active site of the enzyme cytochrome c oxidase and blocks cellular respiration.

6.4.4 Factors Affecting Enzyme Activity

As enzymes are made up of proteins, their actions are sensitive to change in many physio chemical factors such as pH, temperature, substrate concentration, etc.

Enzymes serve a wide variety of functions inside living organisms. They are indispensable for signal transduction and cell regulation, often via kinases and phosphatases. They also generate movement, with myosin hydrolyzing ATP to generate muscle contraction, and also transport cargo around the cell as part of the cytoskeleton. Other ATPases in the cell membrane are ion pumps involved in active transport. Enzymes are also involved in more exotic functions, such as luciferase generating light in fireflies. Viruses can also contain enzymes for infecting cells, such as the HIV integrase and reverse transcriptase, or for viral release from cells, like the influenza virus neuraminidase.

An important function of enzymes is in the digestive systems of animals. Enzymes such as amylases and proteases break down large molecules (starch or proteins, respectively) into smaller ones, so they can be absorbed by the intestines. Starch molecules, for example, are too large to be absorbed from the intestine, but enzymes hydrolyze the starch chains into smaller molecules such as maltose and eventually glucose, which can then be absorbed. Different enzymes digest different food substances. In ruminants, which have herbivorous diets, microorganisms in the gut produce another enzyme, cellulase, to break down the cellulose cell walls of plant fiber.

Several enzymes can work together in a specific order, creating metabolic pathways. In a metabolic pathway, one enzyme takes the product of another enzyme as a substrate. After the catalytic reaction, the product is then passed on to another enzyme. Sometimes more than one enzyme can catalyze the same reaction in parallel; this can allow more complex regulation with, for example, a low constant activity provided by one enzyme but an inducible high activity from a second enzyme.

Enzymes determine what steps occur in these pathways. Without enzymes, metabolism would neither progress through the same steps and could not be regulated to serve the needs of the cell. Most central metabolic pathways are regulated at a few key steps, typically through enzymes whose activity involves the hydrolysis of ATP. Because this reaction releases so much energy, other reactions that are thermodynamically unfavorable can be coupled to ATP hydrolysis, driving the overall series of linked metabolic reactions.

6.4.5 Control of activity

There are five main ways that enzyme activity is controlled in the cell.

Enzymes can be either activated or inhibited by other molecules. For example, the end product(s) of a metabolic pathway are often inhibitors for one of the first enzymes of the pathway (usually the first irreversible step, called committed step), thus regulating the amount of end product made by the pathways. Such a regulatory mechanism is called a negative feedback mechanism, because the amount of the end product produced is regulated by its own concentration. Negative feedback mechanism can effectively adjust the rate of synthesis of intermediate metabolites according to the demands of the cells. This helps with effective allocations of materials and energy economy, and it prevents the excess manufacture of end products. Like other homeostatic devices, the control of enzymatic action helps to maintain a stable internal environment in living organisms.

6.4.6 Post-translational Modification

Examples of post-translational modification include phosphorylation, myristylation and glycosylation. For example, in the response to insulin, the phosphorylation of multiple enzymes, including glycogen synthase, helps control the synthesis or degradation of glycogen and allows the cell to respond to changes in blood sugar. Another example of post-translational modification is the cleavage of the polypeptide chain. Chymotrypsin, a digestive protease, is produced in inactive form as chymotrypsinogen in the pancreas and transported in this form to the stomach where it is activated. This stops the enzyme from digesting the pancreas or other tissues before it enters the gut. This type of inactive precursor to an enzyme is known as a zymogen or proenzyme.

6.4.7 Quantity

Enzyme production (transcription and translation of enzyme genes) can be enhanced or diminished by a cell in response to changes in the cell's environment. This form of gene regulation is called enzyme induction. For example, bacteria may become resistant to antibiotics such as penicillin because enzymes called beta-lactamases are induced that hydrolyse the crucial beta-lactam ring within the penicillin molecule. Another example comes from enzymes in the liver called cytochrome P450 oxidases, which are important in drug metabolism. Induction or inhibition of these enzymes can cause drug interactions. Enzyme levels can also be regulated by changing the rate of enzyme degradation. The opposite of enzyme induction is enzyme repression.

6.4.8 Subcellular Distribution

Enzymes can be compartmentalized, with different metabolic pathways occurring in different cellular compartments. For example, fatty acids are synthesized by one set of enzymes in the cytosol, endoplasmic reticulum and Golgi and used by a different set of enzymes as a source of energy in the mitochondrion, through β -oxidation. In addition, trafficking of the enzyme to different compartments may change the degree of protonation (e.g., the neutral cytoplasm and the acidic lysosome) or oxidative state (e.g., oxidizing periplasm or reducing cytoplasm) which in turn affects enzyme activity. In contrast to partitioning into membrane bound organelles, enzyme subcellular localisation may also be altered through polymerisation of enzymes into macromolecular cytoplasmic filaments.

6.4.9 Organ Specialization

In multicellular eukaryotes, cells in different organs and tissues have different patterns of gene expression and therefore have different sets of enzymes (known as isozymes) available for metabolic reactions. This provides a mechanism for regulating the overall metabolism of the organism. For example, hexokinase, the first enzyme in the glycolysis pathway, has a specialized form called glucokinase expressed in the liver and pancreas that has a lower affinity for glucose yet is more sensitive to glucose concentration. This enzyme is involved in sensing blood sugar and regulating insulin production.

6.4.10 Coenzymes

Metabolism involves a vast array of chemical reactions, but most fall under a few basic types of reactions that involve the transfer of functional groups of atoms and their bonds within molecules. This common chemistry allows cells to use a small set of metabolic intermediates to carry chemical groups between different reactions. These group-transfer intermediates are called coenzymes. Each class of group-transfer reactions is carried out by a particular coenzyme, which is the substrate for a set of enzymes that produce it, and a set of enzymes that consume it. These coenzymes are therefore continuously made, consumed and then recycled.

One central coenzyme is adenosine triphosphate (ATP), the universal energy currency of cells. This nucleotide is used to transfer chemical energy between different chemical reactions. There is only a small amount of ATP in cells, but as it is continuously regenerated, the human body can use about its own weight in ATP per day. ATP acts as a bridge between catabolism and anabolism. Catabolism breaks down molecules, and anabolism puts

them together. Catabolic reactions generate ATP, and anabolic reactions consume it. It also serves as a carrier of phosphate groups in phosphorylation reactions.

A vitamin is an organic compound needed in small quantities that cannot be made in cells. In human nutrition, most vitamins function as coenzymes after modification; for example, all water-soluble vitamins are phosphorylated or are coupled to nucleotides when they are used in cells. Nicotinamide adenine dinucleotide (NAD^+), a derivative of vitamin B3 (niacin), is an important coenzyme that acts as a hydrogen acceptor. Hundreds of separate types of dehydrogenases remove electrons from their substrates and reduce NAD^+ into NADH. This reduced form of the coenzyme is then a substrate for any of the reductases in the cell that need to reduce their substrates. Nicotinamide adenine dinucleotide exists in two related forms in the cell, NADH and NADPH. The NAD^+/NADH form is more important in catabolic reactions, while $\text{NADP}^+/\text{NADPH}$ is used in anabolic reactions.

6.4.11 Mineral Cofactors

Inorganic elements play critical roles in metabolism; some are abundant (e.g. sodium and potassium) while others function at minute concentrations. About 99% of a human's body weight is made up of the elements carbon, nitrogen, calcium, sodium, chlorine, potassium, hydrogen, phosphorus, oxygen and sulfur. Organic compounds (proteins, lipids and carbohydrates) contain the majority of the carbon and nitrogen; most of the oxygen and hydrogen is present as water.

The abundant inorganic elements act as electrolytes. The most important ions are sodium, potassium, calcium, magnesium, chloride, phosphate and the organic ion bicarbonate. The maintenance of precise ion gradients across cell membranes maintains osmotic pressure and pH. Ions are also critical for nerve and muscle function, as action potentials in these tissues are produced by the exchange of electrolytes between the extracellular fluid and the cell's fluid, the cytosol. Electrolytes enter and leave cells through proteins in the cell membrane called ion channels. For example, muscle contraction depends upon the movement of calcium, sodium and potassium through ion channels in the cell membrane and T-tubules.

Transition metals are usually present as trace elements in organisms, with zinc and iron being most abundant of those. These metals are used in some proteins as cofactors and are essential for the activity of enzymes such as catalase and oxygen-carrier proteins such as hemoglobin. Metal cofactors are bound tightly to specific sites in proteins; although enzyme cofactors can be modified during catalysis, they always return to their original state by the end of the reaction catalyzed. Metal micronutrients are

taken up into organisms by specific transporters and bind to storage proteins such as ferritin or metallothionein when not in use.

Chapter 7

Photosynthesis And Cellular Respiration

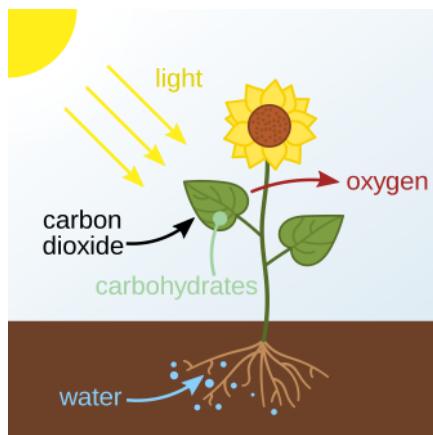


Figure 7.1: Schematic of photosynthesis in plants. The carbohydrates produced are stored in or used by the plant which in turn may provide food for heterotrophic organisms such as animals.¹

Plants acquire energy from sunlight through photosynthesis. Animals get their energy by breaking down chemical bonds in nutrients during cellular respiration. In eukaryotic cells, photosynthesis happens in chloroplasts, cellular respiration in mitochondria. In both photosynthesis and cellular respiration, ATP production involves an electron transfer chain and chemiosmosis.

7.1 The Electron Transport Chain And Chemiosmosis

The electron transport chain (ETC) is a series of complexes that transfer electrons from electron donors to electron acceptors via redox (both reduction and oxidation occurring simultaneously) reactions, and couples this electron transfer with the transfer of protons (H^+ ions) across a membrane. The electron transport chain is built up of peptides, enzymes, and other molecules.

The flow of electrons through the electron transport

chain is an exergonic process. The energy from the redox reactions create an electrochemical proton gradient that drives the synthesis of adenosine triphosphate (ATP). In aerobic respiration, the flow of electrons terminates with molecular oxygen being the final electron acceptor. In anaerobic respiration, other organic or inorganic electron acceptors are used, such as lactic acid and sulfate, for example.

In the electron transport chain, the redox reactions are driven by the Gibbs free energy state of the components. Gibbs free energy is related to a quantity called the redox potential. The complexes in the electron transport chain harvest the energy of the redox reactions that occur when transferring electrons from a low redox potential to a higher redox potential, creating an electrochemical gradient. It is the electrochemical gradient created that drives the synthesis of ATP via coupling with oxidative phosphorylation with ATP synthase.

The electron transport chain, and site of oxidative phosphorylation is found on the inner mitochondrial membrane. The energy stored from the process of respiration in reduced compounds (such as NADH and FADH) is used by the electron transport chain to pump protons into the inter membrane space, generating the electrochemical gradient over the inner mitochondrial membrane.

In photosynthetic eukaryotes, the electron transport chain is found on the thylakoid membrane. Here, light energy drives the reduction of components of the electron transport chain and therefore causes subsequent synthesis of ATP. The electron transport chain, and site of oxidative phosphorylation is found on the inner mitochondrial membrane. The energy stored from the process of respiration in reduced compounds (such as NADH and FADH) is used by the electron transport chain to pump protons into the inter membrane space, generating the electrochemical gradient over the inner mitochondrial membrane.

Hydrogen ions, or protons, will diffuse from an area of high proton concentration to an area of lower proton concentration, and an electrochemical concentration gra-

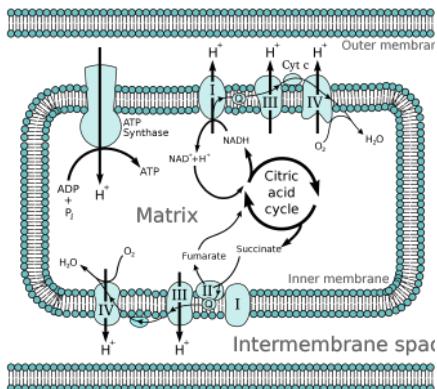


Figure 7.2: The electron transport chain² in the mitochondrion is the site of oxidative phosphorylation in eukaryotes. The NADH and succinate generated in the citric acid cycle are oxidized, providing energy to power ATP synthase.

dient of protons across a membrane can be harnessed to make ATP. This process is related to osmosis, the diffusion of water across a membrane, which is why it is called “chemiosmosis”.

The formation of adenosine triphosphate (ATP) by the movement of hydrogen ions (H⁺) across a membrane during cellular respiration or photosynthesis is an example of chemiosmosis. ATP synthase is the enzyme that makes ATP by chemiosmosis. It allows protons to pass through the membrane and uses the free energy difference to phosphorylate adenosine diphosphate (ADP), making ATP. The generation of ATP by chemiosmosis occurs in mitochondria and chloroplasts, as well as in most bacteria and archaea, an electron transport chain pumps H⁺ ions in the thylakoid spaces through thylakoid membranes to stroma (fluid). The energy from the electron movement through electron transport chains cross through ATP synthase which allows the proton to pass through them and use this free energy difference to photophosphorylate ADP making ATP.

7.2 Photosynthesis

Photosynthesis is a process used by plants and other organisms to convert light energy into chemical energy that can later be released to fuel the organisms' activities. This chemical energy is stored in carbohydrate molecules, such as sugars, which are synthesized from carbon dioxide and water – hence the name photosynthesis, from the Greek φῶς (φῶς), “light”, and σύνθεσις (σύνθεσις), “putting together”. In most cases, oxygen is also released as a waste product. Most plants, most algae, and cyanobacteria perform photosynthesis; such organisms are called photoautotrophs. Photosynthesis is largely responsible for produc-

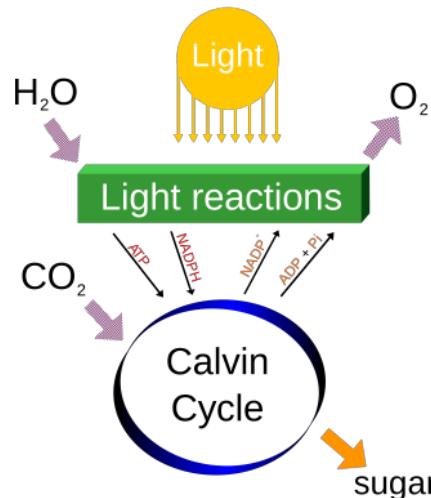


Figure 7.3: Photosynthesis changes sunlight into chemical energy, splits water to liberate O₂, and fixes CO₂ into sugar.³
(#fig:simplephotooverview)

ing and maintaining the oxygen content of the Earth's atmosphere, and supplies most of the energy necessary for life on Earth.

Although photosynthesis is performed differently by different species, the process always begins when energy from light is absorbed by proteins called reaction centres that contain green chlorophyll pigments. In plants, these proteins are held inside organelles called chloroplasts, which are most abundant in leaf cells, while in bacteria they are embedded in the plasma membrane. In these light-dependent reactions, some energy is used to strip electrons from suitable substances, such as water, producing oxygen gas. The hydrogen freed by the splitting of water is used in the creation of two further compounds that serve as short-term stores of energy, enabling its transfer to drive other reactions: these compounds are reduced nicotinamide adenine dinucleotide phosphate (NADPH) and adenosine triphosphate (ATP), the “energy currency” of cells.

In plants, algae and cyanobacteria, long-term energy storage in the form of sugars is produced by a subsequent sequence of light-independent reactions called the Calvin cycle; some bacteria use different mechanisms, such as the reverse Krebs cycle, to achieve the same end. In the Calvin cycle, atmospheric carbon dioxide is incorporated into already existing organic carbon compounds, such as ribulose bisphosphate (RuBP). Using the ATP and NADPH produced by the light-dependent reactions, the resulting compounds are then reduced and removed to form further carbohydrates, such as glucose.

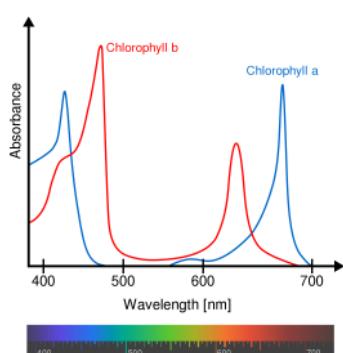


Figure 7.4: Absorbance spectra of free chlorophyll⁴ a (blue) and b (red) in a solvent. The action spectra of chlorophyll molecules are slightly modified *in vivo* depending on specific pigment–protein interactions.

The first photosynthetic organisms probably evolved early in the evolutionary history of life and most likely used reducing agents such as hydrogen or hydrogen sulfide, rather than water, as sources of electrons. Cyanobacteria appeared later; the excess oxygen they produced contributed directly to the oxygenation of the Earth, which rendered the evolution of complex life possible. Today, the average rate of energy capture by photosynthesis globally is approximately 130 terawatts, which is about eight times the current power consumption of human civilization. Photosynthetic organisms also convert around 100–115 billion tons (91–104 petagrams) of carbon into biomass per year.

Photosynthetic organisms are photoautotrophs, which means that they are able to synthesize food directly from carbon dioxide and water using energy from light. However, not all organisms use carbon dioxide as a source of carbon atoms to carry out photosynthesis; photoheterotrophs use organic compounds, rather than carbon dioxide, as a source of carbon. In plants, algae, and cyanobacteria, photosynthesis releases oxygen. This is called oxygenic photosynthesis and is by far the most common type of photosynthesis used by living organisms. Although there are some differences between oxygenic photosynthesis in plants, algae, and cyanobacteria, the overall process is quite similar in these organisms. There are also many varieties of anoxygenic photosynthesis, used mostly by certain types of bacteria, which consume carbon dioxide but do not release oxygen.

Carbon dioxide is converted into sugars in a process called carbon fixation; photosynthesis captures energy from sunlight to convert carbon dioxide into carbohydrate. Carbon fixation is an endothermic redox reaction. In

general outline, photosynthesis is the opposite of cellular respiration: while photosynthesis is a process of reduction of carbon dioxide to carbohydrate, cellular respiration is the oxidation of carbohydrate or other nutrients to carbon dioxide. Nutrients used in cellular respiration include carbohydrates, amino acids and fatty acids. These nutrients are oxidized to produce carbon dioxide and water, and to release chemical energy to drive the organism's metabolism. Photosynthesis and cellular respiration are distinct processes, as they take place through different sequences of chemical reactions and in different cellular compartments.

Photosynthesis occurs in two stages. In the first stage, light-dependent reactions or light reactions capture the energy of light and use it to make the energy-storage molecules ATP and NADPH. During the second stage, the light-independent reactions use these products to capture and reduce carbon dioxide.

Most organisms that utilize oxygenic photosynthesis use visible light for the light-dependent reactions, although at least three use shortwave infrared or, more specifically, far-red radiation.

Some organisms employ even more radical variants of photosynthesis. Some archaea use a simpler method that employs a pigment similar to those used for vision in animals. The bacteriorhodopsin changes its configuration in response to sunlight, acting as a proton pump. This produces a proton gradient more directly, which is then converted to chemical energy. The process does not involve carbon dioxide fixation and does not release oxygen, and seems to have evolved separately from the more common types of photosynthesis.

7.2.1 Photosynthetic Membranes And Organelles

In photosynthetic bacteria, the proteins that gather light for photosynthesis are embedded in cell membranes. In its simplest form, this involves the membrane surrounding the cell itself. However, the membrane may be tightly folded into cylindrical sheets called thylakoids, or bunched up into round vesicles called intracytoplasmic membranes. These structures can fill most of the interior of a cell, giving the membrane a very large surface area and therefore increasing the amount of light that the bacteria can absorb.

In plants and algae, photosynthesis takes place in organelles called chloroplasts. A typical plant cell contains about 10 to 100 chloroplasts. The chloroplast is enclosed by a membrane. This membrane is composed of a phospholipid inner membrane, a phospholipid outer membrane, and an intermembrane space. Enclosed by the membrane is an aqueous fluid called the stroma. Embedded within

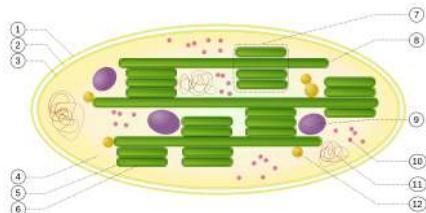


Figure 7.5: Chloroplast ultrastructure⁵: 1. outer membrane 2. intermembrane space 3. inner membrane (1+2+3: envelope) 4. stroma (aqueous fluid) 5. thylakoid lumen (inside of thylakoid) 6. thylakoid membrane 7. granum (stack of thylakoids) 8. thylakoid (lamella) 9. starch 10. ribosome 11. plastidial DNA 12. plastoglobule (drop of lipids)

the stroma are stacks of thylakoids (grana), which are the site of photosynthesis. The thylakoids appear as flattened disks. The thylakoid itself is enclosed by the thylakoid membrane, and within the enclosed volume is a lumen or thylakoid space. Embedded in the thylakoid membrane are integral and peripheral membrane protein complexes of the photosynthetic system.

Plants absorb light primarily using the pigment chlorophyll. The green part of the light spectrum is not absorbed but is reflected which is the reason that most plants have a green color. Besides chlorophyll, plants also use pigments such as carotenes and xanthophylls. Algae also use chlorophyll, but various other pigments are present, such as phycocyanin, carotenes, and xanthophylls in green algae, phycoerythrin in red algae (rhodophytes) and fucoxanthin in brown algae and diatoms resulting in a wide variety of colors.

These pigments are embedded in plants and algae in complexes called antenna proteins. In such proteins, the pigments are arranged to work together. Such a combination of proteins is also called a light-harvesting complex.

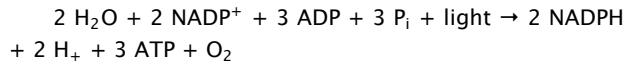
Although all cells in the green parts of a plant have chloroplasts, the majority of those are found in specially adapted structures called leaves. Certain species adapted to conditions of strong sunlight and aridity, such as many Euphorbia and cactus species, have their main photosynthetic organs in their stems. The cells in the interior tissues of a leaf, called the mesophyll, can contain between 450,000 and 800,000 chloroplasts for every square millimeter of leaf. The surface of the leaf is coated with a water-resistant waxy cuticle that protects the leaf from excessive evaporation of water and decreases the absorption of ultraviolet or blue light to reduce heating. The transparent epidermis layer allows light to pass through to the palisade mesophyll cells where most of the photosynthesis takes place.



Figure 7.6: The Z scheme⁶

In the light-dependent reactions, one molecule of the pigment chlorophyll absorbs one photon and loses one electron. This electron is passed to a modified form of chlorophyll called pheophytin, which passes the electron to a quinone molecule, starting the flow of electrons down an electron transport chain that leads to the ultimate reduction of NADP to NADPH. In addition, this creates a proton gradient (energy gradient) across the chloroplast membrane, which is used by ATP synthase in the synthesis of ATP. The chlorophyll molecule ultimately regains the electron it lost when a water molecule is split in a process called photolysis, which releases a dioxygen (O₂) molecule as a waste product.

The overall equation for the light-dependent reactions under the conditions of non-cyclic electron flow in green plants is:



Not all wavelengths of light can support photosynthesis. The photosynthetic action spectrum depends on the type of accessory pigments present. For example, in green plants, the action spectrum resembles the absorption spectrum for chlorophylls and carotenoids with absorption peaks in violet-blue and red light. In red algae, the action spectrum is blue-green light, which allows these algae to use the blue end of the spectrum to grow in the deeper waters that filter out the longer wavelengths (red light) used by above ground green plants. The non-absorbed part of the light spectrum is what gives photosynthetic organisms their color (e.g., green plants, red algae, purple bacteria) and is the least effective for photosynthesis in the respective organisms.

7.2.2 Z Scheme

In plants, light-dependent reactions occur in the thylakoid membranes of the chloroplasts where they drive the synthesis of ATP and NADPH. The light-dependent reactions are of two forms: cyclic and non-cyclic.

In the non-cyclic reaction, the photons are captured in the light-harvesting antenna complexes of photosystem II by chlorophyll and other accessory pigments. The absorption of a photon by the antenna complex frees an electron by a process called photoinduced charge separation. The

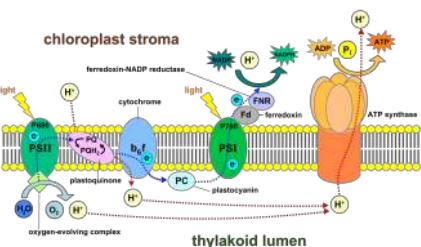


Figure 7.7: Light-dependent reactions of photosynthesis at the thylakoid membrane⁷

antenna system is at the core of the chlorophyll molecule of the photosystem II reaction center. That freed electron is transferred to the primary electron-acceptor molecule, pheophytin. As the electrons are shuttled through an electron transport chain (the so-called Z-scheme shown in the diagram), it initially functions to generate a chemiosmotic potential by pumping proton cations (H^+) across the membrane and into the thylakoid space. An ATP synthase enzyme uses that chemiosmotic potential to make ATP during photophosphorylation, whereas NADPH is a product of the terminal redox reaction in the Z-scheme. The electron enters a chlorophyll molecule in Photosystem I. There it is further excited by the light absorbed by that photosystem. The electron is then passed along a chain of electron acceptors to which it transfers some of its energy. The energy delivered to the electron acceptors is used to move hydrogen ions across the thylakoid membrane into the lumen. The electron is eventually used to reduce the co-enzyme NADP with a H^+ to NADPH (which has functions in the light-independent reaction); at that point, the path of that electron ends.

The cyclic reaction is similar to that of the non-cyclic, but differs in that it generates only ATP, and no reduced NADP (NADPH) is created. The cyclic reaction takes place only at photosystem I. Once the electron is displaced from the photosystem, the electron is passed down the electron acceptor molecules and returns to photosystem I, from where it was emitted, hence the name cyclic reaction.

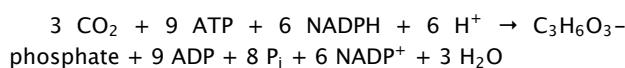
7.2.3 Water Photolysis

Linear electron transport through a photosystem will leave the reaction center of that photosystem oxidized. Elevating another electron will first require re-reduction of the reaction center. The excited electrons lost from the reaction center (P700) of photosystem I are replaced by transfer from plastocyanin, whose electrons come from electron transport through photosystem II. Photosystem II, as the first step of the Z-scheme, requires an external source of electrons to reduce its oxidized chlorophyll a reaction center, called P680. The source of electrons for

photosynthesis in green plants and cyanobacteria is water. Two water molecules are oxidized by four successive charge-separation reactions by photosystem II to yield a molecule of diatomic oxygen and four hydrogen ions. The electrons yielded are transferred to a redox-active tyrosine residue that then reduces the oxidized P680. This resets the ability of P680 to absorb another photon and release another photo-dissociated electron. The oxidation of water is catalyzed in photosystem II by a redox-active structure that contains four manganese ions and a calcium ion; this oxygen-evolving complex binds two water molecules and contains the four oxidizing equivalents that are used to drive the water-oxidizing reaction (Dolai's S-state diagrams). Photosystem II is the only known biological enzyme that carries out this oxidation of water. The hydrogen ions are released in the thylakoid lumen and therefore contribute to the transmembrane chemiosmotic potential that leads to ATP synthesis. Oxygen is a waste product of light-dependent reactions, but the majority of organisms on Earth use oxygen for cellular respiration, including photosynthetic organisms.

7.2.4 Calvin Cycle

In the light-independent (or “dark”) reactions, the enzyme RuBisCO captures CO_2 from the atmosphere and, in a process called the Calvin cycle, it uses the newly formed NADPH and releases three-carbon sugars, which are later combined to form sucrose and starch. The overall equation for the light-independent reactions in green plants is



Carbon fixation produces the intermediate three-carbon sugar product, which is then converted into the final carbohydrate products. The simple carbon sugars produced by photosynthesis are then used in the forming of other organic compounds, such as the building material cellulose, the precursors for lipid and amino acid biosynthesis, or as a fuel in cellular respiration. The latter occurs not only in plants but also in animals when the energy from plants is passed through a food chain.

The fixation or reduction of carbon dioxide is a process in which carbon dioxide combines with a five-carbon sugar, ribulose 1,5-bisphosphate, to yield two molecules of a three-carbon compound, glyceraldehyde 3-phosphate, also known as 3-phosphoglycerate. Glyceraldehyde 3-phosphate, in the presence of ATP and NADPH produced during the light-dependent stages, is reduced to glyceraldehyde 3-phosphate. This product is also referred to as 3-phosphoglyceraldehyde (PGAL) or, more generically, as triose phosphate. Most (5 out of 6 molecules) of

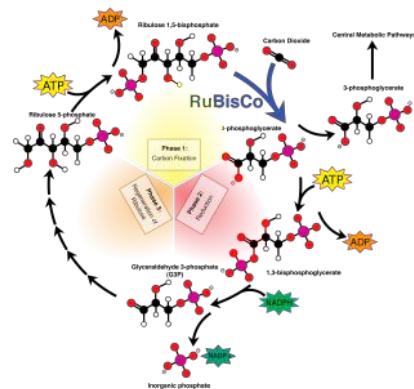


Figure 7.8: Overview of the Calvin cycle and carbon fixation⁸

the glyceraldehyde 3-phosphate produced is used to regenerate ribulose 1,5-bisphosphate so the process can continue. The triose phosphates not thus “recycled” often condense to form hexose phosphates, which ultimately yield sucrose, starch and cellulose. The sugars produced during carbon metabolism yield carbon skeletons that can be used for other metabolic reactions like the production of amino acids and lipids.

7.2.5 Carbon Dioxide Levels And Photorespiration

As carbon dioxide concentrations rise, the rate at which sugars are made by the light-independent reactions increases until limited by other factors. RuBisCO, the enzyme that captures carbon dioxide in the light-independent reactions, has a binding affinity for both carbon dioxide and oxygen. When the concentration of carbon dioxide is high, RuBisCO will fix carbon dioxide. However, if the carbon dioxide concentration is low, RuBisCO will bind oxygen instead of carbon dioxide. This process, called photorespiration, uses energy, but does not produce sugars.

RuBisCO oxygenase activity is disadvantageous to plants for several reasons:

- One product of oxygenase activity is phosphoglycolate (2 carbon) instead of 3-phosphoglycerate (3 carbon). Phosphoglycolate cannot be metabolized by the Calvin-Benson cycle and represents carbon lost from the cycle. A high oxygenase activity, therefore, drains the sugars that are required to recycle ribulose 5-bisphosphate and for the continuation of the Calvin-Benson cycle.
- Phosphoglycolate is quickly metabolized to glycolate that is toxic to a plant at a high concentration; it inhibits photosynthesis.
- Salvaging glycolate is an energetically expensive

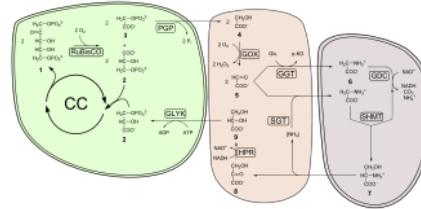
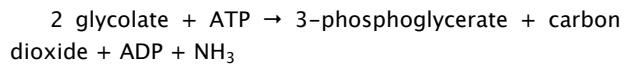


Figure 7.9: Photorespiration⁹ 1. ribulose 1,5-bisphosphate 2. 3-Phosphoglycerate 3. 2-phosphoglycolate 4. glycolate 5. glyoxylate 6. glycine 7. serine 8. hydroxypyruvate 9. glycerate CC Calvin cycle

process that uses the glycolate pathway, and only 75% of the carbon is returned to the Calvin-Benson cycle as 3-phosphoglycerate. The reactions also produce ammonia (NH_3), which is able to diffuse out of the plant, leading to a loss of nitrogen. A highly simplified summary is:



The salvaging pathway for the products of RuBisCO oxygenase activity is more commonly known as photorespiration, since it is characterized by light-dependent oxygen consumption and the release of carbon dioxide.

7.2.6 C4 Carbon Fixation

In hot and dry conditions, plants close their stomata to prevent water loss. Under these conditions, CO_2 will decrease and oxygen gas, produced by the light reactions of photosynthesis, will increase, causing an increase of photorespiration by the oxygenase activity of ribulose-1,5-bisphosphate carboxylase/oxygenase and decrease in carbon fixation. Some plants have evolved mechanisms to increase the CO_2 concentration in the leaves under these conditions.

Plants that use the C4 carbon fixation process chemically fix carbon dioxide in the cells of the mesophyll by adding it to the three-carbon molecule phosphoenolpyruvate (PEP), a reaction catalyzed by an enzyme called PEP carboxylase, creating the four-carbon organic acid oxaloacetic acid. Oxaloacetic acid or malate synthesized by this process is then translocated to specialized bundle sheath cells where the enzyme RuBisCO and other Calvin cycle enzymes are located, and where CO_2 released by decarboxylation of the four-carbon acids is then fixed by RuBisCO activity to the three-carbon 3-phosphoglyceric acids. The physical separation of RuBisCO from the oxygen-generating light reactions reduces photorespiration and increases CO_2 fixation and, thus, the photosynthetic capacity of the leaf. C4 plants

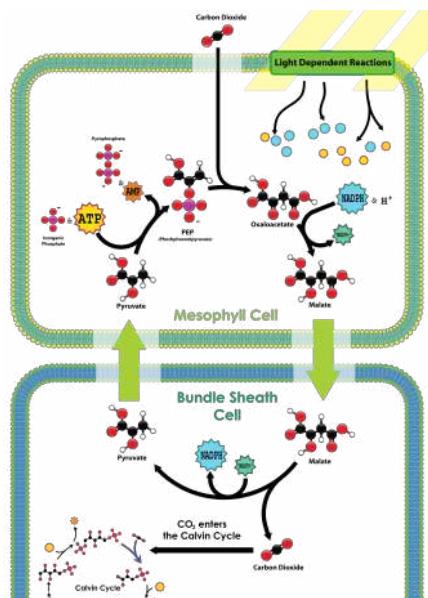


Figure 7.10: Overview of C4 carbon fixation¹⁰

can produce more sugar than C3 plants in conditions of high light and temperature. Many important crop plants are C4 plants, including maize, sorghum, sugarcane, and millet. Plants that do not use PEP-carboxylase in carbon fixation are called C3 plants because the primary carboxylation reaction, catalyzed by RuBisCO, produces the three-carbon 3-phosphoglyceric acids directly in the Calvin-Benson cycle. Over 90% of plants use C3 carbon fixation, compared to 3% that use C4 carbon fixation; however, the evolution of C4 in over 60 plant lineages makes it a striking example of convergent evolution.

7.2.7 CAM Photosynthesis

Xerophytes, such as cacti and most succulents, also use PEP carboxylase to capture carbon dioxide in a process called Crassulacean acid metabolism (CAM). In contrast to C4 metabolism, which spatially separates the CO₂ fixation to PEP from the Calvin cycle, CAM temporally separates these two processes. CAM plants have a different leaf anatomy from C3 plants, and fix the CO₂ at night, when their stomata are open. CAM plants store the CO₂ mostly in the form of malic acid via carboxylation of phosphoenolpyruvate to oxaloacetate, which is then reduced to malate. Decarboxylation of malate during the day releases CO₂ inside the leaves, thus allowing carbon fixation to 3-phosphoglycerate by RuBisCO. Sixteen thousand species of plants use CAM.

Early photosynthetic systems, such as those in green and purple sulfur and green and purple nonsulfur bacteria, are thought to have been anoxygenic, and used various other molecules than water as electron donors. Green and

purple sulfur bacteria are thought to have used hydrogen and sulfur as electron donors. Green nonsulfur bacteria used various amino and other organic acids as an electron donor. Purple nonsulfur bacteria used a variety of non-specific organic molecules. The use of these molecules is consistent with the geological evidence that Earth's early atmosphere was highly reducing at that time.

Fossils of what are thought to be filamentous photosynthetic organisms have been dated at 3.4 billion years old. More recent studies, reported in March 2018, also suggest that photosynthesis may have begun about 3.4 billion years ago.

The main source of oxygen in the Earth's atmosphere derives from oxygenic photosynthesis, and its first appearance is sometimes referred to as the oxygen catastrophe. Geological evidence suggests that oxygenic photosynthesis, such as that in cyanobacteria, became important during the Paleoproterozoic era around 2 billion years ago. Modern photosynthesis in plants and most photosynthetic prokaryotes is oxygenic. Oxygenic photosynthesis uses water as an electron donor, which is oxidized to molecular oxygen (O₂) in the photosynthetic reaction center.

7.3 Symbiosis And The Origin of Chloroplasts

Several groups of animals have formed symbiotic relationships with photosynthetic algae. These are most common in corals, sponges and sea anemones. It is presumed that this is due to the particularly simple body plans and large surface areas of these animals compared to their volumes. In addition, a few marine mollusks *Elysia viridis* and *Elysia chlorotica* also maintain a symbiotic relationship with chloroplasts they capture from the algae in their diet and then store in their bodies (see Kleptoplasty). This allows the mollusks to survive solely by photosynthesis for several months at a time. Some of the genes from the plant cell nucleus have even been transferred to the slugs, so that the chloroplasts can be supplied with proteins that they need to survive.

An even closer form of symbiosis may explain the origin of chloroplasts. Chloroplasts have many similarities with photosynthetic bacteria, including a circular chromosome, prokaryotic-type ribosome, and similar proteins in the photosynthetic reaction center. The endosymbiotic theory suggests that photosynthetic bacteria were acquired (by endocytosis) by early eukaryotic cells to form the first plant cells. Therefore, chloroplasts may be photosynthetic bacteria that adapted to life inside plant cells. Like mitochondria, chloroplasts possess their own DNA, separate from the nuclear DNA of their plant host

cells and the genes in this chloroplast DNA resemble those found in cyanobacteria. DNA in chloroplasts codes for redox proteins such as those found in the photosynthetic reaction centers. The CoRR Hypothesis proposes that this co-location of genes with their gene products is required for redox regulation of gene expression, and accounts for the persistence of DNA in bioenergetic organelles.

7.4 Cyanobacteria And The Evolution of Photosynthesis

The biochemical capacity to use water as the source for electrons in photosynthesis evolved once, in a common ancestor of extant cyanobacteria (formerly called blue-green algae), which are the only prokaryotes performing oxygenic photosynthesis. The geological record indicates that this transforming event took place early in Earth's history, at least 2450-2320 million years ago (Ma), and, it is speculated, much earlier. Because the Earth's atmosphere contained almost no oxygen during the estimated development of photosynthesis, it is believed that the first photosynthetic cyanobacteria did not generate oxygen. Available evidence from geobiological studies of Archean (>2500 Ma) sedimentary rocks indicates that life existed 3500 Ma, but the question of when oxygenic photosynthesis evolved is still unanswered. A clear paleontological window on cyanobacterial evolution opened about 2000 Ma, revealing an already-diverse biota of Cyanobacteria. Cyanobacteria remained the principal primary producers of oxygen throughout the Proterozoic Eon (2500-543 Ma), in part because the redox structure of the oceans favored photoautotrophs capable of nitrogen fixation. Green algae joined cyanobacteria as the major primary producers of oxygen on continental shelves near the end of the Proterozoic, but it was only with the Mesozoic (251-66 Ma) radiations of dinoflagellates, coccolithophorids, and diatoms did the primary production of oxygen in marine shelf waters take modern form. Cyanobacteria remain critical to marine ecosystems as primary producers of oxygen in oceanic gyres, as agents of biological nitrogen fixation, and, in modified form, as the plastids of marine algae.

7.5 Cellular Respiration

Cellular respiration is a set of metabolic reactions and processes that take place in the cells of organisms to convert chemical energy from oxygen molecules or nutrients into adenosine triphosphate (ATP), and then release waste products. The reactions involved in respiration are catabolic reactions, which break large molecules into smaller ones, releasing energy because weak high-energy

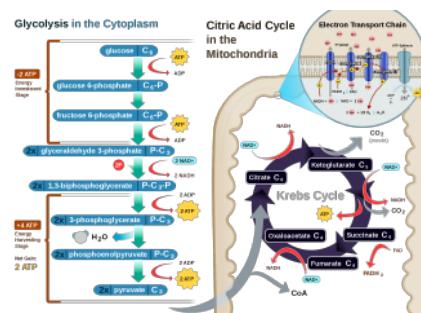


Figure 7.11: A diagram of cellular respiration including glycolysis, Krebs cycle (AKA citric acid cycle), and the electron transport chain¹¹

bonds, in particular in molecular oxygen, are replaced by stronger bonds in the products. Respiration is one of the key ways a cell releases chemical energy to fuel cellular activity. The overall reaction occurs in a series of biochemical steps, some of which are redox reactions. Although cellular respiration is technically a combustion reaction, it clearly does not resemble one when it occurs in a living cell because of the slow, controlled release of energy from the series of reactions.

Nutrients that are commonly used by animal and plant cells in respiration include sugar, amino acids and fatty acids, and the most common oxidizing agent providing most of the chemical energy is molecular oxygen (O_2). The chemical energy stored in ATP (the bond of its third phosphate group to the rest of the molecule can be broken allowing more stable products to form, thereby releasing energy for use by the cell) can then be used to drive processes requiring energy, including biosynthesis, locomotion or transport of molecules across cell membranes.

7.5.1 Aerobic Respiration

Aerobic respiration requires oxygen (O_2) in order to create ATP. Although carbohydrates, fats, and proteins are consumed as reactants, aerobic respiration is the preferred method of pyruvate breakdown in glycolysis, and requires pyruvate to the mitochondria in order to be fully oxidized by the citric acid cycle. The products of this process are carbon dioxide and water, and the energy transferred is used to break bonds in ADP to add a third phosphate group to form ATP (adenosine triphosphate), by substrate-level phosphorylation, NADH and FADH₂

Simplified reaction:



$$\Delta G = -2880 \text{ kJ per mol of } C_6H_{12}O_6$$

The negative ΔG indicates that the reaction can occur spontaneously.

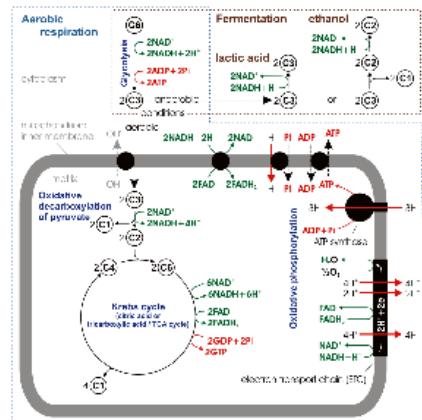


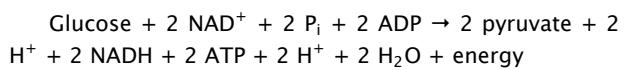
Figure 7.12: Stoichiometry of aerobic respiration¹² and most known fermentation types in eukaryotic cell. Numbers in circles indicate counts of carbon atoms in molecules, C₆ is glucose C₆H₁₂O₆, C₁ carbon dioxide CO₂. Mitochondrial outer membrane is omitted.

The potential of NADH and FADH₂ is converted to more ATP through an electron transport chain with oxygen and protons (hydrogen) as the “terminal electron acceptors”. Most of the ATP produced by aerobic cellular respiration is made by oxidative phosphorylation. The energy of O₂ released is used to create a chemiosmotic potential by pumping protons across a membrane. This potential is then used to drive ATP synthase and produce ATP from ADP and a phosphate group. Biology textbooks often state that 38 ATP molecules can be made per oxidized glucose molecule during cellular respiration (2 from glycolysis, 2 from the Krebs cycle, and about 34 from the electron transport system). However, this maximum yield is never quite reached because of losses due to leaky membranes as well as the cost of moving pyruvate and ADP into the mitochondrial matrix, and current estimates range around 29 to 30 ATP per glucose.

Aerobic metabolism is up to 15 times more efficient than anaerobic metabolism (which yields 2 molecules ATP per 1 molecule glucose) because the double bond in O₂ is of higher energy than other double bonds or pairs of single bonds in other common molecules in the biosphere. However, some anaerobic organisms, such as methanogens are able to continue with anaerobic respiration, yielding more ATP by using other inorganic molecules (not oxygen) of high energy as final electron acceptors in the electron transport chain. They share the initial pathway of glycolysis but aerobic metabolism continues with the Krebs cycle and oxidative phosphorylation. The post-glycolytic reactions take place in the mitochondria in eukaryotic cells, and in the cytoplasm in prokaryotic cells.

7.5.2 Glycolysis

Out of the cytoplasm it goes into the Krebs cycle with the acetyl CoA. It then mixes with CO₂ and makes 2 ATP, NADH, and FADH. From there the NADH and FADH go into the NADH reductase, which produces the enzyme. The NADH pulls the enzyme's electrons to send through the electron transport chain. The electron transport chain pulls H⁺ ions through the chain. From the electron transport chain, the released hydrogen ions make ADP for an end result of 32 ATP. O₂ provides most of the energy for the process and combines with protons and the electrons to make water. Lastly, ATP leaves through the ATP channel and out of the mitochondria. Main article: Glycolysis Glycolysis is a metabolic pathway that takes place in the cytosol of cells in all living organisms. Glycolysis can be literally translated as “sugar splitting”, and occurs with or without the presence of oxygen. In aerobic conditions, the process converts one molecule of glucose into two molecules of pyruvate (pyruvic acid), generating energy in the form of two net molecules of ATP. Four molecules of ATP per glucose are actually produced, however, two are consumed as part of the preparatory phase. The initial phosphorylation of glucose is required to increase the reactivity (decrease its stability) in order for the molecule to be cleaved into two pyruvate molecules by the enzyme aldolase. During the pay-off phase of glycolysis, four phosphate groups are transferred to ADP by substrate-level phosphorylation to make four ATP, and two NADH are produced when the pyruvate is oxidized. The overall reaction can be expressed this way:



Starting with glucose, 1 ATP is used to donate a phosphate to glucose to produce glucose 6-phosphate. Glycogen can be converted into glucose 6-phosphate as well with the help of glycogen phosphorylase. During energy metabolism, glucose 6-phosphate becomes fructose 6-phosphate. An additional ATP is used to phosphorylate fructose 6-phosphate into fructose 1,6-bisphosphate by the help of phosphofructokinase. Fructose 1,6-biphosphate then splits into two phosphorylated molecules with three carbon chains which later degrades into pyruvate.

7.5.3 Oxidative Decarboxylation of Pyruvate

Pyruvate is oxidized to acetyl-CoA and CO₂ by the pyruvate dehydrogenase complex (PDC). The PDC contains multiple copies of three enzymes and is located in the mitochondria of eukaryotic cells and in the cytosol of prokaryotes. In the conversion of pyruvate to acetyl-CoA, one molecule of NADH and one molecule of CO₂ is formed.

7.5.4 The Citric Acid Cycle

This is also called the Krebs cycle or the tricarboxylic acid cycle. When oxygen is present, acetyl-CoA is produced from the pyruvate molecules created from glycolysis. Once acetyl-CoA is formed, aerobic or anaerobic respiration can occur. When oxygen is present, the mitochondria will undergo aerobic respiration which leads to the Krebs cycle. However, if oxygen is not present, fermentation of the pyruvate molecule will occur. In the presence of oxygen, when acetyl-CoA is produced, the molecule then enters the citric acid cycle (Krebs cycle) inside the mitochondrial matrix, and is oxidized to CO_2 while at the same time reducing NAD to NADH. NADH can be used by the electron transport chain to create further ATP as part of oxidative phosphorylation. To fully oxidize the equivalent of one glucose molecule, two acetyl-CoA must be metabolized by the Krebs cycle. Two low-energy waste products, H_2O and CO_2 , are created during this cycle.

The citric acid cycle is an 8-step process involving 18 different enzymes and co-enzymes. During the cycle, acetyl-CoA (2 carbons) + oxaloacetate (4 carbons) yields citrate (6 carbons), which is rearranged to a more reactive form called isocitrate (6 carbons). Isocitrate is modified to become α -ketoglutarate (5 carbons), succinyl-CoA, succinate, fumarate, malate, and, finally, oxaloacetate.

The net gain from one cycle is 3 NADH and 1 FADH_2 as hydrogen- (proton plus electron)-carrying compounds and 1 high-energy GTP, which may subsequently be used to produce ATP. Thus, the total yield from 1 glucose molecule (2 pyruvate molecules) is 6 NADH, 2 FADH_2 , and 2 ATP.

7.5.5 Oxidative Phosphorylation

In eukaryotes, oxidative phosphorylation occurs in the mitochondrial cristae. It comprises the electron transport chain that establishes a proton gradient (chemiosmotic potential) across the boundary of the inner membrane by oxidizing the NADH produced from the Krebs cycle. ATP is synthesized by the ATP synthase enzyme when the chemiosmotic gradient is used to drive the phosphorylation of ADP. The electron transfer is driven by the chemical energy of exogenous oxygen and, with the addition of two protons, water is formed.

7.6 Fermentation

Without oxygen, pyruvate (pyruvic acid) is not metabolized by cellular respiration but undergoes a process of fermentation. The pyruvate is not transported into the mitochondrion, but remains in the cytoplasm, where it is converted to waste products that may be removed from the cell. This serves the purpose of oxidizing the electron carriers so

that they can perform glycolysis again and removing the excess pyruvate. Fermentation oxidizes NADH to NAD^+ so it can be re-used in glycolysis. In the absence of oxygen, fermentation prevents the buildup of NADH in the cytoplasm and provides NAD^+ for glycolysis. This waste product varies depending on the organism. In skeletal muscles, the waste product is lactic acid. This type of fermentation is called lactic acid fermentation. In strenuous exercise, when energy demands exceed energy supply, the respiratory chain cannot process all of the hydrogen atoms joined by NADH. During anaerobic glycolysis, NAD^+ regenerates when pairs of hydrogen combine with pyruvate to form lactate. Lactate formation is catalyzed by lactate dehydrogenase in a reversible reaction. Lactate can also be used as an indirect precursor for liver glycogen. During recovery, when oxygen becomes available, NAD^+ attaches to hydrogen from lactate to form ATP. In yeast, the waste products are ethanol and carbon dioxide. This type of fermentation is known as alcoholic or ethanol fermentation. The ATP generated in this process is made by substrate-level phosphorylation, which does not require oxygen.

Fermentation is less efficient at using the energy from glucose: only 2 ATP are produced per glucose, compared to the 38 ATP per glucose nominally produced by aerobic respiration. This is because most of the energy of aerobic respiration derives from O_2 with its relatively weak, high-energy double bond. Glycolytic ATP, however, is created more quickly. For prokaryotes to continue a rapid growth rate when they are shifted from an aerobic environment to an anaerobic environment, they must increase the rate of the glycolytic reactions. For multicellular organisms, during short bursts of strenuous activity, muscle cells use fermentation to supplement the ATP production from the slower aerobic respiration, so fermentation may be used by a cell even before the oxygen levels are depleted, as is the case in sports that do not require athletes to pace themselves, such as sprinting.

Along with photosynthesis and aerobic respiration, fermentation is a way of extracting energy from molecules, but it is the only one common to all bacteria and eukaryotes. It is therefore considered the oldest metabolic pathway, suitable for an environment that did not yet have oxygen. Yeast, a form of fungus, occurs in almost any environment capable of supporting microbes, from the skins of fruits to the guts of insects and mammals and the deep ocean, and harvests sugar-rich materials to produce ethanol and carbon dioxide.

The basic mechanism for fermentation remains present in all cells of higher organisms. Mammalian muscle carries out fermentation during periods of intense exercise where oxygen supply becomes limited, resulting in the creation of lactic acid. In invertebrates, fermentation also produces

succinate and alanine.

Fermentative bacteria play an essential role in the production of methane in habitats ranging from the rumens of cattle to sewage digesters and freshwater sediments. They produce hydrogen, carbon dioxide, formate and acetate and carboxylic acids; and then consortia of microbes convert the carbon dioxide and acetate to methane. Acetogenic bacteria oxidize the acids, obtaining more acetate and either hydrogen or formate. Finally, methanogens (in the domain Archea) convert acetate to methane.

In ethanol fermentation, one glucose molecule is converted into two ethanol molecules and two carbon dioxide molecules. It is used to make bread dough rise: the carbon dioxide forms bubbles, expanding the dough into a foam. The ethanol is the intoxicating agent in alcoholic beverages such as wine, beer and liquor. Fermentation of feedstocks, including sugarcane, corn, and sugar beets, produces ethanol that is added to gasoline. In some species of fish, including goldfish and carp, it provides energy when oxygen is scarce (along with lactic acid fermentation).

Before fermentation, a glucose molecule breaks down into two pyruvate molecules (Glycolysis). The energy from this exothermic reaction is used to bind inorganic phosphates to ADP, which converts it to ATP, and convert NAD⁺ to NADH. The pyruvates break down into two acetaldehyde molecules and give off two carbon dioxide molecules as waste products. The acetaldehyde is reduced into ethanol using the energy and hydrogen from NADH, and the NADH is oxidized into NAD⁺ so that the cycle may repeat. The reaction is catalyzed by the enzymes pyruvate decarboxylase and alcohol dehydrogenase.

7.7 Anaerobic Respiration

Cellular respiration is the process by which biological fuels are oxidised in the presence of a high-energy inorganic electron acceptor (such as oxygen) to produce large amounts of energy, to drive the bulk production of ATP.

Anaerobic respiration is used by some microorganisms in which neither oxygen (aerobic respiration) nor pyruvate derivatives (fermentation) is the high-energy final electron acceptor. Rather, an inorganic acceptor such as sulfate or nitrate is used. Such organisms are typically found in unusual places such as underwater caves or near hydrothermal vents at the bottom of the ocean.

In July 2019, a scientific study of Kidd Mine in Canada discovered sulfur-breathing organisms which live 7900 feet below the surface, and which breathe sulfur in order to survive. These organisms are also remarkable due to consuming minerals such as pyrite as their food source.

Chapter 8

The Cell Cycle And Cell Division

The cell cycle¹, or cell–division cycle, is the series of events that take place in a cell that cause it to divide into two daughter cells. These events include the duplication of its DNA (DNA replication) and some of its organelles, and subsequently the partitioning of its cytoplasm and other components into two daughter cells in a process called cell division.

In cells with nuclei (eukaryotes), (i.e., animal, plant, fungal, and protist cells), the cell cycle is divided into two main stages: interphase and the mitotic (M) phase (including mitosis and cytokinesis). During interphase, the cell grows, accumulating nutrients needed for mitosis, and replicates its DNA and some of its organelles. During the mitotic phase, the replicated chromosomes, organelles, and cytoplasm separate into two new daughter cells. To ensure the proper replication of cellular components and division, there are control mechanisms known as cell cycle checkpoints after each of the key steps of the cycle that determine if the cell can progress to the next phase.

8.1 The Prokaryotic Cell Cycle

In cells without nuclei (prokaryotes), (i.e., bacteria and archaea), the cell cycle is divided into the B, C, and D periods. The B period extends from the end of cell division to the beginning of DNA replication. DNA replication occurs during the C period. The D period refers to the stage between the end of DNA replication and the splitting of the bacterial cell into two daughter cells. Cell division in prokaryotes (the domains of Archaea and Bacteria) is called binary fission. This form of asexual reproduction and cell division is also used by some organelles within eukaryotic organisms (e.g., mitochondria). Binary fission results in the reproduction of a living prokaryotic cell (or organelle) by dividing the cell into two parts, each with the potential to grow to the size of the original.

The single DNA molecule first replicates, then attaches each copy to a different part of the cell membrane. When

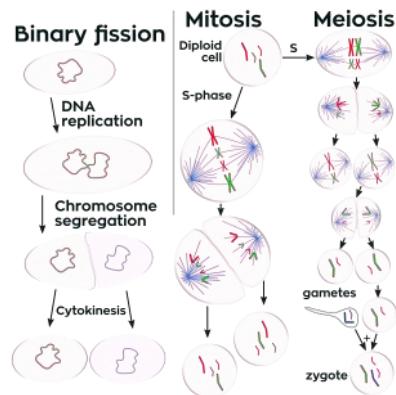


Figure 8.1: Many bacteria reproduce through binary fission, which is compared to mitosis and meiosis in this image.²

the cell begins to pull apart, the replicated and original chromosomes are separated. The consequence of this asexual method of reproduction is that all the cells are genetically identical, meaning that they have the same genetic material (barring random mutations). Unlike the processes of mitosis and meiosis used by eukaryotic cells, binary fission takes place without the formation of a spindle apparatus on the cell. Like in mitosis (and unlike in meiosis), the parental identity is preserved.

8.2 Binary Fission

The process of binary fission in bacteria involves the following steps. First, the cell's DNA is replicated. The replicated DNA copies then move to opposite poles of the cell in an energy-dependent process. The cell lengthens. Then, the equatorial plane of the cell constricts and separates the plasma membrane such that each new cell has exactly the same genetic material.

More specifically, the following steps occur:

¹https://en.wikipedia.org/wiki/Cell_cycle

1. The DNA is tightly coiled.
2. The DNA is unwound and duplicated.
3. The DNA is pulled to the separate poles of the bacterium as it increases the size to prepare for splitting.
4. The growth of a new cell wall begins to separate the bacterium
5. The new cell wall fully develops, resulting in the complete split of the bacterium.
6. The new daughter cells have tightly coiled DNA rods, ribosomes, and plasmids; these are now brand-new organisms.

Binary fission is generally rapid though its speed varies between species. Under optimal conditions, *E. coli*, cells divide about every 20 minutes at 37 °C. Because the new cells will, in turn, undergo binary fission, the time binary fission requires is also the time the bacterial culture requires to double in the number of cells it contains. This time period is, therefore, be referred to as the doubling time. Some strains of *Mycobacterium tuberculosis* have doubling times of nearly 100 hours. Bacterial growth is limited by nutrient availability and density, so binary fission occurs at much lower rates in bacterial cultures once they enter the stationary phase of growth.

Some organelles in eukaryotic cells reproduce using binary fission. Mitochondrial fission occurs frequently within the cell, even when the cell is not actively undergoing mitosis, and is necessary to regulate the cell's metabolism. All chloroplasts and some mitochondria (not in animals), both organelles derived from endosymbiosis of bacteria, also use FtsZ in a bacteria-like fashion. The cell-division cycle is a vital process by which a single-celled fertilized egg develops into a mature organism, as well as the process by which hair, skin, blood cells, and some internal organs are renewed. After cell division, each of the daughter cells begin the interphase of a new cycle. Although the various stages of interphase are not usually morphologically distinguishable, each phase of the cell cycle has a distinct set of specialized biochemical processes that prepare the cell for initiation of the cell division.

Cells that are no longer dividing are said to have entered a state of quiescence called G0 phase.

8.3 The Eukaryotic Cell Cycle

The eukaryotic cell cycle consists of four distinct phases: G1 phase, S phase (synthesis), G2 phase (collectively known as interphase) and M phase (mitosis and cytokinesis). M phase is itself composed of two tightly coupled processes: mitosis, in which the cell's nucleus divides, and cytokinesis, in which the cell's cytoplasm divides forming two daughter cells. Activation of each phase is dependent on the proper progression and completion of the previous one. Cells that have temporarily or reversibly stopped

Table 8.1: Phases of the eukaryotic cell cycle

State	Phase	Abbreviation	Description
Resting	Gap 0	G0	A phase where the cell has left the cycle and has stopped dividing.
	Gap 1	G1	Cells increase in size in Gap 1. The G1 checkpoint control mechanism ensures that everything is ready for DNA synthesis.
Interphase	Synthesis	S	DNA replication occurs during this phase.
	Gap 2	G2	During the gap between DNA synthesis and mitosis, the cell will continue to grow. The G2 checkpoint control mechanism ensures that everything is ready to enter the M (mitosis) phase and divide.
Cell division	Mitosis	M	Cell growth stops at this stage and cellular energy is focused on the orderly division into two daughter cells. A checkpoint in the middle of mitosis (Metaphase Checkpoint) ensures that the cell is ready to complete cell division.

After cell division, each of the daughter cells begin the interphase of a new cycle. Although the various stages of interphase are not usually morphologically distinguishable, each phase of the cell cycle has a distinct set of specialized biochemical processes that prepare the cell for initiation of cell division.

8.3.1 G₀ phase (quiescence)

G₀ is a resting phase where the cell has left the cycle and has stopped dividing. The cell cycle starts with this phase. Non-proliferative (non-dividing) cells in multicellular eukaryotes generally enter the quiescent G₀ state from G₁ and may remain quiescent for long periods of time, possibly indefinitely (as is often the case for neurons). This is very common for cells that are fully differentiated. Some cells enter the G₀ phase semi-permanently and are considered post-mitotic, e.g., some liver, kidney, and stomach cells. Many cells do not enter G₀ and continue to divide throughout an organism's life, e.g., epithelial cells.

The word "post-mitotic" is sometimes used to refer to both quiescent and senescent cells. Cellular senescence occurs in response to DNA damage and external stress and usually constitutes an arrest in G₁. Cellular senescence may make a cell's progeny nonviable; it is often a biochemical alternative to the self-destruction of such a damaged cell by apoptosis.

8.3.2 Interphase

Interphase is a series of changes that takes place in a newly formed cell and its nucleus before it becomes capable of division again. It is also called preparatory phase or intermitosis. Typically interphase lasts for at least 91% of the total time required for the cell cycle.

Interphase proceeds in three stages, G₁, S, and G₂, followed by the cycle of mitosis and cytokinesis. The cell's nuclear DNA contents are duplicated during S phase.

8.3.3 G₁ phase (First growth phase or Post mitotic gap phase)

The first phase within interphase, from the end of the previous M phase until the beginning of DNA synthesis, is called G₁ (G indicating gap). It is also called the growth phase. During this phase, the biosynthetic activities of the cell, which are considerably slowed down during M phase, resume at a high rate. The duration of G₁ is highly variable, even among different cells of the same species. In this phase, the cell increases its supply of proteins, increases the number of organelles (such as mitochondria, ribosomes), and grows in size. In G₁ phase, a cell has three options.

- To continue cell cycle and enter S phase
- Stop cell cycle and enter G₀ phase for undergoing differentiation.
- Become arrested in G₁ phase hence it may enter G₀ phase or re-enter cell cycle.

The deciding point is called check point (Restriction point). This check point is called the restriction point or START and is regulated by G₁/S cyclins, which cause transition from G₁ to S phase. Passage through the G₁ check point commits the cell to division.

8.3.4 S phase (DNA Replication)

The ensuing S phase starts when DNA synthesis commences; when it is complete, all of the chromosomes have been replicated, i.e., each chromosome consists of two sister chromatids. Thus, during this phase, the amount of DNA in the cell has doubled, though the ploidy and number of chromosomes are unchanged. Rates of RNA transcription and protein synthesis are very low during this phase. An exception to this is histone production, most of which occurs during the S phase.

8.3.5 G₂ phase (growth)

G₂ phase occurs after DNA replication and is a period of protein synthesis and rapid cell growth to prepare the cell for mitosis. During this phase microtubules begin to reorganize to form a spindle (preprophase). Before proceeding to mitotic phase, cells must be checked at the G₂ checkpoint for any DNA damage within the chromosomes. The G₂ checkpoint is mainly regulated by the tumor protein p53. If the DNA is damaged, p53 will either repair the DNA or trigger the apoptosis of the cell. If p53 is dysfunctional or mutated, cells with damaged DNA may continue through the cell cycle, leading to the development of cancer.

8.3.6 Mitotic Phase (Chromosome Separation)

The relatively brief M phase consists of nuclear division (karyokinesis). It is a relatively short period of the cell cycle. M phase is complex and highly regulated. The sequence of events is divided into phases, corresponding to the completion of one set of activities and the start of the next. These phases are sequentially known as:

- prometaphase
- metaphase
- anaphase
- telophase

Mitosis is the process by which a eukaryotic cell separates the chromosomes in its cell nucleus into two identical sets in two nuclei. During the process of mitosis the pairs

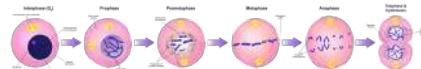


Figure 8.2: A diagram of mitosis stages³. Interphase (G₂): In this substage, the cell prepares for nuclear division and a protein that makes microtubules for cell division is synthesized. Prophase: The longest stage of mitosis. In this stage the chromosomes become visible and the centrioles separate and move to opposite poles of the cell. Prometaphase: The nuclear envelope disintegrates and microtubules can attach to kinetochores. Chromosomes congress toward the metaphase plate of the cell. Metaphase: In this stage the chromosomes line up across the center of the cell and become connected to the spindle fiber at their centromere. Anaphase: In this stage the sister chromatids separate into individual chromosomes and are pulled apart. Telophase & cytokinesis: Chromosomes decondense and are surrounded by a newly formed nuclear envelope. Cytokinesis typically coincides with and telophase.

of chromosomes condense and attach to microtubules that pull the sister chromatids to opposite sides of the cell.

Mitosis occurs exclusively in eukaryotic cells, but occurs in different ways in different species. For example, animal cells undergo an “open” mitosis, where the nuclear envelope breaks down before the chromosomes separate, while fungi such as *Aspergillus nidulans* and *Saccharomyces cerevisiae* (yeast) undergo a “closed” mitosis, where chromosomes divide within an intact cell nucleus.

8.4 Chromosomes: The Vectors of Heredity

A chromosome⁴ is a deoxyribonucleic acid (DNA) molecule with part or all of the genetic material (genome) of an organism. Most eukaryotic chromosomes include packaging proteins which, aided by chaperone proteins, bind to and condense the DNA molecule to prevent it from becoming an unmanageable tangle. This three-dimensional genome structure plays a significant role in transcriptional regulation.

The word chromosome comes from the Greek χρῶμα (chroma, “colour”) and σῶμα (soma, “body”), describing their strong staining by particular dyes. The term was coined by the German scientist Heinrich Wilhelm Gottfried von Waldeyer-Hartz⁵, referring to the term chromatin,



Figure 8.3: The karyotype of a normal human male consisting of 22 pairs of autosomes and 2 sex chromosomes (one x and one y chromosome).⁷ The karyotype is the characteristic chromosome complement of a eukaryote species.

which was itself introduced by Walther Flemming⁶, who discovered cell division.

Chromosomes are normally visible under a light microscope only when the cell is undergoing cell division. Before this happens, every chromosome is copied once (S phase), and the copy is joined to the original by a centromere, resulting either in an X-shaped structure if the centromere is located in the middle of the chromosome or a two-arm structure if the centromere is located near one of the ends. The original chromosome and the copy are now called sister chromatids. During metaphase the X-shape structure is called a metaphase chromosome. In this highly condensed form chromosomes are easiest to distinguish and study. In animal cells, chromosomes reach their highest compaction level in anaphase during chromosome segregation.

The prokaryotes – bacteria and archaea – typically have a single circular chromosome, but many variations exist. The chromosomes of most bacteria, which some authors prefer to call genophores, can range in size from only 130,000 base pairs in the endosymbiotic bacteria *Candidatus Hodgkinia cicadicola* and *Candidatus Tremblaya princeps*, to more than 14,000,000 base pairs in the soil-dwelling bacterium *Sorangium cellulosum*. Spirochaetes of the genus *Borrelia* are a notable exception to this arrangement, with bacteria such as *Borrelia burgdorferi*, the cause of Lyme disease⁸, containing a single linear chromosome.

Chromosomes in eukaryotes are composed of chromatin fiber. Chromatin fiber is made of nucleosomes (histone octamers with part of a DNA strand attached to and wrapped around it). Chromatin fibers are packaged by proteins into a condensed structure called chromatin. Chromatin contains the vast majority of DNA, but a

⁴<https://en.wikipedia.org/wiki/Chromosome>

⁵https://en.wikipedia.org/wiki/Heinrich_Wilhelm_Gottfried_von_Waldeyer-Hartz

⁶https://en.wikipedia.org/wiki/Walther_Flemming

⁸https://en.wikipedia.org/wiki/Lyme_disease

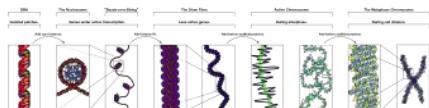


Figure 8.4: The major structures in DNA compaction: DNA, the nucleosome, the 10 nm “beads-on-a-string” fibre, the 30 nm fibre and the metaphase chromosome.⁹

small amount inherited maternally, can be found in the mitochondria. Chromatin is present in most cells, with a few exceptions, for example, red blood cells.

Chromatin allows the very long DNA molecules to fit into the cell nucleus. During cell division chromatin condenses further to form microscopically visible chromosomes. The structure of chromosomes varies through the cell cycle. During cellular division chromosomes are replicated, divided, and passed successfully to their daughter cells so as to ensure the genetic diversity and survival of their progeny. Chromosomes may exist as either duplicated or unduplicated. Unduplicated chromosomes are single double helices, whereas duplicated chromosomes contain two identical copies (called chromatids or sister chromatids) joined by a centromere.

Eukaryotes possess multiple large linear chromosomes contained in the cell’s nucleus. Each chromosome has one centromere, with one or two arms projecting from the centromere, although, under most circumstances, these arms are not visible as such. In addition, most eukaryotes have a small circular mitochondrial genome, and some eukaryotes may have additional small circular or linear cytoplasmic chromosomes.

In the nuclear chromosomes of eukaryotes, the uncondensed DNA exists in a semi-ordered structure, where it is wrapped around histones (structural proteins), forming a composite material called chromatin.

8.4.1 Interphase Chromatin

During interphase (the period of the cell cycle where the cell is not dividing), two types of chromatin can be distinguished:

- Euchromatin, which consists of DNA that is active, e.g., being expressed as protein.
- Heterochromatin, which consists of mostly inactive DNA. It seems to serve structural purposes during the chromosomal stages.

Heterochromatin can be further distinguished into two types

- Constitutive heterochromatin, which is never expressed. It is located around the centromere and

usually contains repetitive sequences.

- Facultative heterochromatin, which is sometimes expressed.

8.4.2 Metaphase chromatin and division

In the early stages of mitosis or meiosis (cell division), the chromatin double helices become more and more condensed. They cease to function as accessible genetic material (transcription stops) and become a compact transportable form. This compact form makes the individual chromosomes visible, and they form the classic four arm structure, a pair of sister chromatids attached to each other at the centromere. The shorter arms are called p arms (from the French petit, small) and the longer arms are called q arms (q follows p in the Latin alphabet; q-g “grande”; alternatively it is sometimes said q is short for queue meaning tail in French). This is the only natural context in which individual chromosomes are visible with an optical microscope.

Mitotic metaphase chromosomes are best described by a linearly organized longitudinally compressed array of consecutive chromatin loops.

During mitosis, microtubules grow from centrosomes located at opposite ends of the cell and also attach to the centromere at specialized structures called kinetochores, one of which is present on each sister chromatid. A special DNA base sequence in the region of the kinetochores provides, along with special proteins, longer-lasting attachment in this region. The microtubules then pull the chromatids apart toward the centrosomes, so that each daughter cell inherits one set of chromatids. Once the cells have divided, the chromatids are uncoiled and DNA can again be transcribed. In spite of their appearance, chromosomes are structurally highly condensed, which enables these giant DNA structures to be contained within a cell nucleus.

8.4.3 Human chromosomes

Chromosomes in humans can be divided into two types: autosomes (body chromosome(s)) and allosomes (sex chromosome(s)). Certain genetic traits are linked to a person’s sex and are passed on through the sex chromosomes. The autosomes contain the rest of the genetic hereditary information. All act in the same way during cell division. Human cells have 23 pairs of chromosomes (22 pairs of autosomes and one pair of sex chromosomes), giving a total of 46 per cell (Figure 8.3). In addition to these, human cells have many hundreds of copies of the mitochondrial genome. Sequencing of the human genome has provided a great deal of information about each of the chromosomes (Table 8.2).

Table 8.2: Human Genome Assembly GRCh38.p12 (nucleotides) and GRCh38.p13 (coding genes)¹⁰. Length of DNA sequence and number of coding genes of each human chromosome. Total lengths are calculated by summing the length of the sequenced bases and estimated gaps.

Chromosome	Total Length	Coding Genes
1	248,956,422	2,057
2	242,193,529	1,303
3	198,295,559	1,078
4	190,214,555	753
5	181,538,259	885
6	170,805,979	1,048
7	159,345,973	999
8	145,138,636	685
9	138,394,717	780
10	133,797,422	733
11	135,086,622	1,317
12	133,275,309	1,034
13	114,364,328	32
14	107,043,718	81
15	101,991,189	61
16	90,338,345	859
17	83,257,441	1,186
18	80,373,285	268
19	58,617,616	1,473
20	64,444,167	546
21	46,709,983	233
22	50,818,468	494
X	156,040,895	852
Y	57,227,415	66
MT	16,569	13
Unplaced	4,485,509	
Genome	3,099,734,149	20,415

8.4.4 Number of Chromosomes in Various Organisms

The following tables (Tables 8.3 and 8.4) give the total number of chromosomes (including sex chromosomes) in a cell nucleus. For example, most eukaryotes are diploid, like humans who have 22 different types of autosomes, each present as homologous pairs (i.e. one chromosome of each type inherited from the mother and one from the father), and two sex chromosomes. This gives 46 chromosomes in total. Other organisms have more than two copies of their chromosome types, such as bread wheat, which is hexaploid and has six copies of seven different chromosome types – 42 chromosomes in total.

Normal members of a particular eukaryotic species all

have the same number of nuclear chromosomes. Other eukaryotic chromosomes, i.e., mitochondrial and plasmid-like small chromosomes, are much more variable in number, and there may be thousands of copies per cell.

Table 8.3: Chromosome numbers in some plants.

Plant	Number of Chromosomes
Adder's tongue fern (polyploid)	approx. 1,200
Arabidopsis thaliana (diploid)	10
Einkorn wheat (diploid)	14
Rye (diploid)	14
Maize (diploid or palaeotetraploid)	20
Durum wheat (tetraploid)	28
Bread wheat (hexaploid)	42
Cultivated tobacco (tetraploid)	48

Asexually reproducing species have one set of chromosomes that are the same in all body cells. However, asexual species can be either haploid or diploid.

Sexually reproducing species have somatic cells (body cells), which are diploid [2n] having two sets of chromosomes (23 pairs in humans with one set of 23 chromosomes from each parent), one set from the mother and one from the father. Gametes, reproductive cells, are haploid [n]: They have one set of chromosomes. Gametes are produced by meiosis of a diploid germ line cell. During meiosis, the matching chromosomes of father and mother can exchange small parts of themselves (crossover), and thus create new chromosomes that are not inherited solely from either parent. When a male and a female gamete merge (fertilization), a new diploid organism is formed.

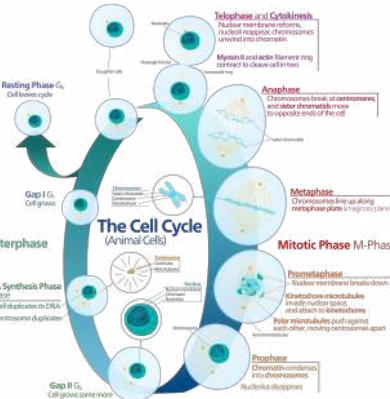
Some animal and plant species are polyploid [Xn]: They have more than two sets of homologous chromosomes. Plants important in agriculture such as tobacco or wheat are often polyploid, compared to their ancestral species. Wheat has a haploid number of seven chromosomes, still seen in some cultivars as well as the wild progenitors. The more-common pasta and bread wheat types are polyploid, having 28 (tetraploid) and 42 (hexaploid) chromosomes, compared to the 14 (diploid) chromosomes in the wild wheat.

Table 8.4: Chromosome numbers ($2n$) in some animals.

Animal	Number of Chromosomes
Indian muntjac	7
Common fruit fly	8
Pill millipede (<i>Arthrosphaera fumosa</i>)	30
Earthworm (<i>Octodrilus complanatus</i>)	36
Tibetan fox	36
Domestic cat	38
Domestic pig	38
Laboratory mouse	40
Laboratory rat	42
Rabbit (<i>Oryctolagus cuniculus</i>)	44
Syrian hamster	44
Guppy (<i>Poecilia reticulata</i>)	46
Human	46
Hares	48
Gorillas, chimpanzees	48
Domestic sheep	54
Garden snail	54
Silkworm	56
Elephants	56
Cow	60
Donkey	62
Guinea pig	64
Horse	64
Dog	78
Hedgehog	90
Goldfish	100–104
Kingfisher	132

8.4.5 Prokaryotic Chromosomes

Prokaryote species generally have one copy of each major chromosome, but most cells can easily survive with multiple copies. For example, *Buchnera*, a symbiont of aphids has multiple copies of its chromosome, ranging from 10–400 copies per cell. However, in some large bacteria, such as *Epulopiscium fishelsoni* up to 100,000 copies of the chromosome can be present. Plasmids and plasmid-like small chromosomes are, as in eukaryotes, highly variable in copy number. The number of plasmids in the cell is almost entirely determined by the rate of division of the plasmid – fast division causes high copy number.

**Figure 8.5:** Mitosis in an animal cell (phases ordered counter-clockwise).¹²

8.5 Mitosis

In cell biology, mitosis¹¹ is a part of the cell cycle when replicated chromosomes are separated into two new nuclei. Cell division gives rise to genetically identical cells in which the number of chromosomes is maintained. In general, mitosis (division of the nucleus) is preceded by the S stage of interphase (during which the DNA is replicated) and is often accompanied or followed by cytokinesis, which divides the cytoplasm, organelles and cell membrane into two new cells containing roughly equal shares of these cellular components (Figure ??). Mitosis and cytokinesis together define the mitotic (M) phase of an animal cell cycle—the division of the mother cell into two daughter cells genetically identical to each other.

The process of mitosis is divided into stages corresponding to the completion of one set of activities and the start of the next. These stages are prophase, prometaphase, metaphase, anaphase, and telophase. During mitosis, the chromosomes, which have already duplicated, condense and attach to spindle fibers that pull one copy of each chromosome to opposite sides of the cell. The result is two genetically identical daughter nuclei. The rest of the cell may then continue to divide by cytokinesis to produce two daughter cells. Producing three or more daughter cells instead of the normal two is a mitotic error called tripolar mitosis or multipolar mitosis (direct cell triplication / multiplication). Other errors during mitosis can induce apoptosis (programmed cell death) or cause mutations. Certain types of cancer can arise from such mutations.

Mitosis varies between organisms. For example, animal cells undergo an “open” mitosis, where the nuclear envelope breaks down before the chromosomes separate,

¹¹<https://en.wikipedia.org/wiki/Mitosis>

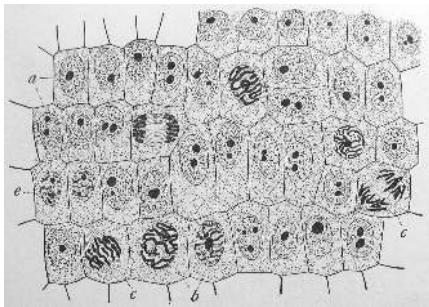


Figure 8.6: Onion (*Allium cepa*) root cells in different phases of the cell cycle (drawn by E. B. Wilson¹³, 1900). General view of cells in the growing root-tip of the onion, from a longitudinal section, enlarged 800 diameters. a. non-dividing cells, with chromatin-network and deeply stained nucleoli; b. nuclei preparing for division (spireme-stage); c. dividing cells showing mitotic figures; e. pair of daughter-cells shortly after division.¹⁴

whereas fungi undergo a “closed” mitosis, where chromosomes divide within an intact cell nucleus. Most animal cells undergo a shape change, known as mitotic cell rounding, to adopt a near spherical morphology at the start of mitosis. Most human cells are produced by mitotic cell division. Important exceptions include the gametes – sperm and egg cells – which are produced by meiosis.

Numerous descriptions of cell division were made during 18th and 19th centuries, with various degrees of accuracy. In 1835, the German botanist Hugo von Mohl¹⁵, described cell division in the green alga *Cladophora glomerata*, stating that multiplication of cells occurs through cell division. In 1838, Matthias Jakob Schleiden affirmed that the formation of new cells in their interior was a general law for cell multiplication in plants, a view later rejected in favour of Mohl model, due to contributions of Robert Remak and others.

The term “mitosis”, coined by Walther Flemming¹⁶ in 1882, is derived from the Greek word μίτος (mitos, “warp thread”). There are some alternative names for the process, e.g., “karyokinesis” (nuclear division), a term introduced by Schleicher in 1878, or “equational division”, proposed by August Weismann¹⁷ in 1887. However, the term “mitosis” is also used in a broad sense by some authors to refer to karyokinesis and cytokinesis together. Presently, “equational division” is more commonly used to refer to meiosis II, the part of meiosis most like mitosis.

The primary result of mitosis and cytokinesis is the transfer of a parent cell’s genome into two daughter cells.

The genome is composed of a number of chromosomes—complexes of tightly coiled DNA that contain genetic information vital for proper cell function. Because each resultant daughter cell should be genetically identical to the parent cell, the parent cell must make a copy of each chromosome before mitosis. This occurs during the S phase of interphase. Chromosome duplication results in two identical sister chromatids bound together by cohesin proteins at the centromere.

When mitosis begins, the chromosomes condense and become visible. In some eukaryotes, for example animals, the nuclear envelope, which segregates the DNA from the cytoplasm, disintegrates into small vesicles. The nucleolus also disappears. Microtubules project from opposite ends of the cell, attach to the centromeres, and align the chromosomes centrally within the cell. The microtubules then contract to pull the sister chromatids of each chromosome apart. Sister chromatids at this point are called daughter chromosomes. As the cell elongates, corresponding daughter chromosomes are pulled toward opposite ends of the cell and condense maximally in late anaphase. A new nuclear envelope forms around the separated daughter chromosomes, which decondense to form interphase nuclei.

During mitotic progression, typically after the anaphase onset, the cell may undergo cytokinesis. In animal cells, a cell membrane pinches inward between the two developing nuclei to produce two new cells. In plant cells, a cell plate forms between the two nuclei. Cytokinesis does not always occur; coenocytic (a type of multinucleate condition) cells undergo mitosis without cytokinesis.

8.5.1 Interphase

The mitotic phase is a relatively short period of the cell cycle. It alternates with the much longer interphase, where the cell prepares itself for the process of cell division. Interphase is divided into three phases: G1 (first gap), S (synthesis), and G2 (second gap). During all three parts of interphase, the cell grows by producing proteins and cytoplasmic organelles. However, chromosomes are replicated only during the S phase. Thus, a cell grows (G1), continues to grow as it duplicates its chromosomes (S), grows more and prepares for mitosis (G2), and finally divides (M) before restarting the cycle. All these phases in the cell cycle are highly regulated by cyclins, cyclin-dependent kinases, and other cell cycle proteins. The phases follow one another in strict order and there are “checkpoints” that give the cell cues to proceed from one phase to another. Cells may also temporarily or permanently leave the cell cycle and enter G0 phase to stop dividing. This can occur when cells become overcrowded (density-dependent inhibition).

¹⁵https://en.wikipedia.org/wiki/Hugo_von_Mohl

¹⁶https://en.wikipedia.org/wiki/Walther_Flemming

¹⁷https://en.wikipedia.org/wiki/August_Weismann

or when they differentiate to carry out specific functions for the organism, as is the case for human heart muscle cells and neurons. Some G0 cells have the ability to re-enter the cell cycle.

8.5.2 Prophase (plant cells)

In plant cells only, prophase is preceded by a pre-prophase stage. In highly vacuolated plant cells, the nucleus has to migrate into the center of the cell before mitosis can begin. This is achieved through the formation of a phragmosome, a transverse sheet of cytoplasm that bisects the cell along the future plane of cell division. In addition to phragmosome formation, preprophase is characterized by the formation of a ring of microtubules and actin filaments (called preprophase band) underneath the plasma membrane around the equatorial plane of the future mitotic spindle. This band marks the position where the cell will eventually divide. The cells of higher plants (such as the flowering plants) lack centrioles; instead, microtubules form a spindle on the surface of the nucleus and are then organized into a spindle by the chromosomes themselves, after the nuclear envelope breaks down. The preprophase band disappears during nuclear envelope breakdown and spindle formation in prometaphase.^{58–67}

8.5.3 Prophase

During prophase, which occurs after G2 interphase, the cell prepares to divide by tightly condensing its chromosomes and initiating mitotic spindle formation. During interphase, the genetic material in the nucleus consists of loosely packed chromatin. At the onset of prophase, chromatin fibers condense into discrete chromosomes that are typically visible at high magnification through a light microscope. In this stage, chromosomes are long, thin and thread-like. Each chromosome has two chromatids. The two chromatids are joined at the centromere.

Gene transcription ceases during prophase and does not resume until late anaphase to early G1 phase. The nucleolus also disappears during early prophase.

Close to the nucleus of animal cells are structures called centrosomes, consisting of a pair of centrioles surrounded by a loose collection of proteins. The centrosome is the coordinating center for the cell's microtubules. A cell inherits a single centrosome at cell division, which is duplicated by the cell before a new round of mitosis begins, giving a pair of centrosomes. The two centrosomes polymerize tubulin to help form a microtubule spindle apparatus. Motor proteins then push the centrosomes along these microtubules to opposite sides of the cell. Although centrosomes help organize microtubule assembly, they are not essential for the formation of the spidele apparatus,

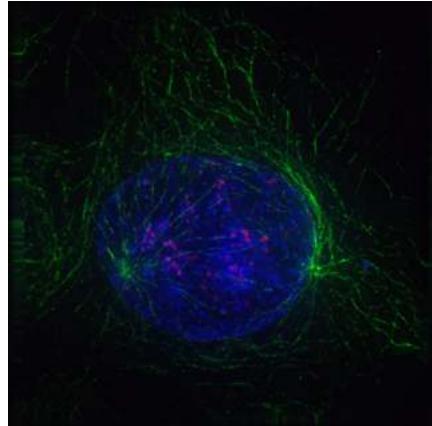


Figure 8.7: Early prophase: Polar microtubules, shown as green strands, have established a matrix around the currently intact nucleus, with the condensing chromosomes in blue. The red nodules are the centromeres.¹⁸

since they are absent from plants, and are not absolutely required for animal cell mitosis.

8.5.4 Prometaphase

At the beginning of prometaphase in animal cells, phosphorylation of nuclear lamins causes the nuclear envelope to disintegrate into small membrane vesicles. As this happens, microtubules invade the nuclear space. This is called open mitosis, and it occurs in some multicellular organisms. Fungi and some protists, such as algae or trichomonads, undergo a variation called closed mitosis where the spindle forms inside the nucleus, or the microtubules penetrate the intact nuclear envelope.

In late prometaphase, kinetochore microtubules begin to search for and attach to chromosomal kinetochores. A kinetochore is a proteinaceous microtubule-binding structure that forms on the chromosomal centromere during late prophase. A number of polar microtubules find and interact with corresponding polar microtubules from the opposite centrosome to form the mitotic spindle. Although the kinetochore structure and function are not fully understood, it is known that it contains some form of molecular motor. When a microtubule connects with the kinetochore, the motor activates, using energy from ATP to "crawl" up the tube toward the originating centrosome. This motor activity, coupled with polymerisation and depolymerisation of microtubules, provides the pulling force necessary to later separate the chromosome's two chromatids.

8.5.5 Metaphase

After the microtubules have located and attached to the kinetochores in prometaphase, the two centrosomes

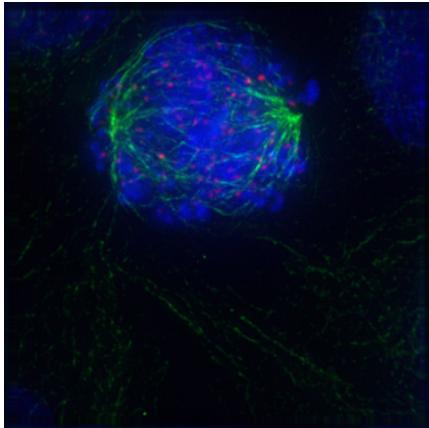


Figure 8.8: Early prometaphase: The nuclear membrane has just disassembled, allowing the microtubules to quickly interact with the kinetochores, which assemble on the centromeres of the condensing chromosomes.¹⁹

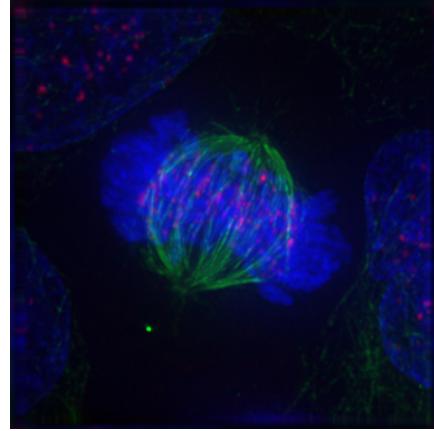


Figure 8.9: Metaphase: The centrosomes have moved to the poles of the cell and have established the mitotic spindle. The chromosomes have congressed at the metaphase plate.²⁰

begin pulling the chromosomes towards opposite ends of the cell. The resulting tension causes the chromosomes to align along the metaphase plate or equatorial plane, an imaginary line that is centrally located between the two centrosomes (at approximately the midline of the cell). To ensure equitable distribution of chromosomes at the end of mitosis, the metaphase checkpoint guarantees that kinetochores are properly attached to the mitotic spindle and that the chromosomes are aligned along the metaphase plate. If the cell successfully passes through the metaphase checkpoint, it proceeds to anaphase.

8.5.6 Anaphase

During anaphase A, the cohesins that bind sister chromatids together are cleaved, forming two identical daughter chromosomes. Shortening of the kinetochore microtubules pulls the newly formed daughter chromosomes to opposite ends of the cell. During anaphase B, polar microtubules push against each other, causing the cell to elongate. In late anaphase, chromosomes also reach their overall maximal condensation level, to help chromosome segregation and the re-formation of the nucleus. In most animal cells, anaphase A precedes anaphase B, but some vertebrate egg cells demonstrate the opposite order of events.

8.5.7 Telophase

Telophase (from the Greek word τελος meaning “end”) is a reversal of prophase and prometaphase events. At telophase, the polar microtubules continue to lengthen, elongating the cell even more. If the nuclear envelope has broken down, a new nuclear envelope forms using

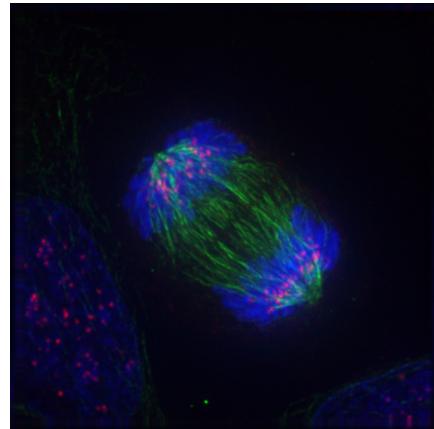


Figure 8.10: Anaphase: Kinetochore microtubules pull the two sets of chromosomes apart, and lengthening polar microtubules push the halves of the dividing cell further apart, while chromosomes are condensed maximally.²¹

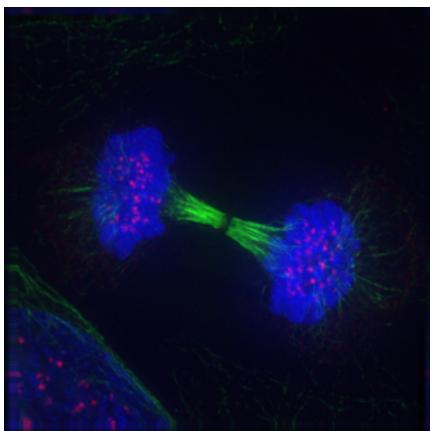


Figure 8.11: Telophase: Reversal of prophase and prometaphase events and thus completing the cell cycle.²²

the membrane vesicles of the parent cell's old nuclear envelope. The new envelope forms around each set of separated daughter chromosomes (though the membrane does not enclose the centrosomes) and the nucleolus reappears. Both sets of chromosomes, now surrounded by new nuclear membrane, begin to "relax" or decondense. Mitosis is complete. Each daughter nucleus has an identical set of chromosomes. Cell division may or may not occur at this time depending on the organism.

8.5.8 Cytokinesis

Mitosis is immediately followed by cytokinesis, which divides the nuclei, cytoplasm, organelles and cell membrane into two cells containing roughly equal shares of these cellular components. Mitosis and cytokinesis together define the division of the mother cell into two daughter cells, genetically identical to each other and to their parent cell. This accounts for approximately 10% of the cell cycle.

Cytokinesis is not a phase of mitosis but rather a separate process, necessary for completing cell division. In animal cells, a cleavage furrow (pinch) containing a contractile ring develops where the metaphase plate used to be, pinching off the separated nuclei. In both animal and plant cells, cell division is also driven by vesicles derived from the Golgi apparatus, which move along microtubules to the middle of the cell. In plants, this structure coalesces into a cell plate at the center of the phragmoplast and develops into a cell wall, separating the two nuclei. Each daughter cell has a complete copy of the genome of its parent cell. The end of cytokinesis marks the end of the M-phase.

There are many cells where mitosis and cytokinesis occur separately, forming single cells with multiple nuclei. The most notable occurrence of this is among the fungi,

slime molds, and coenocytic algae, but the phenomenon is found in various other organisms. Even in animals, cytokinesis and mitosis may occur independently, for instance during certain stages of fruit fly embryonic development.

Mitosis occurs in the following circumstances:

- Development and growth: The number of cells within an organism increases by mitosis. This is the basis of the development of a multicellular body from a single cell, i.e., zygote and also the basis of the growth of a multicellular body.
- Cell replacement: In some parts of the body, e.g. skin and digestive tract, cells are constantly sloughed off and replaced by new ones. New cells are formed by mitosis and so are exact copies of the cells being replaced.
- Regeneration: Some organisms can regenerate body parts. The production of new cells in such instances is achieved by mitosis. For example, starfish regenerate lost arms through mitosis.
- Asexual reproduction: Some organisms produce genetically similar offspring through asexual reproduction. For example, the hydra reproduces asexually by budding. The cells at the surface of hydra undergo mitosis and form a mass called a bud. Mitosis continues in the cells of the bud and this grows into a new individual. The same division happens during asexual reproduction or vegetative propagation in plants.

8.6 Regulation of The Eukaryotic Cell Cycle

Regulation of the cell cycle involves processes crucial to the survival of a cell, including the detection and repair of genetic damage as well as the prevention of uncontrolled cell division. The molecular events that control the cell cycle are ordered and directional; that is, each process occurs in a sequential fashion and it is impossible to "reverse" the cycle.

8.6.1 Role of cyclins and CDKs

Two key classes of regulatory molecules, cyclins and cyclin-dependent kinases (CDKs), determine a cell's progress through the cell cycle. Leland H. Hartwell²³, R. Timothy Hunt²⁴, and Paul M. Nurse²⁵ won the 2001 Nobel Prize in Physiology or Medicine for their discovery of these central molecules. Many of the genes encoding cyclins and CDKs are conserved among all eukaryotes, but in

²³https://en.wikipedia.org/wiki/Leland_H._Hartwell

²⁴https://en.wikipedia.org/wiki/Tim_Hunt

²⁵https://en.wikipedia.org/wiki/Paul_Nurse

general more complex organisms have more elaborate cell cycle control systems that incorporate more individual components. Many of the relevant genes were first identified by studying yeast, especially *Saccharomyces cerevisiae*; genetic nomenclature in yeast dubs many of these genes *cdc* (for "cell division cycle") followed by an identifying number, e.g. *cdc25* or *cdc20*.

Cyclins form the regulatory subunits and CDKs the catalytic subunits of an activated heterodimer; cyclins have no catalytic activity and CDKs are inactive in the absence of a partner cyclin. When activated by a bound cyclin, CDKs perform a common biochemical reaction called phosphorylation that activates or inactivates target proteins to orchestrate coordinated entry into the next phase of the cell cycle. Different cyclin-CDK combinations determine the downstream proteins targeted. CDKs are constitutively expressed in cells whereas cyclins are synthesised at specific stages of the cell cycle, in response to various molecular signals.

General mechanism of cyclin-CDK interaction Upon receiving a pro-mitotic extracellular signal, G1 cyclin-CDK complexes become active to prepare the cell for S phase, promoting the expression of transcription factors that in turn promote the expression of S cyclins and of enzymes required for DNA replication. The G1 cyclin-CDK complexes also promote the degradation of molecules that function as S phase inhibitors by targeting them for ubiquitination. Once a protein has been ubiquitinated, it is targeted for proteolytic degradation by the proteasome. However, results from a recent study of E2F transcriptional dynamics at the single-cell level argue that the role of G1 cyclin-CDK activities, in particular cyclin D-CDK4/6, is to tune the timing rather than the commitment of cell cycle entry.

Active S cyclin-CDK complexes phosphorylate proteins that make up the pre-replication complexes assembled during G1 phase on DNA replication origins. The phosphorylation serves two purposes: to activate each already-assembled pre-replication complex, and to prevent new complexes from forming. This ensures that every portion of the cell's genome will be replicated once and only once. The reason for prevention of gaps in replication is fairly clear, because daughter cells that are missing all or part of crucial genes will die. However, for reasons related to gene copy number effects, possession of extra copies of certain genes is also deleterious to the daughter cells.

Mitotic cyclin-CDK complexes, which are synthesized but inactivated during S and G2 phases, promote the initiation of mitosis by stimulating downstream proteins involved in chromosome condensation and mitotic spindle assembly. A critical complex activated during this process

is a ubiquitin ligase known as the anaphase-promoting complex (APC), which promotes degradation of structural proteins associated with the chromosomal kinetochore. APC also targets the mitotic cyclins for degradation, ensuring that telophase and cytokinesis can proceed.

Specific action of cyclin-CDK complexes Cyclin D is the first cyclin produced in the cells that enter the cell cycle, in response to extracellular signals (e.g. growth factors). Cyclin D levels stay low in resting cells that are not proliferating. Additionally, CDK4/6 and CDK2 are also inactive because CDK4/6 are bound by INK4 family members (e.g., p16), limiting kinase activity. Meanwhile, CDK2 complexes are inhibited by the CIP/KIP proteins such as p21 and p27. When it is time for a cell to enter the cell cycle, which is triggered by a mitogenic stimuli, levels of cyclin D increase. In response to this trigger, cyclin D binds to existing CDK4/6, forming the active cyclin D-CDK4/6 complex. Cyclin D-CDK4/6 complexes in turn mono-phosphorylates the retinoblastoma susceptibility protein (Rb) to pRb. The un-phosphorylated Rb tumour suppressor functions in inducing cell cycle exit and maintaining G0 arrest (senescence).

In the last few decades, a model has been widely accepted whereby pRb proteins are inactivated by cyclin D-Cdk4/6-mediated phosphorylation. Rb has 14+ potential phosphorylation sites. Cyclin D-Cdk 4/6 progressively phosphorylates Rb to hyperphosphorylated state, which triggers dissociation of pRb-E2F complexes, thereby inducing G1/S cell cycle gene expression and progression into S phase.

However, scientific observations from a recent study show that Rb is present in three types of isoforms: (1) unphosphorylated Rb in G0 state; (2) mono-phosphorylated Rb, also referred to as "hypo-phosphorylated" or 'partially' phosphorylated Rb in early G1 state; and (3) inactive hyperphosphorylated Rb in late G1 state. In early G1 cells, mono-phosphorylated Rb exists as 14 different isoforms, one of each has distinct E2F binding affinity. Rb has been found to associate with hundreds of different proteins and the idea that different mono-phosphorylated Rb isoforms have different protein partners was very appealing. A recent report confirmed that mono-phosphorylation controls Rb's association with other proteins and generates functional distinct forms of Rb. All different mono-phosphorylated Rb isoforms inhibit E2F transcriptional program and are able to arrest cells in G1-phase. Importantly, different mono-phosphorylated forms of RB have distinct transcriptional outputs that are extended beyond E2F regulation.

In general, the binding of pRb to E2F inhibits the E2F target gene expression of certain G1/S and S transition genes including E-type cyclins. The partial phosphorylation of RB de-represses the Rb-mediated suppression of

E2F target gene expression, begins the expression of cyclin E. The molecular mechanism that causes the cell switched to cyclin E activation is currently not known, but as cyclin E levels rise, the active cyclin E-CDK2 complex is formed, bringing Rb to be inactivated by hyper-phosphorylation. Hyperphosphorylated Rb is completely dissociated from E2F, enabling further expression of a wide range of E2F target genes are required for driving cells to proceed into S phase . Recently, it has been identified that cyclin D-Cdk4/6 binds to a C-terminal alpha-helix region of Rb that is only distinguishable to cyclin D rather than other cyclins, cyclin E, A and B. This observation based on the structural analysis of Rb phosphorylation supports that Rb is phosphorylated in a different level through multiple Cyclin-Cdk complexes. This also makes feasible the current model of a simultaneous switch-like inactivation of all mono-phosphorylated Rb isoforms through one type of Rb hyper-phosphorylation mechanism. In addition, mutational analysis of the cyclin D- Cdk 4/6 specific Rb C-terminal helix shows that disruptions of cyclin D-Cdk 4/6 binding to Rb prevents Rb phosphorylation, arrests cells in G1, and bolsters Rb's functions in tumor suppressor. This cyclin-Cdk driven cell cycle transitional mechanism governs a cell committed to the cell cycle that allows cell proliferation. A cancerous cell growth often accompanies with deregulation of Cyclin D-Cdk 4/6 activity.

The hyperphosphorylated Rb dissociates from the E2F/DP1/Rb complex (which was bound to the E2F responsive genes, effectively “blocking” them from transcription), activating E2F. Activation of E2F results in transcription of various genes like cyclin E, cyclin A, DNA polymerase, thymidine kinase, etc. Cyclin E thus produced binds to CDK2, forming the cyclin E-CDK2 complex, which pushes the cell from G1 to S phase (G1/S, which initiates the G2/M transition). Cyclin B-cdk1 complex activation causes breakdown of nuclear envelope and initiation of prophase, and subsequently, its deactivation causes the cell to exit mitosis. A quantitative study of E2F transcriptional dynamics at the single-cell level by using engineered fluorescent reporter cells provided a quantitative framework for understanding the control logic of cell cycle entry, challenging the canonical textbook model. Genes that regulate the amplitude of E2F accumulation, such as Myc, determine the commitment in cell cycle and S phase entry. G1 cyclin-CDK activities are not the driver of cell cycle entry. Instead, they primarily tune the timing of E2F increase, thereby modulating the pace of cell cycle progression.

8.6.2 Cell Cycle Checkpoints

Cell cycle checkpoints are used by the cell to monitor and regulate the progress of the cell cycle. Checkpoints pre-

vent cell cycle progression at specific points, allowing verification of necessary phase processes and repair of DNA damage. The cell cannot proceed to the next phase until checkpoint requirements have been met. Checkpoints typically consist of a network of regulatory proteins that monitor and dictate the progression of the cell through the different stages of the cell cycle.

There are several checkpoints to ensure that damaged or incomplete DNA is not passed on to daughter cells. Three main checkpoints exist: the G1/S checkpoint, the G2/M checkpoint and the metaphase (mitotic) checkpoint.

G1/S transition is a rate-limiting step in the cell cycle and is also known as restriction point. This is where the cell checks whether it has enough raw materials to fully replicate its DNA (nucleotide bases, DNA synthase, chromatin, etc.). An unhealthy or malnourished cell will get stuck at this checkpoint.

The G2/M checkpoint is where the cell ensures that it has enough cytoplasm and phospholipids for two daughter cells. But sometimes more importantly, it checks to see if it is the right time to replicate. There are some situations where many cells need to all replicate simultaneously (for example, a growing embryo should have a symmetric cell distribution until it reaches the mid-blastula transition). This is done by controlling the G2/M checkpoint.

The metaphase checkpoint is a fairly minor checkpoint, in that once a cell is in metaphase, it has committed to undergoing mitosis. However that's not to say it isn't important. In this checkpoint, the cell checks to ensure that the spindle has formed and that all of the chromosomes are aligned at the spindle equator before anaphase begins.

While these are the three “main” checkpoints, not all cells have to pass through each of these checkpoints in this order to replicate. Many types of cancer are caused by mutations that allow the cells to speed through the various checkpoints or even skip them altogether. Going from S to M to S phase almost consecutively. Because these cells have lost their checkpoints, any DNA mutations that may have occurred are disregarded and passed on to the daughter cells. This is one reason why cancer cells have a tendency to exponentially accrue mutations. Aside from cancer cells, many fully differentiated cell types no longer replicate so they leave the cell cycle and stay in G0 until their death. Thus removing the need for cellular checkpoints. An alternative model of the cell cycle response to DNA damage has also been proposed, known as the postreplication checkpoint.

Checkpoint regulation plays an important role in an organism's development. In sexual reproduction, when egg fertilization occurs, when the sperm binds to the egg, it releases signalling factors that notify the egg that it has

been fertilized. Among other things, this induces the now fertilized oocyte to return from its previously dormant, G₀, state back into the cell cycle and on to mitotic replication and division.

p53 plays an important role in triggering the control mechanisms at both G₁/S and G₂/M checkpoints. In addition to p53, checkpoint regulators are being heavily researched for their roles in cancer growth and proliferation.

8.6.3 Role in Tumor Formation

A disregulation of the cell cycle components may lead to tumor formation. As mentioned above, when some genes like the cell cycle inhibitors, RB, p53 etc. mutate, they may cause the cell to multiply uncontrollably, forming a tumor. Although the duration of cell cycle in tumor cells is equal to or longer than that of normal cell cycle, the proportion of cells that are in active cell division (versus quiescent cells in G₀ phase) in tumors is much higher than that in normal tissue.[citation needed] Thus there is a net increase in cell number as the number of cells that die by apoptosis or senescence remains the same.

The cells which are actively undergoing cell cycle are targeted in cancer therapy as the DNA is relatively exposed during cell division and hence susceptible to damage by drugs or radiation. This fact is made use of in cancer treatment; by a process known as debulking, a significant mass of the tumor is removed which pushes a significant number of the remaining tumor cells from G₀ to G₁ phase (due to increased availability of nutrients, oxygen, growth factors etc.). Radiation or chemotherapy following the debulking procedure kills these cells which have newly entered the cell cycle.

The fastest cycling mammalian cells in culture, crypt cells in the intestinal epithelium, have a cycle time as short as 9 to 10 hours. Stem cells in resting mouse skin may have a cycle time of more than 200 hours. Most of this difference is due to the varying length of G₁, the most variable phase of the cycle. M and S do not vary much.

In general, cells are most radiosensitive in late M and G₂ phases and most resistant in late S phase.

For cells with a longer cell cycle time and a significantly long G₁ phase, there is a second peak of resistance late in G₁.

The pattern of resistance and sensitivity correlates with the level of sulphhydryl compounds in the cell. Sulphhydryls are natural substances that protect cells from radiation damage and tend to be at their highest levels in S and at their lowest near mitosis.

Homologous recombination (HR) is an accurate process for repairing DNA double-strand breaks. HR is nearly ab-

sent in G₁ phase, is most active in S phase, and declines in G₂/M. Non-homologous end joining, a less accurate and more mutagenic process for repairing double strand breaks, is active throughout the cell cycle.

Chapter 9

Sexual Reproduction And Meiosis

Reproduction¹ (or procreation or breeding) is the biological process by which new individual organisms – “offspring” – are produced from their “parents”. Reproduction is a fundamental feature of all known life; each individual organism exists as the result of reproduction. There are two forms of reproduction: asexual and sexual.

In asexual reproduction, an organism can reproduce without the involvement of another organism. Asexual reproduction is not limited to single-celled organisms. The cloning of an organism is a form of asexual reproduction. By asexual reproduction, an organism creates a genetically similar or identical copy of itself. The evolution of sexual reproduction is a major puzzle for biologists. The two-fold cost of sexual reproduction is that only 50% of organisms reproduce and organisms only pass on 50% of their genes.

9.0.1 Sexual reproduction

Sexual reproduction is a biological process that creates a new organism by combining the genetic material of two organisms in a process that starts with meiosis, a specialized type of cell division. Each of two parent organisms contributes half of the offspring's genetic makeup by creating haploid gametes. Most organisms form two different types of gametes. In these anisogamous species, the two sexes are referred to as male (producing sperm or microspores) and female (producing ova or megasporangia). In isogamous species, the gametes are similar or identical in form (isogametes), but may have separable properties and then may be given other different names. For example, in the green alga, *Chlamydomonas reinhardtii*, there are so-called “plus” and “minus” gametes. A few types of organisms, such as many fungi and the ciliate *Paramecium aurelia*, have more than two “sexes”, called syngens. Most animals (including humans) and plants reproduce sexually. Sexually reproducing organisms have different sets of genes for every trait (called alleles).



Figure 9.1: Sperm fertilizing an egg.²

9.1 Meiosis

Meiosis (from Greek μείωσις, meiosis, which means lessening) is a special type of cell division that reduces the chromosome number by half, creating four haploid cells, each genetically distinct from the parent cell that gave rise to them. This process occurs in all sexually reproducing single-celled and multicellular eukaryotes, including animals, plants, and fungi. Meiotic cell divisions are an essential process during oogenesis and spermatogenesis.

Errors in meiosis resulting in aneuploidy are the leading known cause of miscarriage and the most frequent genetic cause of developmental disabilities.

In meiosis, DNA replication is followed by two rounds of cell division to produce four daughter cells, each with half the number of chromosomes as the original parent cell. The two meiotic divisions are known as Meiosis I and Meiosis II. Before meiosis begins, during S phase of the cell cycle, the DNA of each chromosome is replicated so that it consists of two identical sister chromatids, which remain held together through sister chromatid cohesion. This S-phase can be referred to as “premeiotic S-phase” or “meiotic S-phase”. Immediately following DNA replication, meiotic cells enter a prolonged G2-like stage known as meiotic prophase. During this time, homologous chromosomes pair with each other and undergo genetic recombination, a programmed process in which DNA may be cut

¹<https://en.wikipedia.org/wiki/Reproduction>

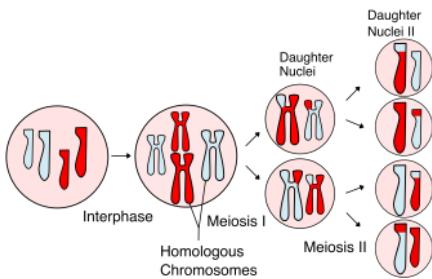


Figure 9.2: In meiosis³, the chromosome or chromosomes duplicate (during interphase) and homologous chromosomes exchange genetic information (chromosomal crossover) during the first division, called meiosis I. The daughter cells divide again in meiosis II, splitting up sister chromatids to form haploid gametes. Two gametes fuse during fertilization, creating a diploid cell with a complete set of paired chromosomes.

and then repaired, which allows them to exchange some of their genetic information. A subset of recombination events results in crossovers, which create physical links known as chiasmata (singular: chiasma, for the Greek letter Chi (X)) between the homologous chromosomes. In most organisms, these links can help direct each pair of homologous chromosomes to segregate away from each other during Meiosis I, resulting in two haploid cells that have half the number of chromosomes as the parent cell. During Meiosis II, the cohesion between sister chromatids is released and they segregate from one another, as during mitosis. In some cases all four of the meiotic products form gametes such as sperm, spores, or pollen. In female animals, three of the four meiotic products are typically eliminated by extrusion into polar bodies, and only one cell develops to produce an ovum. Because the number of chromosomes is halved during meiosis, gametes can fuse (i.e. fertilization) to form a diploid zygote that contains two copies of each chromosome, one from each parent. Thus, alternating cycles of meiosis and fertilization enable sexual reproduction, with successive generations maintaining the same number of chromosomes. For example, diploid human cells contain 23 pairs of chromosomes including 1 pair of sex chromosomes (46 total), half of maternal origin and half of paternal origin. Meiosis produces haploid gametes (ova or sperm) that contain one set of 23 chromosomes. When two gametes (an egg and a sperm) fuse, the resulting zygote is once again diploid, with the mother and father each contributing 23 chromosomes. This same pattern, but not the same number of chromosomes, occurs in all organisms that utilize meiosis.

Meiosis begins with a diploid cell, which contains two copies of each chromosome, termed homologs. First, the cell undergoes DNA replication, so each homolog now con-

sists of two identical sister chromatids. Then each set of homologs pair with each other and exchange genetic information by homologous recombination often leading to physical connections (crossovers) between the homologs. In the first meiotic division, the homologs are segregated to separate daughter cells by the spindle apparatus. The cells then proceed to a second division without an intervening round of DNA replication. The sister chromatids are segregated to separate daughter cells to produce a total of four haploid cells. Female animals employ a slight variation on this pattern and produce one large ovum and two small polar bodies. Because of recombination, an individual chromatid can consist of a new combination of maternal and paternal genetic information, resulting in offspring that are genetically distinct from either parent. Furthermore, an individual gamete can include an assortment of maternal, paternal, and recombinant chromatids. This genetic diversity resulting from sexual reproduction contributes to the variation in traits upon which natural selection can act.

Meiosis uses many of the same mechanisms as mitosis, the type of cell division used by eukaryotes to divide one cell into two identical daughter cells. In some plants, fungi, and protists meiosis results in the formation of spores: haploid cells that can divide vegetatively without undergoing fertilization. Some eukaryotes, like bdelloid rotifers, do not have the ability to carry out meiosis and have acquired the ability to reproduce by parthenogenesis.

Meiosis does not occur in archaea or bacteria, which generally reproduce asexually via binary fission. However, a “sexual” process known as horizontal gene transfer involves the transfer of DNA from one bacterium or archaeon to another and recombination of these DNA molecules of different parental origin.

Meiosis was discovered and described for the first time in sea urchin eggs in 1876 by the German biologist Oscar Hertwig⁴. It was described again in 1883, at the level of chromosomes, by the Belgian zoologist Edouard Van Beneden⁵, in *Ascaris* roundworm eggs. The significance of meiosis for reproduction and inheritance, however, was described only in 1890 by German biologist August Weismann⁶, who noted that two cell divisions were necessary to transform one diploid cell into four haploid cells if the number of chromosomes had to be maintained. In 1911 the American geneticist Thomas Hunt Morgan⁷ detected crossovers in meiosis in the fruit fly *Drosophila melanogaster*, which helped to establish that genetic traits are transmitted on chromosomes.

⁴https://en.wikipedia.org/wiki/Oscar_Hertwig

⁵https://en.wikipedia.org/wiki/Edouard_Van_Beneden

⁶https://en.wikipedia.org/wiki/August_Weismann

⁷https://en.wikipedia.org/wiki/Thomas_Hunt_Morgan

The term “meiosis” (originally spelled “maiosis”) is derived from the Greek word μείωσις, meaning ‘lessening’. It was introduced to biology by J.B. Farmer⁸ and J.E.S. Moore⁹ in 1905:

We propose to apply the terms Maiosis or Maiotic phase to cover the whole series of nuclear changes included in the two divisions that were designated as Heterotype and Homotype by Flemming.”

The term was linguistically corrected to “meiosis” by Koernicke (1905), and by Pantel and De Sinety (1906).

Meiosis is divided into meiosis I and meiosis II which are further divided into Karyokinesis I and Cytokinesis I and Karyokinesis II and Cytokinesis II respectively. The preparatory steps that lead up to meiosis are identical in pattern and name to interphase of the mitotic cell cycle. Interphase is divided into three phases:

- Growth 1 (G1) phase: In this very active phase, the cell synthesizes its vast array of proteins, including the enzymes and structural proteins it will need for growth. In G1, each of the chromosomes consists of a single linear molecule of DNA.
- Synthesis (S) phase: The genetic material is replicated; each of the cell's chromosomes duplicates to become two identical sister chromatids attached at a centromere. This replication does not change the ploidy of the cell since the centromere number remains the same. The identical sister chromatids have not yet condensed into the densely packaged chromosomes visible with the light microscope. This will take place during prophase I in meiosis.
- Growth 2 (G2) phase: G2 phase as seen before mitosis is not present in meiosis. Meiotic prophase corresponds most closely to the G2 phase of the mitotic cell cycle.

Interphase is followed by meiosis I and then meiosis II. Meiosis I separates replicated homologous chromosomes, each still made up of two sister chromatids, into two daughter cells, thus reducing the chromosome number by half. During meiosis II, sister chromatids decouple and the resultant daughter chromosomes are segregated into four daughter cells. For diploid organisms, the daughter cells resulting from meiosis are haploid and contain only one copy of each chromosome. In some species, cells enter a resting phase known as interkinesis between meiosis I and meiosis II.

Meiosis I and II are each divided into prophase, metaphase, anaphase, and telophase stages, similar in purpose to their analogous subphases in the mitotic

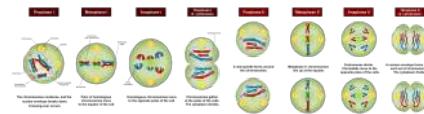


Figure 9.3: Diagram of the meiotic phases¹⁰

cell cycle. Therefore, meiosis includes the stages of meiosis I (prophase I, metaphase I, anaphase I, telophase I) and meiosis II (prophase II, metaphase II, anaphase II, telophase II).

Meiosis generates gamete genetic diversity in two ways: (1) Law of Independent Assortment. The independent orientation of homologous chromosome pairs along the metaphase plate during metaphase I & orientation of sister chromatids in metaphase II, this is the subsequent separation of homologs and sister chromatids during anaphase I & II, it allows a random and independent distribution of chromosomes to each daughter cell (and ultimately to gametes); and (2) Crossing Over. The physical exchange of homologous chromosomal regions by homologous recombination during prophase I results in new combinations of genetic information within chromosomes.

During meiosis, specific genes are more highly transcribed. In addition to strong meiotic stage-specific expression of mRNA, there are also pervasive translational controls (e.g. selective usage of preformed mRNA), regulating the ultimate meiotic stage-specific protein expression of genes during meiosis. Thus, both transcriptional and translational controls determine the broad restructuring of meiotic cells needed to carry out meiosis.

9.1.1 Meiosis I

Meiosis I segregates homologous chromosomes, which are joined as tetrads ($2n$, $4c$), producing two haploid cells (n chromosomes, 23 in humans) which each contain chromatid pairs ($1n$, $2c$). Because the ploidy is reduced from diploid to haploid, meiosis I is referred to as a reductional division. Meiosis II is an equational division analogous to mitosis, in which the sister chromatids are segregated, creating four haploid daughter cells ($1n$, $1c$).

9.1.2 Prophase I

Prophase I is typically the longest phase of meiosis. During prophase I, homologous chromosomes pair and exchange genetic information (homologous recombination). This often results in chromosomal crossover. This process facilitates pairing between homologous chromosomes and hence accurate segregation of the chromosomes at the first meiosis division. The new combinations of DNA created

⁸https://en.wikipedia.org/wiki/John_Bretland_Farmer

⁹https://en.wikipedia.org/wiki/John_Edmund_Sharrock_Moore

during crossover are a significant source of genetic variation, and result in new combinations of alleles, which may be beneficial. The paired and replicated chromosomes are called bivalents or tetrads, which have two chromosomes and four chromatids, with one chromosome coming from each parent. The process of pairing the homologous chromosomes is called synapsis. At this stage, non-sister chromatids may cross-over at points called chiasmata (plural; singular chiasma). Prophase I has historically been divided into a series of substages which are named according to the appearance of chromosomes.

The first stage of prophase I is the leptotene stage, also known as leptonema, from Greek words meaning “thin threads”. In this stage of prophase I, individual chromosomes—each consisting of two sister chromatids—become “individualized” to form visible strands within the nucleus. The two sister chromatids closely associate and are visually indistinguishable from one another. During leptotene, lateral elements of the synaptonemal complex assemble. Leptotene is of very short duration and progressive condensation and coiling of chromosome fibers takes place.

The zygotene stage, also known as zygonema, from Greek words meaning “paired threads”, occurs as the chromosomes approximately line up with each other into homologous chromosome pairs. In some organisms, this is called the bouquet stage because of the way the telomeres cluster at one end of the nucleus. At this stage, the synapsis (pairing/coming together) of homologous chromosomes takes place, facilitated by assembly of central element of the synaptonemal complex. Pairing is brought about in a zipper-like fashion and may start at the centromere (procentric), at the chromosome ends (proterinal), or at any other portion (intermediate). Individuals of a pair are equal in length and in position of the centromere. Thus pairing is highly specific and exact. The paired chromosomes are called bivalent or tetrad chromosomes.

The pachytene stage, also known as pachynema, from Greek words meaning “thick threads”. At this point a tetrad of the chromosomes has formed known as a bivalent. This is the stage when homologous recombination, including chromosomal crossover (crossing over), occurs. Nonsister chromatids of homologous chromosomes may exchange genetic information over regions of homology. Sex chromosomes, however, are not wholly identical, and only exchange information over a small region of homology. At the sites where exchange happens, chiasmata form. The exchange of information between the non-sister chromatids results in a recombination of information; each chromosome has the complete set of information it had before, and there are no gaps formed as a result of the process. Because the chromosomes

cannot be distinguished in the synaptonemal complex, the actual act of crossing over is not perceivable through the microscope, and chiasmata are not visible until the next stage.

During the diplotene stage, also known as diplonema, from Greek words meaning “two threads”, the synaptonemal complex degrades and homologous chromosomes separate from one another a little. The chromosomes themselves uncoil a bit, allowing some transcription of DNA. However, the homologous chromosomes of each bivalent remain tightly bound at chiasmata, the regions where crossing-over occurred. The chiasmata remain on the chromosomes until they are severed at the transition to anaphase I.

In human fetal oogenesis, all developing oocytes develop to this stage and are arrested in prophase I before birth. This suspended state is referred to as the dictyotene stage or dictyate. It lasts until meiosis is resumed to prepare the oocyte for ovulation, which happens at puberty or even later.

Chromosomes condense further during the diakinesis stage, from Greek words meaning “moving through”. This is the first point in meiosis where the four parts of the tetrads are actually visible. Sites of crossing over entangle together, effectively overlapping, making chiasmata clearly visible. Other than this observation, the rest of the stage closely resembles prometaphase of mitosis; the nucleoli disappear, the nuclear membrane disintegrates into vesicles, and the meiotic spindle begins to form.

During these stages, two centrosomes, containing a pair of centrioles in animal cells, migrate to the two poles of the cell. These centrosomes, which were duplicated during S-phase, function as microtubule organizing centers nucleating microtubules, which are essentially cellular ropes and poles. The microtubules invade the nuclear region after the nuclear envelope disintegrates, attaching to the chromosomes at the kinetochore. The kinetochore functions as a motor, pulling the chromosome along the attached microtubule toward the originating centrosome, like a train on a track. There are four kinetochores on each tetrad, but the pair of kinetochores on each sister chromatid fuses and functions as a unit during meiosis I.

9.1.3 Metaphase I

Homologous pairs move together along the metaphase plate: As kinetochore microtubules from both centrosomes attach to their respective kinetochores, the paired homologous chromosomes align along an equatorial plane that bisects the spindle, due to continuous counterbalancing forces exerted on the bivalents by the microtubules emanating from the two kinetochores of

homologous chromosomes. This attachment is referred to as a bipolar attachment. The physical basis of the independent assortment of chromosomes is the random orientation of each bivalent along the metaphase plate, with respect to the orientation of the other bivalents along the same equatorial line. The protein complex cohesin holds sister chromatids together from the time of their replication until anaphase. In mitosis, the force of kinetochore microtubules pulling in opposite directions creates tension. The cell senses this tension and does not progress with anaphase until all the chromosomes are properly bi-oriented. In meiosis, establishing tension ordinarily requires at least one crossover per chromosome pair in addition to cohesin between sister chromatids (see Chromosome segregation).

9.1.4 Anaphase I

Kinetochore microtubules shorten, pulling homologous chromosomes (which each consist of a pair of sister chromatids) to opposite poles. Nonkinetochore microtubules lengthen, pushing the centrosomes farther apart. The cell elongates in preparation for division down the center. Unlike in mitosis, only the cohesin from the chromosome arms is degraded while the cohesin surrounding the centromere remains protected by a protein named Shugoshin (Japanese for “guardian spirit”), what prevents the sister chromatids from separating. This allows the sister chromatids to remain together while homologs are segregated.

9.1.5 Telophase I

The first meiotic division effectively ends when the chromosomes arrive at the poles. Each daughter cell now has half the number of chromosomes but each chromosome consists of a pair of chromatids. The microtubules that make up the spindle network disappear, and a new nuclear membrane surrounds each haploid set. The chromosomes uncoil back into chromatin. Cytokinesis, the pinching of the cell membrane in animal cells or the formation of the cell wall in plant cells, occurs, completing the creation of two daughter cells. Sister chromatids remain attached during telophase I.

Cells may enter a period of rest known as interkinesis or interphase II. No DNA replication occurs during this stage.

9.1.6 Meiosis II

Meiosis II is the second meiotic division, and usually involves equational segregation, or separation of sister chromatids. Mechanically, the process is similar to mitosis, though its genetic results are fundamentally different. The

end result is production of four haploid cells (n chromosomes, 23 in humans) from the two haploid cells (with n chromosomes, each consisting of two sister chromatids) produced in meiosis I. The four main steps of meiosis II are: prophase II, metaphase II, anaphase II, and telophase II.

In prophase II we see the disappearance of the nucleoli and the nuclear envelope again as well as the shortening and thickening of the chromatids. Centrosomes move to the polar regions and arrange spindle fibers for the second meiotic division.

In metaphase II, the centromeres contain two kinetochores that attach to spindle fibers from the centrosomes at opposite poles. The new equatorial metaphase plate is rotated by 90 degrees when compared to meiosis I, perpendicular to the previous plate.

This is followed by anaphase II, in which the remaining centromeric cohesin, not protected by Shugoshin anymore, is cleaved, allowing the sister chromatids to segregate. The sister chromatids by convention are now called sister chromosomes as they move toward opposing poles.

The process ends with telophase II, which is similar to telophase I, and is marked by decondensation and lengthening of the chromosomes and the disassembly of the spindle. Nuclear envelopes re-form and cleavage or cell plate formation eventually produces a total of four daughter cells, each with a haploid set of chromosomes.

Meiosis is now complete and ends up with four new daughter cells.

Meiosis occurs in eukaryotic life cycles involving sexual reproduction, consisting of the constant cyclical process of meiosis and fertilization. This takes place alongside normal mitotic cell division. In multicellular organisms, there is an intermediary step between the diploid and haploid transition where the organism grows. At certain stages of the life cycle, germ cells produce gametes. Somatic cells make up the body of the organism and are not involved in gamete production.

Cycling meiosis and fertilization events produces a series of transitions back and forth between alternating haploid and diploid states. The organism phase of the life cycle can occur either during the diploid state (diplontic life cycle), during the haploid state (haplontic life cycle), or both (haplodiplontic life cycle, in which there are two distinct organism phases, one during the haploid state and the other during the diploid state).

In the diplontic life cycle (with pre-gametic meiosis), of which humans are a part, the organism is diploid, grown from a diploid cell called the zygote. The organism's diploid germ-line stem cells undergo meiosis to create

haploid gametes (the spermatozoa for males and ova for females), which fertilize to form the zygote. The diploid zygote undergoes repeated cellular division by mitosis to grow into the organism.

In the haplontic life cycle (with post-zygotic meiosis), the organism is haploid instead, spawned by the proliferation and differentiation of a single haploid cell called the gamete. Two organisms of opposing sex contribute their haploid gametes to form a diploid zygote. The zygote undergoes meiosis immediately, creating four haploid cells. These cells undergo mitosis to create the organism. Many fungi and many protozoa utilize the haplontic life cycle.

Finally, in the haplodiplontic life cycle (with sporic or intermediate meiosis), the living organism alternates between haploid and diploid states. Consequently, this cycle is also known as the alternation of generations. The diploid organism's germ-line cells undergo meiosis to produce spores. The spores proliferate by mitosis, growing into a haploid organism. The haploid organism's gamete then combines with another haploid organism's gamete, creating the zygote. The zygote undergoes repeated mitosis and differentiation to become a diploid organism again. The haplodiplontic life cycle can be considered a fusion of the diplontic and haplontic life cycles.

9.1.7 Nondisjunction

The normal separation of chromosomes in meiosis I or sister chromatids in meiosis II is termed disjunction. When the segregation is not normal, it is called nondisjunction. This results in the production of gametes which have either too many or too few of a particular chromosome, and is a common mechanism for trisomy or monosomy. Nondisjunction can occur in the meiosis I or meiosis II, phases of cellular reproduction, or during mitosis.

9.2 Chromosomal Disorders

A chromosomal disorder¹¹, anomaly, aberration, or mutation is a missing, extra, or irregular portion of chromosomal DNA. It can be from a typical number of chromosomes or a structural abnormality in one or more chromosomes.

9.2.1 Aberrations

Chromosomal aberrations are disruptions in the normal chromosomal content of a cell and are a major cause of genetic conditions in humans, such as Down syndrome, although most aberrations have little to no effect. Some chromosome abnormalities do not cause disease in carriers, such as translocations, or chromosomal inversions, al-

though they may lead to a higher chance of bearing a child with a chromosome disorder. Abnormal numbers of chromosomes or chromosome sets, called aneuploidy, may be lethal or may give rise to genetic disorders. Genetic counseling is offered for families that may carry a chromosome rearrangement.

9.2.2 Numerical abnormalities

Aneuploidy¹² is the presence of an abnormal number of chromosomes in a cell, for example when an individual either is missing a chromosome from a pair (monosomy) or has more than two chromosomes of a pair (trisomy, tetrasomy, etc.). In the strict sense, a chromosome complement having a number of chromosomes other than 46 (in humans) is considered heteroploid while an exact multiple of the haploid chromosome complement is considered euploid. Thus, a cell with any number of complete chromosome sets is called a euploid cell. An extra or missing chromosome is a common cause of genetic disorders, including some human birth defects. Some cancer cells also have abnormal numbers of chromosomes. About 68% of human solid tumors are aneuploid. Aneuploidy originates during cell division when the chromosomes do not separate properly between the two cells.

¹¹https://en.wikipedia.org/wiki/Chromosome_abnormality

¹²<https://en.wikipedia.org/wiki/Aneuploidy>

Table 9.1: Terminology and examples of heteroploidy in humans.

Number of chromosomes	Name	Description
1	Monosomy	Monosomy refers to lack of one chromosome of the normal complement. Partial monosomy can occur in unbalanced translocations or deletions, in which only a portion of the chromosome is present in a single copy (see deletion (genetics)). Monosomy of the sex chromosomes (45,X) causes Turner syndrome.
2	Disomy	Disomy is the presence of two copies of a chromosome. For organisms such as humans that have two copies of each chromosome (those that are diploid), it is the normal condition. For organisms that normally have three or more copies of each chromosome (those that are triploid or above), disomy is an aneuploid chromosome complement. In uniparental disomy, both copies of a chromosome come from the same parent (with no contribution from the other parent).
3	Trisomy	Trisomy refers to the presence of three copies, instead of the normal two, of a particular chromosome. The presence of an extra chromosome 21, which is found in Down syndrome, is called trisomy 21. Trisomy 18 and Trisomy 13, known as Edwards syndrome and Patau syndrome, respectively, are the two other autosomal trisomies recognized in live-born humans. Trisomy of the sex chromosomes is also possible, for example (47,XXX), (47,XXY), and (47,YY).
4/5	Tetrasomy/pentasomy	Tetrasomy and pentasomy are the presence of four or five copies of a chromosome, respectively. Although rarely seen with autosomes, sex chromosome tetrasomy and pentasomy have been reported in humans, including XXXX, XXXY, XXXXX, XXXXY, and XXXYY.

An example of trisomy in humans is Down syndrome¹³, which is a developmental disorder caused by an extra copy of chromosome 21; the disorder is therefore also called trisomy 21. Having an extra copy of this chromosome means that individuals have three copies of each of its genes instead of two, making it difficult for cells to properly control how much protein is made. Producing too much or too little protein can have serious consequences. Genes on chromosome 21 that specifically contribute to the various symptoms of Down syndrome are now being identified. The frequency of Trisomy 21 has been determined to be a function of advanced maternal age.

An example of monosomy is Turner syndrome¹⁴, where the individual is born with only one sex chromosome, an X.

Other examples include:

- Cri du chat, which is caused by the deletion of part of the short arm of chromosome 5. "Cri du chat" means "cry of the cat" in French; the condition was so-named because affected babies make high-pitched cries that sound like those of a cat. Affected individuals have wide-set eyes, a small head and jaw, moderate to severe mental health problems, and are very short.
- Edwards syndrome, or trisomy-18, the second most common trisomy. Symptoms include motor retardation, developmental disability and numerous congenital anomalies causing serious health problems. Ninety percent of those affected die in infancy. They have characteristic clenched hands and overlapping fingers.
- Isodicentric 15, also called idic(15), partial tetrasomy 15q, or inverted duplication 15 (inv dup 15).
- Jacobsen syndrome, which is very rare. It is also called the terminal 11q deletion disorder. Those affected have normal intelligence or mild developmental disability, with poor expressive language skills. Most have a bleeding disorder called Paris-Trousseau syndrome.
- Klinefelter syndrome (XXY). Men with Klinefelter syndrome are usually sterile and tend to be taller and have longer arms and legs than their peers. Boys with the syndrome are often shy and quiet and have a higher incidence of speech delay and dyslexia. Without testosterone treatment, some may develop gynecomastia during puberty.
- Patau Syndrome, also called D-Syndrome or trisomy-13. Symptoms are somewhat similar to those of trisomy-18, without the characteristic

folded hand.

- Small supernumerary marker chromosome. This means there is an extra, abnormal chromosome. Features depend on the origin of the extra genetic material. Cat-eye syndrome and isodicentric chromosome 15 syndrome (or Idic15) are both caused by a supernumerary marker chromosome, as is Pallister-Killian syndrome.
- Triple-X syndrome (XXX). XXX girls tend to be tall and thin and have a higher incidence of dyslexia.
- Wolf-Hirschhorn syndrome, which is caused by partial deletion of the short arm of chromosome 4. It is characterized by growth retardation, delayed motor skills development, "Greek Helmet" facial features, and mild to profound mental health problems.
- XYY syndrome. XYY boys are usually taller than their siblings. Like XXY boys and XXX girls, they are more likely to have learning difficulties.

Chromosome abnormalities are detected in 1 of 160 live human births. Most cases of aneuploidy in the germline result in miscarriage and the most common extra autosomal chromosomes among live births are 21, 18, and 13.

Most cells in the human body have 23 pairs of chromosomes, or a total of 46 chromosomes. (The sperm and egg, or gametes, each have 23 unpaired chromosomes, and red blood cells have no nucleus and no chromosomes). One copy of each pair is inherited from the mother and the other copy is inherited from the father. The first 22 pairs of chromosomes (called autosomes) are numbered from 1 to 22, from largest to smallest. The 23rd pair of chromosomes are the sex chromosomes. Normal females have two X chromosomes, while normal males have one X chromosome and one Y chromosome. The characteristics of the chromosomes in a cell as they are seen under a light microscope are called the karyotype.

During meiosis, when germ cells divide to create sperm and egg (gametes), each half should have the same number of chromosomes. But sometimes, the whole pair of chromosomes will end up in one gamete, and the other gamete will not get that chromosome at all.

Most embryos cannot survive with a missing or extra autosome (numbered chromosome) and are spontaneously aborted. The most frequent aneuploidy in humans is trisomy 16, although fetuses affected with the full version of this chromosome abnormality do not survive to term (it is possible for surviving individuals to have the mosaic form, where trisomy 16 exists in some cells but not all). The most common aneuploidy that infants can survive with is trisomy 21, which is found in Down syndrome, affecting 1 in 800 births. Trisomy 18 (Edwards syndrome) affects 1 in 6,000 births, and trisomy 13 (Patau syndrome) affects 1 in

¹³https://en.wikipedia.org/wiki/Down_syndrome

¹⁴https://en.wikipedia.org/wiki/Turner_syndrome

10,000 births. 10% of infants with trisomy 18 or 13 reach 1 year of age.

Changes in chromosome number may not necessarily be present in all cells in an individual. When aneuploidy is detected in a fraction of cells in an individual, it is called chromosomal mosaicism. In general, individuals who are mosaic for a chromosomal aneuploidy tend to have a less severe form of the syndrome compared to those with full trisomy. For many of the autosomal trisomies, only mosaic cases survive to term. However, mitotic aneuploidy may be more common than previously recognized in somatic tissues, and aneuploidy is a characteristic of many types of tumorigenesis (see below).

9.2.3 Mechanisms

Nondisjunction usually occurs as the result of a weakened mitotic checkpoint, as these checkpoints tend to arrest or delay cell division until all components of the cell are ready to enter the next phase. If a checkpoint is weakened, the cell may fail to ‘notice’ that a chromosome pair is not lined up on the mitotic plate, for example. In such a case, most chromosomes would separate normally (with one chromatid ending up in each cell), while others could fail to separate at all. This would generate a daughter cell lacking a copy and a daughter cell with an extra copy.

Completely inactive mitotic checkpoints may cause nondisjunction at multiple chromosomes, possibly all. Such a scenario could result in each daughter cell possessing a disjoint set of genetic material.

9.2.4 Diagnosis

Germline aneuploidy is typically detected through karyotyping, a process in which a sample of cells is fixed and stained to create the typical light and dark chromosomal banding pattern and a picture of the chromosomes is analyzed. Other techniques include fluorescence in situ hybridization (FISH), quantitative PCR of short tandem repeats, quantitative fluorescence PCR (QF-PCR), quantitative PCR dosage analysis, Quantitative Mass Spectrometry of Single Nucleotide Polymorphisms, and comparative genomic hybridization (CGH).

These tests can also be performed prenatally to detect aneuploidy in a pregnancy, through either amniocentesis or chorionic villus sampling. Pregnant women of 35 years or older are offered prenatal testing because the chance of chromosomal aneuploidy increases as the mother's age increases.

Recent advances have allowed for less invasive testing methods based on the presence of fetal genetic material in maternal blood.

9.2.5 Structural abnormalities

When the chromosome's structure is altered, this can take several forms:

- Deletions: A portion of the chromosome is missing or deleted. Known disorders in humans include Wolf-Hirschhorn syndrome, which is caused by partial deletion of the short arm of chromosome 4; and Jacobsen syndrome, also called the terminal 11q deletion disorder.
- Duplications: A portion of the chromosome is duplicated, resulting in extra genetic material. Known human disorders include Charcot-Marie-Tooth disease type 1A, which may be caused by duplication of the gene encoding peripheral myelin protein 22 (PMP22) on chromosome 17.
- Translocations: A portion of one chromosome is transferred to another chromosome. There are two main types of translocations:
 - Reciprocal translocation: Segments from two different chromosomes have been exchanged.
 - Robertsonian translocation: An entire chromosome has attached to another at the centromere – in humans these only occur with chromosomes 13, 14, 15, 21, and 22.
- Inversions: A portion of the chromosome has broken off, turned upside down, and reattached, therefore the genetic material is inverted.
- Insertions: A portion of one chromosome has been deleted from its normal place and inserted into another chromosome.
- Rings: A portion of a chromosome has broken off and formed a circle or ring. This can happen with or without loss of genetic material.
- Isochromosome: Formed by the mirror image copy of a chromosome segment including the centromere.

Chromosome instability syndromes are a group of disorders characterized by chromosomal instability and breakage. They often lead to an increased tendency to develop certain types of malignancies.

Most chromosome abnormalities occur as an accident in the egg cell or sperm, and therefore the anomaly is present in every cell of the body. Some anomalies, however, can happen after conception, resulting in mosaicism (where some cells have the anomaly and some do not). Chromosome anomalies can be inherited from a parent or be “de novo”. This is why chromosome studies are often performed on parents when a child is found to have an anomaly. If the parents do not possess the abnormality it was not initially inherited; however it may be transmitted to subsequent generations.

9.2.6 Acquired Chromosomal Abnormalities

Most cancers, if not all, involve chromosome abnormalities, with either the formation of hybrid genes and fusion proteins, deregulation of genes and overexpression of proteins, or loss of tumor suppressor genes.

Chapter 10

Classical Genetics

Genetics¹ is, generally, the study of genes, genetic variation, and heredity. This chapter is about classical genetics². This is the oldest form of genetics that began with Gregor Mendel's³ experiments that formulated and defined a fundamental biological concept now referred to as Mendelian Inheritance. The subsequent chapter is about molecular genetics⁴.

Classical genetics is the part of genetics that is about the transmission of genetic traits via the acts of reproduction⁵. The process by which characteristics are passed down from parents to their offspring is called heredity. In the sense of classical genetics, variation is known as the lack of resemblance in related individuals and can be categorized as discontinuous or continuous. The phenotype⁶ is a general term that defines an individual's visible, physical traits. The genotype⁷ of an offspring is known as its genetic makeup. Mendel was the first to systematically investigate how certain well-defined traits are transmitted from parents to offspring. Mendel's conclusions were largely ignored by the vast majority of scientists at the time. In 1900, however, his work was "re-discovered" by three European scientists, Hugo de Vries⁸, Carl Correns⁹, and Erich von Tschermak¹⁰. In 1905, Wilhelm Johannsen¹¹ introduced the term *gene* and William Bateson¹² the term *genetics*. Our understanding of what a *gene*¹³ is has undergone quite a bit of change. Currently, genes are considered to be pieces of DNA that contain information for synthesis of ribonucleic acids (RNAs) that can be directly functional or serve as the intermediate template for a protein that performs a func-



Figure 10.1: Gregor Mendel, the Moravian Augustinian monk who is credited for having founded the modern science of genetics¹⁹

tion.

10.1 Mendelian Inheritance

Mendelian inheritance¹⁴ is a type of biological inheritance that follows the principles originally proposed by Gregor Mendel¹⁵ in 1865 and 1866, re-discovered in 1900 and popularised by William Bateson¹⁶. These principles were initially controversial. When Mendel's theories were integrated with the Boveri-Sutton chromosome theory of inheritance by Thomas Hunt Morgan¹⁷ in 1915, they became the core of classical genetics. Ronald Fisher¹⁸ combined these ideas with the theory of natural selection in his 1930 book *The Genetical Theory of Natural Selection*, putting evolution onto a mathematical footing and forming the basis for population genetics within the modern evolutionary synthesis.

The principles of Mendelian inheritance were named for and first derived by Gregor Johann Mendel, a nineteenth-century Moravian monk who formulated his ideas after conducting simple hybridisation experiments with pea plants (*Pisum sativum*) he had planted in the garden of his monastery. Between 1856 and 1863, Mendel cultivated and tested some 5,000 pea plants. From these

¹<https://en.wikipedia.org/wiki/Genetics>

²https://en.wikipedia.org/wiki/Classical_genetics

³https://en.wikipedia.org/wiki/Gregor_Mendel

⁴https://en.wikipedia.org/wiki/Molecular_genetics

⁵<https://en.wikipedia.org/wiki/Reproduction>

⁶<https://en.wikipedia.org/wiki/Phenotype>

⁷<https://en.wikipedia.org/wiki/Genotype>

⁸https://en.wikipedia.org/wiki/Hugo_de_Vries

⁹https://en.wikipedia.org/wiki/Carl_Correns

¹⁰https://en.wikipedia.org/wiki/Erich_von_Tschermak

¹¹https://en.wikipedia.org/wiki/Wilhelm_Johannsen

¹²https://en.wikipedia.org/wiki/William_Bateson

¹³https://en.wikipedia.org/wiki/Gene#Discovery_of_discrete_inherited_units

¹⁴https://en.wikipedia.org/wiki/Mendelian_inheritance

¹⁵https://en.wikipedia.org/wiki/Gregor_Mendel

¹⁶https://en.wikipedia.org/wiki/William_Bateson

¹⁷https://en.wikipedia.org/wiki/Thomas_Hunt_Morgan

¹⁸https://en.wikipedia.org/wiki/Ronald_Fisher

experiments, he induced two generalizations which later became known as Mendel's Principles of Heredity or Mendelian inheritance. He described his experiments in a two-part paper, *Versuche über Pflanzen-Hybriden* (Experiments on Plant Hybridization), that he presented to the Natural History Society of Brno on 8 February and 8 March 1865, and which was published in 1866.

Mendel's results were largely ignored. Although they were not completely unknown to biologists of the time, they were not seen as generally applicable, even by Mendel himself, who thought they only applied to certain categories of species or traits. A major block to understanding their significance was the importance attached by 19th-century biologists to the apparent blending of many inherited traits in the overall appearance of the progeny, now known to be due to multi-gene interactions, in contrast to the organ-specific binary characters studied by Mendel. In 1900, however, his work was "re-discovered" by three European scientists, Hugo de Vries²⁰, Carl Correns²¹, and Erich von Tschermak²². The exact nature of the "re-discovery" has been debated: De Vries published first on the subject, mentioning Mendel in a footnote, while Correns pointed out Mendel's priority after having read De Vries' paper and realizing that he himself did not have priority. De Vries may not have acknowledged truthfully how much of his knowledge of the laws came from his own work and how much came only after reading Mendel's paper. Later scholars have accused Von Tschermak of not truly understanding the results at all.

Regardless, the "re-discovery" made Mendelism an important but controversial theory. Its most vigorous promoter in Europe was William Bateson, who coined the terms "genetics" and "allele" to describe many of its tenets. The model of heredity was contested by other biologists because it implied that heredity was discontinuous, in opposition to the apparently continuous variation observable for many traits. Many biologists also dismissed the theory because they were not sure it would apply to all species. However, later work by biologists and statisticians such as Ronald Fisher showed that if multiple Mendelian factors were involved in the expression of an individual trait, they could produce the diverse results observed, and thus showed that Mendelian genetics is compatible with natural selection. Thomas Hunt Morgan and his assistants later integrated Mendel's theoretical model with the chromosome theory of inheritance, in which the chromosomes of cells were thought to hold the actual hereditary material, and created what is now known as classical genetics, a highly successful foundation which eventually cemented Mendel's

place in history.

Mendel's findings allowed scientists such as Fisher and J.B.S. Haldane²³ to predict the expression of traits on the basis of mathematical probabilities. An important aspect of Mendel's success can be traced to his decision to start his crosses only with plants he demonstrated were true-breeding. He only measured discrete (binary) characteristics, such as color, shape, and position of the seeds, rather than quantitatively variable characteristics. He expressed his results numerically and subjected them to statistical analysis. His method of data analysis and his large sample size gave credibility to his data. He had the foresight to follow several successive generations (P, F1, F2, F3) of pea plants and record their variations. Finally, he performed "test crosses" (backcrossing descendants of the initial hybridization to the initial true-breeding lines) to reveal the presence and proportions of recessive characters.

10.1.1 Mendel's Genetic Discoveries

Five parts of Mendel's discoveries were an important divergence from the common theories at the time and were the prerequisite for the establishment of his rules.

- Characters are unitary. That is, they are discrete (purple vs. white, tall vs. dwarf).
- Genetic characteristics have alternate forms, each inherited from one of two parents. Today, we call these alleles.
- One allele is dominant over the other. The phenotype reflects the dominant allele.
- Gametes are created by random segregation. Heterozygotic individuals produce gametes with an equal frequency of the two alleles.
- Different traits have independent assortment. In modern terms, genes are unlinked.

According to customary terminology we refer here to the principles of inheritance discovered by Gregor Mendel as Mendelian laws, although today's geneticists also speak of Mendelian rules or Mendelian principles, as there are many exceptions summarized under the collective term Non-Mendelian inheritance.

Mendel selected for experiment relate the following characters of pea plants:

- Form of the ripe seeds round or roundish, surface shallow or wrinkled
- Colour of the seed-coat white, gray, brown with or without violet spotting
- Colour of the seeds and cotyledons yellow or green
- Flower colour

²⁰https://en.wikipedia.org/wiki/Hugo_de_Vries

²¹https://en.wikipedia.org/wiki/Carl_Correns

²²https://en.wikipedia.org/wiki/Erich_von_Tschermak

²³https://en.wikipedia.org/wiki/J._B._S._Haldane

Characteristics of pea plants Gregor Mendel used in his inheritance experiments						
Seeds	Flower colour	Pod colour	Pod position	Stem	Size	
round wrinkled	yellow white	green yellow	axial constricted between the seeds	axial terminal	long short	
smooth wrinkled	yellow white	green yellow	axial constricted between the seeds	axial terminal	long short	

Figure 10.2: Characteristics of pea plants Gregor Mendel used in his inheritance experiments²⁴

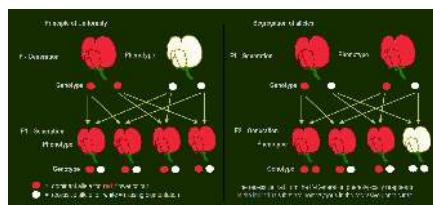


Figure 10.3: P-Generation and F1-Generation: The dominant allele for purple-red flower hides the phenotypic effect of the recessive allele for white flowers. F2-Generation: The recessive trait from the P-Generation phenotypically reappears in the individuals that are homozygous with the recessive genetic trait.²⁵

- Form of the ripe pods simply inflated, not contracted or constricted between the seeds and wrinkled
- Colour of the unripe pods yellow or green
- Position of the flowers axial or terminal
- Length of the stem

When he crossed purebred white flower and purple flower pea plants (the parental or P generation) by artificial pollination, the resulting flower colour was not a blend. Rather than being a mix of the two, the offspring in the first generation (F1-generation) were all purple flowered. Therefore he called this biological trait dominant. When he allowed self-fertilization in the uniform looking F1-generation, he obtained both colours in the F2 generation with a purple flower to white flower ratio of 3 : 1. In some of the other characters also one of the traits was dominant.

He then conceived the idea of heredity units, which he called hereditary "factors". Mendel found that there are alternative forms of factors — now called genes — that account for variations in inherited characteristics. For example, the gene for flower color in pea plants exists in two forms, one for purple and the other for white. The alternative "forms" are now called alleles. For each trait, an organism inherits two alleles, one from each parent. These alleles may be the same or different. An organism that has two identical alleles for a gene is said to be homozygous for that gene (and is called a homozygote). An organism that has two different alleles for a gene is said to be heterozy-

gous for that gene (and is called a heterozygote).

Mendel hypothesized that allele pairs separate randomly, or segregate, from each other during the production of the gametes in the seed plant (egg cell) and the pollen plant (sperm). Because allele pairs separate during gamete production, a sperm or egg carries only one allele for each inherited trait. When sperm and egg unite at fertilization, each contributes its allele, restoring the paired condition in the offspring. Mendel also found that each pair of alleles segregates independently of the other pairs of alleles during gamete formation.

The genotype of an individual is made up of the many alleles it possesses. The phenotype is the result of the expression of all characteristics that are genetically determined by its alleles as well as by its environment. The presence of an allele does not mean that the trait will be expressed in the individual that possesses it. If the two alleles of an inherited pair differ (the heterozygous condition), then one determines the organism's appearance and is called the dominant allele; the other has no noticeable effect on the organism's appearance and is called the recessive allele.

Table 10.1: Mendel's Laws of Inheritance.

Law	Definition
Law of dominance	Some alleles are dominant while others are recessive; an organism with at least one dominant allele will display the effect of the dominant allele.
Law of segregation	During gamete formation, the alleles for each gene segregate from each other so that each gamete carries only one allele for each gene.
Law of independent assortment	Genes for different traits can segregate independently during the formation of gametes.

10.1.2 Law of Dominance and Uniformity

If two parents are mated with each other who differ in one genetic characteristic for which they are both homozygous (each pure-bred), all offspring in the first generation (F1) are equal to the examined characteristic in genotype and phenotype showing the dominant trait. This uniformity rule or reciprocity rule applies to all individuals of the F1-generation.

The dominant inheritance discovered by Mendel states that in a heterozygote the recessive allele will be masked in

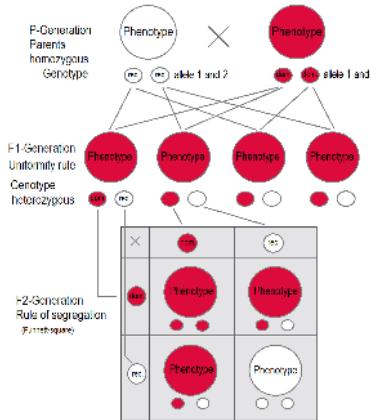


Figure 10.4: In the F1 generation all individuals have the same genotype and same phenotype expressing the dominant trait (red). In the F2 generation, the phenotypes show a 3:1 ratio. In the genotype 25% are homozygous with the dominant trait, 50% are heterozygous genetic carriers of the recessive trait, 25% are homozygous with the recessive genetic trait and expressing the recessive character.²⁶

the phenotype by the dominant allele. Only if the individual owns the recessive allele homozygous the recessive trait gets expressed. Therefore, a cross between a homozygous dominant and a homozygous recessive will always show the dominant trait in the phenotype, while still having a heterozygous genotype.

The F1 offspring of Mendel's pea crosses always looked like one of the two parental varieties. In this situation of "complete dominance," the dominant allele had the same phenotypic effect whether present in one or two copies.

But for some characteristics, the F1 hybrids have an appearance in between the phenotypes of the two parental varieties. A cross between two four o'clock (*Mirabilis jalapa*) plants shows an exception to Mendel's principle, called incomplete dominance. Flowers of heterozygous plants have less pigment than the homozygous, therefore there is a third phenotype. The phenotype lies somewhere between the two homozygous genotype. In cases of intermediate inheritance (incomplete dominance) in the F1-generation Mendel's principle of uniformity in genotype and phenotype applies as well. Research about intermediate inheritance was done by other scientists. The first was Carl Correns with his studies about *Mirabilis jalapa*.

10.1.3 Law of Segregation of genes

The Law of Segregation of genes applies when two individuals, both heterozygous for a certain trait are crossed, for example hybrids of the F1-generation. The offspring in the F2-generation differ in genotype and phenotype, so

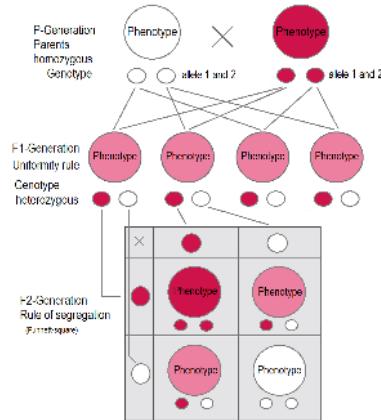


Figure 10.5: In *Mirabilis jalapa* and *Antirrhinum majus* are examples for intermediate inheritance. As seen in the F1-generation, heterozygous plants have "light pink" flowers—a mix of "red" and "white". The F2-generation shows a 1:2:1 ratio of red : light pink : white²⁷

that the characteristics of the grandparents (P-generation) regularly occur again. In a dominant-recessive inheritance an average of 25% are homozygous with the dominant trait, 50% are heterozygous showing the dominant trait in the phenotype (genetic carriers), 25% are homozygous with the recessive trait and therefore express the recessive trait in the phenotype. The genotypic ratio is 1 : 2 : 1, the phenotypic ratio is 3 : 1.

In the pea plant example, the capital "B" represents the dominant allele for purple blossom and lowercase "b" represents the recessive allele for white blossom. The pistil plant and the pollen plant are both F1-hybrids with genotype "B b". Each has one allele for purple and one allele for white. In the offspring, in the F2-plants in the Punnett-square, three combinations are possible. The genotypic ratio is 1 BB : 2 Bb : 1 bb. But the phenotypic ratio of plants with purple blossoms to those with white blossoms is 3 : 1 due to the dominance of the allele for purple. Plants with homozygous "b b" are white flowered like one of the grandparents in the P-generation.

In cases of incomplete dominance the same segregation of alleles takes place in the F2-generation, but here also the phenotypes show a ratio of 1 : 2 : 1, as the heterozygous are different in phenotype from the homozygous because the genetic expression of one allele compensates the missing expression of the other allele only partially. This results in an intermediate inheritance which was later described by other scientists.

In some literature sources the principle of segregation is cited as "first law". Nevertheless, Mendel did his crossing experiments with heterozygous plants after obtaining

these hybrids by crossing two purebred plants, discovering the principle of dominance and uniformity at first.

Molecular proof of segregation of genes was subsequently found through observation of meiosis by two scientists independently, the German botanist Oscar Hertwig in 1876, and the Belgian zoologist Edouard Van Beneden in 1883. Most alleles are located in chromosomes in the cell nucleus. Paternal and maternal chromosomes get separated in meiosis, because during spermatogenesis the chromosomes are segregated on the four sperm cells that arise from one mother sperm cell, and during oogenesis the chromosomes are distributed between the polar bodies and the egg cell. Every individual organism contains two alleles for each trait. They segregate (separate) during meiosis such that each gamete contains only one of the alleles. When the gametes unite in the zygote the alleles – one from the mother one from the father – get passed on to the offspring. An offspring thus receives a pair of alleles for a trait by inheriting homologous chromosomes from the parent organisms: one allele for each trait from each parent. Heterozygous individuals with the dominant trait in the phenotype are genetic carriers of the recessive trait.

10.1.4 Law of Independent Assortment

The Law of Independent Assortment states that alleles for separate traits are passed independently of one another. That is, the biological selection of an allele for one trait has nothing to do with the selection of an allele for any other trait. Mendel found support for this law in his dihybrid cross experiments. In his monohybrid crosses, an idealized 3:1 ratio between dominant and recessive phenotypes resulted. In dihybrid crosses, however, he found a 9:3:3:1 ratios. This shows that each of the two alleles is inherited independently from the other, with a 3:1 phenotypic ratio for each.

Independent assortment occurs in eukaryotic organisms during meiotic metaphase I, and produces a gamete with a mixture of the organism's chromosomes. The physical basis of the independent assortment of chromosomes is the random orientation of each bivalent chromosome along the metaphase plate with respect to the other bivalent chromosomes. Along with crossing over, independent assortment increases genetic diversity by producing novel genetic combinations.

There are many deviations from the principle of independent assortment due to genetic linkage.

Of the 46 chromosomes in a normal diploid human cell, half are maternally derived (from the mother's egg) and half are paternally derived (from the father's sperm). This occurs as sexual reproduction involves the fusion of two

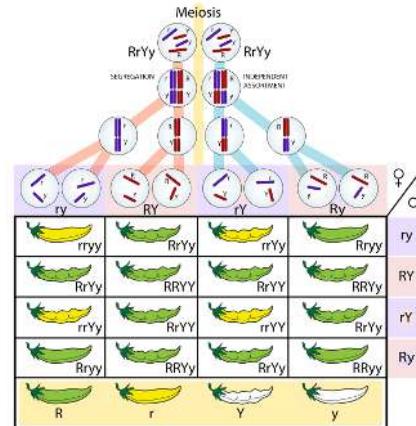


Figure 10.6: Segregation and independent assortment are consistent with the chromosome theory of inheritance.²⁸

haploid gametes (the egg and sperm) to produce a zygote and a new organism, in which every cell has two sets of chromosomes (diploid). During gametogenesis the normal complement of 46 chromosomes needs to be halved to 23 to ensure that the resulting haploid gamete can join with another haploid gamete to produce a diploid organism.

In independent assortment, the chromosomes that result are randomly sorted from all possible maternal and paternal chromosomes. Because zygotes end up with a mix instead of a pre-defined "set" from either parent, chromosomes are therefore considered assorted independently. As such, the zygote can end up with any combination of paternal or maternal chromosomes. For human gametes, with 23 chromosomes, the number of possibilities is 2^{23} or 8,388,608 possible combinations. This contributes to the genetic variability of progeny. Generally the recombination of genes has important implications for many evolutionary processes.

10.1.5 Mendelian Trait

A Mendelian trait is one that is controlled by a single locus in an inheritance pattern. In such cases, a mutation in a single gene can cause a disease that is inherited according to Mendel's principles. Dominant diseases manifest in heterozygous individuals. Recessive ones are sometimes inherited unnoticed by genetic carriers. Examples include sickle-cell anemia, Tay-Sachs disease, cystic fibrosis and xeroderma pigmentosa. A disease controlled by a single gene contrasts with a multi-factorial disease, like heart disease, which is affected by several loci (and the environment) as well as those diseases inherited in a non-Mendelian fashion.

10.1.6 Zygosity

Zygosity refers to the grade of similarity between the alleles that determine one specific trait in an organism. In its simplest form, a pair of alleles can be either homozygous or heterozygous. Homozygosity, with homo relating to same while zygous pertains to a zygote, is seen when a combination of either two dominant or two recessive alleles code for the same trait. For example, using 'A' as the representative character for each allele, a homozygous dominant pair's genotype would be depicted as 'AA', while homozygous recessive is shown as 'aa'. Heterozygosity, with hetero associated with different, can only be 'Aa' (the capital letter is always presented first by convention). The phenotype of a homozygous dominant pair is 'A', or dominant, while the opposite is true for homozygous recessive. Heterozygous pairs always have a dominant phenotype. To a lesser degree, hemizygosity and nullizygosity can also be seen in gene pairs.

10.1.7 Monohybrid Cross

"Mono" means "one"; this cross indicates that the examination of a single trait. For example, when a strain of corn producing pure purple kernels (RR) is crossed with a strain producing pure yellow kernels (rr). In this cross, both parents are homozygous, one carrying two copies of the dominant allele (R; purple), the other two copies of the recessive (r; yellow) allele.

10.1.8 Punnett Square

The Punnett square²⁹ (Figures 10.7 and 10.8) is a visual representation of Mendelian inheritance and used to predict an outcome of a particular cross or breeding experiment. It is named after Reginald C. Punnett³⁰, who devised the approach. The diagram is used by biologists to determine the probability of an offspring having a particular genotype. The Punnett square is a tabular summary of possible combinations of maternal alleles with paternal alleles. These tables can be used to examine the genotypical outcome probabilities of the offspring of a single trait (allele), or when crossing multiple traits from the parents.

Figure 10.7 shows a Punnett square for a monohybrid cross. A strain of corn producing pure purple kernels (RR) is crossed with a strain producing pure yellow kernels (rr). In this cross, both parents are homozygous, one carrying two copies of the dominant allele (R; purple), the other two copies of the recessive (r; yellow) allele. Each parent can only make gametes that have either the R or r allele. The squares containing the single letters represent the possible gametes. The squares with two letters represent the

	Yellow	
Purple	r	r
R	Rr	Rr
R	Rr	Rr

Figure 10.7: Punnett square for homozygous cross.

	Maternal gametes	
Paternal gametes	R	r
R	RR	Rr
r	Rr	rr

Figure 10.8: Punnett square for heterozygous cross.

zygotes resulting from the combination of the respective gametes. It can be easily seen that all offspring will be heterozygous (Rr) and therefore purple. Purple is dominant with the resulting F1 ears all bearing purple kernels. These plants that are heterozygous for a single trait are called monohybrids. When the F1 is self-pollinated, the resulting F2 ears bear both purple and yellow kernels (Figure 10.9). The Punnett square for the F1 cross is depicted in Figure 10.8



Figure 10.9: Monohybrid cross

²⁹ https://en.wikipedia.org/wiki/Punnett_square

³⁰ https://en.wikipedia.org/wiki/Reginald_Punnett



Figure 10.10: Dihybrid cross

10.1.9 Dihybrid Cross

More complicated crosses can be made by looking at two or more genes. The Punnett square works, however, only if the genes are independent of each other, which means that having a particular allele of gene "A" does not alter the probability of possessing an allele of gene "B". This is equivalent to stating that the genes are not linked, so that the two genes do not tend to sort together during meiosis.

A dihybrid cross is a cross between two different lines (varieties, strains) that differ in two observed traits. In the name "Dihybrid cross", the "di" indicates that there are two traits involved (in our example designated R and Su), the "hybrid" means that each trait has two different alleles (in our example R and r, or Su and su), and "cross" means that there are two individuals who are combining or "crossing" their genetic information. In our example, a pure strain of corn producing purple-starchy kernels (RR SuSu) is crossed with a pure strain producing yellow-sweet (rr susu). The starchy seeds are smooth, the sweet seeds are wrinkled. The resulting F1 ears all bear purple-starchy (smooth) kernels. Plants that are heterozygous for two traits are called dihybrids. When the F1 is self-pollinated, the resulting F2 generation contains various combinations (Figure 10.10).

The rules of meiosis, as they apply to the dihybrid, are codified in Mendel's first law and Mendel's second law, which are also called the Law of Segregation and the Law of Independent Assortment, respectively (Table ??). For genes on separate chromosomes, each allele pair showed independent segregation. If the first filial generation (F1 generation) produces four identical offspring, the second filial generation, which occurs by crossing the members of the first filial generation, shows a phenotypic (appearance) ratio of 9:3:3:1, where:

- the 9 represents the proportion of individuals displaying both dominant traits
- the first 3 represents the individuals displaying the first dominant trait and the second recessive trait
- the second 3 represents those displaying the first recessive trait and second dominant trait
- the 1 represents the homozygous, displaying both recessive traits.

10.2 Mendelian Traits in Humans

Mendelian traits in humans concerns how, in Mendelian inheritance, a child receiving a dominant allele from either parent will have the dominant form of the phenotypic trait or characteristic. Only those that received the recessive allele from both parents, known as zygosity, will have the recessive phenotype. Those that receive a dominant allele from one parent and a recessive allele from the other parent will have the dominant form of the trait. Purely Mendelian traits are a tiny minority of all traits, since most phenotypic traits exhibit incomplete dominance, codominance, and contributions from many genes.

The recessive phenotype may theoretically skip any number of generations, lying dormant in heterozygous "carrier" individuals until they have children with someone who also has the recessive allele and both pass it on to their child.

Genes that do not follow Mendelian genetics include the human Y chromosome which is passed virtually unchanged from father to son. Similarly, the mitochondrial DNA (mtDNA) comes only from the mother and is given to both male and female children. Epigenetic modifications, linked genes, and duplicated genes elsewhere in the genome will also lead to a non-mendelian inheritance of traits.

Examples

- Albinism (recessive)
- Achondroplasia
- Alkaptonuria
- Ataxia telangiectasia
- Brachydactyly (shortness of fingers and toes)
- Colour blindness (monochromatism, dichromatism, anomalous trichromatism, tritanopia, deutanopia, protanopia)
- Cystic fibrosis
- Duchenne muscular dystrophy
- Ectrodactyly
- Ehlers-Danlos syndrome
- Fabry disease
- Galactosemia
- Gaucher's disease
- Haemophilia
- Hereditary breast-ovarian cancer syndrome
- Hereditary nonpolyposis colorectal cancer
- HFE hereditary haemochromatosis
- Huntington's disease
- Hypercholesterolemia
- Krabbe disease
- Lactase persistence (dominant)
- Leber's hereditary optic neuropathy
- Lesch-Nyhan syndrome

- Marfan syndrome
- Niemann-Pick disease
- Phenylketonuria
- Porphyria
- Retinoblastoma
- Sickle-cell disease
- Sanfilippo syndrome
- Tay-Sachs disease
- Wet (dominant) or dry (recessive) earwax – dry is found mostly in Asians and Native Americans

10.3 Non-Mendelian Inheritance

After Mendel's studies and discoveries more and more new discoveries about genetics were made. Mendel himself has said that the regularities he discovered apply only to the organisms and characteristics he consciously chose for his experiments. Mendel explained inheritance in terms of discrete factors—genes—that are passed along from generation to generation according to the rules of probability. Mendel's laws are valid for all sexually reproducing organisms, including garden peas and human beings. However, Mendel's laws stop short of explaining some patterns of genetic inheritance. For most sexually reproducing organisms, cases where Mendel's laws can strictly account for all patterns of inheritance are relatively rare. Often the inheritance patterns are more complex.

In cases of codominance the phenotypes produced by both alleles are clearly expressed. Mendel chose genetic traits in plants that are determined by only two alleles, such as "A" and "a". In nature, genes often exist in several different forms with multiple alleles. Furthermore, many traits are produced by the interaction of several genes. Traits controlled by two or more genes are said to be polygenic traits.

Chapter 11

Molecular Biology of The Gene

Molecular biology¹ is the branch of biology that concerns the molecular basis of biological activity in and between cells, including molecular synthesis, modification, mechanisms and interactions. Molecular biology arose as an attempt to answer the questions regarding the mechanisms of genetic inheritance and the structure of a gene. In 1953, James Watson² and Francis Crick³ published the double helical structure of DNA courtesy of the X-ray crystallography work done by Rosalind Franklin⁴ and Maurice Wilkins⁵. Watson and Crick described the structure of DNA and the interactions within the molecule. This publication jump-started research into molecular biology and increased interest in the subject.

Nucleic acids⁶ are biopolymers that are essential to all known forms of life. The term nucleic acid is the overall name for DNA and RNA. They are composed of nucleotides, which are the monomers made of three components: a 5-carbon sugar, a phosphate group and a nitrogenous base. If the sugar is a compound ribose, the polymer is RNA (ribonucleic acid); if the sugar is derived from ribose as deoxyribose, the polymer is DNA (deoxyribonucleic acid).

In an influential published in 1941 paper, George Beadle⁷ and Edward Tatum⁸ proposed the idea that genes act through the production of enzymes, with each gene responsible for producing a single enzyme that in turn affects a single step in a metabolic pathway. The concept arose from work on genetic mutations in the mold *Neurospora crassa*, and subsequently was dubbed the "one gene-one enzyme hypothesis" by their collaborator Norman Horowitz⁹. In 2004 Norman Horowitz reminisced that "these experiments founded the science of what Beadle and Tatum called 'biochemical genetics.' These experiments

are by some considered to constitute the beginning of what became molecular genetics and the development of the one gene-one enzyme hypothesis is often considered the first significant result in what came to be called molecular biology. Although it has been extremely influential, the hypothesis was recognized soon after its proposal to be an oversimplification. Even the subsequent reformulation of the "one gene-one polypeptide" hypothesis is now considered too simple to describe the relationship between genes and proteins. In attributing an instructional role to genes, Beadle and Tatum implicitly accorded genes an informational capability. This insight provided the foundation for the concept of a genetic code. However, it was not until the experiments were performed showing that DNA was the genetic material, that proteins consist of a defined linear sequence of amino acids, and that DNA structure contained a linear sequence of base pairs, was there a clear basis for solving the genetic code.

Although genes were known to exist on chromosomes, chromosomes are composed of both protein and DNA, and scientists did not know which of the two is responsible for inheritance. In 1928, Frederick Griffith¹⁰ discovered the phenomenon of transformation: dead bacteria could transfer genetic material to "transform" other still-living bacteria. Sixteen years later, in 1944, the Avery-MacLeod-McCarty experiment identified DNA as the molecule responsible for transformation. The role of the nucleus as the repository of genetic information in eukaryotes had been established by Hämmerling in 1943 in his work on the single celled alga Acetabularia. The Hershey-Chase experiment in 1952 confirmed that DNA (rather than protein) is the genetic material of the viruses that infect bacteria, providing further evidence that DNA is the molecule responsible for inheritance.

James Watson¹¹ and Francis Crick¹² determined the structure of DNA in 1953, using the X-ray crystallography

¹https://en.wikipedia.org/wiki/Molecular_biology

²https://en.wikipedia.org/wiki/James_Watson

³https://en.wikipedia.org/wiki/Francis_Crick

⁴https://en.wikipedia.org/wiki/Rosalind_Franklin

⁵https://en.wikipedia.org/wiki/Maurice_Wilkins

⁶https://en.wikipedia.org/wiki/Nucleic_acid

⁷https://en.wikipedia.org/wiki/George_Beadle

⁸https://en.wikipedia.org/wiki/Edward_Tatum

⁹https://en.wikipedia.org/wiki/Norman_Horowitz

¹⁰https://en.wikipedia.org/wiki/Frederick_Griffith

¹¹https://en.wikipedia.org/wiki/James_Watson

¹²https://en.wikipedia.org/wiki/Francis_Crick

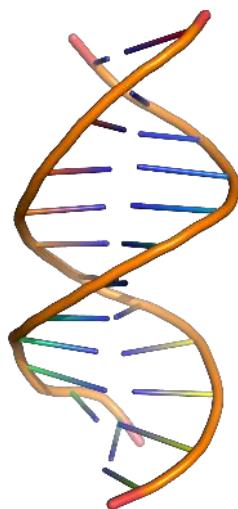


Figure 11.1: A cartoon representation of DNA based on atomic coordinates of PDB 1BNA¹⁵, rendered with open source molecular visualization tool PyMol.

work of Rosalind Franklin¹³ and Maurice Wilkins¹⁴ that indicated DNA has a helical structure (i.e., shaped like a corkscrew). Their double-helix model had two strands of DNA with the nucleotides pointing inward, each matching a complementary nucleotide on the other strand to form what look like rungs on a twisted ladder. This structure showed that genetic information exists in the sequence of nucleotides on each strand of DNA. The structure also suggested a simple method for replication: if the strands are separated, new partner strands can be reconstructed for each based on the sequence of the old strand. This property is what gives DNA its semi-conservative nature where one strand of new DNA is from an original parent strand.

Although the structure of DNA showed how inheritance works, it was still not known how DNA influences the behavior of cells. In the following years, scientists tried to understand how DNA controls the process of protein production. It was discovered that the cell uses DNA as a template to create matching messenger RNA, molecules with nucleotides very similar to DNA. The nucleotide sequence of a messenger RNA is used as a template by ribosomes to create an amino acid sequence in protein; this correspondence between nucleotide sequences and amino acid sequences is known as the genetic code.

With the newfound molecular understanding of inheritance came an explosion of research. One important development was chain-termination DNA sequencing in 1977 by Frederick Sanger¹⁶. This technology allows scientists to

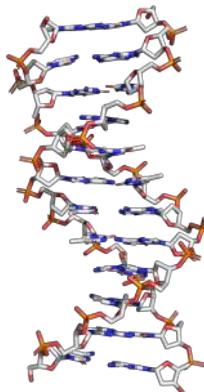


Figure 11.2: The structure of the DNA double helix. A section of DNA. The bases lie horizontally between the two spiraling strands. The atoms in the structure are colour-coded by element (based on atomic coordinates of PDB 1bna¹⁹ rendered with open source molecular visualization tool PyMol.)

read the nucleotide sequence of a DNA molecule. In 1983, Kary Banks Mullis¹⁷ developed the polymerase chain reaction, providing a quick way to isolate and amplify a specific section of DNA from a mixture. The efforts of the Human Genome Project, Department of Energy, NIH, and parallel private efforts by Celera Genomics led to the sequencing of the human genome in 2003.

11.1 Structure And Function of Deoxyribonucleic Acid

Deoxyribonucleic acid¹⁸ (DNA) is a molecule composed of two chains that coil around each other to form a double helix carrying genetic instructions for the development, functioning, growth and reproduction of all known organisms and many viruses. DNA and ribonucleic acid (RNA) are nucleic acids; alongside proteins, lipids and complex carbohydrates (polysaccharides), nucleic acids are one of the four major types of macromolecules that are essential for all known forms of life.

The two DNA strands are also known as polynucleotides as they are composed of simpler monomeric units called nucleotides. Each nucleotide is composed of one of four nitrogen-containing nucleobases (cytosine [C], guanine [G], adenine [A] or thymine [T]), a sugar called deoxyribose, and a phosphate group. The nucleotides are joined to one another in a chain by covalent bonds between the sugar of one nucleotide and the phosphate

¹³https://en.wikipedia.org/wiki/Rosalind_Franklin

¹⁴https://en.wikipedia.org/wiki/Maurice_Wilkins

¹⁶https://en.wikipedia.org/wiki/Frederick_Sanger

¹⁷https://en.wikipedia.org/wiki/Kary_Mullis

¹⁸<https://en.wikipedia.org/wiki/DNA>

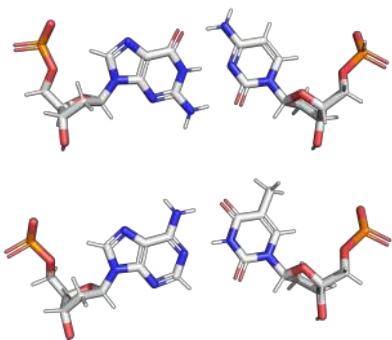


Figure 11.3: The structure of the four nucleotides and their base pairing in the DNA double helix. The atoms in the structure are colour-coded by element (based on atomic coordinates of PDB 1bna²⁰ rendered with open source molecular visualization tool PyMol.)

of the next, resulting in an alternating sugar-phosphate backbone. The nitrogenous bases of the two separate polynucleotide strands are bound together, according to base pairing rules (A with T and C with G), with hydrogen bonds to make double-stranded DNA. The complementary nitrogenous bases are divided into two groups, pyrimidines and purines. In DNA, the pyrimidines are thymine and cytosine; the purines are adenine and guanine.

Both strands of DNA store biological information. This information is replicated as and when the two strands separate. A large part of DNA (more than 98% for humans) is non-coding, meaning that these sections do not serve as patterns for protein sequences. The two strands of DNA run in opposite directions to each other and are thus antiparallel. Attached to each sugar is one of four types of nucleobases (informally, bases). It is the sequence of these four nucleobases along the backbone that encodes genetic information. RNA strands are created using DNA strands as a template in a process called transcription, where DNA bases are exchanged for their corresponding bases except in the case of thymine (T), which RNA substitutes for uracil (U). Under the genetic code, these RNA strands specify the sequence of amino acids within proteins in a process called translation.

Within eukaryotic cells, DNA is organized into long structures called chromosomes. Before typical cell division, these chromosomes are duplicated in the process of DNA replication, providing a complete set of chromosomes for each daughter cell. Eukaryotic organisms (animals, plants, fungi and protists) store most of their DNA inside the cell nucleus as nuclear DNA, and some in the mitochondria as mitochondrial DNA or in chloroplasts as chloroplast DNA. In contrast, prokaryotes (bacteria and archaea) store their DNA only in the cytoplasm, in circular

chromosomes. Within eukaryotic chromosomes, chromatin proteins, such as histones, compact and organize DNA. These compacting structures guide the interactions between DNA and other proteins, helping control which parts of the DNA are transcribed.

DNA was first isolated by Friedrich Miescher²¹ in 1869. Its molecular structure was first identified by Francis Crick²² and James Watson²³ at the Cavendish Laboratory within the University of Cambridge in 1953, whose model-building efforts were guided by X-ray diffraction data acquired by Raymond Gosling²⁴, who was a post-graduate student of Rosalind Franklin²⁵.

DNA is a long polymer made from repeating units called nucleotides, each of which is usually symbolized by a single letter: either A, T, C, or G. The structure of DNA is dynamic along its length, being capable of coiling into tight loops and other shapes. In all species it is composed of two helical chains, bound to each other by hydrogen bonds. Both chains are coiled around the same axis, and have the same pitch of 34 angstroms (\AA) (3.4 nanometres). The pair of chains has a radius of 10 angstroms (1.0 nanometre). Although each individual nucleotide is very small, a DNA polymer can be very large and contain hundreds of millions, such as in chromosome 1. Chromosome 1 is the largest human chromosome with approximately 220 million base pairs, and would be 85 mm long if straightened.

DNA does not usually exist as a single strand, but instead as a pair of strands that are held tightly together. These two long strands coil around each other, in the shape of a double helix. The nucleotide contains both a segment of the backbone of the molecule (which holds the chain together) and a nucleobase (which interacts with the other DNA strand in the helix). A nucleobase linked to a sugar is called a nucleoside, and a base linked to a sugar and to one or more phosphate groups is called a nucleotide. A biopolymer comprising multiple linked nucleotides (as in DNA) is called a polynucleotide.

The backbone of the DNA strand is made from alternating phosphate and sugar residues. The sugar in DNA is 2-deoxyribose, which is a pentose (five-carbon) sugar. The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar rings. These are known as the 3'-end (three prime end), and 5'-end (five prime end) carbons, the prime symbol being used to distinguish these carbon atoms from those of the base to which the de-

²¹https://en.wikipedia.org/wiki/Friedrich_Miescher

²²https://en.wikipedia.org/wiki/Francis_Crick

²³https://en.wikipedia.org/wiki/James_Watson

²⁴https://en.wikipedia.org/wiki/Raymond_Gosling

²⁵https://en.wikipedia.org/wiki/Rosalind_Franklin

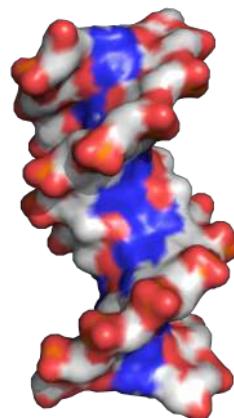


Figure 11.4: DNA major and minor grooves. PDB 1bna²⁶ rendered with open source molecular visualization tool PyMol.)

oxyribose forms a glycosidic bond. When imagining DNA, each phosphoryl is normally considered to “belong” to the nucleotide whose 5' carbon forms a bond therewith. Any DNA strand therefore normally has one end at which there is a phosphoryl attached to the 5' carbon of a ribose (the 5' phosphoryl) and another end at which there is a free hydroxyl attached to the 3' carbon of a ribose (the 3' hydroxyl). The orientation of the 3' and 5' carbons along the sugar-phosphate backbone confers directionality (sometimes called polarity) to each DNA strand. In a nucleic acid double helix, the direction of the nucleotides in one strand is opposite to their direction in the other strand: the strands are antiparallel. The asymmetric ends of DNA strands are said to have a directionality of five prime end (5'), and three prime end (3'), with the 5' end having a terminal phosphate group and the 3' end a terminal hydroxyl group. One major difference between DNA and RNA is the sugar, with the 2-deoxyribose in DNA being replaced by the alternative pentose sugar ribose in RNA.

Twin helical strands form the DNA backbone. Another double helix may be found tracing the spaces, or grooves, between the strands (Figure 11.4). These voids are adjacent to the base pairs and may provide a binding site. As the strands are not symmetrically located with respect to each other, the grooves are unequally sized. One groove, the major groove, is 22 angstroms (\AA) wide and the other, the minor groove, is 12 \AA wide. The width of the major groove means that the edges of the bases are more accessible in the major groove than in the minor groove. As a result, proteins such as transcription factors that can bind to specific sequences in double-stranded DNA usually make contact with the sides of the bases exposed in the major groove.

In a DNA double helix, each type of nucleobase on one

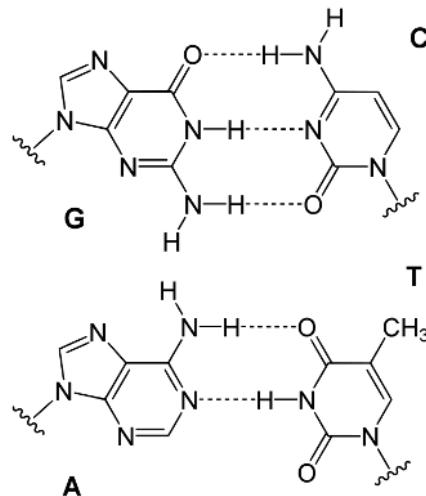


Figure 11.5: Top, a GC base pair²⁷ with three hydrogen bonds. Bottom, an AT base pair²⁸ with two hydrogen bonds. Non-covalent hydrogen bonds between the pairs are shown as dashed lines. The two types of base pairs form different numbers of hydrogen bonds, AT forming two hydrogen bonds, and GC forming three hydrogen bonds (see figures, right). DNA with high GC-content is more stable than DNA with low GC-content.

strand bonds with just one type of nucleobase on the other strand. This is called complementary base pairing. Here, purines form hydrogen bonds to pyrimidines, with adenine bonding only to thymine in two hydrogen bonds, and cytosine bonding only to guanine in three hydrogen bonds (Figure 11.5). This arrangement of two nucleotides binding together across the double helix is called a Watson-Crick base pair. As hydrogen bonds are not covalent, they can be broken and rejoined relatively easily. The two strands of DNA in a double helix can thus be pulled apart like a zipper, either by a mechanical force or high temperature. As a result of this base pair complementarity, all the information in the double-stranded sequence of a DNA helix is duplicated on each strand, which is vital in DNA replication. This reversible and specific interaction between complementary base pairs is critical for all the functions of DNA in organisms.

As noted above, most DNA molecules are actually two polymer strands, bound together in a helical fashion by noncovalent bonds; this double-stranded (dsDNA) structure is maintained largely by the intrastrand base stacking interactions, which are strongest for G,C stacks. The two strands can come apart—a process known as melting—to form two single-stranded DNA (ssDNA) molecules. Melting occurs at high temperature, low salt and high pH (low pH also melts DNA, but since DNA is unstable due to acid depurination, low pH is rarely used).

The stability of the dsDNA form depends not only on the GC-content (% G,C basepairs) but also on sequence (since stacking is sequence specific) and also length (longer molecules are more stable). The stability can be measured in various ways; a common way is the “melting temperature”, which is the temperature at which 50% of the ds molecules are converted to ss molecules; melting temperature is dependent on ionic strength and the concentration of DNA. As a result, it is both the percentage of GC base pairs and the overall length of a DNA double helix that determines the strength of the association between the two strands of DNA. Long DNA helices with a high GC-content have stronger-interacting strands, while short helices with high AT content have weaker-interacting strands. In biology, parts of the DNA double helix that need to separate easily, such as the TATAAT Pribnow box²⁹ in some promoters, tend to have a high AT content, making the strands easier to pull apart.

In the laboratory, the strength of this interaction can be measured by finding the temperature necessary to break the hydrogen bonds, their melting temperature (also called T_m value). When all the base pairs in a DNA double helix melt, the strands separate and exist in solution as two entirely independent molecules. These single-stranded DNA molecules have no single common shape, but some conformations are more stable than others.

A DNA sequence is called a “sense” sequence if it is the same as that of a messenger RNA copy that is translated into protein. The sequence on the opposite strand is called the “antisense” sequence. Both sense and antisense sequences can exist on different parts of the same strand of DNA (i.e. both strands can contain both sense and antisense sequences). In both prokaryotes and eukaryotes, antisense RNA sequences are produced, but the functions of these RNAs are not entirely clear. One proposal is that antisense RNAs are involved in regulating gene expression through RNA-RNA base pairing.

A few DNA sequences in prokaryotes and eukaryotes, and more in plasmids and viruses, blur the distinction between sense and antisense strands by having overlapping genes. In these cases, some DNA sequences do double duty, encoding one protein when read along one strand, and a second protein when read in the opposite direction along the other strand. In bacteria, this overlap may be involved in the regulation of gene transcription, while in viruses, overlapping genes increase the amount of information that can be encoded within the small viral genome.

DNA can be twisted like a rope in a process called DNA supercoiling. With DNA in its “relaxed” state, a strand usually circles the axis of the double helix once every 10.4

base pairs, but if the DNA is twisted the strands become more tightly or more loosely wound. If the DNA is twisted in the direction of the helix, this is positive supercoiling, and the bases are held more tightly together. If they are twisted in the opposite direction, this is negative supercoiling, and the bases come apart more easily. In nature, most DNA has slight negative supercoiling that is introduced by enzymes called topoisomerases. These enzymes are also needed to relieve the twisting stresses introduced into DNA strands during processes such as transcription and DNA replication.

The expression of genes is influenced by how the DNA is packaged in chromosomes, in a structure called chromatin. Base modifications can be involved in packaging, with regions that have low or no gene expression usually containing high levels of methylation of cytosine bases. DNA packaging and its influence on gene expression can also occur by covalent modifications of the histone protein core around which DNA is wrapped in the chromatin structure or else by remodeling carried out by chromatin remodeling complexes. There is, further, crosstalk between DNA methylation and histone modification, so they can coordinately affect chromatin and gene expression.

For one example, cytosine methylation produces 5-methylcytosine, which is important for X-inactivation of chromosomes. The average level of methylation varies between organisms—the worm *Caenorhabditis elegans* lacks cytosine methylation, while vertebrates have higher levels, with up to 1% of their DNA containing 5-methylcytosine. Despite the importance of 5-methylcytosine, it can deaminate to leave a thymine base, so methylated cytosines are particularly prone to mutations. Other base modifications include adenine methylation in bacteria, the presence of 5-hydroxymethylcytosine in the brain, and the glycosylation of uracil to produce the “J-base” in kinetoplastids.

11.1.1 DNA Damage

DNA can be damaged by many sorts of mutagens³⁰, which change the DNA sequence. Mutagens include oxidizing agents, alkylating agents and also high-energy electromagnetic radiation such as ultraviolet light and X-rays. The type of DNA damage produced depends on the type of mutagen. For example, UV light can damage DNA by producing thymine dimers, which are cross-links between pyrimidine bases. On the other hand, oxidants such as free radicals or hydrogen peroxide produce multiple forms of damage, including base modifications, particularly of guanosine, and double-strand breaks. A typical human cell contains about 150,000 bases that have suffered oxidative damage. Of these oxidative lesions, the most dangerous are double-strand breaks, as these are difficult

²⁹https://en.wikipedia.org/wiki/Pribnow_box

³⁰<https://en.wikipedia.org/wiki/Mutagen>

to repair and can produce point mutations, insertions, deletions from the DNA sequence, and chromosomal translocations. These mutations can cause cancer. DNA damage that is naturally occurring, due to normal cellular processes that produce reactive oxygen species, the hydrolytic activities of cellular water, etc., also occurs frequently. Although most of this damage is repaired, in any cell some DNA damage may remain despite the action of repair processes. This DNA damage accumulates with age in mammalian postmitotic tissues. This accumulation appears to be an important underlying cause of aging.

Many mutagens fit into the space between two adjacent base pairs, this is called intercalation. Most intercalators are aromatic and planar molecules; examples include ethidium bromide, acridines, daunomycin, and doxorubicin. For an intercalator to fit between base pairs, the bases must separate, distorting the DNA strands by unwinding of the double helix. This inhibits both transcription and DNA replication, causing toxicity and mutations. As a result, DNA intercalators may be carcinogens, and in the case of thalidomide, a teratogen. Others such as benzo[a]pyrene diol epoxide and aflatoxin form DNA adducts that induce errors in replication. Nevertheless, due to their ability to inhibit DNA transcription and replication, other similar toxins are also used in chemotherapy to inhibit rapidly growing cancer cells.

DNA usually occurs as linear chromosomes in eukaryotes, and circular chromosomes in prokaryotes. The set of chromosomes in a cell makes up its genome; the human genome has approximately 3 billion base pairs of DNA arranged into 46 chromosomes. Transmission of genetic information in genes is achieved via complementary base pairing. For example, in transcription, the DNA sequence is copied into a complementary RNA sequence. Usually, this RNA copy is then used to make a matching protein sequence in a process called translation. In alternative fashion, a cell may simply copy its genetic information in a process called DNA replication.

11.1.2 Genes And Genomes

Genomic DNA is tightly and orderly packed in the process called DNA condensation, to fit the small available volumes of the cell. In eukaryotes, DNA is located in the cell nucleus, with small amounts in mitochondria and chloroplasts. In prokaryotes, the DNA is held within an irregularly shaped body in the cytoplasm called the nucleoid.

In many species, only a small fraction of the total sequence of the genome encodes protein. For example, only about 1.5% of the human genome consists of protein-coding exons, with over 50% of human DNA consisting of non-coding repetitive sequences. The reasons for the presence of so much noncoding DNA in

eukaryotic genomes and the extraordinary differences in genome size, or C-value, among species, represent a long-standing puzzle known as the “C-value enigma”. However, some DNA sequences that do not code protein may still encode functional non-coding RNA molecules, which are involved in the regulation of gene expression.

Some noncoding DNA sequences play structural roles in chromosomes. Telomeres and centromeres typically contain few genes but are important for the function and stability of chromosomes. An abundant form of noncoding DNA in humans are pseudogenes, which are copies of genes that have been disabled by mutation. These sequences are usually just molecular fossils, although they can occasionally serve as raw genetic material for the creation of new genes through the process of gene duplication and divergence.

11.1.3 Transcription And Translation

A gene is a sequence of DNA that contains genetic information and can influence the phenotype of an organism. Within a gene, the sequence of bases along a DNA strand defines a messenger RNA sequence, which then defines one or more protein sequences. The relationship between the nucleotide sequences of genes and the amino-acid sequences of proteins is determined by the rules of translation, known collectively as the genetic code. The genetic code consists of three-letter ‘words’ called codons formed from a sequence of three nucleotides (e.g. ACT, CAG, TTT).

In transcription, the codons of a gene are copied into messenger RNA by RNA polymerase. This RNA copy is then decoded by a ribosome that reads the RNA sequence by base-pairing the messenger RNA to transfer RNA, which carries amino acids. Since there are 4 bases in 3-letter combinations, there are 64 possible codons. These encode the twenty standard amino acids, giving most amino acids more than one possible codon. There are also three ‘stop’ or ‘nonsense’ codons signifying the end of the coding region; these are the TAA, TGA, and TAG codons.

11.1.4 DNA Replication

Cell division is essential for an organism to grow, but, when a cell divides, it must replicate the DNA in its genome so that the two daughter cells have the same genetic information as their parent. The double-stranded structure of DNA provides a simple mechanism for DNA replication. Here, the two strands are separated and then each strand’s complementary DNA sequence is recreated by an enzyme called DNA polymerase. This enzyme makes the complementary strand by finding the correct base through complementary base pairing and bonding it onto the original strand. As DNA polymerases can only extend a DNA strand in a 5’ to

3' direction, different mechanisms are used to copy the antiparallel strands of the double helix. In this way, the base on the old strand dictates which base appears on the new strand, and the cell ends up with a perfect copy of its DNA.

In molecular biology, DNA replication is the biological process of producing two identical replicas of DNA from one original DNA molecule. DNA replication occurs in all living organisms acting as the basis for biological inheritance.

DNA is made up of a double helix of two complementary strands. During replication, these strands are separated. Each strand of the original DNA molecule then serves as a template for the production of its counterpart, a process referred to as semi-conservative replication. As a result of semi-conservative replication, the new helix will be composed of an original DNA strand as well as a newly synthesized strand. Cellular proofreading and error-checking mechanisms ensure near perfect fidelity for DNA replication.

In a cell, DNA replication begins at specific locations, or origins of replication, in the genome. Unwinding of DNA at the origin and synthesis of new strands, accommodated by an enzyme known as helicase, results in replication forks growing bi-directionally from the origin. A number of proteins are associated with the replication fork to help in the initiation and continuation of DNA synthesis. Most prominently, DNA polymerase synthesizes the new strands by adding nucleotides that complement each (template) strand. DNA replication occurs during the S-stage of interphase.

DNA replication (DNA amplification) can also be performed *in vitro* (artificially, outside a cell). DNA polymerases isolated from cells and artificial DNA primers can be used to start DNA synthesis at known sequences in a template DNA molecule. Polymerase chain reaction (PCR), ligase chain reaction (LCR), and transcription-mediated amplification (TMA) are examples.

The replisome³² is a complex molecular machine that carries out replication of DNA. The replisome first unwinds double stranded DNA into two single strands. For each of the resulting single strands, a new complementary sequence of DNA is synthesized. The net result is formation of two new double stranded DNA sequences that are exact copies of the original double stranded DNA sequence.

In terms of structure, the replisome is composed of two replicative polymerase complexes, one of which synthesizes the leading strand, while the other synthesizes the lagging strand. The replisome is composed of a number of proteins including helicase, RFC, PCNA, gy-

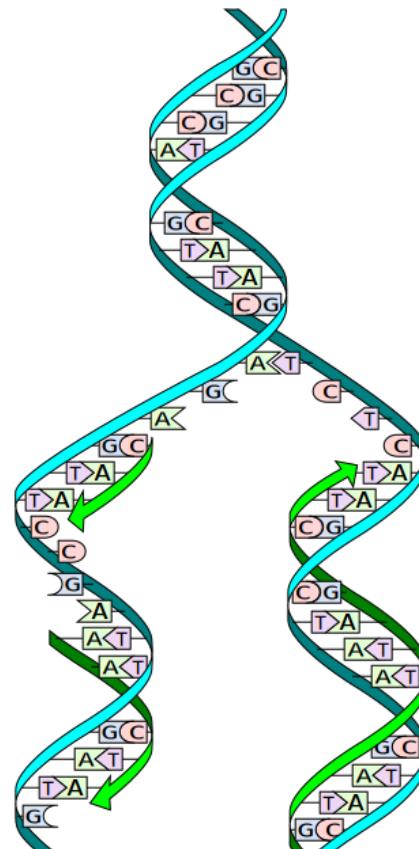


Figure 11.6: DNA polymerases adds nucleotides to the 3' end of a strand of DNA. If a mismatch is accidentally incorporated, the polymerase is inhibited from further extension. Proofreading removes the mismatched nucleotide and extension continues.³¹

³²<https://en.wikipedia.org/wiki/Replisome>

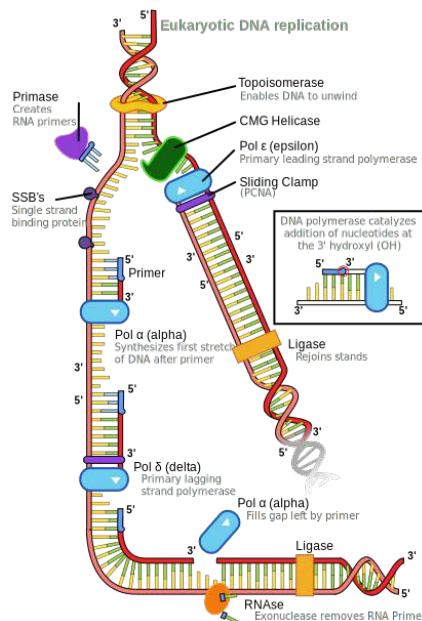


Figure 11.7: Many enzymes are involved in forming the DNA replication fork and DNA polymerization.³³

rase/topoisomerase, SSB/RPA, primase, DNA polymerase III, RNase H, and ligase.

For prokaryotes, each dividing nucleoid (region containing genetic material which is not a nucleus) requires two replisomes for bidirectional replication. The two replisomes continue replication at both forks in the middle of the cell. Finally, as the termination site replicates, the two replisomes separate from the DNA. The replisome remains at a fixed, midcell location in the cell, attached to the membrane, and the template DNA threads through it. DNA is fed through the stationary pair of replisomes located at the cell membrane.

For eukaryotes, numerous replication bubbles form at origins of replication throughout the chromosome. As with prokaryotes, two replisomes are required, one at each replication fork located at the terminus of the replication bubble. Because of significant differences in chromosome size, and the associated complexities of highly condensed chromosomes, various aspects of the DNA replication process in eukaryotes, including the terminal phases, are less well-characterised than for prokaryotes.

The replisome is responsible for copying the entirety of genomic DNA in each proliferative cell. This process allows for the high-fidelity passage of hereditary/genetic information from parental cell to daughter cell and is thus essential to all organisms. Much of the cell cycle is built around ensuring that DNA replication occurs without errors.

In G1 phase of the cell cycle, many of the DNA replication regulatory processes are initiated. In eukaryotes, the vast majority of DNA synthesis occurs during S phase of the cell cycle, and the entire genome must be unwound and duplicated to form two daughter copies. During G2, any damaged DNA or replication errors are corrected. Finally, one copy of the genomes is segregated to each daughter cell at mitosis or M phase. These daughter copies each contain one strand from the parental duplex DNA and one nascent antiparallel strand.

11.1.5 Eukaryotic DNA Replication

Eukaryotic DNA replication is a conserved mechanism that restricts DNA replication to once per cell cycle. Eukaryotic DNA replication of chromosomal DNA is central for the duplication of a cell and is necessary for the maintenance of the eukaryotic genome.

DNA replication is the action of DNA polymerases synthesizing a DNA strand complementary to the original template strand. To synthesize DNA, the double-stranded DNA is unwound by DNA helicases ahead of polymerases, forming a replication fork containing two single-stranded templates. Replication processes permit the copying of a single DNA double helix into two DNA helices, which are divided into the daughter cells at mitosis. The major enzymatic functions carried out at the replication fork are well conserved from prokaryotes to eukaryotes, but the replication machinery in eukaryotic DNA replication is a much larger complex, coordinating many proteins at the site of replication, forming the replisome.

After the replicative helicase has unwound the parental DNA duplex, exposing two single-stranded DNA templates, replicative polymerases are needed to generate two copies of the parental genome. DNA polymerase function is highly specialized and accomplish replication on specific templates and in narrow localizations. At the eukaryotic replication fork, there are three distinct replicative polymerase complexes that contribute to DNA replication: Polymerase α , Polymerase δ , and Polymerase ϵ . These three polymerases are essential for viability of the cell.

Because DNA polymerases require a primer on which to begin DNA synthesis, polymerase α (Pol α) acts as a replicative primase. Pol α is associated with an RNA primase and this complex accomplishes the priming task by synthesizing a primer that contains a short 10 nucleotide stretch of RNA followed by 10 to 20 DNA bases. Importantly, this priming action occurs at replication initiation at origins to begin leading-strand synthesis and also at the 5' end of each Okazaki fragment on the lagging strand.

However, Pol α is not able to continue DNA replica-

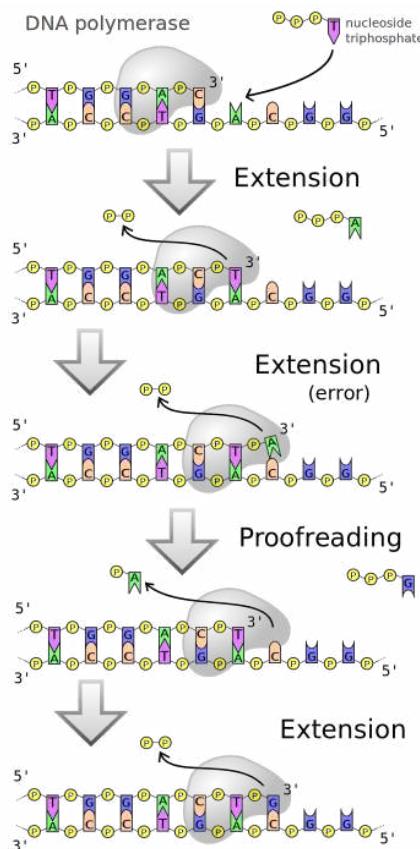


Figure 11.8: DNA polymerases adds nucleotides to the 3' end of a strand of DNA. If a mismatch is accidentally incorporated, the polymerase is inhibited from further extension. Proofreading removes the mismatched nucleotide and extension continues.³⁴

tion and must be replaced with another polymerase to continue DNA synthesis. Polymerase switching requires clamp loaders and it has been proven that normal DNA replication requires the coordinated actions of all three DNA polymerases: Pol α for priming synthesis, Pol ϵ for leading-strand replication, and the Pol δ , which is constantly loaded, for generating Okazaki fragments during lagging-strand synthesis.

- **Polymerase α (Pol α):** Forms a complex with a small catalytic subunit (PriS) and a large noncatalytic (PriL) subunit. First, synthesis of an RNA primer allows DNA synthesis by DNA polymerase alpha. Occurs once at the origin on the leading strand and at the start of each Okazaki fragment on the lagging strand. Pri subunits act as a primase, synthesizing an RNA primer. DNA Pol α elongates the newly formed primer with DNA nucleotides. After around 20 nucleotides, elongation is taken over by Pol ϵ on the leading strand and Pol δ on the lagging strand.
- **Polymerase δ (Pol δ):** Highly processive and has proofreading, 3'->5' exonuclease activity. In vivo, it is the main polymerase involved in both lagging strand and leading strand synthesis.
- **Polymerase ϵ (Pol ϵ):** Highly processive and has proofreading, 3'->5' exonuclease activity. Highly related to pol δ , in vivo it functions mainly in error checking of pol δ .

DNA replication, like all biological polymerization processes, proceeds in three enzymatically catalyzed and co-ordinated steps: initiation, elongation and termination.

11.1.6 Initiation

For a cell to divide, it must first replicate its DNA. DNA replication is an all-or-none process; once replication begins, it proceeds to completion. Once replication is complete, it does not occur again in the same cell cycle. This is made possible by the division of initiation into two temporally distinct steps: formation of the pre-replication complex and the preinitiation complex.

11.1.7 Pre-replication complex

In late mitosis and early G1 phase, a large complex of initiator proteins assembles into the pre-replication complex at particular points in the DNA, known as "origins". In *E. coli* the primary initiator protein is DnaA; in yeast, this is the origin recognition complex. Sequences used by initiator proteins tend to be "AT-rich" (rich in adenine and thymine bases), because A-T base pairs have two hydrogen bonds (rather than the three formed in a C-G pair) and thus are easier to strand-separate. In eukaryotes, the origin recognition complex catalyzes the assembly of initiator proteins

into the pre-replication complex. Cdc6 and Cdt1 then associate with the bound origin recognition complex at the origin in order to form a larger complex necessary to load the Mcm complex onto the DNA. The Mcm complex is the helicase that will unravel the DNA helix at the replication origins and replication forks in eukaryotes. The Mcm complex is recruited at late G1 phase and loaded by the ORC-Cdc6-Cdt1 complex onto the DNA via ATP-dependent protein remodeling. The loading of the Mcm complex onto the origin DNA marks the completion of pre-replication complex formation.

If environmental conditions are right in late G1 phase, the G1 and G1/S cyclin-Cdk complexes are activated, which stimulate expression of genes that encode components of the DNA synthetic machinery. G1/S-Cdk activation also promotes the expression and activation of S-Cdk complexes, which may play a role in activating replication origins depending on species and cell type. Control of these Cdks vary depending cell type and stage of development.

In a similar manner, Cdc7 is also required through S phase to activate replication origins. Cdc7 is not active throughout the cell cycle, and its activation is strictly timed to avoid premature initiation of DNA replication. In late G1, Cdc7 activity rises abruptly as a result of association with the regulatory subunit Dbf4, which binds Cdc7 directly and promotes its protein kinase activity. Cdc7 has been found to be a rate-limiting regulator of origin activity. Together, the G1/S-Cdks and/or S-Cdks and Cdc7 collaborate to directly activate the replication origins, leading to initiation of DNA synthesis.

11.1.8 Preinitiation complex

In early S phase, S-Cdk and Cdc7 activation lead to the assembly of the preinitiation complex, a massive protein complex formed at the origin. Formation of the preinitiation complex displaces Cdc6 and Cdt1 from the origin replication complex, inactivating and disassembling the pre-replication complex. Loading the preinitiation complex onto the origin activates the Mcm helicase, causing unwinding of the DNA helix. The preinitiation complex also loads α -primase and other DNA polymerases onto the DNA.

After α -primase synthesizes the first primers, the primer-template junctions interact with the clamp loader, which loads the sliding clamp onto the DNA to begin DNA synthesis. The components of the preinitiation complex remain associated with replication forks as they move out from the origin.

11.1.9 Elongation

DNA polymerase has 5'-3' activity. All known DNA replication systems require a free 3' hydroxyl group before synthesis can be initiated (note: the DNA template is read in 3' to 5' direction whereas a new strand is synthesized in the 5' to 3' direction—this is often confused). Four distinct mechanisms for DNA synthesis are recognized:

All cellular life forms and many DNA viruses, phages and plasmids use a primase to synthesize a short RNA primer with a free 3' OH group which is subsequently elongated by a DNA polymerase. The 5' end of the nicked strand is transferred to a tyrosine residue on the nuclease and the free 3' OH group is then used by the DNA polymerase to synthesize the new strand. The first is the best known of these mechanisms and is used by the cellular organisms. In this mechanism, once the two strands are separated, primase adds RNA primers to the template strands. The leading strand receives one RNA primer while the lagging strand receives several. The leading strand is continuously extended from the primer by a DNA polymerase with high processivity, while the lagging strand is extended discontinuously from each primer forming Okazaki fragments. RNase removes the primer RNA fragments, and a low processivity DNA polymerase distinct from the replicative polymerase enters to fill the gaps. When this is complete, a single nick on the leading strand and several nicks on the lagging strand can be found. Ligase works to fill these nicks in, thus completing the newly replicated DNA molecule.

Multiple DNA polymerases take on different roles in the DNA replication process. In *E. coli*, DNA Pol III is the polymerase enzyme primarily responsible for DNA replication. It assembles into a replication complex at the replication fork that exhibits extremely high processivity, remaining intact for the entire replication cycle. In contrast, DNA Pol I is the enzyme responsible for replacing RNA primers with DNA. DNA Pol I has a 5' to 3' exonuclease activity in addition to its polymerase activity, and uses its exonuclease activity to degrade the RNA primers ahead of it as it extends the DNA strand behind it, in a process called nick translation. Pol I is much less processive than Pol III because its primary function in DNA replication is to create many short DNA regions rather than a few very long regions.

In eukaryotes, the low-processivity enzyme, Pol α , helps to initiate replication because it forms a complex with primase. In eukaryotes, leading strand synthesis is thought to be conducted by Pol ϵ ; however, this view has recently been challenged, suggesting a role for Pol δ . Primer removal is completed by Pol δ while repair of DNA during replication is completed by Pol ϵ .

As DNA synthesis continues, the original DNA strands

continue to unwind on each side of the bubble, forming a replication fork with two prongs. In bacteria, which have a single origin of replication on their circular chromosome, this process creates a “theta structure” (resembling the Greek letter theta: θ). In contrast, eukaryotes have longer linear chromosomes and initiate replication at multiple origins within these.

11.1.10 Replication Fork

The replication fork is a structure that forms within the long helical DNA during DNA replication. It is created by helicases, which break the hydrogen bonds holding the two DNA strands together in the helix. The resulting structure has two branching “prongs”, each one made up of a single strand of DNA. These two strands serve as the template for the leading and lagging strands, which will be created as DNA polymerase matches complementary nucleotides to the templates; the templates may be properly referred to as the leading strand template and the lagging strand template.

DNA is always synthesized by adding nucleotides to the 3' end of a strand. Since the leading and lagging strand templates are oriented in opposite directions at the replication fork, a major issue is how to achieve synthesis of nascent (new) lagging strand DNA, whose direction of synthesis is opposite to the direction of the growing replication fork.

11.1.11 Replication of The Leading Strand

The leading strand is the strand of nascent DNA which is synthesized in the same direction as the growing replication fork. This sort of DNA replication is continuous.

11.1.12 Replication of The Lagging Strand

The lagging strand is the strand of nascent DNA whose direction of synthesis is opposite to the direction of the growing replication fork. Because of its orientation, replication of the lagging strand is more complicated as compared to that of the leading strand. As a consequence, the DNA polymerase on this strand is seen to “lag behind” the other strand.

The lagging strand is synthesized in short, separated segments. On the lagging strand template, a primase “reads” the template DNA and initiates synthesis of a short complementary RNA primer. A DNA polymerase extends the primed segments, forming Okazaki fragments. The RNA primers are then removed and replaced with DNA, and the fragments of DNA are joined together by DNA ligase.

In all cases the helicase is composed of six polypeptides that wrap around only one strand of the DNA being replicated. The two polymerases are bound to the helicase hexamer. In eukaryotes the helicase wraps around the leading strand, and in prokaryotes it wraps around the lagging strand.

As helicase unwinds DNA at the replication fork, the DNA ahead is forced to rotate. This process results in a build-up of twists in the DNA ahead. This build-up forms a torsional resistance that would eventually halt the progress of the replication fork. Topoisomerases are enzymes that temporarily break the strands of DNA, relieving the tension caused by unwinding the two strands of the DNA helix; topoisomerases (including DNA gyrase) achieve this by adding negative supercoils to the DNA helix.

Bare single-stranded DNA tends to fold back on itself forming secondary structures; these structures can interfere with the movement of DNA polymerase. To prevent this, single-strand binding proteins bind to the DNA until a second strand is synthesized, preventing secondary structure formation.

Clamp proteins form a sliding clamp around DNA, helping the DNA polymerase maintain contact with its template, thereby assisting with processivity. The inner face of the clamp enables DNA to be threaded through it. Once the polymerase reaches the end of the template or detects double-stranded DNA, the sliding clamp undergoes a conformational change that releases the DNA polymerase. Clamp-loading proteins are used to initially load the clamp, recognizing the junction between template and RNA primers.

11.1.13 DNA Replication Proteins

At the replication fork, many replication enzymes assemble on the DNA into a complex molecular machine called the replisome. The following is a list of major DNA replication enzymes that participate in the replisome:

Table 11.1: A list of major DNA replication enzymes that participate in the replisome

Enzymes	Function in DNA replication
DNA helicase	Also known as helix destabilizing enzyme. Helicase separates the two strands of DNA at the Replication Fork behind the topoisomerase.
DNA polymerase	The enzyme responsible for catalyzing the addition of nucleotide substrates to DNA in the 5' to 3' direction during DNA replication. Also performs proof-reading and error correction. There exist many different types of DNA Polymerase, each of which perform different functions in different types of cells.
DNA clamp	A protein which prevents elongating DNA polymerases from dissociating from the DNA parent strand.
Single-strand DNA-binding protein	Bind to ssDNA and prevent the DNA double helix from re-annealing after DNA helicase unwinds it; thus maintaining the strand separation, and facilitating the synthesis of the nascent strand.
Topoisomerase	Relaxes the DNA from its super-coiled nature.
DNA gyrase	Relieves strain of unwinding by DNA helicase; this is a specific type of topoisomerase
DNA ligase	Re-anneals the semi-conservative strands and joins Okazaki Fragments of the lagging strand.
Primase	Provides a starting point of RNA (or DNA) for DNA polymerase to begin synthesis of the new DNA strand.
Telomerase	Lengthens telomeric DNA by adding repetitive nucleotide sequences to the ends of eukaryotic chromosomes. This allows germ cells and stem cells to avoid the Hayflick limit on cell division.

11.1.14 Termination

Eukaryotes initiate DNA replication at multiple points in the chromosome, so replication forks meet and terminate at many points in the chromosome. Because eukaryotes have linear chromosomes, DNA replication is unable to reach the very end of the chromosomes. Due to this problem, DNA is lost in each replication cycle from the end of the chromosome. Telomeres are regions of repetitive DNA close to the ends and help prevent loss of genes due to this shortening. Shortening of the telomeres is a normal process in somatic cells. This shortens the telomeres of the daughter DNA chromosome. As a result, cells can only divide a certain number of times before the DNA loss prevents further division. (This is known as the Hayflick³⁵ limit.) Within the germ cell line, which passes DNA to the next generation, telomerase extends the repetitive sequences of the telomere region to prevent degradation. Telomerase can become mistakenly active in somatic cells, sometimes leading to cancer formation. Increased telomerase activity is one of the hallmarks of cancer.

Termination requires that the progress of the DNA replication fork must stop or be blocked. Termination at a specific locus, when it occurs, involves the interaction between two components: (1) a termination site sequence in the DNA, and (2) a protein which binds to this sequence to physically stop DNA replication. In various bacterial species, this is named the DNA replication terminus site-binding protein, or Ter protein.

Because bacteria have circular chromosomes, termination of replication occurs when the two replication forks meet each other on the opposite end of the parental chromosome. *E. coli* regulates this process through the use of termination sequences that, when bound by the Tus protein, enable only one direction of replication fork to pass through. As a result, the replication forks are constrained to always meet within the termination region of the chromosome.

11.1.15 Regulation of DNA Replication

Within eukaryotes, DNA replication is controlled within the context of the cell cycle. As the cell grows and divides, it progresses through stages in the cell cycle; DNA replication takes place during the S phase (synthesis phase). The progress of the eukaryotic cell through the cycle is controlled by cell cycle checkpoints. Progression through checkpoints is controlled through complex interactions between various proteins, including cyclins and cyclin-dependent kinases.

The G1/S checkpoint (or restriction checkpoint) regulates whether eukaryotic cells enter the process of DNA

replication and subsequent division. Cells that do not proceed through this checkpoint remain in the G0 stage and do not replicate their DNA.

After passing through the G1/S checkpoint, DNA must be replicated only once in each cell cycle. When the Mcm complex moves away from the origin, the pre-replication complex is dismantled. Because a new Mcm complex cannot be loaded at an origin until the pre-replication subunits are reactivated, one origin of replication can not be used twice in the same cell cycle.

Activation of S-Cdk's in early S phase promotes the destruction or inhibition of individual pre-replication complex components, preventing immediate reassembly. S and M-Cdk's continue to block pre-replication complex assembly even after S phase is complete, ensuring that assembly cannot occur again until all Cdk activity is reduced in late mitosis.

Replication of chloroplast and mitochondrial genomes occurs independently of the cell cycle, through the process of D-loop replication.

11.1.16 Interactions of DNA with Proteins

All the functions of DNA depend on interactions with proteins. These protein interactions can be non-specific, or the protein can bind specifically to a single DNA sequence. Enzymes can also bind to DNA and of these, the polymerases that copy the DNA base sequence in transcription and DNA replication are particularly important.

11.1.17 DNA-binding proteins

Structural proteins that bind DNA are well-understood examples of non-specific DNA-protein interactions. Within chromosomes, DNA is held in complexes with structural proteins. These proteins organize the DNA into a compact structure called chromatin. In eukaryotes, this structure involves DNA binding to a complex of small basic proteins called histones, while in prokaryotes multiple types of proteins are involved. The histones form a disk-shaped complex called a nucleosome, which contains two complete turns of double-stranded DNA wrapped around its surface. These non-specific interactions are formed through basic residues in the histones, making ionic bonds to the acidic sugar-phosphate backbone of the DNA, and are thus largely independent of the base sequence. Chemical modifications of these basic amino acid residues include methylation, phosphorylation, and acetylation. These chemical changes alter the strength of the interaction between the DNA and the histones, making the DNA more or less accessible to transcription factors and changing the rate of transcription. Other non-specific DNA-binding proteins in chromatin include

³⁵https://en.wikipedia.org/wiki/Leonard_Hayflick

the high-mobility group proteins, which bind to bent or distorted DNA. These proteins are important in bending arrays of nucleosomes and arranging them into the larger structures that make up chromosomes.

A distinct group of DNA-binding proteins is the DNA-binding proteins that specifically bind single-stranded DNA. In humans, replication protein A is the best-understood member of this family and is used in processes where the double helix is separated, including DNA replication, recombination, and DNA repair. These binding proteins seem to stabilize single-stranded DNA and protect it from forming stem-loops or being degraded by nucleases.

In contrast, other proteins have evolved to bind to particular DNA sequences. The most intensively studied of these are the various transcription factors, which are proteins that regulate transcription. Each transcription factor binds to one particular set of DNA sequences and activates or inhibits the transcription of genes that have these sequences close to their promoters. The transcription factors do this in two ways. Firstly, they can bind the RNA polymerase responsible for transcription, either directly or through other mediator proteins; this locates the polymerase at the promoter and allows it to begin transcription. Alternatively, transcription factors can bind enzymes that modify the histones at the promoter. This changes the accessibility of the DNA template to the polymerase.

As these DNA targets can occur throughout an organism's genome, changes in the activity of one type of transcription factor can affect thousands of genes. Consequently, these proteins are often the targets of the signal transduction processes that control responses to environmental changes or cellular differentiation and development. The specificity of these transcription factors' interactions with DNA come from the proteins making multiple contacts to the edges of the DNA bases, allowing them to "read" the DNA sequence. Most of these base-interactions are made in the major groove, where the bases are most accessible.

11.1.18 DNA-modifying Enzymes

11.1.19 Nucleases And Ligases

Nucleases are enzymes that cut DNA strands by catalyzing the hydrolysis of the phosphodiester bonds. Nucleases that hydrolyse nucleotides from the ends of DNA strands are called exonucleases, while endonucleases cut within strands. The most frequently used nucleases in molecular biology are the restriction endonucleases, which cut DNA at specific sequences. For instance, the EcoRI enzyme recognizes the 6-base sequence 5'-GAATTC-3' and cuts each

strand after the G creating 4 nucleotide sticky ends with a 5' end overhang of AATT. In nature, these enzymes protect bacteria against phage infection by digesting the phage DNA when it enters the bacterial cell, acting as part of the restriction modification system. These sequence-specific nucleases are used in molecular cloning and DNA finger-printing.

Enzymes called DNA ligases can rejoin cut or broken DNA strands. Ligases are particularly important in lagging strand DNA replication, as they join together the short segments of DNA produced at the replication fork into a complete copy of the DNA template. They are also used in DNA repair and genetic recombination.

11.1.20 Topoisomerases And Helicases

Topoisomerases are enzymes with both nuclease and ligase activity. These proteins change the amount of supercoiling in DNA. Some of these enzymes work by cutting the DNA helix and allowing one section to rotate, thereby reducing its level of supercoiling; the enzyme then seals the DNA break. Other types of these enzymes are capable of cutting one DNA helix and then passing a second strand of DNA through this break, before rejoining the helix. Topoisomerases are required for many processes involving DNA, such as DNA replication and transcription.

Helicases are proteins that are a type of molecular motor. They use the chemical energy in nucleoside triphosphates, predominantly adenosine triphosphate (ATP), to break hydrogen bonds between bases and unwind the DNA double helix into single strands. These enzymes are essential for most processes where enzymes need to access the DNA bases.

11.1.21 Polymerases

Polymerases are enzymes that synthesize polynucleotide chains from nucleoside triphosphates. The sequence of their products is created based on existing polynucleotide chains—which are called templates. These enzymes function by repeatedly adding a nucleotide to the 3' hydroxyl group at the end of the growing polynucleotide chain. As a consequence, all polymerases work in a 5' to 3' direction. In the active site of these enzymes, the incoming nucleoside triphosphate base-pairs to the template: this allows polymerases to accurately synthesize the complementary strand of their template. Polymerases are classified according to the type of template that they use.

In DNA replication, DNA-dependent DNA polymerases make copies of DNA polynucleotide chains. To preserve biological information, it is essential that the sequence of bases in each copy are precisely complementary to the sequence of bases in the template strand. Many

DNA polymerases have a proofreading activity. Here, the polymerase recognizes the occasional mistakes in the synthesis reaction by the lack of base pairing between the mismatched nucleotides. If a mismatch is detected, a 3' to 5' exonuclease activity is activated and the incorrect base removed. In most organisms, DNA polymerases function in a large complex called the replisome³⁶ that contains multiple accessory subunits, such as the DNA clamp or helicases.

RNA-dependent DNA polymerases are a specialized class of polymerases that copy the sequence of an RNA strand into DNA. They include reverse transcriptase, which is a viral enzyme involved in the infection of cells by retroviruses, and telomerase, which is required for the replication of telomeres. For example, HIV reverse transcriptase is an enzyme for AIDS virus replication. Telomerase is an unusual polymerase because it contains its own RNA template as part of its structure. It synthesizes telomeres at the ends of chromosomes. Telomeres prevent fusion of the ends of neighboring chromosomes and protect chromosome ends from damage.

Transcription is carried out by a DNA-dependent RNA polymerase that copies the sequence of a DNA strand into RNA. To begin transcribing a gene, the RNA polymerase binds to a sequence of DNA called a promoter and separates the DNA strands. It then copies the gene sequence into a messenger RNA transcript until it reaches a region of DNA called the terminator, where it halts and detaches from the DNA. As with human DNA-dependent DNA polymerases, RNA polymerase II, the enzyme that transcribes most of the genes in the human genome, operates as part of a large protein complex with multiple regulatory and accessory subunits.

11.1.22 DNA Recombination

A DNA helix usually does not interact with other segments of DNA, and in human cells, the different chromosomes even occupy separate areas in the nucleus called "chromosome territories". This physical separation of different chromosomes is important for the ability of DNA to function as a stable repository for information, as one of the few times chromosomes interact is in chromosomal crossover which occurs during sexual reproduction, when genetic recombination occurs. Chromosomal crossover is when two DNA helices break, swap a section and then rejoin.

Recombination allows chromosomes to exchange genetic information and produces new combinations of genes, which increases the efficiency of natural selection and can be important in the rapid evolution of new proteins. Genetic recombination can also be involved in DNA

repair, particularly in the cell's response to double-strand breaks.

The most common form of chromosomal crossover is homologous recombination, where the two chromosomes involved share very similar sequences. Non-homologous recombination can be damaging to cells, as it can produce chromosomal translocations and genetic abnormalities. The recombination reaction is catalyzed by enzymes known as recombinases, such as RAD51. The first step in recombination is a double-stranded break caused by either an endonuclease or damage to the DNA. A series of steps catalyzed in part by the recombinase then leads to joining of the two helices by at least one Holliday junction, in which a segment of a single strand in each helix is annealed to the complementary strand in the other helix. The Holliday junction is a tetrahedral junction structure that can be moved along the pair of chromosomes, swapping one strand for another. The recombination reaction is then halted by cleavage of the junction and re-ligation of the released DNA. Only strands of like polarity exchange DNA during recombination. There are two types of cleavage: east-west cleavage and north-south cleavage. The north-south cleavage nicks both strands of DNA, while the east-west cleavage has one strand of DNA intact.

11.1.23 Evolutionary History of DNA

DNA contains the genetic information that allows all forms of life to function, grow and reproduce. However, it is unclear how long in the 4-billion-year history of life DNA has performed this function, as it has been proposed that the earliest forms of life may have used RNA as their genetic material. RNA may have acted as the central part of early cell metabolism as it can both transmit genetic information and carry out catalysis as part of ribozymes. This ancient RNA world where nucleic acid would have been used for both catalysis and genetics may have influenced the evolution of the current genetic code based on four nucleotide bases. This would occur, since the number of different bases in such an organism is a trade-off between a small number of bases increasing replication accuracy and a large number of bases increasing the catalytic efficiency of ribozymes. However, there is no direct evidence of ancient genetic systems, as recovery of DNA from most fossils is impossible because DNA survives in the environment for less than one million years, and slowly degrades into short fragments in solution.

Building blocks of DNA (adenine, guanine, and related organic molecules) may have been formed extraterrestrially in outer space. Complex DNA and RNA organic compounds of life, including uracil, cytosine, and thymine, have also been formed in the laboratory under conditions mimicking those found in outer space, using starting

³⁶<https://en.wikipedia.org/wiki/Replisome>

chemicals, such as pyrimidine, found in meteorites. Pyrimidine, like polycyclic aromatic hydrocarbons (PAHs), the most carbon-rich chemical found in the universe, may have been formed in red giants or in interstellar cosmic dust and gas clouds.

11.1.24 Genetic engineering

Methods have been developed to purify DNA from organisms, such as phenol-chloroform extraction, and to manipulate it in the laboratory, such as restriction digests and the polymerase chain reaction. Modern biology and biochemistry make intensive use of these techniques in recombinant DNA technology. Recombinant DNA is a man-made DNA sequence that has been assembled from other DNA sequences. They can be transformed into organisms in the form of plasmids or in the appropriate format, by using a viral vector. The genetically modified organisms produced can be used to produce products such as recombinant proteins, used in medical research, or be grown in agriculture.

11.1.25 DNA profiling

Forensic scientists can use DNA in blood, semen, skin, saliva or hair found at a crime scene to identify a matching DNA of an individual, such as a perpetrator. This process is formally termed DNA profiling, also called DNA fingerprinting. In DNA profiling, the lengths of variable sections of repetitive DNA, such as short tandem repeats and minisatellites, are compared between people. This method is usually an extremely reliable technique for identifying a matching DNA. However, identification can be complicated if the scene is contaminated with DNA from several people. DNA profiling was developed in 1984 by British geneticist Sir Alec Jeffreys, and first used in forensic science to convict Colin Pitchfork in the 1988 Enderby murders case.

DNA profiling is also used in DNA paternity testing to determine if someone is the biological parent or grandparent of a child with the probability of parentage is typically 99.99% when the alleged parent is biologically related to the child. Normally, paternity testing is performed after birth, but recently developed methods allow isolation and sequencing of fetal DNA from the blood of the mother.

11.2 Structure And Function of Ribonucleic Acid

Ribonucleic acid (RNA)³⁷ is a polymeric molecule essential in various biological roles in coding, decoding, regulation and expression of genes. Like DNA, RNA is assembled as a

chain of nucleotides, but unlike DNA it is more often found in nature as a single-strand folded onto itself, rather than a paired double-strand. Cellular organisms use messenger RNA (mRNA) to convey genetic information (using the nitrogenous bases of guanine, uracil, adenine, and cytosine, denoted by the letters G, U, A, and C) and direct synthesis of proteins by ribosomes. Many viruses encode their genetic information using an RNA genome.

Some RNA molecules play an active role within cells by catalyzing biological reactions, controlling gene expression, or sensing and communicating responses to cellular signals. One of these active processes is protein synthesis, a universal function in which RNA molecules direct the synthesis of proteins on ribosomes. This process uses transfer RNA (tRNA) molecules to deliver amino acids to the ribosome, where ribosomal RNA (rRNA) then links amino acids together to form coded proteins.

11.2.1 Comparison with DNA

The chemical structure of RNA is very similar to that of DNA, but differs in three primary ways:

- Unlike double-stranded DNA, RNA is a single-stranded molecule in many of its biological roles and consists of much shorter chains of nucleotides. However, a single RNA molecule can, by complementary base pairing, form intrastrand double helices, as in tRNA.
- While the sugar-phosphate “backbone” of DNA contains deoxyribose, RNA contains ribose instead. Ribose has a hydroxyl group attached to the pentose ring in the 2' position, whereas deoxyribose does not. The hydroxyl groups in the ribose backbone make RNA more chemically labile than DNA by lowering the activation energy of hydrolysis.
- The complementary base to adenine in DNA is thymine, whereas in RNA, it is uracil, which is an unmethylated form of thymine.

Like DNA, most biologically active RNAs, including mRNA, tRNA, rRNA, snRNAs, and other non-coding RNAs, contain self-complementary sequences that allow parts of the RNA to fold and pair with itself to form double helices. Analysis of these RNAs has revealed that they are highly structured. Unlike DNA, their structures do not consist of long double helices, but rather collections of short helices packed together into structures akin to proteins. In this fashion, RNAs can achieve chemical catalysis (like enzymes). For instance, determination of the structure of the ribosome—an RNA-protein complex that catalyzes peptide bond formation—revealed that its active site is composed entirely of RNA.

³⁷<https://en.wikipedia.org/wiki/RNA>

11.2.2 Structure of RNA

Each nucleotide in RNA contains a ribose sugar, with carbons numbered 1' through 5'. A base is attached to the 1' position, in general, adenine (A), cytosine (C), guanine (G), or uracil (U). Adenine and guanine are purines, cytosine and uracil are pyrimidines. A phosphate group is attached to the 3' position of one ribose and the 5' position of the next. The phosphate groups have a negative charge each, making RNA a charged molecule (polyanion). The bases form hydrogen bonds between cytosine and guanine, between adenine and uracil and between guanine and uracil. However, other interactions are possible, such as a group of adenine bases binding to each other in a bulge, or the GNRA tetraloop that has a guanine–adenine base-pair.

An important structural component of RNA that distinguishes it from DNA is the presence of a hydroxyl group at the 2' position of the ribose sugar.

RNA is transcribed with only four bases (adenine, cytosine, guanine and uracil), but these bases and attached sugars can be modified in numerous ways as the RNAs mature. Pseudouridine (Ψ), in which the linkage between uracil and ribose is changed from a C–N bond to a C–C bond, and ribothymidine (T) are found in various places (the most notable ones being in the TΨC loop of tRNA). Another notable modified base is hypoxanthine, a deaminated adenine base whose nucleoside is called inosine (I). Inosine plays a key role in the wobble hypothesis of the genetic code.

There are more than 100 other naturally occurring modified nucleosides. The greatest structural diversity of modifications can be found in tRNA, while pseudouridine and nucleosides with 2'-O-methylribose often present in rRNA are the most common. The specific roles of many of these modifications in RNA are not fully understood. However, it is notable that, in ribosomal RNA, many of the post-transcriptional modifications occur in highly functional regions, such as the peptidyl transferase center and the subunit interface, implying that they are important for normal function.

The functional form of single-stranded RNA molecules, just like proteins, frequently requires a specific tertiary structure. The scaffold for this structure is provided by secondary structural elements that are hydrogen bonds within the molecule. This leads to several recognizable "domains" of secondary structure like hairpin loops, bulges, and internal loops. Since RNA is charged, metal ions such as Mg²⁺ are needed to stabilise many secondary and tertiary structures.

The naturally occurring enantiomer of RNA is D-RNA composed of D-ribonucleotides. All chirality centers are located in the D-ribose. By the use of L-ribose or rather

L-ribonucleotides, L-RNA can be synthesized. L-RNA is much more stable against degradation by RNase.

11.2.3 Synthesis of RNA

Synthesis of RNA is usually catalyzed by an enzyme—RNA polymerase—using DNA as a template, a process known as transcription. Initiation of transcription begins with the binding of the enzyme to a promoter sequence in the DNA (usually found "upstream" of a gene). The DNA double helix is unwound by the helicase activity of the enzyme. The enzyme then progresses along the template strand in the 3' to 5' direction, synthesizing a complementary RNA molecule with elongation occurring in the 5' to 3' direction. The DNA sequence also dictates where termination of RNA synthesis will occur.

Primary transcript RNAs are often modified by enzymes after transcription. For example, a poly(A) tail and a 5' cap are added to eukaryotic pre-mRNA and introns are removed by the spliceosome.

There are also a number of RNA-dependent RNA polymerases that use RNA as their template for synthesis of a new strand of RNA. For instance, a number of RNA viruses (such as poliovirus) use this type of enzyme to replicate their genetic material. Also, RNA-dependent RNA polymerase is part of the RNA interference pathway in many organisms.

11.2.4 Coding and non-coding RNA

Messenger RNA (mRNA) is the RNA that carries information from DNA to the ribosome, the sites of protein synthesis (translation) in the cell. The coding sequence of the mRNA determines the amino acid sequence in the protein that is produced. However, many RNAs do not code for protein (about 97% of the transcriptional output is non-protein-coding in eukaryotes).

These so-called non-coding RNAs ("ncRNA") can be encoded by their own genes (RNA genes), but can also derive from mRNA introns. The most prominent examples of non-coding RNAs are transfer RNA (tRNA) and ribosomal RNA (rRNA), both of which are involved in the process of translation. There are also non-coding RNAs involved in gene regulation, RNA processing and other roles. Certain RNAs are able to catalyse chemical reactions such as cutting and ligating other RNA molecules, and the catalysis of peptide bond formation in the ribosome; these are known as ribozymes.

According to the length of RNA chain, RNA includes small RNA and long RNA. Usually, small RNAs are shorter than 200 nt in length, and long RNAs are greater than 200 nt long. Long RNAs, also called large RNAs, mainly

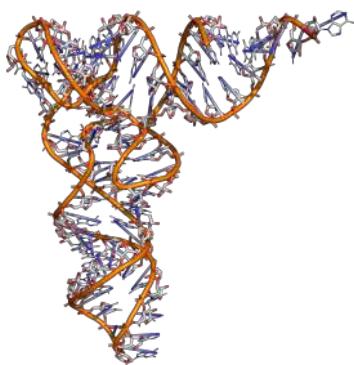


Figure 11.9: Tertiary structure of tRNA (based on atomic coordinates of PDB 1ehz³⁸ rendered with open source molecular visualization tool PyMol.)

include long non-coding RNA (lncRNA) and mRNA. Small RNAs mainly include 5.8S ribosomal RNA (rRNA), 5S rRNA, transfer RNA (tRNA), microRNA (miRNA), small interfering RNA (siRNA), small nucleolar RNA (snoRNAs), Piwi-interacting RNA (piRNA), tRNA-derived small RNA (tsRNA) and small rDNA-derived RNA (srRNA).

Messenger RNA (mRNA) carries genetic information from the nucleus to the cytoplasm and serves as template for the synthesis of protein by ribosomes, a process called translation. In eukaryotic cells, once precursor mRNA (pre-mRNA) has been transcribed from DNA, it is processed to mature mRNA. This removes its introns (non-coding stretches of DNA sequence) from the pre-mRNA. The mRNA is then exported from the nucleus to the cytoplasm, where ribosomes bind to it and translate its corresponding protein form with the help of tRNA. In prokaryotic cells, which do not have nucleus and cytoplasm compartments, mRNA can bind to ribosomes while it is being transcribed from DNA. After a certain amount of time, the message degrades into its component nucleotides with the assistance of ribonucleases.

Transfer RNA (tRNA) is a small RNA chain of about 80 nucleotides that transfers a specific amino acid to a growing polypeptide chain at the ribosomal site of protein synthesis during translation. It has sites for amino acid attachment and an anticodon region for codon recognition that binds to a specific sequence on the messenger RNA chain through hydrogen bonding.

Ribosomal RNA (rRNA) is the catalytic component of the ribosomes. Eukaryotic ribosomes contain four different rRNA molecules: 18S, 5.8S, 28S and 5S rRNA. Three of the rRNA molecules are synthesized in the nucleolus, and one is synthesized elsewhere. In the cytoplasm, ribosomal

RNA and protein combine to form a nucleoprotein called a ribosome. The ribosome binds mRNA and carries out protein synthesis. Several ribosomes may be attached to a single mRNA at any time. Nearly all the RNA found in a typical eukaryotic cell is rRNA.

Transfer-messenger RNA (tmRNA) is found in many bacteria and plastids. It tags proteins encoded by mRNAs that lack stop codons for degradation and prevents the ribosome from stalling.

11.2.5 Regulatory RNA

The earliest known regulators of gene expression were proteins known as repressors and activators, regulators with specific short binding sites within enhancer regions near the genes to be regulated. More recently, RNAs have been found to regulate genes as well. There are several kinds of RNA-dependent processes in eukaryotes regulating the expression of genes at various points, such as RNAi repressing genes post-transcriptionally, long non-coding RNAs shutting down blocks of chromatin epigenetically, and enhancer RNAs inducing increased gene expression. In addition to these mechanisms in eukaryotes, both bacteria and archaea have been found to use regulatory RNAs extensively. Bacterial small RNA and the CRISPR system are examples of such prokaryotic regulatory RNA systems.

11.2.6 RNA interference by miRNAs

Post-transcriptional expression levels of many genes can be controlled by RNA interference, in which miRNAs, specific short RNA molecules, pair with mRNA regions and target them for degradation. This antisense-based process involves steps that first process the RNA so that it can base-pair with a region of its target mRNAs. Once the base pairing occurs, other proteins direct the mRNA to be destroyed by nucleases.

Next to be linked to regulation were Xist and other long noncoding RNAs associated with X chromosome inactivation. Their roles, at first mysterious, were shown to be the silencing of blocks of chromatin via recruitment of Polycomb complex so that messenger RNA could not be transcribed from them. Additional lncRNAs, currently defined as RNAs of more than 200 base pairs that do not appear to have coding potential, have been found associated with regulation of stem cell pluripotency and cell division.

The third major group of regulatory RNAs is called enhancer RNAs. It is not clear at present whether they are a unique category of RNAs of various lengths or constitute a distinct subset of lncRNAs. In any case, they are transcribed from enhancers, which are known regulatory sites in the DNA near genes they regulate. They up-regulate the

transcription of the gene(s) under control of the enhancer from which they are transcribed.

At first, regulatory RNA was thought to be a eukaryotic phenomenon, a part of the explanation for why so much more transcription in higher organisms was seen than had been predicted. But as soon as researchers began to look for possible RNA regulators in bacteria, they turned up there as well, termed as small RNA (sRNA). Bacterial small RNAs generally act via antisense pairing with mRNA to down-regulate its translation, either by affecting stability or affecting cis-binding ability. Riboswitches have also been discovered. They are cis-acting regulatory RNA sequences acting allosterically. They change shape when they bind metabolites so that they gain or lose the ability to bind chromatin to regulate expression of genes.

Archaea also have systems of regulatory RNA. The CRISPR system, recently being used to edit DNA in situ, acts via regulatory RNAs in archaea and bacteria to provide protection against virus invaders.

Many RNAs are involved in modifying other RNAs. Introns are spliced out of pre-mRNA by spliceosomes, which contain several small nuclear RNAs (snRNA), or the introns can be ribozymes that are spliced by themselves. RNA can also be altered by having its nucleotides modified to nucleotides other than A, C, G and U. In eukaryotes, modifications of RNA nucleotides are in general directed by small nucleolar RNAs (snoRNA; 60–300 nt), found in the nucleolus and cajal bodies. snoRNAs associate with enzymes and guide them to a spot on an RNA by basepairing to that RNA. These enzymes then perform the nucleotide modification. rRNAs and tRNAs are extensively modified, but snRNAs and mRNAs can also be the target of base modification. RNA can also be methylated.

RNA viruses have genomes composed of RNA that encodes a number of proteins. The viral genome is replicated by some of those proteins, while other proteins protect the genome as the virus particle moves to a new host cell. Viroids are another group of pathogens, but they consist only of RNA, do not encode any protein and are replicated by a host plant cell's polymerase.

Reverse transcribing viruses replicate their genomes by reverse transcribing DNA copies from their RNA; these DNA copies are then transcribed to new RNA. Retrotransposons also spread by copying DNA and RNA from one another, and telomerase contains an RNA that is used as template for building the ends of eukaryotic chromosomes.

Research on RNA has led to many important biological discoveries and numerous Nobel Prizes. Nucleic acids were discovered in 1868 by Friedrich Miescher, who called the material 'nuclein' since it was found in the nucleus. It was later discovered that prokaryotic cells, which do not have

a nucleus, also contain nucleic acids. The role of RNA in protein synthesis was suspected already in 1939. Severo Ochoa³⁹ won the 1959 Nobel Prize in Medicine (shared with Arthur Kornberg⁴⁰) after he discovered an enzyme that can synthesize RNA in the laboratory. However, the enzyme discovered by Ochoa (polynucleotide phosphorylase) was later shown to be responsible for RNA degradation, not RNA synthesis. In 1956 Alex Rich⁴¹ and David Davies hybridized two separate strands of RNA to form the first crystal of RNA whose structure could be determined by X-ray crystallography.

The sequence of the 77 nucleotides of a yeast tRNA was found by Robert W. Holley⁴² in 1965, winning Holley the 1968 Nobel Prize in Medicine (shared with Har Gobind Khorana⁴³ and Marshall Nirenberg⁴⁴).

During the early 1970s, retroviruses and reverse transcriptase were discovered, showing for the first time that enzymes could copy RNA into DNA (the opposite of the usual route for transmission of genetic information). For this work, David Baltimore⁴⁵, Renato Dulbecco⁴⁶ and Howard Temin⁴⁷ were awarded a Nobel Prize in 1975. In 1976, Walter Fiers⁴⁸ and his team determined the first complete nucleotide sequence of an RNA virus genome, that of bacteriophage MS2.

In 1977, introns and RNA splicing were discovered in both mammalian viruses and in cellular genes, resulting in a 1993 Nobel to Philip Sharp⁴⁹ and Richard Roberts⁵⁰. Catalytic RNA molecules (ribozymes) were discovered in the early 1980s, leading to a 1989 Nobel award to Thomas Cech⁵¹ and Sidney Altman⁵². In 1990, it was found in Petunia that introduced genes can silence similar genes of the plant's own, now known to be a result of RNA interference.

At about the same time, 22 nt long RNAs, now called microRNAs, were found to have a role in the development of *C. elegans*. Studies on RNA interference gleaned a Nobel Prize for Andrew Fire⁵³ and Craig Mello⁵⁴ in 2006, and another Nobel was awarded for studies on the transcription of RNA to Roger Kornberg⁵⁵ in the same year. The dis-

³⁹

⁴⁰

⁴¹https://en.wikipedia.org/wiki/Alexander_Rich

⁴²https://en.wikipedia.org/wiki/Robert_W._Holley

⁴³https://en.wikipedia.org/wiki/Har_Gobind_Khorana

⁴⁴https://en.wikipedia.org/wiki/Marshall_Warren_Nirenberg

⁴⁵https://en.wikipedia.org/wiki/David_Baltimore

⁴⁶https://en.wikipedia.org/wiki/Renato_Dulbecco

⁴⁷https://en.wikipedia.org/wiki/Howard_Martin_Temin

⁴⁸

⁴⁹https://en.wikipedia.org/wiki/Phillip_Allen_Sharp

⁵⁰https://en.wikipedia.org/wiki/Richard_J._Roberts

⁵¹https://en.wikipedia.org/wiki/Thomas_Cech

⁵²https://en.wikipedia.org/wiki/Sidney_Altman

⁵³https://en.wikipedia.org/wiki/Andrew_Fire

⁵⁴https://en.wikipedia.org/wiki/Craig_Mello

⁵⁵

covery of gene regulatory RNAs has led to attempts to develop drugs made of RNA, such as siRNA, to silence genes. Adding to the Nobel prizes awarded for research on RNA in 2009 it was awarded for the elucidation of the atomic structure of the ribosome to Venki Ramakrishnan⁵⁶, Tom Steitz⁵⁷, and Ada Yonath⁵⁸.

In 1967, Carl Woese⁵⁹ hypothesized that RNA might be catalytic and suggested that the earliest forms of life (self-replicating molecules) could have relied on RNA both to carry genetic information and to catalyze biochemical reactions—an RNA world.

11.3 Gene Expression

Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product. These products are often proteins, but in non-protein-coding genes such as transfer RNA (tRNA) or small nuclear RNA (snRNA) genes, the product is a functional RNA. Gene expression is summarized in the Central Dogma first formulated by Francis Crick in 1958, further developed in his 1970 article, and expanded by the subsequent discoveries of reverse transcription and RNA replication.

The process of gene expression is used by all known life—eukaryotes (including multicellular organisms), prokaryotes (bacteria and archaea), and utilized by viruses—to generate the macromolecular machinery for life.

In genetics, gene expression is the most fundamental level at which the genotype gives rise to the phenotype, i.e. observable trait. The genetic information stored in DNA represents the genotype, whereas the phenotype results from the “interpretation” of that information. Such phenotypes are often expressed by the synthesis of proteins that control the organism’s structure and development, or that act as enzymes catalyzing specific metabolic pathways.

All steps in the gene expression process may be modulated (regulated), including the transcription, RNA splicing, translation, and post-translational modification of a protein. Regulation of gene expression gives control over the timing, location, and amount of a given gene product (protein or ncRNA) present in a cell and can have a profound effect on the cellular structure and function. Regulation of gene expression is the basis for cellular differentiation, development, morphogenesis and the versatility and adaptability of any organism. Gene regulation may therefore serve as a substrate for evolutionary change.

⁵⁶https://en.wikipedia.org/wiki/Venki_Ramakrishnan

⁵⁷https://en.wikipedia.org/wiki/Thomas_A._Steitz

⁵⁸https://en.wikipedia.org/wiki/Ada_Yonath

⁵⁹https://en.wikipedia.org/wiki/Carl_Woese

11.3.1 Transcription And RNA Processing

Transcription⁶⁰ is the first of several steps of DNA based gene expression, in which a particular segment of DNA is copied into RNA (especially mRNA) by the enzyme RNA polymerase. During transcription, a DNA sequence is read by an RNA polymerase, which produces a complementary, antiparallel RNA strand called a primary transcript.

Transcription proceeds in the following general steps:

1. RNA polymerase, together with one or more general transcription factors, binds to promoter DNA.
2. RNA polymerase creates a transcription bubble, which separates the two strands of the DNA helix. This is done by breaking the hydrogen bonds between complementary DNA nucleotides.
3. RNA polymerase adds RNA nucleotides (which are complementary to the nucleotides of one DNA strand).
4. RNA sugar-phosphate backbone forms with assistance from RNA polymerase to form an RNA strand.
5. Hydrogen bonds of the RNA-DNA helix break, freeing the newly synthesized RNA strand.
6. If the cell has a nucleus, the RNA may be further processed. This may include polyadenylation, capping, and splicing.
7. The RNA may remain in the nucleus or exit to the cytoplasm through the nuclear pore complex.

The stretch of DNA transcribed into an RNA molecule is called a transcription unit. If the DNA encodes a protein, the transcription produces messenger RNA (mRNA); the mRNA, in turn, serves as a template for the protein’s synthesis through translation. Alternatively, the transcribed DNA may encode for non-coding RNA such as microRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), or enzymatic RNA molecules called ribozymes.

A DNA transcription unit encoding for a protein may contain both a coding sequence, which will be translated into the protein, and regulatory sequences, which direct and regulate the synthesis of that protein. The regulatory sequence before (“upstream” from) the coding sequence is called the five prime untranslated region (5’UTR); the sequence after (“downstream” from) the coding sequence is called the three prime untranslated region (3’UTR).

As opposed to DNA replication, transcription results in an RNA complement that includes the nucleotide uracil (U) in all instances where thymine (T) would have occurred in a DNA complement.

Only one of the two DNA strands serve as a template for transcription. The antisense strand of DNA is read by RNA

⁶⁰[https://en.wikipedia.org/wiki/Transcription_\(biology\)](https://en.wikipedia.org/wiki/Transcription_(biology))

polymerase from the 3' end to the 5' end during transcription ($3' \rightarrow 5'$). The complementary RNA is created in the opposite direction, in the $5' \rightarrow 3'$ direction, matching the sequence of the sense strand with the exception of switching uracil for thymine. This directionality is because RNA polymerase can only add nucleotides to the 3' end of the growing mRNA chain. This use of only the $3' \rightarrow 5'$ DNA strand eliminates the need for the Okazaki fragments that are seen in DNA replication. This also removes the need for an RNA primer to initiate RNA synthesis, as is the case in DNA replication.

The non-template (sense) strand of DNA is called the coding strand, because its sequence is the same as the newly created RNA transcript (except for the substitution of uracil for thymine). This is the strand that is used by convention when presenting a DNA sequence.

Transcription has some proofreading mechanisms, but they are fewer and less effective than the controls for copying DNA. As a result, transcription has a lower copying fidelity than DNA replication.

Transcription is divided into initiation, promoter escape, elongation, and termination.

11.3.2 Initiation

Transcription begins with the binding of RNA polymerase, together with one or more general transcription factors, to a specific DNA sequence referred to as a "promoter" to form an RNA polymerase-promoter "closed complex". In the "closed complex" the promoter DNA is still fully double-stranded.

RNA polymerase, assisted by one or more general transcription factors, then unwinds approximately 14 base pairs of DNA to form an RNA polymerase-promoter "open complex". In the "open complex" the promoter DNA is partly unwound and single-stranded. The exposed, single-stranded DNA is referred to as the "transcription bubble."

RNA polymerase, assisted by one or more general transcription factors, then selects a transcription start site in the transcription bubble, binds to an initiating NTP and an extending NTP (or a short RNA primer and an extending NTP) complementary to the transcription start site sequence, and catalyzes bond formation to yield an initial RNA product.

In bacteria, RNA polymerase holoenzyme consists of five subunits: 2 α subunits, 1 β subunit, 1 β' subunit, and 1 ω subunit. In bacteria, there is one general RNA transcription factor known as a sigma factor. RNA polymerase core enzyme binds to the bacterial general transcription (sigma) factor to form RNA polymerase holoenzyme and then binds

to a promoter. (RNA polymerase is called a holoenzyme when sigma subunit is attached to the core enzyme which is consist of 2 α subunits, 1 β subunit, 1 β' subunit only).

In archaea and eukaryotes, RNA polymerase contains subunits homologous to each of the five RNA polymerase subunits in bacteria and also contains additional subunits. In archaea and eukaryotes, the functions of the bacterial general transcription factor sigma are performed by multiple general transcription factors that work together. In archaea, there are three general transcription factors: TBP, TFB, and TFE. In eukaryotes, in RNA polymerase II-dependent transcription, there are six general transcription factors: TFIIA, TFIIB (an ortholog of archaeal TFB), TFIID (a multisubunit factor in which the key subunit, TBP, is an ortholog of archaeal TBP), TFIIE (an ortholog of archaeal TFE), TFIIF, and TFIIH. The TFIID is the first component to bind to DNA due to binding of TBP, while TFIIH is the last component to be recruited. In archaea and eukaryotes, the RNA polymerase-promoter closed complex is usually referred to as the "preinitiation complex."

Transcription initiation is regulated by additional proteins, known as activators and repressors, and, in some cases, associated coactivators or corepressors, which modulate formation and function of the transcription initiation complex.

After the first bond is synthesized, the RNA polymerase must escape the promoter. During this time there is a tendency to release the RNA transcript and produce truncated transcripts. This is called abortive initiation, and is common for both eukaryotes and prokaryotes. Abortive initiation continues to occur until an RNA product of a threshold length of approximately 10 nucleotides is synthesized, at which point promoter escape occurs and a transcription elongation complex is formed.

Mechanistically, promoter escape occurs through DNA scrunching, providing the energy needed to break interactions between RNA polymerase holoenzyme and the promoter.

In eukaryotes, at an RNA polymerase II-dependent promoter, upon promoter clearance, TFIIH phosphorylates serine 5 on the carboxy terminal domain of RNA polymerase II, leading to the recruitment of capping enzyme (CE). The exact mechanism of how CE induces promoter clearance in eukaryotes is not yet known.

11.3.3 Elongation

One strand of the DNA, the template strand (or noncoding strand), is used as a template for RNA synthesis. As transcription proceeds, RNA polymerase traverses the template strand and uses base pairing complementarity

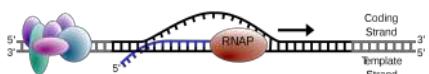


Figure 11.10: RNA polymerase (RNAP) at work. Note the coding and template strands. The resulting RNA is a synthesized from the template strand and identical in sequence to the coding strand.⁶¹

with the DNA template to create an RNA copy (which elongates during the traversal). Although RNA polymerase traverses the template strand from 3' → 5', the coding (non-template) strand and newly formed RNA can also be used as reference points, so transcription can be described as occurring 5' → 3'. This produces an RNA molecule from 5' → 3', an exact copy of the coding strand (except that thymines are replaced with uracils, and the nucleotides are composed of a ribose (5-carbon) sugar where DNA has deoxyribose (one fewer oxygen atom) in its sugar-phosphate backbone).

mRNA transcription can involve multiple RNA polymerases on a single DNA template and multiple rounds of transcription (amplification of particular mRNA), so many mRNA molecules can be rapidly produced from a single copy of a gene. The characteristic elongation rates in prokaryotes and eukaryotes are about 10–100 nts/sec. In eukaryotes, however, nucleosomes act as major barriers to transcribing polymerases during transcription elongation. In these organisms, the pausing induced by nucleosomes can be regulated by transcription elongation factors such as TFIIS.

Elongation also involves a proofreading mechanism that can replace incorrectly incorporated bases. In eukaryotes, this may correspond with short pauses during transcription that allow appropriate RNA editing factors to bind. These pauses may be intrinsic to the RNA polymerase or due to chromatin structure.

11.3.4 Termination

Bacteria use two different strategies for transcription termination - Rho-independent termination and Rho-dependent termination. In Rho-independent transcription termination, RNA transcription stops when the newly synthesized RNA molecule forms a G-C-rich hairpin loop followed by a run of Us. When the hairpin forms, the mechanical stress breaks the weak rU-dA bonds, now filling the DNA-RNA hybrid. This pulls the poly-U transcript out of the active site of the RNA polymerase, terminating transcription. In the "Rho-dependent" type of termination, a protein factor called "Rho" destabilizes the interaction between the template and the mRNA, thus releasing the newl synthesized mRNA from the elongation

complex.

Transcription termination in eukaryotes is less well understood than in bacteria, but involves cleavage of the new transcript followed by template-independent addition of adenines at its new 3' end, in a process called polyadenylation.

11.3.5 Inhibitors of Transcription

Transcription inhibitors can be used as antibiotics against, for example, pathogenic bacteria (antibacterials) and fungi (antifungals). An example of such an antibacterial is rifampicin, which inhibits bacterial transcription of DNA into mRNA by inhibiting DNA-dependent RNA polymerase by binding its beta-subunit, while 8-hydroxyquinoline is an antifungal transcription inhibitor. The effects of histone methylation may also work to inhibit the action of transcription.

In vertebrates, the majority of gene promoters contain a CpG island with numerous CpG sites. When many of a gene's promoter CpG sites are methylated the gene becomes inhibited (silenced).

11.3.6 Transcription Factors

Active transcription units are clustered in the nucleus, in discrete sites called transcription factories or euchromatin. Such sites can be visualized by allowing engaged polymerases to extend their transcripts in tagged precursors (Br-UTP or Br-U) and immuno-labeling the tagged nascent RNA. Transcription factories can also be localized using fluorescence in situ hybridization or marked by antibodies directed against polymerases. There are ~10,000 factories in the nucleoplasm of a HeLa cell, among which are ~8,000 polymerase II factories and ~2,000 polymerase III factories. Each polymerase II factory contains ~8 polymerases. As most active transcription units are associated with only one polymerase, each factory usually contains ~8 different transcription units. These units might be associated through promoters and/or enhancers, with loops forming a "cloud" around the factor.

A molecule that allows the genetic material to be realized as a protein was first hypothesized by François Jacob and Jacques Monod. Severo Ochoa won a Nobel Prize in Physiology or Medicine in 1959 for developing a process for synthesizing RNA in vitro with polynucleotide phosphorylase, which was useful for cracking the genetic code. RNA synthesis by RNA polymerase was established in vitro by several laboratories by 1965.

Roger D. Kornberg won the 2006 Nobel Prize in Chemistry "for his studies of the molecular basis of eukaryotic transcription".

11.3.7 RNA Processing

Post-transcriptional modification⁶² or co-transcriptional modification is a set of biological processes common to most eukaryotic cells by which an RNA primary transcript is chemically altered following transcription from a gene to produce a mature, functional RNA molecule that can then leave the nucleus and perform any of a variety of different functions in the cell. There are many types of post-transcriptional modifications achieved through a diverse class of molecular mechanisms.

Perhaps the most notable example is the conversion of precursor messenger RNA transcripts into mature messenger RNA that is subsequently capable of being translated into protein. This process includes three major steps that significantly modify the chemical structure of the RNA molecule: the addition of a 5' cap, the addition of a 3' polyadenylated tail, and RNA splicing. Such processing is vital for the correct translation of eukaryotic genomes because the initial precursor mRNA produced by transcription often contains both exons (coding sequences) and introns (non-coding sequences); splicing removes the introns and links the exons directly, while the cap and tail facilitate the transport of the mRNA to a ribosome and protect it from molecular degradation.

Post-transcriptional modifications may also occur during the processing of other transcripts which ultimately become transfer RNA, ribosomal RNA, or any of the other types of RNA used by the cell.

11.3.8 mRNA processing

The pre-mRNA molecule undergoes three main modifications. These modifications are 5' capping, 3' polyadenylation, and RNA splicing, which occur in the cell nucleus before the RNA is translated.

11.3.9 5' processing

Capping of the pre-mRNA involves the addition of 7-methylguanosine (m7G) to the 5' end. To achieve this, the terminal 5' phosphate requires removal, which is done with the aid of a phosphatase enzyme. The enzyme guanosyl transferase then catalyses the reaction, which produces the diphosphate 5' end. The diphosphate 5' end then attacks the alpha phosphorus atom of a GTP molecule in order to add the guanine residue in a 5'5' triphosphate link. The enzyme (guanine-N7)-methyltransferase ("cap MTase") transfers a methyl group from S-adenosyl methionine to the guanine ring. This type of cap, with just the (m7G) in position is called a cap 0 structure. The ribose of the adjacent nucleotide may also be methylated to give

a cap 1. Methylation of nucleotides downstream of the RNA molecule produce cap 2, cap 3 structures and so on. In these cases the methyl groups are added to the 2' OH groups of the ribose sugar. The cap protects the 5' end of the primary RNA transcript from attack by ribonucleases that have specificity to the 3'5' phosphodiester bonds.

11.3.10 3' processing

The pre-mRNA processing at the 3' end of the RNA molecule involves cleavage of its 3' end and then the addition of about 250 adenine residues to form a poly(A) tail. The cleavage and adenylation reactions occur primarily if a polyadenylation signal sequence (5'- AAUAAA-3') is located near the 3' end of the pre-mRNA molecule, which is followed by another sequence, which is usually (5'-CA-3') and is the site of cleavage. A GU-rich sequence is also usually present further downstream on the pre-mRNA molecule. More recently, it has been demonstrated that alternate signal sequences such as UGUA upstream off the cleavage site can also direct cleavage and polyadenylation in the absence of the AAUAAA signal. It is important to understand that these two signals are not mutually independent and often coexist. After the synthesis of the sequence elements, several multi-subunit proteins are transferred to the RNA molecule. The transfer of these sequence specific binding proteins cleavage and polyadenylation specificity factor (CPSF), Cleavage Factor I (CF I) and cleavage stimulation factor (CStF) occurs from RNA Polymerase II. The three factors bind to the sequence elements. The AAUAAA signal is directly bound by CPSF. For UGUA dependent processing sites, binding of the multi protein complex is done by Cleavage Factor I (CF I). The resultant protein complex formed contains additional cleavage factors and the enzyme Polyadenylate Polymerase (PAP). This complex cleaves the RNA between the polyadenylation sequence and the GU-rich sequence at the cleavage site marked by the (5'-CA-3') sequences. Poly(A) polymerase then adds about 200 adenine units to the new 3' end of the RNA molecule using ATP as a precursor. As the poly(A) tail is synthesised, it binds multiple copies of poly(A) binding protein, which protects the 3'end from ribonuclease digestion.

11.3.11 Intron Splicing

RNA splicing is the process by which introns, regions of RNA that do not code for proteins, are removed from the pre-mRNA and the remaining exons connected to re-form a single continuous molecule. Exons are sections of mRNA which become "expressed" or translated into a protein. They are the coding portions of a mRNA molecule. Although most RNA splicing occurs after the complete synthesis and end-capping of the pre-mRNA, transcripts

⁶²https://en.wikipedia.org/wiki/Post-transcriptional_modification

with many exons can be spliced co-transcriptionally. The splicing reaction is catalyzed by a large protein complex called the spliceosome assembled from proteins and small nuclear RNA molecules that recognize splice sites in the pre-mRNA sequence. Many pre-mRNAs, including those encoding antibodies, can be spliced in multiple ways to produce different mature mRNAs that encode different protein sequences. This process is known as alternative splicing, and allows production of a large variety of proteins from a limited amount of DNA.

11.3.12 Reverse Transcription

Some viruses (such as HIV, the cause of AIDS), have the ability to transcribe RNA into DNA. HIV has an RNA genome that is reverse transcribed into DNA. The resulting DNA can be merged with the DNA genome of the host cell. The main enzyme responsible for synthesis of DNA from an RNA template is called reverse transcriptase.

In the case of HIV, reverse transcriptase is responsible for synthesizing a complementary DNA strand (cDNA) to the viral RNA genome. The enzyme ribonuclease H then digests the RNA strand, and reverse transcriptase synthesizes a complementary strand of DNA to form a double helix DNA structure ("cDNA"). The cDNA is integrated into the host cell's genome by the enzyme integrase, which causes the host cell to generate viral proteins that reassemble into new viral particles. In HIV, subsequent to this, the host cell undergoes programmed cell death, or apoptosis of T cells. However, in other retroviruses, the host cell remains intact as the virus buds out of the cell.

Some eukaryotic cells contain an enzyme with reverse transcription activity called telomerase. Telomerase is a reverse transcriptase that lengthens the ends of linear chromosomes. Telomerase carries an RNA template from which it synthesizes a repeating sequence of DNA, or "junk" DNA. This repeated sequence of DNA is called a telomere and can be thought of as a "cap" for a chromosome. It is important because every time a linear chromosome is duplicated, it is shortened. With this "junk" DNA or "cap" at the ends of chromosomes, the shortening eliminates some of the non-essential, repeated sequence rather than the protein-encoding DNA sequence, that is farther away from the chromosome end.

Telomerase is often activated in cancer cells to enable cancer cells to duplicate their genomes indefinitely without losing important protein-coding DNA sequence. Activation of telomerase could be part of the process that allows cancer cells to become immortal. The immortalizing factor of cancer via telomere lengthening due to telomerase has been proven to occur in 90% of all carcinogenic tumors *in vivo* with the remaining 10% using an alternative telomere

maintenance route called ALT or Alternative Lengthening of Telomeres.

11.3.13 RNA Export

In eukaryotes most mature RNA must be exported to the cytoplasm from the nucleus. While some RNAs function in the nucleus, many RNAs are transported through the nuclear pores and into the cytosol. Export of RNAs requires association with specific proteins known as exportins. Specific exportin molecules are responsible for the export of a given RNA type. mRNA transport also requires the correct association with Exon Junction Complex (EJC), which ensures that correct processing of the mRNA is completed before export. In some cases RNAs are additionally transported to a specific part of the cytoplasm, such as a synapse; they are then towed by motor proteins that bind through linker proteins to specific sequences (called "zipcodes") on the RNA.

11.3.14 Translation And The Genetic Code

In molecular biology and genetics, translation⁶³ is the process in which ribosomes in the cytoplasm or endoplasmic reticulum (ER) synthesize proteins after the process of transcription of DNA to RNA in the cell's nucleus. The entire process is called gene expression.

11.3.15 Ribosomes

Ribosomes⁶⁴ are complex macromolecular machines, found within all living cells, that serve as the site of biological protein synthesis. Ribosomes link amino acids together in the order specified by messenger RNA (RNA) molecules. Ribosomes consist of two major components: the small ribosomal subunits, which read the mRNA, and the large subunits, which join amino acids to form a polypeptide chain. Each subunit consists of one or more ribosomal RNA (rRNA) molecules and a variety of ribosomal proteins. The ribosomes and associated molecules are also known as the translational apparatus.

The sequence of DNA, which encodes the sequence of the amino acids in a protein, is copied into a messenger RNA chain. It may be copied many times into RNA chains. Ribosomes can bind to a messenger RNA chain and use its sequence for determining the correct sequence of amino acids for generating a given protein. Amino acids are selected and collected and carried to the ribosome by transfer RNA (tRNA) molecules, which enter one part of the ribosome and bind to the messenger RNA chain. It is during this binding that the correct translation of nucleic acid

⁶³[https://en.wikipedia.org/wiki/Translation_\(biology\)](https://en.wikipedia.org/wiki/Translation_(biology))

⁶⁴<https://en.wikipedia.org/wiki/Ribosome>

sequence to amino acid sequence occurs. For each coding triplet in the messenger RNA there is a distinct transfer RNA that matches and which carries the correct amino acid for that coding triplet. The attached amino acids are then linked together by another part of the ribosome. Once the protein is produced, it can then fold to produce a specific functional three-dimensional structure although during synthesis some proteins start folding into their correct form.

A ribosome is made from complexes of RNAs and proteins and is therefore a ribonucleoprotein. Each ribosome is divided into two subunits:

1. a smaller subunit which binds to a larger subunit and the mRNA pattern, and
2. a larger subunit which binds to the tRNA, the amino acids, and the smaller subunit.

When a ribosome finishes reading an mRNA molecule, these two subunits split apart. Ribosomes are ribozymes, because the catalytic peptidyl transferase activity that links amino acids together is performed by the ribosomal RNA. Ribosomes are often associated with the intracellular membranes that make up the rough endoplasmic reticulum.

Ribosomes from bacteria, archaea and eukaryotes in the three-domain system, resemble each other to a remarkable degree, evidence of a common origin. They differ in their size, sequence, structure, and the ratio of protein to RNA. The differences in structure allow some antibiotics to kill bacteria by inhibiting their ribosomes, while leaving human ribosomes unaffected. In bacteria and archaea, more than one ribosome may move along a single mRNA chain at one time, each "reading" its sequence and producing a corresponding protein molecule.

The mitochondrial ribosomes of eukaryotic cells, are produced from mitochondrial genes, and functionally resemble many features of those in bacteria, reflecting the likely evolutionary origin of mitochondria.

Ribosomes were first observed in the mid-1950s by Romanian-American cell biologist George Emil Palade⁶⁵, using an electron microscope, as dense particles or granules. The term "ribosome" was proposed by scientist Richard B. Roberts in the end of 1950s.

Albert Claude⁶⁶, Christian de Duve⁶⁷, and George Emil Palade were jointly awarded the Nobel Prize in Physiology or Medicine, in 1974, for the discovery of the ribosome. The Nobel Prize in Chemistry 2009 was awarded to Venkatraman Ramakrishnan⁶⁸, Thomas A. Steitz⁶⁹ and

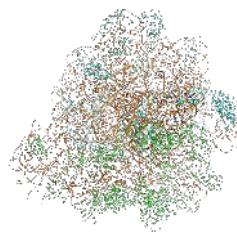


Figure 11.11: Crystal structure of the bacterial 70S ribosome of the bacterium *Thermus thermophilus*. The 30S (small) ribosomal subunit proteins are colored in green, the 50S (large) subunit proteins are colored in blue, the ribosomal RNA is colored orange. The 30S subunits contain 3 tRNA molecules (based on atomic coordinates of PDB 1JGQ⁷¹ and PDB 1GIY⁷² rendered with open source molecular visualization tool PyMol.)

Ada E. Yonath⁷⁰ for determining the detailed structure and mechanism of the ribosome.

11.3.16 Bacterial Ribosomes

Prokaryotic ribosomes are around 20 nm (200 Å) in diameter and are composed of 65% rRNA and 35% ribosomal proteins. Eukaryotic ribosomes are between 25 and 30 nm (250–300 Å) in diameter with an rRNA-to-protein ratio that is close to 1. Crystallographic work has shown that there are no ribosomal proteins close to the reaction site for polypeptide synthesis. This suggests that the protein components of ribosomes do not directly participate in peptide bond formation catalysis, but rather that these proteins act as a scaffold that may enhance the ability of rRNA to synthesize protein

The unit of measurement used to describe the ribosomal subunits and the rRNA fragments is the Svedberg unit, a measure of the rate of sedimentation in centrifugation rather than size. This accounts for why fragment names do not add up: for example, bacterial 70S ribosomes are made of 50S and 30S subunits.

Bacteria have 70S ribosomes, each consisting of a small (30S) and a large (50S) subunit. *Escherichia coli*, for example, has a 16S RNA subunit (consisting of 1540 nucleotides) that is bound to 21 proteins. The large subunit is composed of a 5S RNA subunit (120 nucleotides), a 23S RNA subunit (2900 nucleotides) and 31 proteins.

11.3.17 Eukaryotic Ribosomes

Eukaryotes have 80S ribosomes located in their cytosol, each consisting of a small (40S) and large (60S) subunit. Their 40S subunit has an 18S RNA (1900 nucleotides) and

⁶⁵https://en.wikipedia.org/wiki/George_Emil_Palade

⁶⁶https://en.wikipedia.org/wiki/Albert_Claude

⁶⁷https://en.wikipedia.org/wiki/Christian_de_Duve

⁶⁸https://en.wikipedia.org/wiki/Venki_Ramakrishnan

⁶⁹https://en.wikipedia.org/wiki/Thomas_A._Steitz

⁷⁰https://en.wikipedia.org/wiki/Ada_Yonath

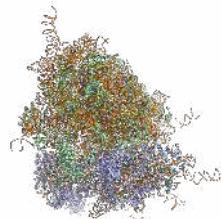


Figure 11.12: Crystal structure of the human 80S ribosome (based on atomic coordinates of PDB 4V6X⁷³) rendered with open source molecular visualization tool PyMol). The 40S (small) ribosomal subunit proteins are shown in lightblue, the 60S (large) subunit proteins in palegreen, the ribosomal RNA in orange.

33 proteins. The large subunit is composed of a 5S RNA (120 nucleotides), 28S RNA (4700 nucleotides), a 5.8S RNA (160 nucleotides) subunits and 46 proteins.

The differences between the bacterial and eukaryotic ribosomes are exploited by pharmaceutical chemists to create antibiotics that can kill bacteria without harming eukaryotic cells. Due to the differences in their structures, the bacterial 70S ribosomes are vulnerable to these antibiotics while the eukaryotic 80S ribosomes are not.

The various ribosomes share a core structure, which is quite similar despite the large differences in size. Much of the RNA is highly organized into various tertiary structural motifs, for example pseudoknots that exhibit coaxial stacking. The extra RNA in the larger ribosomes is in several long continuous insertions, such that they form loops out of the core structure without disrupting or changing it. All of the catalytic activity of the ribosome is carried out by the RNA; the proteins reside on the surface and seem to stabilize the structure.

Aminoacyl tRNA synthetases (enzymes) catalyze the bonding between specific tRNAs and the amino acids that their anticodon sequences call for. The product of this reaction is an aminoacyl-tRNA. In prokaryotes, this aminoacyl-tRNA is carried to the ribosome by EF-Tu, where mRNA codons are matched through complementary base pairing to specific tRNA anticodons.

The ribosome has three sites for tRNA to bind. They are the aminoacyl site (abbreviated A), the peptidyl site (abbreviated P) and the exit site (abbreviated E). With respect to the mRNA, the three sites are oriented 5' to 3' E-P-A, because ribosomes move toward the 3' end of mRNA. The A-site binds the incoming tRNA with the complementary codon on the mRNA. The P-site holds the tRNA with the growing polypeptide chain. The E-site holds the tRNA without its amino acid, and the tRNA is then released. When an aminoacyl-tRNA initially binds to its corresponding codon

on the mRNA, it is in the A site. Then, a peptide bond forms between the amino acid of the tRNA in the A site and the amino acid of the charged tRNA in the P site. The growing polypeptide chain is transferred to the tRNA in the A site. Translocation occurs, moving the tRNA in the P site, now without an amino acid, to the E site; the tRNA that was in the A site, now charged with the polypeptide chain, is moved to the P site. The tRNA in the E site leaves and another aminoacyl-tRNA enters the A site to repeat the process.

After the new amino acid is added to the chain, and after the mRNA is released out of the nucleus and into the ribosome's core, the energy provided by the hydrolysis of an ATP bound to the translocase EF-G (in prokaryotes) and eEF-2 (in eukaryotes) moves the ribosome down one codon towards the 3' end. The energy required for translation of proteins is significant. For a protein containing n amino acids, the number of high-energy phosphate bonds required to translate it is $4n+1$. The rate of translation varies; it is significantly higher in prokaryotic cells (up to 17–21 amino acid residues per second) than in eukaryotic cells (up to 6–9 amino acid residues per second).

Even though the ribosomes are usually considered accurate, processive machines, the translation process is subject to errors that can lead either to the synthesis of erroneous proteins or to the premature abandonment of translation. The rate of error in synthesizing proteins has been estimated to be between $1/10^5$ and $1/10^3$ misincorporated amino acids, depending on the experimental conditions. The rate of premature translation abandonment, instead, has been estimated to be of the order of magnitude of 10^{-4} events per translated codon. The correct amino acid is covalently bonded to the correct transfer RNA (tRNA) by amino acyl transferases. The amino acid is joined by its carboxyl group to the 3' OH of the tRNA by an ester bond. When the tRNA has an amino acid linked to it, the tRNA is termed "charged". Initiation involves the small subunit of the ribosome binding to the 5' end of mRNA with the help of initiation factors (IF). In prokaryotes, initiation of protein synthesis involves the recognition of a purine-rich initiation sequence on the mRNA called the Shine-Dalgarno sequence. The Shine-Dalgarno sequence binds to a complementary pyrimidine-rich sequence on the 3' end of the 16S rRNA part of the 30S ribosomal subunit. The binding of these complementary sequences ensures that the 30S ribosomal subunit is bound to the mRNA and is aligned such that the initiation codon is placed in the 30S portion of the P-site. Once the mRNA and 30S subunit are properly bound, an initiation factor brings the initiator tRNA-amino acid complex, f-Met-tRNA, to the 30S P site. The initiation phase is completed once a 50S subunit joins the 30S subunit, forming an active 70S ribosome. Termination of

the polypeptide occurs when the A site of the ribosome is occupied by a stop codon (UAA, UAG, or UGA) on the mRNA. mRNA usually cannot recognize or bind to stop codons. Instead, the stop codon induces the binding of a release factor protein (RF1 & RF2) that prompts the disassembly of the entire ribosome/mRNA complex by the hydrolysis of the polypeptide chain from the peptidyl transferase center of the ribosome. Drugs or special sequence motifs on the mRNA can change the ribosomal structure so that near-cognate tRNAs are bound to the stop codon instead of the release factors. In such cases of 'translational readthrough', translation continues until the ribosome encounters the next stop codon.

The process of translation is highly regulated in prokaryotic and eukaryotic organisms. Regulation of translation can impact the global rate of protein synthesis which is closely coupled to the metabolic and proliferative state of a cell. In addition, recent work has revealed that genetic differences and their subsequent expression as mRNAs can also impact translation rate in an RNA-specific manner.

11.3.18 Translation

In translation, messenger RNA (mRNA) is decoded in the ribosome decoding center to produce a specific amino acid chain, or polypeptide. The polypeptide later folds into an active protein and performs its functions in the cell. The ribosome facilitates decoding by inducing the binding of complementary tRNA anticodon sequences to mRNA codons. The tRNAs carry specific amino acids that are chained together into a polypeptide as the mRNA passes through and is read by the ribosome.

Translation proceeds in three phases:

- Initiation:** The ribosome assembles around the target mRNA. The first tRNA is attached at the start codon.
- Elongation:** The tRNA transfers an amino acid to the tRNA corresponding to the next codon. The ribosome then moves (translocates) to the next mRNA codon to continue the process, creating an amino acid chain.
- Termination:** When a peptidyl tRNA encounters a stop codon, the ribosome detaches.

In prokaryotes, translation occurs in the cytosol, where the medium and small subunits of the ribosome bind to the tRNA. In eukaryotes, translation occurs in the cytosol or across the membrane of the endoplasmic reticulum in a process called co-translational translocation. In co-translational translocation, the entire ribosome/mRNA complex binds to the outer membrane of the rough endoplasmic reticulum (ER) and the new protein is synthesized

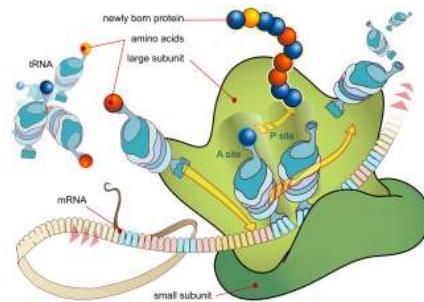


Figure 11.13: Diagram showing the translation of mRNA and the synthesis of proteins by a ribosome.⁷⁴

and released into the ER.

Many types of transcribed RNA, such as transfer RNA, ribosomal RNA, and small nuclear RNA, do not undergo translation into proteins.

A number of antibiotics act by inhibiting translation. These include clindamycin, anisomycin, cycloheximide, chloramphenicol, tetracycline, streptomycin, erythromycin, and puromycin. Prokaryotic ribosomes have a different structure from that of eukaryotic ribosomes, and thus antibiotics can specifically target bacterial infections without any harm to a eukaryotic host's cells.

In 1954, Zamecnik and Hoagland discovered tRNA. In 1955, George E. Palade discovered ribosomes.

11.3.19 Eukaryotic Translation

11.3.20 Initiation

Initiation of translation usually involves the interaction of certain key proteins, the initiation factors, with a special tag bound to the 5'-end of an mRNA molecule, the 5' cap, as well as with the 5' UTR. These proteins bind the small (40S) ribosomal subunit and hold the mRNA in place. eIF3 is associated with the 40S ribosomal subunit and plays a role in keeping the large (60S) ribosomal subunit from prematurely binding. eIF3 also interacts with the eIF4F complex, which consists of three other initiation factors: eIF4A, eIF4E, and eIF4G. eIF4G is a scaffolding protein that directly associates with both eIF3 and the other two components. eIF4E is the cap-binding protein. Binding of the cap by eIF4E is often considered the rate-limiting step of cap-dependent initiation, and the concentration of eIF4E is a regulatory nexus of translational control. Certain viruses cleave a portion of eIF4G that binds eIF4E, thus preventing cap-dependent translation to hijack the host machinery in favor of the viral (cap-independent) messages. eIF4A is an ATP-dependent RNA helicase that aids the ribosome by resolving certain secondary structures formed along the mRNA transcript. The poly(A)-binding protein (PABP) also

associates with the eIF4F complex via eIF4G, and binds the poly-A tail of most eukaryotic mRNA molecules. This 43S preinitiation complex (43S PIC) accompanied by the protein factors moves along the mRNA chain toward its 3'-end, in a process known as 'scanning', to reach the start codon (typically AUG). In eukaryotes and archaea, the amino acid encoded by the start codon is methionine. The Met-charged initiator tRNA (Met-tRNAA_iMet) is brought to the P-site of the small ribosomal subunit by eukaryotic initiation factor 2 (eIF2). It hydrolyzes GTP, and signals for the dissociation of several factors from the small ribosomal subunit, eventually leading to the association of the large subunit (or the 60S subunit). The complete ribosome (80S) then commences translation elongation. Regulation of protein synthesis is partly influenced by phosphorylation of eIF2 (via the α subunit), which is a part of the eIF2-GTP-Met-tRNAA_iMet ternary complex (eIF2-TC). When large numbers of eIF2 are phosphorylated, protein synthesis is inhibited. This occurs under amino acid starvation or after viral infection. However, a small fraction of this initiation factor is naturally phosphorylated. Another regulator is 4EBP, which binds to the initiation factor eIF4E and inhibits its interactions with eIF4G, thus preventing cap-dependent initiation. To oppose the effects of 4EBP, growth factors phosphorylate 4EBP, reducing its affinity for eIF4E and permitting protein synthesis. While protein synthesis is globally regulated by modulating the expression of key initiation factors as well as the number of ribosomes, individual mRNAs can have different translation rates due to the presence of regulatory sequence elements. This has been shown to be important in a variety of settings including yeast meiosis and ethylene response in plants. In addition, recent work in yeast and humans suggest that evolutionary divergence in cis-regulatory sequences can impact translation regulation. Additionally, RNA helicases such as DHX29 and Ded1/DDX3 may participate in the process of translation initiation, especially for mRNAs with structured 5'UTRs.

Cap-independent initiation The best-studied example of cap-independent translation initiation in eukaryotes is that by the Internal ribosome entry site (IRES). What differentiates cap-independent translation from cap-dependent translation is that cap-independent translation does not require the 5' cap to initiate scanning from the 5' end of the mRNA until the start codon. The ribosome can be trafficked to the start site by direct binding, initiation factors, and/or ITRAFs (IRES trans-acting factors) bypassing the need to scan the entire 5' UTR. This method of translation has been found important in conditions that require the translation of specific mRNAs during cellular stress, when overall translation is reduced. Examples include factors responding to apoptosis and stress-induced responses.

11.3.21 Elongation

Elongation depends on eukaryotic elongation factors. At the end of the initiation step, the mRNA is positioned so that the next codon can be translated during the elongation stage of protein synthesis. The initiator tRNA occupies the P site in the ribosome, and the A site is ready to receive an aminoacyl-tRNA. During chain elongation, each additional amino acid is added to the nascent polypeptide chain in a three-step microcycle. The steps in this microcycle are (1) positioning the correct aminoacyl-tRNA in the A site of the ribosome, (2) forming the peptide bond and (3) shifting the mRNA by one codon relative to the ribosome. Unlike bacteria, in which translation initiation occurs as soon as the 5' end of an mRNA is synthesized, in eukaryotes such tight coupling between transcription and translation is not possible because transcription and translation are carried out in separate compartments of the cell (the nucleus and cytoplasm). Eukaryotic mRNA precursors must be processed in the nucleus (e.g., capping, polyadenylation, splicing) before they are exported to the cytoplasm for translation.

11.3.22 Termination

Termination of elongation depends on eukaryotic release factors. The process is similar to that of prokaryotic termination, but unlike prokaryotic termination, there is a universal release factor, eRF1, that recognizes all three stop codons. Upon termination, the ribosome is disassembled and the completed polypeptide is released. eRF3 is a ribosome-dependent GTPase that helps eRF1 release the completed polypeptide.

11.3.23 The Genetic Code

Whereas other aspects such as the 3D structure, called tertiary structure, of protein can only be predicted using sophisticated algorithms, the amino acid sequence, called primary structure, can be determined solely from the nucleic acid sequence with the aid of a translation table.

Table 11.2: The genetic code: RNA codons.

	U	C	A	G	
U	UUU Phenylalanine (Phe)	UCU Serine (Ser)	UAU Tyrosine (Tyr)	UGU Cysteine (Cys)	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leucine (Leu)	UCA Ser	UAA STOP	UGA STOP	A
	UUG Leu	UCG Ser	UAG STOP	UGG Tryptophan (Trp)	G
C	CUU Leucine (Leu)	CCU Proline (Pro)	CAU Histidine (His)	CGU Arginine (Arg)	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Glutamine (Gln)	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Isoleucine (Ile)	ACU Threonine (Thr)	AAU Asparagine (Asn)	AGU Serine (Ser)	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lysine (Lys)	AGA Arginine (Arg)	A
	AUG Methionine (Met) or START	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Valine Val	GCU Alanine (Ala)	GAU Aspartic acid (Asp)	GGU Glycine (Gly)	U
	GUC (Val)	GCC Ala	GAC Asp	GGC Gly	C
	GUU Val	GCA Ala	GAA Glutamic acid (Glu)	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

There are many computer programs capable of translating a DNA/RNA sequence into a protein sequence. Normally this is performed using the Standard Genetic Code (Table 11.2), however, few programs can handle all the "special" cases, such as the use of the alternative initiation codons. For instance, the rare alternative start codon CTG codes for Methionine when used as a start codon, and for Leucine in all other positions.

11.3.24 Protein Folding

Each protein exists as an unfolded polypeptide or random coil when translated from a sequence of mRNA into a linear chain of amino acids. This polypeptide lacks any developed three-dimensional structure (the left hand side of the neighboring figure). The polypeptide then folds into its characteristic and functional three-dimensional structure from a random coil. Amino acids interact with each other to produce a well-defined three-dimensional structure, the folded protein (the right hand side of the figure) known as the native state. The resulting three-dimensional structure is determined by the amino acid sequence (Anfinsen's dogma).

The correct three-dimensional structure is essential to function, although some parts of functional proteins may remain unfolded. Failure to fold into the intended shape usually produces inactive proteins with different properties including toxic prions. Several neurodegenerative and other diseases are believed to result from the accumulation of misfolded proteins. Many allergies are caused by the folding of the proteins, for the immune system does not produce antibodies for certain protein structures.

Enzymes called chaperones assist the newly formed protein to attain (fold into) the 3-dimensional structure it needs to function. Similarly, RNA chaperones help RNAs attain their functional shapes. Assisting protein folding is one of the main roles of the endoplasmic reticulum in eukaryotes.

11.3.25 Protein Translocation

Secretory proteins of eukaryotes or prokaryotes must be translocated to enter the secretory pathway. Newly synthesized proteins are directed to the eukaryotic Sec61 or prokaryotic SecYEG translocation channel by signal peptides. The efficiency of protein secretion in eukaryotes is very dependent on the signal peptide which has been used.

11.3.26 Protein Transport

Many proteins are destined for other parts of the cell than the cytosol and a wide range of signalling sequences or (signal peptides) are used to direct proteins to where they

are supposed to be. In prokaryotes this is normally a simple process due to limited compartmentalisation of the cell. However, in eukaryotes there is a great variety of different targeting processes to ensure the protein arrives at the correct organelle.

Not all proteins remain within the cell and many are exported, for example, digestive enzymes, hormones and extracellular matrix proteins. In eukaryotes the export pathway is well developed and the main mechanism for the export of these proteins is translocation to the endoplasmic reticulum, followed by transport via the Golgi apparatus.

11.4 Techniques of Molecular Genetics

11.4.1 Polymerase chain reaction

Polymerase chain reaction⁷⁵ (PCR) is a method widely used in molecular biology to make copies of a specific DNA segment. Using PCR, copies of DNA sequences are exponentially amplified to generate thousands to millions of more copies of that particular DNA segment. PCR is now a common and often indispensable technique used in medical laboratory and clinical laboratory research for a broad variety of applications including biomedical research and criminal forensics. The vast majority of PCR methods rely on thermal cycling. Thermal cycling exposes reactants to repeated cycles of heating and cooling to permit different temperature-dependent reactions – specifically, DNA melting and enzyme-driven DNA replication.

A basic PCR set-up requires several components and reagents, including a DNA template that contains the DNA target region to amplify; a DNA polymerase; an enzyme that polymerizes new DNA strands; heat-resistant Taq polymerase is especially common, as it is more likely to remain intact during the high-temperature DNA denaturation process; two DNA primers that are complementary to the 3' (three prime) ends of each of the sense and anti-sense strands of the DNA target (DNA polymerase can only bind to and elongate from a double-stranded region of DNA; without primers there is no double-stranded initiation site at which the polymerase can bind); specific primers that are complementary to the DNA target region are selected beforehand, and are often custom-made in a laboratory or purchased from commercial biochemical suppliers; deoxynucleoside triphosphates, or dNTPs (sometimes called "deoxynucleotide triphosphates"; nucleotides containing triphosphate groups), the building blocks from which the DNA polymerase synthesizes a new DNA strand; a buffer solution providing a suitable chemical environment for optimum activity and stability of the DNA

⁷⁵ https://en.wikipedia.org/wiki/Polymerase_chain_reaction

polymerase; bivalent cations, typically magnesium (Mg) or manganese (Mn) ions; Mg^{2+} is the most common, but Mn^{2+} can be used for PCR-mediated DNA mutagenesis, as a higher Mn^{2+} concentration increases the error rate during DNA synthesis; and monovalent cations, typically potassium (K) ions.

The reaction is commonly carried out in a volume of 10–200 μL in small reaction tubes (0.2–0.5 mL volumes) in a thermal cycler. The thermal cycler heats and cools the reaction tubes to achieve the temperatures required at each step of the reaction (see below). Many modern thermal cyclers make use of the Peltier effect, which permits both heating and cooling of the block holding the PCR tubes simply by reversing the electric current. Thin-walled reaction tubes permit favorable thermal conductivity to allow for rapid thermal equilibration. Most thermal cyclers have heated lids to prevent condensation at the top of the reaction tube. Older thermal cyclers lacking a heated lid require a layer of oil on top of the reaction mixture or a ball of wax inside the tube.

Almost all PCR applications employ a heat-stable DNA polymerase, such as Taq polymerase, an enzyme originally isolated from the thermophilic bacterium *Thermus aquaticus*. If the polymerase used was heat-susceptible, it would denature under the high temperatures of the denaturation step. Before the use of Taq polymerase, DNA polymerase had to be manually added every cycle, which was a tedious and costly process.

Applications of the technique include DNA cloning for sequencing, gene cloning and manipulation, gene mutagenesis; construction of DNA-based phylogenies, or functional analysis of genes; diagnosis and monitoring of hereditary diseases; amplification of ancient DNA; analysis of genetic fingerprints for DNA profiling (for example, in forensic science and parentage testing); and detection of pathogens in nucleic acid tests for the diagnosis of infectious diseases.

PCR amplifies a specific region of a DNA strand (the DNA target). Most PCR methods amplify DNA fragments of between 0.1 and 10 kilo base pairs (kbp) in length, although some techniques allow for amplification of fragments up to 40 kbp. The amount of amplified product is determined by the available substrates in the reaction, which become limiting as the reaction progresses.

11.4.2 DNA Sequencing

DNA sequencing is the process of determining the nucleic acid sequence – the order of nucleotides in DNA. It includes any method or technology that is used to determine the order of the four bases: adenine, guanine, cytosine, and thymine. The advent of rapid DNA sequencing meth-

ods has greatly accelerated biological and medical research and discovery.

Knowledge of DNA sequences has become indispensable for basic biological research, and in numerous applied fields such as medical diagnosis, biotechnology, forensic biology, virology and biological systematics. Comparing healthy and mutated DNA sequences can diagnose different diseases including various cancers, characterize antibody repertoire, and can be used to guide patient treatment. Having a quick way to sequence DNA allows for faster and more individualized medical care to be administered, and for more organisms to be identified and catalogued.

The rapid speed of sequencing attained with modern DNA sequencing technology has been instrumental in the sequencing of complete DNA sequences, or genomes, of numerous types and species of life, including the human genome and other complete DNA sequences of many animal, plant, and microbial species.

The first full DNA genome to be sequenced was that of bacteriophage ϕ X174 in 1977. Medical Research Council scientists deciphered the complete DNA sequence of the Epstein–Barr virus in 1984, finding it contained 172,282 nucleotides. Completion of the sequence marked a significant turning point in DNA sequencing because it was achieved with no prior genetic profile knowledge of the virus.

The first method for determining DNA sequences involved a location-specific primer extension strategy established by Ray Wu at Cornell University in 1970. DNA polymerase catalysis and specific nucleotide labeling, both of which figure prominently in current sequencing schemes, were used to sequence the cohesive ends of lambda phage DNA. Between 1970 and 1973, Wu, R Padmanabhan and colleagues demonstrated that this method can be employed to determine any DNA sequence using synthetic location-specific primers. Frederick Sanger then adopted this primer-extension strategy to develop more rapid DNA sequencing methods at the MRC Centre, Cambridge, UK and published a method for “DNA sequencing with chain-terminating inhibitors” in 1977. Walter Gilbert and Allan Maxam at Harvard also developed sequencing methods, including one for “DNA sequencing by chemical degradation”. In 1973, Gilbert and Maxam reported the sequence of 24 basepairs using a method known as wandering-spot analysis. Advancements in sequencing were aided by the concurrent development of recombinant DNA technology, allowing DNA samples to be isolated from sources other than viruses.

11.4.3 DNA Microarray

A DNA microarray (also commonly known as DNA chip or biochip) is a collection of microscopic DNA spots attached to a solid surface. Scientists use DNA microarrays to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome. Each DNA spot contains picomoles (10–12 moles) of a specific DNA sequence, known as probes (or reporters or oligos). These can be a short section of a gene or other DNA element that are used to hybridize a cDNA or cRNA (also called anti-sense RNA) sample (called target) under high-stringency conditions. Probe-target hybridization is usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labeled targets to determine relative abundance of nucleic acid sequences in the target. The original nucleic acid arrays were macro arrays approximately 9 cm × 12 cm and the first computerized image based analysis was published in 1981. It was invented by Patrick O. Brown.

The core principle behind microarrays is hybridization between two DNA strands, the property of complementary nucleic acid sequences to specifically pair with each other by forming hydrogen bonds between complementary nucleotide base pairs. A high number of complementary base pairs in a nucleotide sequence means tighter non-covalent bonding between the two strands. After washing off non-specific bonding sequences, only strongly paired strands will remain hybridized. Fluorescently labeled target sequences that bind to a probe sequence generate a signal that depends on the hybridization conditions (such as temperature), and washing after hybridization. Total strength of the signal, from a spot (feature), depends upon the amount of target sample binding to the probes present on that spot. Microarrays use relative quantitation in which the intensity of a feature is compared to the intensity of the same feature under a different condition, and the identity of the feature is known by its position.

11.4.4 Recombinant DNA

Recombinant DNA (rDNA) molecules are DNA molecules formed by laboratory methods of genetic recombination (such as molecular cloning) to bring together genetic material from multiple sources, creating sequences that would not otherwise be found in the genome.

Recombinant DNA is the general name for a piece of DNA that has been created by combining at least two strands. Recombinant DNA is possible because DNA molecules from all organisms share the same chemical structure, and differ only in the nucleotide sequence within that identical overall structure. Recombinant DNA molecules are sometimes called chimeric DNA, because they can be made of material from two different species,

like the mythical chimera. R-DNA technology uses palindromic sequences and leads to the production of sticky and blunt ends.

The DNA sequences used in the construction of recombinant DNA molecules can originate from any species. For example, plant DNA may be joined to bacterial DNA, or human DNA may be joined with fungal DNA. In addition, DNA sequences that do not occur anywhere in nature may be created by the chemical synthesis of DNA, and incorporated into recombinant molecules. Using recombinant DNA technology and synthetic DNA, literally any DNA sequence may be created and introduced into any of a very wide range of living organisms.

Proteins that can result from the expression of recombinant DNA within living cells are termed recombinant proteins. When recombinant DNA encoding a protein is introduced into a host organism, the recombinant protein is not necessarily produced. Expression of foreign proteins requires the use of specialized expression vectors and often necessitates significant restructuring by foreign coding sequences.

Recombinant DNA differs from genetic recombination in that the former results from artificial methods in the test tube, while the latter is a normal biological process that results in the remixing of existing DNA sequences in essentially all organisms.

Molecular cloning is the laboratory process used to create recombinant DNA. It is one of two most widely used methods, along with polymerase chain reaction (PCR), used to direct the replication of any specific DNA sequence chosen by the experimentalist. There are two fundamental differences between the methods. One is that molecular cloning involves replication of the DNA within a living cell, while PCR replicates DNA in the test tube, free of living cells. The other difference is that cloning involves cutting and pasting DNA sequences, while PCR amplifies by copying an existing sequence.

Formation of recombinant DNA requires a cloning vector, a DNA molecule that replicates within a living cell. Vectors are generally derived from plasmids or viruses, and represent relatively small segments of DNA that contain necessary genetic signals for replication, as well as additional elements for convenience in inserting foreign DNA, identifying cells that contain recombinant DNA, and, where appropriate, expressing the foreign DNA. The choice of vector for molecular cloning depends on the choice of host organism, the size of the DNA to be cloned, and whether and how the foreign DNA is to be expressed. The DNA segments can be combined by using a variety of methods, such as restriction enzyme/ligase cloning or Gibson assembly.

In standard cloning protocols, the cloning of any DNA

fragment essentially involves seven steps: (1) Choice of host organism and cloning vector, (2) Preparation of vector DNA, (3) Preparation of DNA to be cloned, (4) Creation of recombinant DNA, (5) Introduction of recombinant DNA into the host organism, (6) Selection of organisms containing recombinant DNA, and (7) Screening for clones with desired DNA inserts and biological properties.

Following transplantation into the host organism, the foreign DNA contained within the recombinant DNA construct may or may not be expressed. That is, the DNA may simply be replicated without expression, or it may be transcribed and translated and a recombinant protein is produced. Generally speaking, expression of a foreign gene requires restructuring the gene to include sequences that are required for producing an mRNA molecule that can be used by the host's translational apparatus (e.g. promoter, translational initiation signal, and transcriptional terminator). Specific changes to the host organism may be made to improve expression of the ectopic gene. In addition, changes may be needed to the coding sequences as well, to optimize translation, make the protein soluble, direct the recombinant protein to the proper cellular or extracellular location, and stabilize the protein from degradation.

In most cases, organisms containing recombinant DNA have apparently normal phenotypes. That is, their appearance, behavior and metabolism are usually unchanged, and the only way to demonstrate the presence of recombinant sequences is to examine the DNA itself, typically using a polymerase chain reaction (PCR) test. Significant exceptions exist, and are discussed below.

If the rDNA sequences encode a gene that is expressed, then the presence of RNA and/or protein products of the recombinant gene can be detected, typically using RT-PCR or western hybridization methods. Gross phenotypic changes are not the norm, unless the recombinant gene has been chosen and modified so as to generate biological activity in the host organism. Additional phenotypes that are encountered include toxicity to the host organism induced by the recombinant gene product, especially if it is over-expressed or expressed within inappropriate cells or tissues.

In some cases, recombinant DNA can have deleterious effects even if it is not expressed. One mechanism by which this happens is insertional inactivation, in which the rDNA becomes inserted into a host cell's gene. In some cases, researchers use this phenomenon to "knock out" genes to determine their biological function and importance. Another mechanism by which rDNA insertion into chromosomal DNA can affect gene expression is by inappropriate activation of previously unexpressed host cell genes. This can happen, for example, when a recombinant DNA fragment containing an active promoter

becomes located next to a previously silent host cell gene, or when a host cell gene that functions to restrain gene expression undergoes insertional inactivation by recombinant DNA.

Recombinant DNA is widely used in biotechnology, medicine and research. Today, recombinant proteins and other products that result from the use of DNA technology are found in essentially every western pharmacy, physician or veterinarian office, medical testing laboratory, and biological research laboratory. In addition, organisms that have been manipulated using recombinant DNA technology, as well as products derived from those organisms, have found their way into many farms, supermarkets, home medicine cabinets, and even pet shops, such as those that sell GloFish and other genetically modified animals.

The most common application of recombinant DNA is in basic research, in which the technology is important to most current work in the biological and biomedical sciences. Recombinant DNA is used to identify, map and sequence genes, and to determine their function. rDNA probes are employed in analyzing gene expression within individual cells, and throughout the tissues of whole organisms. Recombinant proteins are widely used as reagents in laboratory experiments and to generate antibody probes for examining protein synthesis within cells and organisms.

Many additional practical applications of recombinant DNA are found in industry, food production, human and veterinary medicine, agriculture, and bioengineering. Some specific examples are:

- Recombinant human insulin: almost completely replaced insulin obtained from animal sources (e.g. pigs and cattle) for the treatment of insulin-dependent diabetes. A variety of different recombinant insulin preparations are in widespread use. Recombinant insulin is synthesized by inserting the human insulin gene into *E. coli*, or yeast (*Saccharomyces cerevisiae*) which then produces insulin for human use. Administered to patients whose pituitary glands generate insufficient quantities to support normal growth and development. Before recombinant HGH became available, HGH for therapeutic use was obtained from pituitary glands of cadavers. This unsafe practice led to some patients developing Creutzfeldt-Jakob disease. Recombinant HGH eliminated this problem, and is now used therapeutically. It has also been misused as a performance-enhancing drug by athletes and others. DrugBank entry
- Recombinant blood clotting factor VIII: a blood-clotting protein that is administered to patients with

forms of the bleeding disorder hemophilia, who are unable to produce factor VIII in quantities sufficient to support normal blood coagulation. Before the development of recombinant factor VIII, the protein was obtained by processing large quantities of human blood from multiple donors, which carried a very high risk of transmission of blood borne infectious diseases, for example HIV and hepatitis B. DrugBank entry

- Recombinant hepatitis B vaccine: Hepatitis B infection is controlled through the use of a recombinant hepatitis B vaccine, which contains a form of the hepatitis B virus surface antigen that is produced in yeast cells. The development of the recombinant subunit vaccine was an important and necessary development because hepatitis B virus, unlike other common viruses such as polio virus, cannot be grown *in vitro*. Vaccine information from Hepatitis B Foundation
- Diagnosis of infection with HIV: each of the three widely used methods for diagnosing HIV infection has been developed using recombinant DNA. The antibody test (ELISA or western blot) uses a recombinant HIV protein to test for the presence of antibodies that the body has produced in response to an HIV infection. The DNA test looks for the presence of HIV genetic material using reverse transcription polymerase chain reaction (RT-PCR). Development of the RT-PCR test was made possible by the molecular cloning and sequence analysis of HIV genomes. HIV testing page from US Centers for Disease Control (CDC)
- Golden rice: a recombinant variety of rice that has been engineered to express the enzymes responsible for β -carotene biosynthesis. This variety of rice holds substantial promise for reducing the incidence of vitamin A deficiency in the world's population. Golden rice is not currently in use, pending the resolution of regulatory and intellectual property issues.
- Herbicide-resistant crops: commercial varieties of important agricultural crops (including soy, maize/corn, sorghum, canola, alfalfa and cotton) have been developed that incorporate a recombinant gene that results in resistance to the herbicide glyphosate (trade name Roundup), and simplifies weed control by glyphosate application. These crops are in common commercial use in several countries.
- Insect-resistant crops: *bacillus thuringiensis* is a bacterium that naturally produces a protein (Bt toxin) with insecticidal properties. The bacterium has been applied to crops as an insect-control strategy for many years, and this practice has

been widely adopted in agriculture and gardening. Recently, plants have been developed that express a recombinant form of the bacterial protein, which may effectively control some insect predators. Environmental issues associated with the use of these transgenic crops have not been fully resolved.

The first publications describing the successful production and intracellular replication of recombinant DNA appeared in 1972 and 1973, from Stanford and UCSF. In 1980, Paul Berg⁷⁶, a professor in the Biochemistry Department at Stanford and an author on one of the first papers was awarded the Nobel Prize in Chemistry for his work on nucleic acids "with particular regard to recombinant DNA". Werner Arber⁷⁷, Hamilton Smith⁷⁸, and Daniel Nathans⁷⁹ shared the 1978 Nobel Prize in Physiology or Medicine for the discovery of restriction endonucleases which are used rDNA technology.

Stanford University applied for a US patent on recombinant DNA in 1974, listing the inventors as Herbert W. Boyer⁸⁰ and Stanley N. Cohen⁸¹; this patent was awarded in 1980. The first licensed drug generated using recombinant DNA technology was human insulin, developed by Genentech⁸² and licensed by Eli Lilly and Company.

Scientists associated with the initial development of recombinant DNA methods recognized that the potential existed for organisms containing recombinant DNA to have undesirable or dangerous properties. At the 1975 Asilomar Conference on Recombinant DNA⁸³, these concerns were discussed and a voluntary moratorium on recombinant DNA research was initiated for experiments that were considered particularly risky. This moratorium was widely observed until the National Institutes of Health (USA) developed and issued formal guidelines for rDNA work. Today, recombinant DNA molecules and recombinant proteins are usually not regarded as dangerous. However, concerns remain about some organisms that express recombinant DNA, particularly when they leave the laboratory and are introduced into the environment or food chain.

11.4.5 Gene Synthesis

As the cost of DNA oligonucleotide synthesis falls, artificial synthesis of a complete gene is now a viable method for introducing mutations into a gene. This method allows

⁷⁶https://en.wikipedia.org/wiki/Paul_Berg

⁷⁷https://en.wikipedia.org/wiki/Werner_Arber

⁷⁸https://en.wikipedia.org/wiki/Hamilton_O._Smith

⁷⁹https://en.wikipedia.org/wiki/Daniel_Nathans

⁸⁰https://en.wikipedia.org/wiki/Herbert_Boyer

⁸¹https://en.wikipedia.org/wiki/Stanley_Norman_Cohen

⁸²<https://en.wikipedia.org/wiki/Genentech>
⁸³https://en.wikipedia.org/wiki/Asilomar_Conference_on_Recombinant_DNA

for extensive mutation at multiple sites, including the complete redesign of the codon usage of a gene to optimise it for a particular organism.

11.4.6 Gene Knockout

A gene knockout (abbreviation: KO) is a genetic technique in which one of an organism's genes is made inoperative ("knocked out" of the organism). However, KO can also refer to the gene that is knocked out or the organism that carries the gene knockout. Knockout organisms or simply knockouts are used to study gene function, usually by investigating the effect of gene loss. Researchers draw inferences from the difference between the knockout organism and normal individuals.

The KO technique is essentially the opposite of a gene knock-in. Knocking out two genes simultaneously in an organism is known as a double knockout (DKO). Similarly the terms triple knockout (TKO) and quadruple knockouts (QKO) are used to describe three or four knocked out genes, respectively. However, one needs to distinguish between heterozygous and homozygous KOs. In the former, only one of two gene copies (alleles) is knocked out, in the latter both are knocked out.

Knockouts are accomplished through a variety of techniques. Originally, naturally occurring mutations were identified and then gene loss or inactivation had to be established by DNA sequencing or other methods.

Traditionally, homologous recombination was the main method for causing a gene knockout. This method involves creating a DNA construct containing the desired mutation. For knockout purposes, this typically involves a drug resistance marker in place of the desired knockout gene. The construct will also contain a minimum of 2kb of homology to the target sequence. The construct can be delivered to stem cells either through microinjection or electroporation. This method then relies on the cell's own repair mechanisms to recombine the DNA construct into the existing DNA. This results in the sequence of the gene being altered, and most cases the gene will be translated into a nonfunctional protein, if it is translated at all. However, this is an inefficient process, as homologous recombination accounts for only 10–2 to 10–3 of DNA integrations. Often, the drug selection marker on the construct is used to select for cells in which the recombination event has occurred.^e

These stem cells now lacking the gene could be used *in vivo*, for instance in mice, by inserting them into early embryos. If the resulting chimeric mouse contained the genetic change in their germline, this could then be passed on offspring.

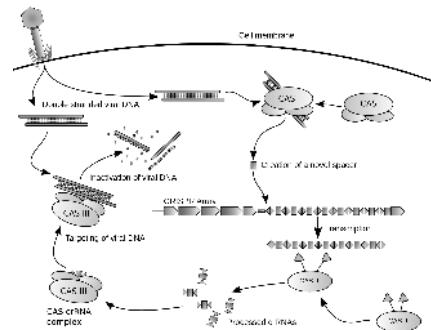


Figure 11.14: Diagram of the CRISPR prokaryotic antiviral defense mechanism.⁸⁴

11.4.7 CRISPR Gene Editing

CRISPR gene editing is a method by which the genomes of living organisms may be edited. It is based on a simplified version of the bacterial CRISPR/Cas (CRISPR–Cas9) antiviral defense system. By delivering the Cas9 nuclease complexed with a synthetic guide RNA (gRNA) into a cell, the cell's genome can be cut at a desired location, allowing existing genes to be removed and/or new ones added.

While genomic editing in eukaryotic cells has been possible using various methods since the 1980s, the methods employed had proved to be inefficient and impractical to implement on a larger scale. Genomic editing leads to irreversible changes to the gene. Working like genetic scissors, the Cas9 nuclease opens both strands of the targeted sequence of DNA to introduce the modification by one of two methods. Knock-in mutations, facilitated via Homology Directed Repair (HDR), is the traditional pathway of targeted genomic editing approaches. This allows for the introduction of targeted DNA damage and repair. HDR employs the use of similar DNA sequences to drive the repair of the break via the incorporation of exogenous DNA to function as the repair template. This method relies on the periodic and isolated occurrence of DNA damage at the target site in order for a repair to commence. Knock-out mutations caused by CRISPR–Cas9 results in the repair of the double-strand break by means of NHEJ (Non-Homologous End Joining). NHEJ can often result in random deletions or insertions at the repair site disrupting or altering gene functionality. Therefore, genomic engineering by CRISPR–Cas9 allows researchers the ability to generate targeted random gene disruption.

Because of this, the precision of genomic editing is a great concern. With the discovery of CRISPR and specifically the Cas9 nuclease molecule, efficient and highly selective editing is now a reality. Cas9 allows for a reliable method of creating a targeted break at a specific location as designated by the crRNA and tracrRNA guide strands.

Cas9 derived from *Streptococcus pyogenes* bacteria has facilitated the targeted genomic modification in eukaryotic cells. The ease with which researchers can insert Cas9 and template RNA in order to silence or cause point mutations on specific loci has proved invaluable to the quick and efficient mapping of genomic models and biological processes associated with various genes in a variety of eukaryotes. Newly engineered variants of the Cas9 nuclease have been developed that significantly reduce off-target activity.

CRISPR-Cas9 genome editing techniques have many potential applications, including medicine and crop seed enhancement. The use of CRISPR-Cas9-gRNA complex for genome editing was the AAAS's choice for breakthrough of the year in 2015. Bioethical concerns have been raised about the prospect of using CRISPR for germline editing.

11.4.8 Human Germline Modification

As of March 2015, multiple groups had announced ongoing research with the intention of laying the foundations for applying CRISPR to human embryos for human germline engineering, including labs in the US, China, and the UK, as well as US biotechnology company OvaScience. Scientists, including a CRISPR co-discoverer, urged a worldwide moratorium on applying CRISPR to the human germline, especially for clinical use. They said "scientists should avoid even attempting, in lax jurisdictions, germline genome modification for clinical application in humans" until the full implications "are discussed among scientific and governmental organizations". These scientists support further low-level research on CRISPR and do not see CRISPR as developed enough for any clinical use in making heritable changes to humans.

In April 2015, Chinese scientists reported results of an attempt to alter the DNA of non-viable human embryos using CRISPR to correct a mutation that causes beta thalassemia, a lethal heritable disorder. The study had previously been rejected by both Nature and Science in part because of ethical concerns. The experiments resulted in successfully changing only some of the intended genes, and had off-target effects on other genes. The researchers stated that CRISPR is not ready for clinical application in reproductive medicine. In April 2016, Chinese scientists were reported to have made a second unsuccessful attempt to alter the DNA of non-viable human embryos using CRISPR – this time to alter the CCR5 gene to make the embryo HIV resistant.

In December 2015, an International Summit on Human Gene Editing took place in Washington under the guidance of David Baltimore. Members of national scientific academies of America, Britain and China discussed the ethics of germline modification. They agreed to support basic and clinical research under certain

legal and ethical guidelines. A specific distinction was made between somatic cells, where the effects of edits are limited to a single individual, versus germline cells, where genome changes could be inherited by descendants. Heritable modifications could have unintended and far-reaching consequences for human evolution, genetically (e.g. gene/environment interactions) and culturally (e.g. Social Darwinism). Altering of gametocytes and embryos to generate inheritable changes in humans was defined to be irresponsible. The group agreed to initiate an international forum to address such concerns and harmonize regulations across countries.

In November 2018, Jiankui He announced that he had edited two human embryos, to attempt to disable the gene for CCR5, which codes for a receptor that HIV uses to enter cells. He said that twin girls, Lulu and Nana, had been born a few weeks earlier. He said that the girls still carried functional copies of CCR5 along with disabled CCR5 (mosaicism) and were still vulnerable to HIV. The work was widely condemned as unethical, dangerous, and premature. An international group of scientists called for a global moratorium on genetically editing human embryos.

Policy regulations for the CRISPR-Cas9 system vary around the globe. In February 2016, British scientists were given permission by regulators to genetically modify human embryos by using CRISPR-Cas9 and related techniques. However, researchers were forbidden from implanting the embryos and the embryos were to be destroyed after seven days.

The US has an elaborate, interdepartmental regulatory system to evaluate new genetically modified foods and crops. For example, the Agriculture Risk Protection Act of 2000 gives the USDA the authority to oversee the detection, control, eradication, suppression, prevention, or retardation of the spread of plant pests or noxious weeds to protect the agriculture, environment and economy of the US. The act regulates any genetically modified organism that utilizes the genome of a predefined "plant pest" or any plant not previously categorized. In 2015, Yinong Yang successfully deactivated 16 specific genes in the white button mushroom, to make them non-browning. Since he had not added any foreign-species (transgenic) DNA to his organism, the mushroom could not be regulated by the USDA under Section 340.2. Yang's white button mushroom was the first organism genetically modified with the CRISPR-Cas9 protein system to pass US regulation. In 2016, the USDA sponsored a committee to consider future regulatory policy for upcoming genetic modification techniques. With the help of the US National Academies of Sciences, Engineering and Medicine, special interests groups met on April 15 to contemplate the possible advancements in genetic engineering within the

next five years and any new regulations that might be needed as a result. The FDA in 2017 proposed a rule that would classify genetic engineering modifications to animals as “animal drugs”, subjecting them to strict regulation if offered for sale, and reducing the ability for individuals and small businesses to make them profitably.

Chapter 12

Regulation of Gene Expression

Regulation of gene expression¹, or gene regulation, includes a wide range of mechanisms that are used by cells to increase or decrease the production of specific gene products (protein or RNA). Sophisticated programs of gene expression are widely observed in biology, for example to trigger developmental pathways, respond to environmental stimuli, or adapt to new food sources. Virtually any step of gene expression can be modulated, from transcriptional initiation, to RNA processing, and to the post-translational modification of a protein. Often, one gene regulator controls another, and so on, in a gene regulatory network.

Gene regulation is essential for viruses, prokaryotes and eukaryotes as it increases the versatility and adaptability of an organism by allowing the cell to express protein when needed.

In multicellular organisms, gene regulation drives cellular differentiation and morphogenesis in the embryo, leading to the creation of different cell types that possess different gene expression profiles from the same genome sequence. Although this does not explain how gene regulation originated, evolutionary biologists include it as a partial explanation of how evolution works at a molecular level, and it is central to the science of evolutionary developmental biology ("evo-devo").

Although as early as 1951, Barbara McClintock² showed interaction between two genetic loci, Activator (Ac) and Dissociator (Ds), in the color formation of maize seeds, the first discovery of a gene regulation system is widely considered to be the identification in 1961 of the *lac* operon, discovered by François Jacob³ and Jacques Monod⁴, in which some enzymes involved in lactose metabolism are expressed by *Escherichia coli* only in the presence of lactose and absence of glucose.

12.1 The *lac* operon

The *lac*⁵ operon (lactose operon) is an operon required for the transport and metabolism of lactose in *Escherichia coli* and many other enteric bacteria.

An operon is a functioning unit of DNA containing a cluster of genes under the control of a single promoter. The genes are transcribed together into an mRNA strand and either translated together in the cytoplasm, or undergo splicing to create monocistronic mRNAs that are translated separately, i.e. several strands of mRNA that each encode a single gene product. The result of this is that the genes contained in the operon are either expressed together or not at all. Several genes must be co-transcribed to define an operon.

Although glucose is the preferred carbon source for most bacteria, the *lac* operon allows for the effective digestion of lactose when glucose is not available through the activity of beta-galactosidase. Gene regulation of the *lac* operon was the first genetic regulatory mechanism to be understood clearly, so it has become a foremost example of prokaryotic gene regulation. It is discussed here in detail for this reason. This lactose metabolism system was used by François Jacob and Jacques Monod to determine how a biological cell knows which enzyme to synthesize. Their work on the *lac* operon won them the Nobel Prize in Physiology in 1965.

Bacterial operons are polycistronic transcripts that are able to produce multiple proteins from one mRNA transcript. In this case, when lactose is required as a sugar source for the bacterium, the three genes of the *lac* operon can be expressed and their subsequent proteins translated: *lacZ*, *lacY*, and *lacA*. The gene product of *lacZ* is β-galactosidase which cleaves lactose, a disaccharide, into glucose and galactose. *lacY* encodes Beta-galactoside permease, a membrane protein which becomes embedded in the cytoplasmic membrane to enable the cellular transport of lactose into the cell. Finally, *lacA* encodes

¹https://en.wikipedia.org/wiki/Regulation_of_gene_expression

²https://en.wikipedia.org/wiki/Barbara_McClintock

³https://en.wikipedia.org/wiki/Fran%C3%A7ois_Jacob

⁴https://en.wikipedia.org/wiki/Jacques_Monod

⁵https://en.wikipedia.org/wiki/Lac_operon

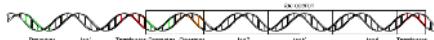


Figure 12.1: ⁶

Galactoside acetyltransferase.

It would be wasteful to produce enzymes when no lactose were available or if a preferable energy source such as glucose were available. The *lac* operon uses a two-part control mechanism to ensure that the cell expends energy producing the enzymes encoded by the *lac* operon only when necessary. In the absence of lactose, the *lac* repressor, lacI, halts production of the enzymes encoded by the *lac* operon. The *lac* repressor is always expressed unless a co-inducer binds to it. In other words, it is transcribed only in the presence of small molecule co-inducer. In the presence of glucose, the catabolite activator protein (CAP), required for production of the enzymes, remains inactive, and EIAGlc shuts down lactose permease to prevent transport of lactose into the cell. This dual control mechanism causes the sequential utilization of glucose and lactose in two distinct growth phases, known as diauxie.

Three-letter abbreviations are used to describe phenotypes in bacteria including *E. coli*.

Examples include:

- Lac (the ability to use lactose),
- His (the ability to synthesize the amino acid histidine)
- Mot (swimming motility)
- SmR (resistance to the antibiotic streptomycin)

In the case of Lac, wild type cells are Lac+ and are able to use lactose as a carbon and energy source, while Lac- mutant derivatives cannot use lactose. The same three letters are typically used (lower-case, italicized) to label the genes involved in a particular phenotype, where each different gene is additionally distinguished by an extra letter. The *lac* genes encoding enzymes are *lacZ*, *lacY*, and *lacA*. The fourth *lac* gene is *lacI*, encoding the lactose repressor—“I” stands for inducibility.

One may distinguish between structural genes encoding enzymes, and regulatory genes encoding proteins that affect gene expression. Current usage expands the phenotypic nomenclature to apply to proteins: thus, LacZ is the protein product of the *lacZ* gene, β-galactosidase. Various short sequences that are not genes also affect gene expression, including the *lac* promoter, *lac p*, and the *lac* operator, *lac o*. Although it is not strictly standard usage, mutations affecting *lac o* are referred to as *lac oc*, for historical reasons.

The *lac* operon consists of 3 structural genes, and a promoter, a terminator, regulator, and an operator. The

three structural genes are: *lacZ*, *lacY*, and *lacA*.

- *lacZ* encodes β-galactosidase (LacZ), an intracellular enzyme that cleaves the disaccharide lactose into glucose and galactose.
- *lacY* encodes Beta-galactoside permease (LacY), a transmembrane symporter that pumps β-galactosides including lactose into the cell using a proton gradient in the same direction. Permease increases the permeability of the cell to β-galactosides.
- *lacA* encodes β-galactoside transacetylase (LacA), an enzyme that transfers an acetyl group from acetyl-CoA to β-galactosides.

Only *lacZ* and *lacY* appear to be necessary for lactose catabolism.

Specific control of the *lac* genes depends on the availability of the substrate lactose to the bacterium. The proteins are not produced by the bacterium when lactose is unavailable as a carbon source. The *lac* genes are organized into an operon; that is, they are oriented in the same direction immediately adjacent on the chromosome and are co-transcribed into a single polycistronic mRNA molecule. Transcription of all genes starts with the binding of the enzyme RNA polymerase (RNAP), a DNA-binding protein, which binds to a specific DNA binding site, the promoter, immediately upstream of the genes. Binding of RNA polymerase to the promoter is aided by the cAMP-bound catabolite activator protein (CAP, also known as the cAMP receptor protein). However, the *lacI* gene (regulatory gene for *lac* operon) produces a protein that blocks RNAP from binding to the promoter of the operon. This protein can only be removed when allolactose binds to it, and inactivates it. The protein that is formed by the *lacI* gene is known as the *lac* repressor. The type of regulation that the *lac* operon undergoes is referred to as negative inducible, meaning that the gene is turned off by the regulatory factor (*lac* repressor) unless some molecule (lactose) is added. Because of the presence of the *lac* repressor protein, genetic engineers who replace the *lacZ* gene with another gene will have to grow the experimental bacteria on agar with lactose available on it. If they do not, the gene they are trying to express will not be expressed as the repressor protein is still blocking RNAP from binding to the promoter and transcribing the gene. Once the repressor is removed, RNAP then proceeds to transcribe all three genes (*lacZYA*) into mRNA. Each of the three genes on the mRNA strand has its own Shine-Dalgarno sequence, so the genes are independently translated. The DNA sequence of the *E. coli* *lac* operon, the *lacZYA* mRNA, and the *lacI* genes are available from GenBank (view).

The first control mechanism is the regulatory response

to lactose, which uses an intracellular regulatory protein called the lactose repressor to hinder production of β -galactosidase in the absence of lactose. The *lacI* gene coding for the repressor lies nearby the *lac* operon and is always expressed (constitutive). If lactose is missing from the growth medium, the repressor binds very tightly to a short DNA sequence just downstream of the promoter near the beginning of *lacZ* called the *lac* operator. The repressor binding to the operator interferes with binding of RNAP to the promoter, and therefore mRNA encoding LacZ and LacY is only made at very low levels. When cells are grown in the presence of lactose, however, a lactose metabolite called allolactose, made from lactose by the product of the *lacZ* gene, binds to the repressor, causing an allosteric shift. Thus altered, the repressor is unable to bind to the operator, allowing RNAP to transcribe the *lac* genes and thereby leading to higher levels of the encoded proteins.

The second control mechanism is a response to glucose, which uses the catabolite activator protein (CAP) homodimer to greatly increase production of β -galactosidase in the absence of glucose. Cyclic adenosine monophosphate (cAMP) is a signal molecule whose prevalence is inversely proportional to that of glucose. It binds to the CAP, which in turn allows the CAP to bind to the CAP binding site (a 16 bp DNA sequence upstream of the promoter on the left in the diagram below, about 60 bp upstream of the transcription start site), which assists the RNAP in binding to the DNA. In the absence of glucose, the cAMP concentration is high and binding of CAP-cAMP to the DNA significantly increases the production of β -galactosidase, enabling the cell to hydrolyse lactose and release galactose and glucose.

More recently inducer exclusion was shown to block expression of the *lac* operon when glucose is present. Glucose is transported into the cell by the PEP-dependent phosphotransferase system. The phosphate group of phosphoenolpyruvate is transferred via a phosphorylation cascade consisting of the general PTS (phosphotransferase system) proteins HPr and EIA and the glucose-specific PTS proteins EIAGlc and EIIBGlc, the cytoplasmic domain of the EII glucose transporter. Transport of glucose is accompanied by its phosphorylation by EIIBGlc, draining the phosphate group from the other PTS proteins, including EIAGlc. The unphosphorylated form of EIAGlc binds to the *lac* permease and prevents it from bringing lactose into the cell. Therefore, if both glucose and lactose are present, the transport of glucose blocks the transport of the inducer of the *lac* operon.

The *lac* repressor is a four-part protein, a tetramer, with identical subunits (Figure 12.2). Each subunit contains a helix-turn-helix (HTH) motif capable of binding to DNA.

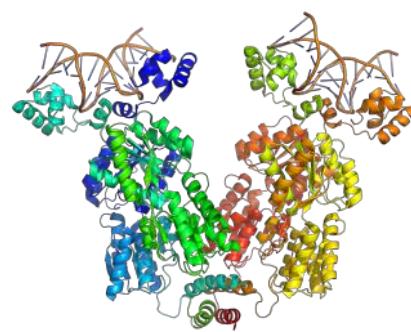


Figure 12.2: A simulated structural model⁷ of a complex between the *lac* repressor protein (LacI) and a 107-bp-long DNA segment. Two dimeric LacI functional subunits (green + blue and yellow + orange) each bind a DNA operator sequence (top). The mobility of the DNA-binding head groups coupled to the stable body of the body of LacI provide the force for the looping of the DNA (Ville et al.⁸).

The operator site where repressor binds is a DNA sequence with inverted repeat symmetry. The two DNA half-sites of the operator together bind to two of the subunits of the repressor. Although the other two subunits of repressor are not doing anything in this model, this property was not understood for many years.

Eventually it was discovered that two additional operators are involved in *lac* regulation. One (O3) lies about -90 bp upstream of O1 in the end of the *lacI* gene, and the other (O2) is about +410 bp downstream of O1 in the early part of *lacZ*. These two sites were not found in the early work because they have redundant functions and individual mutations do not affect repression very much. Single mutations to either O2 or O3 have only 2 to 3-fold effects. However, their importance is demonstrated by the fact that a double mutant defective in both O2 and O3 is dramatically de-repressed (by about 70-fold).

In the current model, *lac* repressor is bound simultaneously to both the main operator O1 and to either O2 or O3. The intervening DNA loops out from the complex. The redundant nature of the two minor operators suggests that it is not a specific looped complex that is important. One idea is that the system works through tethering; if bound repressor releases from O1 momentarily, binding to a minor operator keeps it in the vicinity, so that it may rebind quickly. This would increase the affinity of repressor for O1.

The repressor is an allosteric protein, i.e. it can assume either one of two slightly different shapes, which are in

equilibrium with each other. In one form the repressor will bind to the operator DNA with high specificity, and in the other form it has lost its specificity. According to the classical model of induction, binding of the inducer, either allolactose or IPTG, to the repressor affects the distribution of repressor between the two shapes. Thus, repressor with inducer bound is stabilized in the non-DNA-binding conformation. However, this simple model cannot be the whole story, because repressor is bound quite stably to DNA, yet it is released rapidly by addition of inducer. Therefore, it seems clear that an inducer can also bind to the repressor when the repressor is already bound to DNA. It is still not entirely known what the exact mechanism of binding is.

Non-specific binding of the repressor to DNA plays a crucial role in the repression and induction of the Lac-operon. The specific binding site for the Lac-repressor protein is the operator. The non-specific interaction is mediated mainly by charge-charge interactions while binding to the operator is reinforced by hydrophobic interactions. Additionally, there is an abundance of non-specific DNA sequences to which the repressor can bind. Essentially, any sequence that is not the operator, is considered non-specific. Studies have shown, that without the presence of non-specific binding, induction (or unrepresion) of the Lac-operon could not occur even with saturated levels of inducer. It had been demonstrated that, without non-specific binding, the basal level of induction is ten thousand times smaller than observed normally. This is because the non-specific DNA acts as sort of a "sink" for the repressor proteins, distracting them from the operator. The non-specific sequences decrease the amount of available repressor in the cell. This in turn reduces the amount of inducer required to unrepres the system.

A number of lactose derivatives or analogs have been described that are useful for work with the *lac* operon. These compounds are mainly substituted galactosides, where the glucose moiety of lactose is replaced by another chemical group. Isopropyl- β -D-thiogalactopyranoside (IPTG) is frequently used as an inducer of the *lac* operon for physiological work. IPTG binds to repressor and inactivates it, but is not a substrate for β -galactosidase. One advantage of IPTG for in vivo studies is that since it cannot be metabolized by *E. coli* its concentration remains constant and the rate of expression of *lac* p/o-controlled genes, is not a variable in the experiment. IPTG intake is dependent on the action of lactose permease in *P. fluorescens*, but not in *E. coli*.

The experimental microorganism used by François Jacob and Jacques Monod was the common laboratory bacterium, *E. coli*, but many of the basic regulatory concepts that were discovered by Jacob and Monod are fundamental

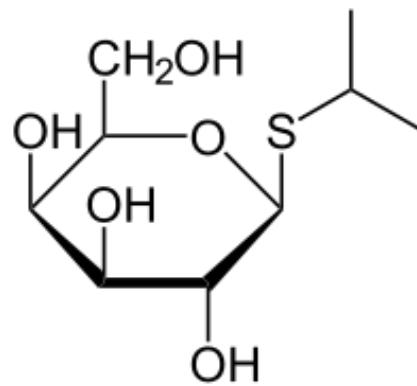


Figure 12.3: Structure of isopropyl β -D-thiogalactopyranoside (IPTG)⁹

to cellular regulation in all organisms. The key idea is that proteins are not synthesized when they are not needed—*E. coli* conserves cellular resources and energy by not making the three Lac proteins when there is no need to metabolize lactose, such as when other sugars like glucose are available. The following section discusses how *E. coli* controls certain genes in response to metabolic needs.

During World War II, Monod was testing the effects of combinations of sugars as nutrient sources for *E. coli* and *B. subtilis*. Monod was following up on similar studies that had been conducted by other scientists with bacteria and yeast. He found that bacteria grown with two different sugars often displayed two phases of growth. For example, if glucose and lactose were both provided, glucose was metabolized first (growth phase I, see Figure 2) and then lactose (growth phase II). Lactose was not metabolized during the first part of the diauxic growth curve because β -galactosidase was not made when both glucose and lactose were present in the medium. Monod named this phenomenon diauxie.

Monod then focused his attention on the induction of β -galactosidase formation that occurred when lactose was the sole sugar in the culture medium.

A conceptual breakthrough of Jacob and Monod was to recognize the distinction between regulatory substances and sites where they act to change gene expression. A former soldier, Jacob used the analogy of a bomber that would release its lethal cargo upon receipt of a special radio transmission or signal. A working system requires both a ground transmitter and a receiver in the airplane. Now, suppose that the usual transmitter is broken. This system can be made to work by introduction of a second, functional transmitter. In contrast, he said, consider a bomber with a defective receiver. The behavior of this bomber cannot be changed by introduction of a second, functional aeroplane.

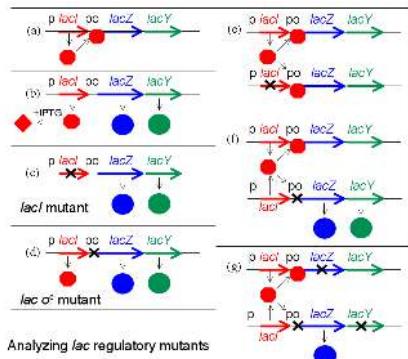


Figure 12.4: Analysis of *lac* regulatory mutants by Lac complementation.¹⁰

To analyze regulatory mutants of the *lac* operon, Jacob developed a system by which a second copy of the *lac* genes (*lacI* with its promoter, and *lacZYA* with promoter and operator) could be introduced into a single cell. A culture of such bacteria, which are diploid for the *lac* genes but otherwise normal, is then tested for the regulatory phenotype. In particular, it is determined whether LacZ and LacY are made even in the absence of IPTG (due to the lac-tose repressor produced by the mutant gene being non-functional). This experiment, in which genes or gene clusters are tested pairwise, is called a complementation test.

This test is illustrated in Figure 12.4 (*lacA* is omitted for simplicity). First, certain haploid states are shown (i.e. the cell carries only a single copy of the *lac* genes). Panel (a) shows repression, (b) shows induction by IPTG, and (c) and (d) show the effect of a mutation to the *lacI* gene or to the operator, respectively. In panel (e) the complementation test for repressor is shown. If one copy of the *lac* genes carries a mutation in *lacI*, but the second copy is wild type for *lacI*, the resulting phenotype is normal—but *lacZ* is expressed when exposed to inducer IPTG. Mutations affecting repressor are said to be recessive to wild type (and that wild type is dominant), and this is explained by the fact that repressor is a small protein which can diffuse in the cell. The copy of the *lac* operon adjacent to the defective *lacI* gene is effectively shut off by protein produced from the second copy of *lacI*.

If the same experiment is carried out using an operator mutation, a different result is obtained (panel (f)). The phenotype of a cell carrying one mutant and one wild type operator site is that LacZ and LacY are produced even in the absence of the inducer IPTG; because the damaged operator site, does not permit binding of the repressor to inhibit transcription of the structural genes. The operator mutation is dominant. When the operator site where repressor must bind is damaged by mutation, the presence of a second functional site in the same cell makes no difference to

expression of genes controlled by the mutant site.

A more sophisticated version of this experiment uses marked operons to distinguish between the two copies of the *lac* genes and show that the unregulated structural gene(s) is(are) the one(s) next to the mutant operator (panel (g)). For example, suppose that one copy is marked by a mutation inactivating *lacZ* so that it can only produce the LacY protein, while the second copy carries a mutation affecting *lacY* and can only produce LacZ. In this version, only the copy of the *lac* operon that is adjacent to the mutant operator is expressed without IPTG. We say that the operator mutation is *cis*-dominant, it is dominant to wild type but affects only the copy of the operon which is immediately adjacent to it.

This explanation is misleading in an important sense, because it proceeds from a description of the experiment and then explains the results in terms of a model. But in fact, it is often true that the model comes first, and an experiment is fashioned specifically to test the model. Jacob and Monod first imagined that there must be a site in DNA with the properties of the operator, and then designed their complementation tests to show this.

The dominance of operator mutants also suggests a procedure to select them specifically. If regulatory mutants are selected from a culture of wild type using phenyl-Gal, as described above, operator mutations are rare compared to repressor mutations because the target-size is so small. But if instead we start with a strain which carries two copies of the whole *lac* region (that is diploid for *lac*), the repressor mutations (which still occur) are not recovered because complementation by the second, wild type *lacI* gene confers a wild type phenotype. In contrast, mutation of one copy of the operator confers a mutant phenotype because it is dominant to the second, wild type copy.

Explanation of diauxie depended on the characterization of additional mutations affecting the *lac* genes other than those explained by the classical model. Two other genes, *cya* and *crp*, subsequently were identified that mapped far from *lac*, and that, when mutated, result in a decreased level of expression in the presence of IPTG and even in strains of the bacterium lacking the repressor or operator. The discovery of cAMP in *E. coli* led to the demonstration that mutants defective the *cya* gene but not the *crp* gene could be restored to full activity by the addition of cAMP to the medium.

The *cya* gene encodes adenylate cyclase, which produces cAMP. In a *cya* mutant, the absence of cAMP makes the expression of the *lacZYA* genes more than ten times lower than normal. Addition of cAMP corrects the low Lac expression characteristic of *cya* mutants. The second gene, *crp*, encodes a protein called catabolite activator protein

(CAP) or cAMP receptor protein (CRP).

However the lactose metabolism enzymes are made in small quantities in the presence of both glucose and lactose (sometimes called leaky expression) due to the fact that the LacI repressor rapidly associates/dissociates from the DNA rather than tightly binding to it, which can allow time for RNAP to bind and transcribe mRNAs of lacZYA. Leaky expression is necessary in order to allow for metabolism of some lactose after the glucose source is expended, but before *lac* expression is fully activated.

In summary:

- When lactose is absent then there is very little Lac enzyme production (the operator has Lac repressor bound to it).
- When lactose is present but a preferred carbon source (like glucose) is also present then a small amount of enzyme is produced (Lac repressor is not bound to the operator).
- When glucose is absent, CAP-cAMP binds to a specific DNA site upstream of the promoter and makes a direct protein-protein interaction with RNAP that facilitates the binding of RNAP to the promoter.

The delay between growth phases reflects the time needed to produce sufficient quantities of lactose-metabolizing enzymes. First, the CAP regulatory protein has to assemble on the *lac* promoter, resulting in an increase in the production of *lac* mRNA. More available copies of the *lac* mRNA results in the production (see translation) of significantly more copies of LacZ (β -galactosidase, for lactose metabolism) and LacY (lactose permease to transport lactose into the cell). After a delay needed to increase the level of the lactose metabolizing enzymes, the bacteria enter into a new rapid phase of cell growth.

Two puzzles of catabolite repression relate to how cAMP levels are coupled to the presence of glucose, and secondly, why the cells should even bother. After lactose is cleaved it actually forms glucose and galactose (easily converted to glucose). In metabolic terms, lactose is just as good a carbon and energy source as glucose. The cAMP level is related not to intracellular glucose concentration but to the rate of glucose transport, which influences the activity of adenylate cyclase. (In addition, glucose transport also leads to direct inhibition of the lactose permease.) As to why *E. coli* works this way, one can only speculate. All enteric bacteria ferment glucose, which suggests they encounter it frequently. It is possible that a small difference in efficiency of transport or metabolism of glucose v. lactose makes it advantageous for cells to regulate the *lac* operon in this way.

The *lac* gene and its derivatives are amenable to

use as a reporter gene in a number of bacterial-based selection techniques such as two hybrid analysis, in which the successful binding of a transcriptional activator to a specific promoter sequence must be determined. In LB plates containing X-gal, the colour change from white colonies to a shade of blue corresponds to about 20–100 β -galactosidase units, while tetrazolium lactose and MacConkey lactose media have a range of 100–1000 units, being most sensitive in the high and low parts of this range respectively. Since MacConkey lactose and tetrazolium lactose media both rely on the products of lactose breakdown, they require the presence of both *lacZ* and *lacY* genes. The many *lac* fusion techniques which include only the *lacZ* gene are thus suited to X-gal plates or ONPG¹¹ liquid broths. ONPG (ortho-Nitrophenyl- β -galactoside) is a colorimetric and spectrophotometric substrate for detection of β -galactosidase activity. This compound is normally colorless. However, if β -galactosidase is present, it hydrolyzes the ONPG molecule into galactose and ortho-nitrophenol. The latter compound has a yellow color that can be used to check for enzyme activity by means of a colorimetric assay (at 420 nm wavelength). β -Galactosidase is required for lactose utilization, so the intensity of the color produced can be used as a measure of the enzymatic rate. Though ONPG mimics lactose and is hydrolyzed by β -galactosidase, it is unable to act as an inducer for the *lac* operon. Without another lactose analog that can act as an inducer, such as isopropyl β -D-1-thiogalactopyranoside (IPTG), β -galactosidase will not be transcribed and ONPG will not be hydrolyzed.

12.2 The *trp* operon

Discovered in 1953 by Jacques Monod and colleagues, the *trp* operon in *E. coli* was the first repressible operon to be discovered. While the *lac* operon can be activated by a chemical (allolactose), the tryptophan (Trp) operon is inhibited by a chemical (tryptophan). This operon contains five structural genes: *trp E*, *trp D*, *trp C*, *trp B*, and *trp A*, which encodes tryptophan synthetase. It also contains a promoter which binds to RNA polymerase and an operator which blocks transcription when bound to the protein synthesized by the repressor gene (*trp R*) that binds to the operator. In the *lac* operon, lactose binds to the repressor protein and prevents it from repressing gene transcription, while in the *trp* operon, tryptophan binds to the repressor protein and enables it to repress gene transcription. Also unlike the *lac* operon, the *trp* operon contains a leader peptide and an attenuator sequence which allows for graded regulation. This is an example of the corepressible model.

¹¹<https://en.wikipedia.org/wiki/Ortho-Nitrophenyl-%CE% B2-galactoside>

12.3 Regulated stages of gene expression

Any step of gene expression may be modulated, from the DNA–RNA transcription step to post-translational modification of a protein. The following is a list of stages where gene expression is regulated, the most extensively utilised point is Transcription Initiation:

- Chromatin domains
- Transcription
- Post-transcriptional modification
- RNA transport
- Translation
- mRNA degradation
- Modification of DNA

In eukaryotes, the accessibility of large regions of DNA can depend on its chromatin structure, which can be altered as a result of histone modifications directed by DNA methylation, ncRNA, or DNA–binding protein. Hence these modifications may up or down regulate the expression of a gene. Some of these modifications that regulate gene expression are inheritable and are referred to as epigenetic regulation.

12.3.1 Regulation of transcription by DNA packing

Transcription of DNA is dictated by its structure. In general, the density of its packing is indicative of the frequency of transcription. Octameric protein complexes called nucleosomes are responsible for the amount of supercoiling of DNA, and these complexes can be temporarily modified by processes such as phosphorylation or more permanently modified by processes such as methylation. Such modifications are considered to be responsible for more or less permanent changes in gene expression levels.

12.3.2 Regulation of transcription by DNA modification

Methylation of DNA is a common method of gene silencing. DNA is typically methylated by methyltransferase enzymes on cytosine nucleotides in a CpG dinucleotide sequence (also called “CpG islands” when densely clustered). Analysis of the pattern of methylation in a given region of DNA (which can be a promoter) can be achieved through a method called bisulfite mapping. Methylated cytosine residues are unchanged by the treatment, whereas unmethylated ones are changed to uracil. The differences are analyzed by DNA sequencing or by methods developed to quantify SNPs, such as Pyrosequencing (Biotage) or MassArray (Sequenom), measuring the relative amounts

of C/T at the CG dinucleotide. Abnormal methylation patterns are thought to be involved in oncogenesis.

Histone acetylation is also an important process in transcription. Histone acetyltransferase enzymes (HATs) such as CREB-binding protein also dissociate the DNA from the histone complex, allowing transcription to proceed. Often, DNA methylation and histone deacetylation work together in gene silencing. The combination of the two seems to be a signal for DNA to be packed more densely, lowering gene expression.[citation needed]

12.3.3 Regulation of RNA polymerase

Regulation of transcription thus controls when transcription occurs and how much RNA is created. Transcription of a gene by RNA polymerase can be regulated by several mechanisms. Specificity factors alter the specificity of RNA polymerase for a given promoter or set of promoters, making it more or less likely to bind to them (i.e., sigma factors used in prokaryotic transcription). Repressors bind to the Operator, coding sequences on the DNA strand that are close to or overlapping the promoter region, impeding RNA polymerase's progress along the strand, thus impeding the expression of the gene. The image to the right demonstrates regulation by a repressor in the *lac* operon. General transcription factors position RNA polymerase at the start of a protein-coding sequence and then release the polymerase to transcribe the mRNA. Activators enhance the interaction between RNA polymerase and a particular promoter, encouraging the expression of the gene. Activators do this by increasing the attraction of RNA polymerase for the promoter, through interactions with subunits of the RNA polymerase or indirectly by changing the structure of the DNA. Enhancers are sites on the DNA helix that are bound by activators in order to loop the DNA bringing a specific promoter to the initiation complex. Enhancers are much more common in eukaryotes than prokaryotes, where only a few examples exist (to date). Silencers are regions of DNA sequences that, when bound by particular transcription factors, can silence expression of the gene.

12.4 Post-transcriptional regulation

After the DNA is transcribed and mRNA is formed, there must be some sort of regulation on how much the mRNA is translated into proteins. Cells do this by modulating the capping, splicing, addition of a Poly(A) Tail, the sequence-specific nuclear export rates, and, in several contexts, sequestration of the RNA transcript. These processes occur in eukaryotes but not in prokaryotes. This modulation is a result of a protein or transcript that, in turn, is regulated and may have an affinity for certain sequences.

12.4.1 Three prime untranslated regions and microRNAs

Three prime untranslated regions (3'-UTRs) of messenger RNAs (mRNAs) often contain regulatory sequences that post-transcriptionally influence gene expression. Such 3'-UTRs often contain both binding sites for microRNAs (miRNAs) as well as for regulatory proteins. By binding to specific sites within the 3'-UTR, miRNAs can decrease gene expression of various mRNAs by either inhibiting translation or directly causing degradation of the transcript. The 3'-UTR also may have silencer regions that bind repressor proteins that inhibit the expression of a mRNA.

The 3'-UTR often contains miRNA response elements (MREs). MREs are sequences to which miRNAs bind. These are prevalent motifs within 3'-UTRs. Among all regulatory motifs within the 3'-UTRs (e.g. including silencer regions), MREs make up about half of the motifs.

As of 2014, the miRBase web site, an archive of miRNA sequences and annotations, listed 28,645 entries in 233 biologic species. Of these, 1,881 miRNAs were in annotated human miRNA loci. miRNAs were predicted to have an average of about four hundred target mRNAs (affecting expression of several hundred genes). Freidman et al. estimate that >45,000 miRNA target sites within human mRNA 3'-UTRs are conserved above background levels, and >60% of human protein-coding genes have been under selective pressure to maintain pairing to miRNAs.

Direct experiments show that a single miRNA can reduce the stability of hundreds of unique mRNAs. Other experiments show that a single miRNA may repress the production of hundreds of proteins, but that this repression often is relatively mild (less than 2-fold).

The effects of miRNA dysregulation of gene expression seem to be important in cancer. For instance, in gastrointestinal cancers, a 2015 paper identified nine miRNAs as epigenetically altered and effective in down-regulating DNA repair enzymes.

The effects of miRNA dysregulation of gene expression also seem to be important in neuropsychiatric disorders, such as schizophrenia, bipolar disorder, major depressive disorder, Parkinson's disease, Alzheimer's disease and autism spectrum disorders.

binding, or protein binding. In both prokaryotes and eukaryotes, a large number of RNA binding proteins exist, which often are directed to their target sequence by the secondary structure of the transcript, which may change depending on certain conditions, such as temperature or presence of a ligand (aptamer). Some transcripts act as ribozymes and self-regulate their expression.

12.5 Regulation of translation

The translation of mRNA can also be controlled by a number of mechanisms, mostly at the level of initiation. Recruitment of the small ribosomal subunit can indeed be modulated by mRNA secondary structure, antisense RNA

Chapter 13

Bacteria, Archaea And Viruses

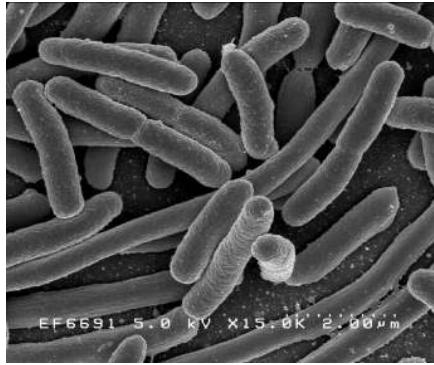


Figure 13.1: Scanning electron micrograph of *Escherichia coli* bacteria.³



Figure 13.2: *Halobacterium sp. strain NRC-1*, each cell about 5 μm long⁴

Bacteria¹ (singular bacterium) and Archaea² (singular archaeon) constitute two domains of single-celled organisms. These microorganisms lack cell nuclei and are therefore prokaryotes.

For much of the 20th century, prokaryotes were regarded as a single group of organisms and classified based on their biochemistry, morphology and metabolism. Microbiologists tried to classify microorganisms based on the structures of their cell walls, their shapes, and the substances they consume. In 1965, Emile Zuckerkandl and Linus Pauling instead proposed using the sequences of the genes in different prokaryotes to work out how they are related to each other. This phylogenetic approach is the main method used today.

Archaea – at that time only the methanogens were known – were first classified separately from bacteria in 1977 by Carl Woese and George E. Fox based on their ribosomal RNA (rRNA) genes. They called these groups the Urkingdoms of Archaeabacteria and Eubacteria, though other researchers treated them as kingdoms or subkingdoms. Woese and Fox gave the first evidence for Archaeabacteria as a separate “line of descent”: 1. lack

of peptidoglycan in their cell walls, 2. two unusual coenzymes, 3. results of 16S ribosomal RNA gene sequencing. To emphasize this difference, Woese, Otto Kandler and Mark Wheelis later proposed reclassifying organisms into three natural domains known as the three-domain system: the Eukarya, the Bacteria and the Archaea, in what is now known as “The Woesian Revolution”.

A virus⁵ is a submicroscopic infectious agent that replicates only inside the living cells of an organism. Viruses infect all types of life forms, from animals and plants to microorganisms, including bacteria and archaea. Since Dmitri Ivanovsky's 1892 article describing a non-bacterial pathogen infecting tobacco plants, and the discovery of the tobacco mosaic virus by Martinus Beijerinck in 1898, more than 6,000 virus species have been described in detail, of the millions of types of viruses in the environment. Viruses are found in almost every ecosystem on Earth and are the most numerous type of biological entity. The study of viruses is known as virology, a subspecialty of microbiology.

When infected, a host cell is forced to rapidly produce thousands of identical copies of the original virus. When

¹<https://en.wikipedia.org/wiki/Bacteria>

²<https://en.wikipedia.org/wiki/Archaea>

⁵<https://en.wikipedia.org/wiki/Virus>

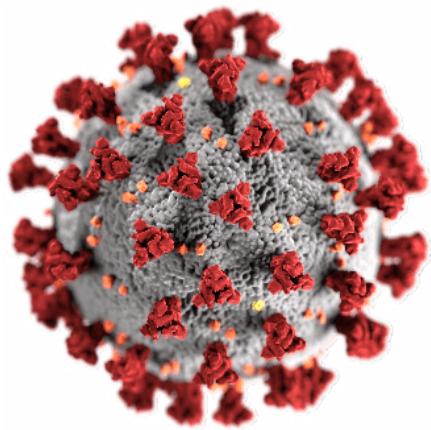


Figure 13.3: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the strain of coronavirus that causes coronavirus disease 2019 (COVID-19), the respiratory illness responsible for the COVID-19 pandemic.⁶

not inside an infected cell or in the process of infecting a cell, viruses exist in the form of independent particles, or virions, consisting of: (i) the genetic material, i.e. long molecules of DNA or RNA that encode the structure of the proteins by which the virus acts; (ii) a protein coat, the capsid, which surrounds and protects the genetic material; and in some cases (iii) an outside envelope of lipids. The shapes of these virus particles range from simple helical and icosahedral forms to more complex structures. Most virus species have virions too small to be seen with an optical microscope as they are one hundredth the size of most bacteria.

The origins of viruses in the evolutionary history of life are unclear: some may have evolved from plasmids—pieces of DNA that can move between cells—while others may have evolved from bacteria. In evolution, viruses are an important means of horizontal gene transfer, which increases genetic diversity in a way analogous to sexual reproduction. Viruses are considered by some biologists to be a life form, because they carry genetic material, reproduce, and evolve through natural selection, although they lack the key characteristics such as cell structure that are generally considered necessary criteria for life. Because they possess some but not all such qualities, viruses have been described as “organisms at the edge of life”, and as self-replicators.

13.1 Bacteria

Bacteria (plural of the New Latin *bacterium*, which is the latinisation of the Greek βακτήριον (*bakterion*), the diminutive of βακτηρία (*bakteria*), meaning “staff, cane”, because the first ones to be discovered were rod-shaped) consti-

tute a large domain of prokaryotic microorganisms. Typically a few micrometres in length, bacteria have a number of shapes, ranging from spheres to rods and spirals. Bacteria were among the first life forms to appear on Earth, and are present in most of its habitats. Bacteria inhabit soil, water, acidic hot springs, radioactive waste, and the deep biosphere of the earth's crust. Bacteria also live in symbiotic and parasitic relationships with plants and animals. Most bacteria have not been characterised, and only about 27 percent of the bacterial phyla have species that can be grown in the laboratory. The study of bacteria is known as bacteriology, a branch of microbiology.

Nearly all animal life is dependent on bacteria for survival as only bacteria and some archaea possess the genes and enzymes necessary to synthesize vitamin B12, also known as cobalamin, and provide it through the food chain. Vitamin B12 is a water-soluble vitamin that is involved in the metabolism of every cell of the human body. It is a co-factor in DNA synthesis, and in both fatty acid and amino acid metabolism. It is particularly important in the normal functioning of the nervous system via its role in the synthesis of myelin.

There are typically 40 million bacterial cells in a gram of soil and a million bacterial cells in a millilitre of fresh water. There are approximately 5×10^{30} bacteria on Earth, forming a biomass which exceeds that of all plants and animals. Bacteria are vital in many stages of the nutrient cycle by recycling nutrients such as the fixation of nitrogen from the atmosphere. The nutrient cycle includes the decomposition of dead bodies; bacteria are responsible for the putrefaction stage in this process. In the biological communities surrounding hydrothermal vents and cold seeps, extremophile bacteria provide the nutrients needed to sustain life by converting dissolved compounds, such as hydrogen sulphide and methane, to energy.

In humans and most animals the largest number of bacteria exist in the gut, and a large number on the skin. The vast majority of the bacteria in the body are rendered harmless by the protective effects of the immune system, though many are beneficial, particularly in the gut flora. However, several species of bacteria are pathogenic and cause infectious diseases, including cholera, syphilis, anthrax, leprosy, and bubonic plague. The most common fatal bacterial diseases are respiratory infections. Tuberculosis alone kills about 2 million people per year, mostly in sub-Saharan Africa. Antibiotics are used to treat bacterial infections and are also used in farming, making antibiotic resistance a growing problem. In industry, bacteria are important in sewage treatment and the breakdown of oil spills, the production of cheese and yogurt through fermentation, the recovery of gold, palladium, copper and other metals in the mining sector, as well as in biotechnol-

ogy, and the manufacture of antibiotics and other chemicals.

Once regarded as plants constituting the class Schizomycetes ("fission fungi"), bacteria are now classified as prokaryotes. Unlike cells of animals and other eukaryotes, bacterial cells do not contain a nucleus and rarely harbour membrane-bound organelles. Although the term bacteria traditionally included all prokaryotes, the scientific classification changed after the discovery in the 1990s that prokaryotes consist of two very different groups of organisms that evolved from an ancient common ancestor. These evolutionary domains are called Bacteria and Archaea.

The ancestors of modern bacteria were unicellular microorganisms that were the first forms of life to appear on Earth, about 4 billion years ago. For about 3 billion years, most organisms were microscopic, and bacteria and archaea were the dominant forms of life. Although bacterial fossils exist, such as stromatolites, their lack of distinctive morphology prevents them from being used to examine the history of bacterial evolution, or to date the time of origin of a particular bacterial species. However, gene sequences can be used to reconstruct the bacterial phylogeny, and these studies indicate that bacteria diverged first from the archaeal/eukaryotic lineage. The most recent common ancestor of bacteria and archaea was probably a hyperthermophile that lived about 2.5 billion–3.2 billion years ago. The earliest life on land may have been bacteria some 3.22 billion years ago.

Bacteria were also involved in the second great evolutionary divergence, that of the archaea and eukaryotes. Here, eukaryotes resulted from the entering of ancient bacteria into endosymbiotic associations with the ancestors of eukaryotic cells, which were themselves possibly related to the Archaea. This involved the engulfment by proto-eukaryotic cells of alphaproteobacterial symbionts to form either mitochondria or hydrogenosomes, which are still found in all known Eukarya (sometimes in highly reduced form, e.g. in ancient "amitochondrial" protzoa). Later, some eukaryotes that already contained mitochondria also engulfed cyanobacteria-like organisms, leading to the formation of chloroplasts in algae and plants. This is known as primary endosymbiosis.

Bacteria display a wide diversity of shapes and sizes, called morphologies. Bacterial cells are about one-tenth the size of eukaryotic cells and are typically 0.5–5.0 micrometres in length. However, a few species are visible to the unaided eye—for example, *Thiomargarita namibiensis* is up to half a millimetre long and *Epulopiscium fishelsoni* reaches 0.7 mm. Among the smallest bacteria are members of the genus *Mycoplasma*, which measure only 0.3 micrometres, as small as the largest viruses. Some bacte-

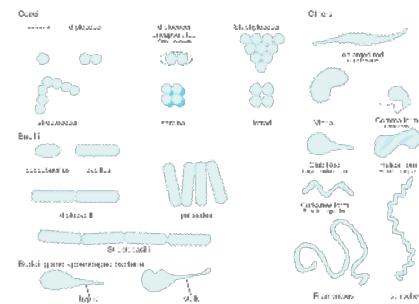


Figure 13.4: Bacteria display various cell morphologies and arrangements⁷

ria may be even smaller, but these ultramicrobacteria are not well-studied.

Most bacterial species are either spherical, called cocci (singular coccus, from Greek κόκκος, grain, seed), or rod-shaped, called bacilli (sing. bacillus, from Latin baculus, stick). Some bacteria, called vibrio, are shaped like slightly curved rods or comma-shaped; others can be spiral-shaped, called spirilla, or tightly coiled, called spirochaetes. A small number of other unusual shapes have been described, such as star-shaped bacteria. This wide variety of shapes is determined by the bacterial cell wall and cytoskeleton, and is important because it can influence the ability of bacteria to acquire nutrients, attach to surfaces, swim through liquids and escape predators.

Many bacterial species exist simply as single cells, others associate in characteristic patterns: *Neisseria* form diploids (pairs), *Streptococcus* form chains, and *Staphylococcus* group together in "bunch of grapes" clusters. Bacteria can also group to form larger multicellular structures, such as the elongated filaments of *Actinobacteria*, the aggregates of *Myxobacteria*, and the complex hyphae of *Streptomyces*. These multicellular structures are often only seen in certain conditions. For example, when starved of amino acids, *Myxobacteria* detect surrounding cells in a process known as quorum sensing, migrate towards each other, and aggregate to form fruiting bodies up to 500 micrometres long and containing approximately 100,000 bacterial cells. In these fruiting bodies, the bacteria perform separate tasks; for example, about one in ten cells migrate to the top of a fruiting body and differentiate into a specialised dormant state called a myxospore, which is more resistant to drying and other adverse environmental conditions.

Bacteria often attach to surfaces and form dense aggregations called biofilms, and larger formations known as microbial mats. These biofilms and mats can range from a few micrometres in thickness to up to half a metre in depth, and may contain multiple

species of bacteria, protists and archaea. Bacteria living in biofilms display a complex arrangement of cells and extracellular components, forming secondary structures, such as microcolonies, through which there are networks of channels to enable better diffusion of nutrients. In natural environments, such as soil or the surfaces of plants, the majority of bacteria are bound to surfaces in biofilms. Biofilms are also important in medicine, as these structures are often present during chronic bacterial infections or in infections of implanted medical devices, and bacteria protected within biofilms are much harder to kill than individual isolated bacteria.

13.1.1 Intracellular Structure

The bacterial cell is surrounded by a cell membrane, which is made primarily of phospholipids. This membrane encloses the contents of the cell and acts as a barrier to hold nutrients, proteins and other essential components of the cytoplasm within the cell. Unlike eukaryotic cells, bacteria usually lack large membrane-bound structures in their cytoplasm such as a nucleus, mitochondria, chloroplasts and the other organelles present in eukaryotic cells. However, some bacteria have protein-bound organelles in the cytoplasm which compartmentalize aspects of bacterial metabolism, such as the carboxysome. Additionally, bacteria have a multi-component cytoskeleton to control the localisation of proteins and nucleic acids within the cell, and to manage the process of cell division.

Many important biochemical reactions, such as energy generation, occur due to concentration gradients across membranes, creating a potential difference analogous to a battery. The general lack of internal membranes in bacteria means these reactions, such as electron transport, occur across the cell membrane between the cytoplasm and the outside of the cell or periplasm. However, in many photosynthetic bacteria the plasma membrane is highly folded and fills most of the cell with layers of light-gathering membrane. These light-gathering complexes may even form lipid-enclosed structures called chlorosomes in green sulfur bacteria.

Bacteria do not have a membrane-bound nucleus, and their genetic material is typically a single circular bacterial chromosome of DNA located in the cytoplasm in an irregularly shaped body called the nucleoid. The nucleoid contains the chromosome with its associated proteins and RNA. Like all other organisms, bacteria contain ribosomes for the production of proteins, but the structure of the bacterial ribosome is different from that of eukaryotes and Archaea.

Some bacteria produce intracellular nutrient storage granules, such as glycogen, polyphosphate, sulfur or polyhydroxyalkanoates. Bacteria such as the photosynthetic

cyanobacteria, produce internal gas vacuoles, which they use to regulate their buoyancy, allowing them to move up or down into water layers with different light intensities and nutrient levels.

13.1.2 Extracellular Structures

Around the outside of the cell membrane is the cell wall. Bacterial cell walls are made of peptidoglycan (also called murein), which is made from polysaccharide chains cross-linked by peptides containing D-amino acids. Bacterial cell walls are different from the cell walls of plants and fungi, which are made of cellulose and chitin, respectively. The cell wall of bacteria is also distinct from that of Archaea, which do not contain peptidoglycan. The cell wall is essential to the survival of many bacteria, and the antibiotic penicillin (produced by a fungus called Penicillium) is able to kill bacteria by inhibiting a step in the synthesis of peptidoglycan.

There are broadly speaking two different types of cell wall in bacteria, that classify bacteria into Gram-positive bacteria and Gram-negative bacteria. The names originate from the reaction of cells to the Gram stain, a long-standing test for the classification of bacterial species.

Gram-positive bacteria possess a thick cell wall containing many layers of peptidoglycan and teichoic acids. In contrast, Gram-negative bacteria have a relatively thin cell wall consisting of a few layers of peptidoglycan surrounded by a second lipid membrane containing lipopolysaccharides and lipoproteins. Most bacteria have the Gram-negative cell wall, and only the Firmicutes and Actinobacteria (previously known as the low G+C and high G+C Gram-positive bacteria, respectively) have the alternative Gram-positive arrangement. These differences in structure can produce differences in antibiotic susceptibility; for instance, vancomycin can kill only Gram-positive bacteria and is ineffective against Gram-negative pathogens, such as *Haemophilus influenzae* or *Pseudomonas aeruginosa*. Some bacteria have cell wall structures that are neither classically Gram-positive or Gram-negative. This includes clinically important bacteria such as Mycobacteria which have a thick peptidoglycan cell wall like a Gram-positive bacterium, but also a second outer layer of lipids.

In many bacteria, an S-layer of rigidly arrayed protein molecules covers the outside of the cell. This layer provides chemical and physical protection for the cell surface and can act as a macromolecular diffusion barrier. S-layers have diverse but mostly poorly understood functions, but are known to act as virulence factors in *Campylobacter* and contain surface enzymes in *Bacillus stearothermophilus*.

Flagella are rigid protein structures, about 20 nanome-

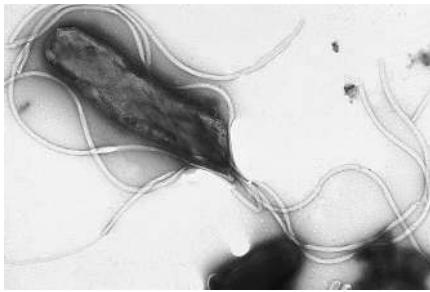


Figure 13.5: *Helicobacter pylori* electron micrograph⁸, showing multiple flagella on the cell surface

tres in diameter and up to 20 micrometres in length, that are used for motility. Flagella are driven by the energy released by the transfer of ions down an electrochemical gradient across the cell membrane.

Fimbriae (sometimes called “attachment pili”) are fine filaments of protein, usually 2–10 nanometres in diameter and up to several micrometres in length. They are distributed over the surface of the cell, and resemble fine hairs when seen under the electron microscope. Fimbriae are believed to be involved in attachment to solid surfaces or to other cells, and are essential for the virulence of some bacterial pathogens. Pili (sing. pilus) are cellular appendages, slightly larger than fimbriae, that can transfer genetic material between bacterial cells in a process called conjugation where they are called conjugation pili or sex pili (see bacterial genetics, below). They can also generate movement where they are called type IV pili.

Glycocalyx is produced by many bacteria to surround their cells, and varies in structural complexity: ranging from a disorganized slime layer of extracellular polymeric substances to a highly structured capsule. These structures can protect cells from engulfment by eukaryotic cells such as macrophages (part of the human immune system). They can also act as antigens and be involved in cell recognition, as well as aiding attachment to surfaces and the formation of biofilms.

The assembly of these extracellular structures is dependent on bacterial secretion systems. These transfer proteins from the cytoplasm into the periplasm or into the environment around the cell. Many types of secretion systems are known and these structures are often essential for the virulence of pathogens, so are intensively studied.

13.1.3 Endospores

Certain genera of Gram-positive bacteria, such as *Bacillus*, *Clostridium*, *Sporohalobacter*, *Anaerobacter*, and *Helio bacterium*, can form highly resistant, dormant structures called endospores. Endospores develop within

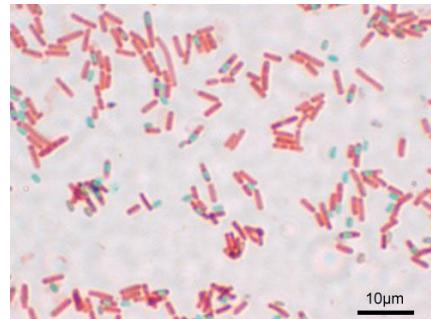


Figure 13.6: A stained preparation of the cell *Bacillus subtilis* showing endospores as green and the vegetative cell as red.⁹

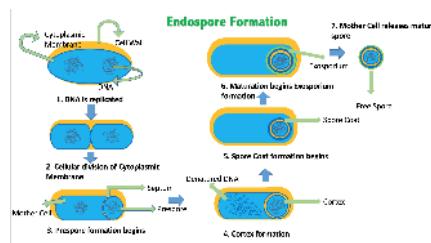


Figure 13.7: Formation of an endospore through the process of sporulation.¹⁰

the cytoplasm of the cell; generally a single endospore develops in each cell. Each endospore contains a core of DNA and ribosomes surrounded by a cortex layer and protected by a multilayer rigid coat composed of peptidoglycan and a variety of proteins.

Endospores show no detectable metabolism and can survive extreme physical and chemical stresses, such as high levels of UV light, gamma radiation, detergents, disinfectants, heat, freezing, pressure, and desiccation. In this dormant state, these organisms may remain viable for millions of years, and endospores even allow bacteria to survive exposure to the vacuum and radiation in space, possibly bacteria could be distributed throughout the Universe by space dust, meteoroids, asteroids, comets, planetoids or via directed panspermia. Endospore-forming bacteria can also cause disease: for example, anthrax can be contracted by the inhalation of *Bacillus anthracis* endospores, and contamination of deep puncture wounds with *Clostridium tetani* endospores causes tetanus.

13.1.4 Metabolism

Bacteria exhibit an extremely wide variety of metabolic types. The distribution of metabolic traits within a group of bacteria has traditionally been used to define their taxonomy, but these traits often do not correspond with modern genetic classifications. Bacterial metabolism is

classified into nutritional groups on the basis of three major criteria: the source of energy, the electron donors used, and the source of carbon used for growth.

Bacteria either derive energy from light using photosynthesis (called phototrophy), or by breaking down chemical compounds using oxidation (called chemotrophy). Chemotrophs use chemical compounds as a source of energy by transferring electrons from a given electron donor to a terminal electron acceptor in a redox reaction. This reaction releases energy that can be used to drive metabolism. Chemotrophs are further divided by the types of compounds they use to transfer electrons. Bacteria that use inorganic compounds such as hydrogen, carbon monoxide, or ammonia as sources of electrons are called lithotrophs, while those that use organic compounds are called organotrophs. The compounds used to receive electrons are also used to classify bacteria: aerobic organisms use oxygen as the terminal electron acceptor, while anaerobic organisms use other compounds such as nitrate, sulfate, or carbon dioxide.

Many bacteria get their carbon from other organic carbon, called heterotrophy. Others such as cyanobacteria and some purple bacteria are autotrophic, meaning that they obtain cellular carbon by fixing carbon dioxide. In unusual circumstances, the gas methane can be used by methanotrophic bacteria as both a source of electrons and a substrate for carbon anabolism.

Table 13.1: Nutritional types in bacterial metabolism

Nutritional Type	Source of Energy	Source of Carbon	Examples
Phototroph	Sunlight	Organic compounds (photoheterotrophs) or carbon fixation (photoautotrophs)	Cyanobacteria, Green sulfur bacteria, Chloroflexi, or Purple bacteria
Lithotrophs	Inorganic compounds	Organic compounds (lithoheterotrophs) or carbon fixation (lithoautotrophs)	Thermodesulfobacteria, Hydrogenophilaceae, or Nitrospirae
Organotroph	Organic compounds	Organic compounds (chemoheterotrophs) or carbon fixation (chemoautotrophs)	Bacillus, Clostridium or Enterobacteriaceae

In many ways, bacterial metabolism provides traits that are useful for ecological stability and for human society. One example is that some bacteria have the ability to fix nitrogen gas using the enzyme nitrogenase. This environmentally important trait can be found in bacteria of most metabolic types listed above. This leads to the ecologically important processes of denitrification, sulfate reduction, and acetogenesis, respectively. Bacterial metabolic processes are also important in biological responses to pollution; for example, sulfate-reducing bacteria are largely responsible for the production of the highly toxic forms of mercury (methyl- and dimethylmercury) in the environment. Non-respiratory anaerobes use fermentation to generate energy and reducing power, secreting metabolic by-products (such as ethanol in brewing) as waste. Facultative anaerobes can switch between fermentation and different terminal electron acceptors depending on the environmental conditions in which they find themselves.

13.1.5 Growth and reproduction

Many bacteria reproduce through binary fission, which is compared to mitosis and meiosis in this image.

Unlike in multicellular organisms, increases in cell size (cell growth) and reproduction by cell division are tightly linked in unicellular organisms. Bacteria grow to a fixed size and then reproduce through binary fission, a form of asexual reproduction. Under optimal conditions, bacteria can grow and divide extremely rapidly, and bacterial populations can double as quickly as every 9.8 minutes. In cell division, two identical clone daughter cells are produced. Some bacteria, while still reproducing asexually, form more complex reproductive structures that help disperse the newly formed daughter cells. Examples include fruiting body formation by Myxobacteria and aerial hyphae formation by Streptomyces, or budding. Budding involves a cell forming a protrusion that breaks away and produces a daughter cell.

In the laboratory, bacteria are usually grown using solid or liquid media. Solid growth media, such as agar plates, are used to isolate pure cultures of a bacterial strain. However, liquid growth media are used when the measurement of growth or large volumes of cells are required. Growth in stirred liquid media occurs as an even cell suspension, making the cultures easy to divide and transfer, although isolating single bacteria from liquid media is difficult. The use of selective media (media with specific nutrients added or deficient, or with antibiotics added) can help identify specific organisms.

Most laboratory techniques for growing bacteria use high levels of nutrients to produce large amounts of cells cheaply and quickly. However, in natural environments, nutrients are limited, meaning that bacteria cannot con-

tinue to reproduce indefinitely. This nutrient limitation has led the evolution of different growth strategies (see r/K selection theory). Some organisms can grow extremely rapidly when nutrients become available, such as the formation of algal (and cyanobacterial) blooms that often occur in lakes during the summer. Other organisms have adaptations to harsh environments, such as the production of multiple antibiotics by Streptomyces that inhibit the growth of competing microorganisms. In nature, many organisms live in communities (e.g., biofilms) that may allow for increased supply of nutrients and protection from environmental stresses. These relationships can be essential for growth of a particular organism or group of organisms (syntrophy).

Bacterial growth follows four phases. When a population of bacteria first enter a high-nutrient environment that allows growth, the cells need to adapt to their new environment. The first phase of growth is the lag phase, a period of slow growth when the cells are adapting to the high-nutrient environment and preparing for fast growth. The lag phase has high biosynthesis rates, as proteins necessary for rapid growth are produced. The second phase of growth is the logarithmic phase, also known as the exponential phase. The log phase is marked by rapid exponential growth. The rate at which cells grow during this phase is known as the growth rate (k), and the time it takes the cells to double is known as the generation time (g). During log phase, nutrients are metabolised at maximum speed until one of the nutrients is depleted and starts limiting growth. The third phase of growth is the stationary phase and is caused by depleted nutrients. The cells reduce their metabolic activity and consume non-essential cellular proteins. The stationary phase is a transition from rapid growth to a stress response state and there is increased expression of genes involved in DNA repair, antioxidant metabolism and nutrient transport. The final phase is the death phase where the bacteria run out of nutrients and die.

13.1.6 Genetics

Most bacteria have a single circular chromosome that can range in size from only 160,000 base pairs in the endosymbiotic bacteria *Carsonella ruddii*, to 12,200,000 base pairs (12.2 Mbp) in the soil-dwelling bacteria *Sorangium cellulosum*. There are many exceptions to this, for example some *Streptomyces* and *Borrelia* species contain a single linear chromosome, while some *Vibrio* species contain more than one chromosome. Bacteria can also contain plasmids, small extra-chromosomal molecules of DNA that may contain genes for various useful functions such as antibiotic resistance, metabolic capabilities, or various virulence factors.

Bacteria genomes usually encode a few hundred to a few thousand genes. The genes in bacterial genomes are usually a single continuous stretch of DNA and although several different types of introns do exist in bacteria, these are much rarer than in eukaryotes.

Bacteria, as asexual organisms, inherit an identical copy of the parent's genomes and are clonal. However, all bacteria can evolve by selection on changes to their genetic material DNA caused by genetic recombination or mutations. Mutations come from errors made during the replication of DNA or from exposure to mutagens. Mutation rates vary widely among different species of bacteria and even among different clones of a single species of bacteria. Genetic changes in bacterial genomes come from either random mutation during replication or "stress-directed mutation", where genes involved in a particular growth-limiting process have an increased mutation rate.

Some bacteria also transfer genetic material between cells. This can occur in three main ways. First, bacteria can take up exogenous DNA from their environment, in a process called transformation. Many bacteria can naturally take up DNA from the environment, while others must be chemically altered in order to induce them to take up DNA. The development of competence in nature is usually associated with stressful environmental conditions, and seems to be an adaptation for facilitating repair of DNA damage in recipient cells. The second way bacteria transfer genetic material is by transduction, when the integration of a bacteriophage introduces foreign DNA into the chromosome. Many types of bacteriophage exist, some simply infect and lyse their host bacteria, while others insert into the bacterial chromosome. Bacteria resist phage infection through restriction modification systems that degrade foreign DNA, and a system that uses CRISPR sequences to retain fragments of the genomes of phage that the bacteria have come into contact with in the past, which allows them to block virus replication through a form of RNA interference. The third method of gene transfer is conjugation, whereby DNA is transferred through direct cell contact. In ordinary circumstances, transduction, conjugation, and transformation involve transfer of DNA between individual bacteria of the same species, but occasionally transfer may occur between individuals of different bacterial species and this may have significant consequences, such as the transfer of antibiotic resistance. In such cases, gene acquisition from other bacteria or the environment is called horizontal gene transfer and may be common under natural conditions.

13.1.7 Movement

Many bacteria are motile (able to move themselves) and do so using a variety of mechanisms. The best studied of these are flagella, long filaments that are turned by a motor at the base to generate propeller-like movement. The bacterial flagellum is made of about 20 proteins, with approximately another 30 proteins required for its regulation and assembly. The flagellum is a rotating structure driven by a reversible motor at the base that uses the electrochemical gradient across the membrane for power.

Bacteria can use flagella in different ways to generate different kinds of movement. Many bacteria (such as *E. coli*) have two distinct modes of movement: forward movement (swimming) and tumbling. The tumbling allows them to reorient and makes their movement a three-dimensional random walk. Bacterial species differ in the number and arrangement of flagella on their surface; some have a single flagellum (monotrichous), a flagellum at each end (amphitrichous), clusters of flagella at the poles of the cell (lophotrichous), while others have flagella distributed over the entire surface of the cell (peritrichous). The flagella of a unique group of bacteria, the spirochaetes, are found between two membranes in the periplasmic space. They have a distinctive helical body that twists about as it moves.

Two other types of bacterial motion are called twitching motility that relies on a structure called the type IV pilus, and gliding motility, that uses other mechanisms. In twitching motility, the rod-like pilus extends out from the cell, binds some substrate, and then retracts, pulling the cell forward.

Motile bacteria are attracted or repelled by certain stimuli in behaviours called taxes: these include chemotaxis, phototaxis, energy taxis, and magnetotaxis. In one peculiar group, the myxobacteria, individual bacteria move together to form waves of cells that then differentiate to form fruiting bodies containing spores. The myxobacteria move only when on solid surfaces, unlike *E. coli*, which is motile in liquid or solid media.

Several *Listeria* and *Shigella* species move inside host cells by usurping the cytoskeleton, which is normally used to move organelles inside the cell. By promoting actin polymerisation at one pole of their cells, they can form a kind of tail that pushes them through the host cell's cytoplasm.

13.1.8 Communication

Bacteria often function as multicellular aggregates known as biofilms, exchanging a variety of molecular signals for inter-cell communication, and engaging in coordinated multicellular behaviour.

The communal benefits of multicellular cooperation

include a cellular division of labour, accessing resources that cannot effectively be used by single cells, collectively defending against antagonists, and optimising population survival by differentiating into distinct cell types. For example, bacteria in biofilms can have more than 500 times increased resistance to antibacterial agents than individual "planktonic" bacteria of the same species.

One type of inter-cellular communication by a molecular signal is called quorum sensing, which serves the purpose of determining whether there is a local population density that is sufficiently high that it is productive to invest in processes that are only successful if large numbers of similar organisms behave similarly, as in excreting digestive enzymes or emitting light.

Quorum sensing allows bacteria to coordinate gene expression, and enables them to produce, release and detect autoinducers or pheromones which accumulate with the growth in cell population.

13.1.9 Classification and identification

Classification seeks to describe the diversity of bacterial species by naming and grouping organisms based on similarities. Bacteria can be classified on the basis of cell structure, cellular metabolism or on differences in cell components, such as DNA, fatty acids, pigments, antigens and quinones. While these schemes allowed the identification and classification of bacterial strains, it was unclear whether these differences represented variation between distinct species or between strains of the same species. This uncertainty was due to the lack of distinctive structures in most bacteria, as well as lateral gene transfer between unrelated species. Due to lateral gene transfer, some closely related bacteria can have very different morphologies and metabolisms. To overcome this uncertainty, modern bacterial classification emphasises molecular systematics, using genetic techniques such as guanine cytosine ratio determination, genome–genome hybridisation, as well as sequencing genes that have not undergone extensive lateral gene transfer, such as the rRNA gene. Classification of bacteria is determined by publication in the International Journal of Systematic Bacteriology, and Bergey's Manual of Systematic Bacteriology. The International Committee on Systematic Bacteriology (ICSB) maintains international rules for the naming of bacteria and taxonomic categories and for the ranking of them in the International Code of Nomenclature of Bacteria.

The term "bacteria" was traditionally applied to all microscopic, single-cell prokaryotes. However, molecular systematics showed prokaryotic life to consist of two separate domains, originally called Eubacteria and Archaeabacteria, but now called Bacteria and Archaea that evolved independently from an ancient common ancestor. The

archaea and eukaryotes are more closely related to each other than either is to the bacteria. These two domains, along with Eukarya, are the basis of the three-domain system, which is currently the most widely used classification system in microbiology. However, due to the relatively recent introduction of molecular systematics and a rapid increase in the number of genome sequences that are available, bacterial classification remains a changing and expanding field. For example, Cavalier-Smith argued that the Archaea and Eukaryotes evolved from Gram-positive bacteria.

The identification of bacteria in the laboratory is particularly relevant in medicine, where the correct treatment is determined by the bacterial species causing an infection. Consequently, the need to identify human pathogens was a major impetus for the development of techniques to identify bacteria.

The Gram stain, developed in 1884 by Hans Christian Gram, characterises bacteria based on the structural characteristics of their cell walls. The thick layers of peptidoglycan in the "Gram-positive" cell wall stain purple, while the thin "Gram-negative" cell wall appears pink. By combining morphology and Gram-staining, most bacteria can be classified as belonging to one of four groups (Gram-positive cocci, Gram-positive bacilli, Gram-negative cocci and Gram-negative bacilli). Some organisms are best identified by stains other than the Gram stain, particularly mycobacteria or Nocardia, which show acid-fastness on Ziehl-Neelsen or similar stains. Other organisms may need to be identified by their growth in special media, or by other techniques, such as serology.

Culture techniques are designed to promote the growth and identify particular bacteria, while restricting the growth of the other bacteria in the sample. Often these techniques are designed for specific specimens; for example, a sputum sample will be treated to identify organisms that cause pneumonia, while stool specimens are cultured on selective media to identify organisms that cause diarrhoea, while preventing growth of non-pathogenic bacteria. Specimens that are normally sterile, such as blood, urine or spinal fluid, are cultured under conditions designed to grow all possible organisms. Once a pathogenic organism has been isolated, it can be further characterised by its morphology, growth patterns (such as aerobic or anaerobic growth), patterns of hemolysis, and staining.

As with bacterial classification, identification of bacteria is increasingly using molecular methods. Diagnostics using DNA-based tools, such as polymerase chain reaction, are increasingly popular due to their specificity and speed, compared to culture-based methods. These methods also allow the detection and identification of "viable but noncul-

turable" cells that are metabolically active but non-dividing. However, even using these improved methods, the total number of bacterial species is not known and cannot even be estimated with any certainty. Following present classification, there are a little less than 9,300 known species of prokaryotes, which includes bacteria and archaea; but attempts to estimate the true number of bacterial diversity have ranged from 10⁷ to 10⁹ total species—and even these diverse estimates may be off by many orders of magnitude.

13.1.10 Interactions With Other Organisms

Despite their apparent simplicity, bacteria can form complex associations with other organisms. These symbiotic associations can be divided into parasitism, mutualism and commensalism. Due to their small size, commensal bacteria are ubiquitous and grow on animals and plants exactly as they will grow on any other surface. However, their growth can be increased by warmth and sweat, and large populations of these organisms in humans are the cause of body odour.

Some species of bacteria kill and then consume other microorganisms, these species are called predatory bacteria. These include organisms such as *Myxococcus xanthus*, which forms swarms of cells that kill and digest any bacteria they encounter. Other bacterial predators either attach to their prey in order to digest them and absorb nutrients, such as *Vampirovibrio chlorellavorus*, or invade another cell and multiply inside the cytosol, such as *Daptobacter*. These predatory bacteria are thought to have evolved from saprophages that consumed dead microorganisms, through adaptations that allowed them to entrap and kill other organisms.

Certain bacteria form close spatial associations that are essential for their survival. One such mutualistic association, called interspecies hydrogen transfer, occurs between clusters of anaerobic bacteria that consume organic acids, such as butyric acid or propionic acid, and produce hydrogen, and methanogenic Archaea that consume hydrogen. The bacteria in this association are unable to consume the organic acids as this reaction produces hydrogen that accumulates in their surroundings. Only the intimate association with the hydrogen-consuming Archaea keeps the hydrogen concentration low enough to allow the bacteria to grow.

In soil, microorganisms that reside in the rhizosphere (a zone that includes the root surface and the soil that adheres to the root after gentle shaking) carry out nitrogen fixation, converting nitrogen gas to nitrogenous compounds. This serves to provide an easily absorbable form of nitrogen for many plants, which cannot fix nitrogen themselves. Many other bacteria are found as symbionts in humans and other organisms. For example, the presence

of over 1,000 bacterial species in the normal human gut flora of the intestines can contribute to gut immunity, synthesise vitamins, such as folic acid, vitamin K and biotin, convert sugars to lactic acid (see *Lactobacillus*), as well as fermenting complex undigestible carbohydrates. The presence of this gut flora also inhibits the growth of potentially pathogenic bacteria (usually through competitive exclusion) and these beneficial bacteria are consequently sold as probiotic dietary supplements.

13.1.11 Bacteria as Pathogens

If bacteria form a parasitic association with other organisms, they are classed as pathogens. Pathogenic bacteria are a major cause of human death and disease and cause infections such as tetanus (Caused by *Clostridium tetani*), typhoid fever, diphtheria, syphilis, cholera, foodborne illness, leprosy (caused by *Micobacterium leprae*) and tuberculosis (Caused by *Mycobacterium tuberculosis*). A pathogenic cause for a known medical disease may only be discovered many years after, as was the case with *Helicobacter pylori* and peptic ulcer disease. Bacterial diseases are also important in agriculture, with bacteria causing leaf spot, fire blight and wilts in plants, as well as Johne's disease, mastitis, salmonella and anthrax in farm animals.

Each species of pathogen has a characteristic spectrum of interactions with its human hosts. Some organisms, such as *Staphylococcus* or *Streptococcus*, can cause skin infections, pneumonia, meningitis and even overwhelming sepsis, a systemic inflammatory response producing shock, massive vasodilation and death. Yet these organisms are also part of the normal human flora and usually exist on the skin or in the nose without causing any disease at all. Other organisms invariably cause disease in humans, such as the *Rickettsia*, which are obligate intracellular parasites able to grow and reproduce only within the cells of other organisms. One species of *Rickettsia* causes typhus, while another causes Rocky Mountain spotted fever. Chlamydia, another phylum of obligate intracellular parasites, contains species that can cause pneumonia, or urinary tract infection and may be involved in coronary heart disease. Finally, some species, such as *Pseudomonas aeruginosa*, *Burkholderia cenocepacia*, and *Mycobacterium avium*, are opportunistic pathogens and cause disease mainly in people suffering from immunosuppression or cystic fibrosis.

Bacterial infections may be treated with antibiotics, which are classified as bacteriocidal if they kill bacteria, or bacteriostatic if they just prevent bacterial growth. There are many types of antibiotics and each class inhibits a process that is different in the pathogen from that found in the host. An example of how antibiotics produce selective toxicity are chloramphenicol and puromycin, which inhibit

the bacterial ribosome, but not the structurally different eukaryotic ribosome. Antibiotics are used both in treating human disease and in intensive farming to promote animal growth, where they may be contributing to the rapid development of antibiotic resistance in bacterial populations. Infections can be prevented by antiseptic measures such as sterilising the skin prior to piercing it with the needle of a syringe, and by proper care of indwelling catheters. Surgical and dental instruments are also sterilised to prevent contamination by bacteria. Disinfectants such as bleach are used to kill bacteria or other pathogens on surfaces to prevent contamination and further reduce the risk of infection.

13.1.12 Significance in Technology And Industry

Bacteria, often lactic acid bacteria, such as *Lactobacillus* and *Lactococcus*, in combination with yeasts and moulds, have been used for thousands of years in the preparation of fermented foods, such as cheese, pickles, soy sauce, sauerkraut, vinegar, wine and yogurt.

The ability of bacteria to degrade a variety of organic compounds is remarkable and has been used in waste processing and bioremediation. Bacteria capable of digesting the hydrocarbons in petroleum are often used to clean up oil spills. Fertiliser was added to some of the beaches in Prince William Sound in an attempt to promote the growth of these naturally occurring bacteria after the 1989 Exxon Valdez oil spill. These efforts were effective on beaches that were not too thickly covered in oil. Bacteria are also used for the bioremediation of industrial toxic wastes. In the chemical industry, bacteria are most important in the production of enantiomerically pure chemicals for use as pharmaceuticals or agrochemicals.

Bacteria can also be used in the place of pesticides in the biological pest control. This commonly involves *Bacillus thuringiensis* (also called BT), a Gram-positive, soil dwelling bacterium. Subspecies of this bacteria are used as a Lepidopteran-specific insecticides under trade names such as Dipel and Thuricide. Because of their specificity, these pesticides are regarded as environmentally friendly, with little or no effect on humans, wildlife, pollinators and most other beneficial insects.

Because of their ability to quickly grow and the relative ease with which they can be manipulated, bacteria are the workhorses for the fields of molecular biology, genetics and biochemistry. By making mutations in bacterial DNA and examining the resulting phenotypes, scientists can determine the function of genes, enzymes and metabolic pathways in bacteria, then apply this knowledge to more complex organisms. This aim of understanding

the biochemistry of a cell reaches its most complex expression in the synthesis of huge amounts of enzyme kinetic and gene expression data into mathematical models of entire organisms. This is achievable in some well-studied bacteria, with models of *Escherichia coli* metabolism now being produced and tested. This understanding of bacterial metabolism and genetics allows the use of biotechnology to bioengineer bacteria for the production of therapeutic proteins, such as insulin, growth factors, or antibodies.

Because of their importance for research in general, samples of bacterial strains are isolated and preserved in Biological Resource Centers. This ensures the availability of the strain to scientists worldwide.

13.2 Archaea

The word archaea comes from the Ancient Greek ἀρχαῖα, meaning “ancient things”, as the first representatives of the domain Archaea were methanogens and it was assumed that their metabolism reflected Earth’s primitive atmosphere and the organisms’ antiquity, but as new habitats were studied, more organisms were discovered. Extreme halophilic and hyperthermophilic microbes were also included in Archaea. For a long time, archaea were seen as extremophiles that exist only in extreme habitats such as hot springs and salt lakes, but by the end of the 20th century, archaea had been identified in non-extreme environments as well. Today, they are known to be a large and diverse group of organisms abundantly distributed throughout nature. This new appreciation of the importance and ubiquity of archaea came from using polymerase chain reaction (PCR) to detect prokaryotes from environmental samples (such as water or soil) by multiplying their ribosomal genes. This allows the detection and identification of organisms that have not been cultured in the laboratory.

Archaeal cells have unique properties separating them from the other two domains, Bacteria and Eukaryota. Archaea are further divided into multiple recognized phyla. Classification is difficult because most have not been isolated in the laboratory and have been detected only by analysis of their nucleic acids in samples from their environment.

Archaea and bacteria are generally similar in size and shape, although a few archaea have very different shapes, such as the flat and square cells of *Haloquadratum walsbyi*. Despite this morphological similarity to bacteria, archaea possess genes and several metabolic pathways that are more closely related to those of eukaryotes, notably for the enzymes involved in transcription and translation. Other aspects of archaeal biochemistry are unique, such as their reliance on ether lipids in their cell membranes, in-

cluding archaeols. Archaea use more energy sources than eukaryotes: these range from organic compounds, such as sugars, to ammonia, metal ions or even hydrogen gas. Salt-tolerant archaea (the Haloarchaea) use sunlight as an energy source, and other species of archaea fix carbon, but unlike plants and cyanobacteria, no known species of archaea does both. Archaea reproduce asexually by binary fission, fragmentation, or budding; unlike bacteria, no known species of Archaea form endospores. The first observed archaea were extremophiles, living in extreme environments, such as hot springs and salt lakes with no other organisms. Improved detection tools led to the discovery of archaea in almost every habitat, including soil, oceans, and marshlands. Archaea are particularly numerous in the oceans, and the archaea in plankton may be one of the most abundant groups of organisms on the planet.

Archaea are a major part of Earth's life. They are part of the microbiota of all organisms. In the human microbiota, they are important in the gut, mouth, and on the skin. Their morphological, metabolic, and geographical diversity permits them to play multiple ecological roles: carbon fixation; nitrogen cycling; organic compound turnover; and maintaining microbial symbiotic and syntrophic communities, for example.

No clear examples of archaeal pathogens or parasites are known. Instead they are often mutualists or commensals, such as the methanogens (methane-producing strains) that inhabit the gastrointestinal tract in humans and ruminants, where their vast numbers aid digestion. Methanogens are also used in biogas production and sewage treatment, and biotechnology exploits enzymes from extremophile archaea that can endure high temperatures and organic solvents.

13.2.1 Origin and evolution

The age of the Earth is about 4.54 billion years. Scientific evidence suggests that life began on Earth at least 3.5 billion years ago. The earliest evidence for life on Earth is graphite found to be biogenic in 3.7-billion-year-old metasedimentary rocks discovered in Western Greenland and microbial mat fossils found in 3.48-billion-year-old sandstone discovered in Western Australia. In 2015, possible remains of biotic matter were found in 4.1-billion-year-old rocks in Western Australia.

Although probable prokaryotic cell fossils date to almost 3.5 billion years ago, most prokaryotes do not have distinctive morphologies, and fossil shapes cannot be used to identify them as archaea. Instead, chemical fossils of unique lipids are more informative because such compounds do not occur in other organisms. Some publications suggest that archaeal or eukaryotic lipid remains are present in shales dating from 2.7 billion

years ago; though such data have since been questioned. These lipids have also been detected in even older rocks from west Greenland. The oldest such traces come from the Isua district, which includes Earth's oldest known sediments, formed 3.8 billion years ago. The archaeal lineage may be the most ancient that exists on Earth.

Woese argued that the Bacteria, Archaea, and Eukaryotes represent separate lines of descent that diverged early on from an ancestral colony of organisms. One possibility is that this occurred before the evolution of cells, when the lack of a typical cell membrane allowed unrestricted lateral gene transfer, and that the common ancestors of the three domains arose by fixation of specific subsets of genes. It is possible that the last common ancestor of bacteria and archaea was a thermophile, which raises the possibility that lower temperatures are "extreme environments" for archaea, and organisms that live in cooler environments appeared only later. Since archaea and bacteria are no more related to each other than they are to eukaryotes, the term prokaryote may suggest a false similarity between them. However, structural and functional similarities between lineages often occur because of shared ancestral traits or evolutionary convergence. These similarities are known as a grade, and prokaryotes are best thought of as a grade of life, characterized by such features as an absence of membrane-bound organelles.

The following table (13.2 compares some major characteristics of the three domains, to illustrate their similarities and differences.

Table 13.2: A comparison of major characteristics of the domains of Bacteria, Archaea and Eukaryia

Property	Bacteria	Archaea	Eukarya
Cell membrane	Ether-linked lipids	Ester-linked lipids	Ester-linked lipids
Cell wall	Pseudopeptidoglycan; glycoprotein, or S-layer	Peptidoglycan, S-layer, or no cell wall	Various structures
Gene structure	Circular chromosomes, similar translation and transcription to Eukarya	Circular chromosomes, unique translation and transcription	Multiple, linear chromosomes, but translation and transcription similar to Archaea
Internal cell structure	No membrane-bound organelles ([62]) or nucleus	No membrane-bound organelles or nucleus	Membrane-bound organelles and nucleus
Metabolism[63]	Various, including diazotrophy, with methanogenesis unique to Archaea	Various, including photosynthesis, aerobic and anaerobic respiration, fermentation, diazotrophy, and autotrophy	Photosynthesis, cellular respiration, and fermentation; no diazotrophy
Reproduction	Asexual reproduction, horizontal gene transfer	Asexual reproduction, horizontal gene transfer	Sexual and asexual reproduction
Protein synthesis initiation	Methionine	Formylmethionine	Methionine
RNA polymerase	Many	One	Many
Toxin	Sensitive to diphtheria toxin	Resistant to diphtheria toxin	Sensitive to diphtheria toxin

Archaea were split off as a third domain because of the large differences in their ribosomal RNA structure. The particular molecule 16S rRNA is key to the production of proteins in all organisms. Because this function is so central to life, organisms with mutations in their 16S rRNA are unlikely to survive, leading to great (but not absolute) stability in the structure of this nucleotide over generations. 16S rRNA is large enough to show organism-specific variations, but still small enough to be compared quickly. In 1977, Carl Woese, a microbiologist studying the genetic sequences of organisms, developed a new comparison method that involved splitting the RNA into fragments that could be sorted and compared with other fragments from other organisms. The more similar the patterns between species, the more closely they are related.

Woese used his new rRNA comparison method to categorize and contrast different organisms. He compared a variety of species and happened upon a group of methanogens with rRNA vastly different from any known prokaryotes or eukaryotes. These methanogens were much more similar to each other than to other organisms, leading Woese to propose the new domain of Archaea. His experiments showed that the archaea were genetically more similar to eukaryotes than prokaryotes, even though they were more similar to prokaryotes in structure. This led to the conclusion that Archaea and Eukarya shared a common ancestor more recent than Eukarya and Bacteria. The development of the nucleus occurred after the split between Bacteria and this common ancestor.

One property unique to archaea is the abundant use of ether-linked lipids in their cell membranes. Ether linkages are more chemically stable than the ester linkages found in bacteria and eukarya, which may be a contributing factor to the ability of many archaea to survive in extreme environments that place heavy stress on cell membranes, such as extreme heat and salinity. Comparative analysis of archaeal genomes has also identified several molecular conserved signature indels and signature proteins uniquely present in either all archaea or different main groups within archaea. Another unique feature of archaea, found in no other organisms, is methanogenesis (the metabolic production of methane). Methanogenic archaea play a pivotal role in ecosystems with organisms that derive energy from oxidation of methane, many of which are bacteria, as they are often a major source of methane in such environments and can play a role as primary producers. Methanogens also play a critical role in the carbon cycle, breaking down organic carbon into methane, which is also a major greenhouse gas.

13.2.2 Relationship to Bacteria

The relationships among the three domains are of central importance for understanding the origin of life. Most of the metabolic pathways, which are the object of the majority of an organism's genes, are common between Archaea and Bacteria, while most genes involved in genome expression are common between Archaea and Eukarya. Within prokaryotes, archaeal cell structure is most similar to that of gram-positive bacteria, largely because both have a single lipid bilayer and usually contain a thick sacculus (exoskeleton) of varying chemical composition. In some phylogenetic trees based upon different gene/protein sequences of prokaryotic homologs, the archaeal homologs are more closely related to those of gram-positive bacteria. Archaea and gram-positive bacteria also share conserved indels in a number of important proteins, such as Hsp70 and glutamine synthetase I; but the phylogeny of these genes was interpreted to reveal interdomain gene transfer, and might not reflect the organismal relationship(s).

It has been proposed that the archaea evolved from gram-positive bacteria in response to antibiotic selection pressure. This is suggested by the observation that archaea are resistant to a wide variety of antibiotics that are produced primarily by gram-positive bacteria, and that these antibiotics act primarily on the genes that distinguish archaea from bacteria. The proposal is that the selective pressure towards resistance generated by the gram-positive antibiotics was eventually sufficient to cause extensive changes in many of the antibiotics' target genes, and that these strains represented the common ancestors of present-day Archaea. The evolution of Archaea in response to antibiotic selection, or any other competitive selective pressure, could also explain their adaptation to extreme environments (such as high temperature or acidity) as the result of a search for unoccupied niches to escape from antibiotic-producing organisms; Cavalier-Smith has made a similar suggestion. This proposal is also supported by other work investigating protein structural relationships and studies that suggest that gram-positive bacteria may constitute the earliest branching lineages within the prokaryotes.

13.2.3 Relation to Eukaryotes

The evolutionary relationship between archaea and eukaryotes remains unclear. Aside from the similarities in cell structure and function that are discussed below, many genetic trees group the two.

Complicating factors include claims that the relationship between eukaryotes and the archaeal phylum Crenarchaeota is closer than the relationship between the Euryarchaeota and the phylum Crenarchaeota and the

presence of archaea-like genes in certain bacteria, such as *Thermotoga maritima*, from horizontal gene transfer. The standard hypothesis states that the ancestor of the eukaryotes diverged early from the Archaea, and that eukaryotes arose through fusion of an archaean and eubacterium, which became the nucleus and cytoplasm; this hypothesis explains various genetic similarities but runs into difficulties explaining cell structure. An alternative hypothesis, the eocyte hypothesis, posits that Eukaryota emerged relatively late from the Archaea.

A lineage of archaea discovered in 2015, *Lokiarchaeum* (of proposed new Phylum “*Lokiarchaeota*”), named for a hydrothermal vent called Loki’s Castle in the Arctic Ocean, was found to be the most closely related to eukaryotes known at that time. It has been called a transitional organism between prokaryotes and eukaryotes.

Several sister phyla of “*Lokiarchaeota*” have since been found (“*Thorarchaeota*”, “*Odinarchaeota*”, “*Heimdallarchaeota*”), all together comprising a newly proposed supergroup Asgard, which may appear as a sister taxon to *Proteoarchaeota*.

Details of the relation of Asgard members and eukaryotes are still under consideration, although, in January 2020, scientists reported that *Candidatus Prometheoarchaeum syntrophicum*, a type of Asgard archaea, may be a possible link between simple prokaryotic and complex eukaryotic microorganisms about two billion years ago.

13.2.4 Morphology

Individual archaea range from 0.1 micrometers (μm) to over 15 μm in diameter, and occur in various shapes, commonly as spheres, rods, spirals or plates. Other morphologies in the Crenarchaeota include irregularly shaped lobed cells in *Sulfolobus*, needle-like filaments that are less than half a micrometer in diameter in *Thermofilum*, and almost perfectly rectangular rods in *Thermoproteus* and *Pyrobaculum*. Archaea in the genus *Haloquadratum* such as *Haloquadratum walsbyi* are flat, square specimens that live in hypersaline pools. These unusual shapes are probably maintained by both their cell walls and a prokaryotic cytoskeleton. Proteins related to the cytoskeleton components of other organisms exist in archaea, and filaments form within their cells, but in contrast with other organisms, these cellular structures are poorly understood. In *Thermoplasma* and *Ferroplasma* the lack of a cell wall means that the cells have irregular shapes, and can resemble amoebae.

Some species form aggregates or filaments of cells up to 200 μm long. These organisms can be prominent in biofilms. Notably, aggregates of *Thermococcus coalescens* cells fuse together in culture, forming single giant cells. Ar-

chaea in the genus *Pyrodictium* produce an elaborate multi-cell colony involving arrays of long, thin hollow tubes called cannulae that stick out from the cells’ surfaces and connect them into a dense bush-like agglomeration. The function of these cannulae is not settled, but they may allow communication or nutrient exchange with neighbors. Multi-species colonies exist, such as the “string-of-pearls” community that was discovered in 2001 in a German swamp. Round whitish colonies of a novel *Euryarchaeota* species are spaced along thin filaments that can range up to 15 centimetres (5.9 in) long; these filaments are made of a particular bacteria species.

13.2.5 Structure, Composition Development, And Operation

Archaea and bacteria have generally similar cell structure, but cell composition and organization set the archaea apart. Like bacteria, archaea lack interior membranes and organelles. Like bacteria, the cell membranes of archaea are usually bounded by a cell wall and they swim using one or more flagella. Structurally, archaea are most similar to gram-positive bacteria. Most have a single plasma membrane and cell wall, and lack a periplasmic space; the exception to this general rule is *Ignicoccus*, which possess a particularly large periplasm that contains membrane-bound vesicles and is enclosed by an outer membrane.

13.2.6 Cell Wall And Flagella

Most archaea (but not *Thermoplasma* and *Ferroplasma*) possess a cell wall. In most archaea the wall is assembled from surface-layer proteins, which form an S-layer. An S-layer is a rigid array of protein molecules that cover the outside of the cell (like chain mail). This layer provides both chemical and physical protection, and can prevent macromolecules from contacting the cell membrane. Unlike bacteria, archaea lack peptidoglycan in their cell walls. *Methanobacteriales* do have cell walls containing pseudopeptidoglycan, which resembles eubacterial peptidoglycan in morphology, function, and physical structure, but pseudopeptidoglycan is distinct in chemical structure; it lacks D-amino acids and N-acetylmuramic acid, substituting the latter with N-Acetyltaulosaminuronic acid.

Archaeal flagella are known as archaella, that operate like bacterial flagella – their long stalks are driven by rotary motors at the base. These motors are powered by a proton gradient across the membrane, but archaella are notably different in composition and development. The two types of flagella evolved from different ancestors. The bacterial flagellum shares a common ancestor with the type III secretion system, while archaeal flagella appear to have

evolved from bacterial type IV pili. In contrast with the bacterial flagellum, which is hollow and assembled by sub-units moving up the central pore to the tip of the flagella, archaeal flagella are synthesized by adding subunits at the base.

13.2.7 Membranes

Archaeal membranes are made of molecules that are distinctly different from those in all other life forms, showing that archaea are related only distantly to bacteria and eukaryotes. In all organisms, cell membranes are made of molecules known as phospholipids. These molecules possess both a polar part that dissolves in water (the phosphate "head"), and a "greasy" non-polar part that does not (the lipid tail). These dissimilar parts are connected by a glycerol moiety. In water, phospholipids cluster, with the heads facing the water and the tails facing away from it. The major structure in cell membranes is a double layer of these phospholipids, which is called a lipid bilayer.

The phospholipids of archaea are unusual in four ways:

- They have membranes composed of glycerol-ether lipids, whereas bacteria and eukaryotes have membranes composed mainly of glycerol-ester lipids. The difference is the type of bond that joins the lipids to the glycerol moiety; the two types are shown in yellow in the figure at the right. In ester lipids this is an ester bond, whereas in ether lipids this is an ether bond.
- The stereochemistry of the archaeal glycerol moiety is the mirror image of that found in other organisms. The glycerol moiety can occur in two forms that are mirror images of one another, called enantiomers. Just as a right hand does not fit easily into a left-handed glove, enantiomers of one type generally cannot be used or made by enzymes adapted for the other. The archaeal phospholipids are built on a backbone of sn-glycerol-1-phosphate, which is an enantiomer of sn-glycerol-3-phosphate, the phospholipid backbone found in bacteria and eukaryotes. This suggests that archaea use entirely different enzymes for synthesizing phospholipids as compared to bacteria and eukaryotes. Such enzymes developed very early in life's history, indicating an early split from the other two domains.
- Archaeal lipid tails differ from those of other organisms in that they are based upon long isoprenoid chains with multiple side-branches, sometimes with cyclopropane or cyclohexane rings. By contrast, the fatty acids in the membranes of other organisms have straight chains without side branches or rings. Although isoprenoids play an important role in the biochemistry of many organisms, only the archaea

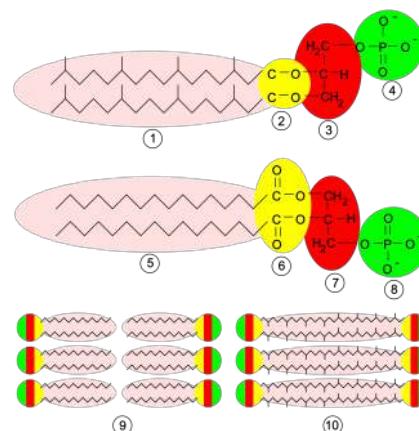


Figure 13.8: A comparison of the membrane structures of Archaea, Bacteria and Eukarya. Top, an archaeal phospholipid: 1, isoprene chains; 2, ether linkages; 3, L-glycerol moiety; 4, phosphate group. Middle, a bacterial or eukaryotic phospholipid: 5, fatty acid chains; 6, ester linkages; 7, D-glycerol moiety; 8, phosphate group. Bottom: 9, lipid bilayer of bacteria and eukaryotes; 10, lipid monolayer of some archaea.¹¹

use them to make phospholipids. These branched chains may help prevent archaeal membranes from leaking at high temperatures.

- In some archaea, the lipid bilayer is replaced by a monolayer. In effect, the archaea fuse the tails of two phospholipid molecules into a single molecule with two polar heads (a bolaamphiphile); this fusion may make their membranes more rigid and better able to resist harsh environments. For example, the lipids in Ferroplasma are of this type, which is thought to aid this organism's survival in its highly acidic habitat.

13.2.8 Metabolism

Archaea exhibit a great variety of chemical reactions in their metabolism and use many sources of energy. These reactions are classified into nutritional groups, depending on energy and carbon sources. Some archaea obtain energy from inorganic compounds such as sulfur or ammonia (they are chemotrophs). These include nitrifiers, methanogens and anaerobic methane oxidisers. In these reactions one compound passes electrons to another (in a redox reaction), releasing energy to fuel the cell's activities. One compound acts as an electron donor and one as an electron acceptor. The energy released is used to generate adenosine triphosphate (ATP) through chemiosmosis, the same basic process that happens in the mitochondrion of eukaryotic cells.

Other groups of archaea use sunlight as a source of en-

ergy (they are phototrophs), but oxygen-generating photosynthesis does not occur in any of these organisms. Many basic metabolic pathways are shared among all forms of life; for example, archaea use a modified form of glycolysis (the Entner-Doudoroff pathway) and either a complete or partial citric acid cycle. These similarities to other organisms probably reflect both early origins in the history of life and their high level of efficiency.

Table 13.3: Nutritional types in archaeal metabolism

Nutritional Type	Source of Energy	Source of Carbon	Examples
Phototroph	Sunlight	Organic compounds	Halobacterium
Lithotrophs	Inorganic compounds	Organic compounds or carbon fixation	Ferroglobus, Methanobacteria or Pyrolobus
Organotroph	Organic compounds	Organic compounds or carbon fixation	Pyrococcus, Sulfolobus or Methanococcales

Some Euryarchaeota are methanogens (archaea that produce methane as a result of metabolism) living in anaerobic environments, such as swamps. This form of metabolism evolved early, and it is even possible that the first free-living organism was a methanogen. A common reaction involves the use of carbon dioxide as an electron acceptor to oxidize hydrogen. Methanogenesis involves a range of coenzymes that are unique to these archaea, such as coenzyme M and methanofuran. Other organic compounds such as alcohols, acetic acid or formic acid are used as alternative electron acceptors by methanogens. These reactions are common in gut-dwelling archaea. Acetic acid is also broken down into methane and carbon dioxide directly, by acetotrophic archaea. These acetotrophs are archaea in the order Methanosarcinales, and are a major part of the communities of microorganisms that produce biogas.

Other archaea use CO₂ in the atmosphere as a source of carbon, in a process called carbon fixation (they are autotrophs). This process involves either a highly modified form of the Calvin cycle or another metabolic pathway called the 3-hydroxypropionate/ 4-hydroxybutyrate cycle. The Crenarchaeota also use the reverse Krebs cycle while the Euryarchaeota also use the reductive acetyl-CoA pathway. Carbon fixation is powered by inorganic energy sources. No known archaea carry out photosynthesis. Archaeal energy sources are extremely diverse, and range from the oxidation of ammonia by the Nitrosopumilales to the oxidation of hydrogen sulfide or elemental sulfur by species of *Sulfolobus*, using either oxygen or metal ions as electron acceptors.

Phototrophic archaea use light to produce chemical energy in the form of ATP. In the Halobacteria, light-activated ion pumps like bacteriorhodopsin and halorhodopsin generate ion gradients by pumping ions out of and into the cell across the plasma membrane. The energy stored in these electrochemical gradients is then converted into ATP by ATP synthase. This process is a form of photophosphorylation. The ability of these light-driven pumps to move ions across membranes depends on light-driven changes in the structure of a retinol cofactor buried in the center of the protein.

13.2.9 Genetics

Archaea usually have a single circular chromosome, with as many as 5,751,492 base pairs in *Methanosarcina acetivorans*, the largest known archaeal genome. The tiny 490,885 base-pair genome of *Nanoarchaeum equitans* is one-tenth of this size and the smallest archaeal genome known; it is estimated to contain only 537 protein-encoding genes. Smaller independent pieces of DNA, called plasmids, are also found in archaea. Plasmids

may be transferred between cells by physical contact, in a process that may be similar to bacterial conjugation.

Archaea are genetically distinct from bacteria and eukaryotes, with up to 15% of the proteins encoded by any one archaeal genome being unique to the domain, although most of these unique genes have no known function. Of the remainder of the unique proteins that have an identified function, most belong to the Euryarchaeota and are involved in methanogenesis. The proteins that archaea, bacteria and eukaryotes share form a common core of cell function, relating mostly to transcription, translation, and nucleotide metabolism. Other characteristic archaeal features are the organization of genes of related function – such as enzymes that catalyze steps in the same metabolic pathway into novel operons, and large differences in tRNA genes and their aminoacyl tRNA synthetases.

Transcription in archaea more closely resembles eukaryotic than bacterial transcription, with the archaeal RNA polymerase being very close to its equivalent in eukaryotes, while archaeal translation shows signs of both bacterial and eukaryotic equivalents. Although archaea have only one type of RNA polymerase, its structure and function in transcription seems to be close to that of the eukaryotic RNA polymerase II, with similar protein assemblies (the general transcription factors) directing the binding of the RNA polymerase to a gene's promoter, but other archaeal transcription factors are closer to those found in bacteria. Post-transcriptional modification is simpler than in eukaryotes, since most archaeal genes lack introns, although there are many introns in their transfer RNA and ribosomal RNA genes, and introns may occur in a few protein-encoding genes.

13.2.10 Ecology

Archaea exist in a broad range of habitats, and are now recognized as a major part of global ecosystems, and may represent about 20% of microbial cells in the oceans. However, the first-discovered archaeans were extremophiles. Indeed, some archaea survive high temperatures, often above 100 °C (212 °F), as found in geysers, black smokers, and oil wells. Other common habitats include very cold habitats and highly saline, acidic, or alkaline water, but archaea include mesophiles that grow in mild conditions, in swamps and marshland, sewage, the oceans, the intestinal tract of animals, and soils.

Extremophile archaea are members of four main physiological groups. These are the halophiles, thermophiles, alkaliphiles, and acidophiles. These groups are not comprehensive or phylum-specific, nor are they mutually exclusive, since some archaea belong to several

groups. Nonetheless, they are a useful starting point for classification.

Halophiles, including the genus *Halobacterium*, live in extremely saline environments such as salt lakes and outnumber their bacterial counterparts at salinities greater than 20–25%. Thermophiles grow best at temperatures above 45 °C (113 °F), in places such as hot springs; hyperthermophilic archaea grow optimally at temperatures greater than 80 °C (176 °F). The archaeal *Methanopyrus kandleri* Strain 116 can even reproduce at 122 °C (252 °F), the highest recorded temperature of any organism.

Other archaea exist in very acidic or alkaline conditions. For example, one of the most extreme archaeal acidophiles is *Picrophilus torridus*, which grows at pH 0, which is equivalent to thriving in 1.2 molar sulfuric acid.

This resistance to extreme environments has made archaea the focus of speculation about the possible properties of extraterrestrial life. Some extremophile habitats are not dissimilar to those on Mars, leading to the suggestion that viable microbes could be transferred between planets in meteorites.

Recently, several studies have shown that archaea exist not only in mesophilic and thermophilic environments but are also present, sometimes in high numbers, at low temperatures as well. For example, archaea are common in cold oceanic environments such as polar seas. Even more significant are the large numbers of archaea found throughout the world's oceans in non-extreme habitats among the plankton community (as part of the picoplankton). Although these archaea can be present in extremely high numbers (up to 40% of the microbial biomass), almost none of these species have been isolated and studied in pure culture. Consequently, our understanding of the role of archaea in ocean ecology is rudimentary, so their full influence on global biogeochemical cycles remains largely unexplored. Some marine Crenarchaeota are capable of nitrification, suggesting these organisms may affect the oceanic nitrogen cycle, although these oceanic Crenarchaeota may also use other sources of energy.

Vast numbers of archaea are also found in the sediments that cover the sea floor, with these organisms making up the majority of living cells at depths over 1 meter below the ocean bottom. It has been demonstrated that in all oceanic surface sediments (from 1000– to 10,000-m water depth), the impact of viral infection is higher on archaea than on bacteria and virus-induced lysis of archaea accounts for up to one-third of the total microbial biomass killed, resulting in the release of ~0.3 to 0.5 gigatons of carbon per year globally.

13.2.11 Significance in Technology And Industry

Extremophile archaea, particularly those resistant either to heat or to extremes of acidity and alkalinity, are a source of enzymes that function under these harsh conditions. These enzymes have found many uses. For example, thermostable DNA polymerases, such as the Pfu DNA polymerase from *Pyrococcus furiosus*, revolutionized molecular biology by allowing the polymerase chain reaction to be used in research as a simple and rapid technique for cloning DNA. In industry, amylases, galactosidases and pullulanases in other species of *Pyrococcus* that function at over 100 °C (212 °F) allow food processing at high temperatures, such as the production of low lactose milk and whey. Enzymes from these thermophilic archaea also tend to be very stable in organic solvents, allowing their use in environmentally friendly processes in green chemistry that synthesize organic compounds. This stability makes them easier to use in structural biology. Consequently, the counterparts of bacterial or eukaryotic enzymes from extremophile archaea are often used in structural studies.

In contrast with the range of applications of archaeal enzymes, the use of the organisms themselves in biotechnology is less developed. Methanogenic archaea are a vital part of sewage treatment, since they are part of the community of microorganisms that carry out anaerobic digestion and produce biogas. In mineral processing, acidophilic archaea display promise for the extraction of metals from ores, including gold, cobalt and copper.

Archaea host a new class of potentially useful antibiotics. A few of these archaeocins have been characterized, but hundreds more are believed to exist, especially within *Haloarchaea* and *Sulfolobus*. These compounds differ in structure from bacterial antibiotics, so they may have novel modes of action. In addition, they may allow the creation of new selectable markers for use in archaeological molecular biology.

13.3 Viruses

Viruses are found wherever there is life and have probably existed since living cells first evolved. The origin of viruses is unclear because they do not form fossils, so molecular techniques are used to investigate how they arose. In addition, viral genetic material occasionally integrates into the germline of the host organisms, by which they can be passed on vertically to the offspring of the host for many generations. This provides an invaluable source of information for paleovirologists to trace back ancient viruses that have existed up to millions of years ago. There are three main hypotheses that aim to explain the origins of

viruses:

The word is from the Latin neuter *vīrus* referring to poison and other noxious liquids, from the same Indo-European base as Sanskrit *vīṣa*, Avestan *vīšā*, and ancient Greek *ἰός* (all meaning “poison”), first attested in English in 1398 in John Trevisa’s translation of Bartholomeus Anglicus’s *De Proprietatibus Rerum*. Virulent, from Latin *virulentus* (poisonous), dates to c. 1400. A meaning of “agent that causes infectious disease” is first recorded in 1728, long before the discovery of viruses by Dmitri Ivanovsky in 1892. The English plural is viruses (sometimes also *vira*) whereas the Latin word is a mass noun, which has no classically attested plural (*vīra* is used in Neo-Latin). The adjective viral dates to 1948. The term virion (plural virions), which dates from 1959, is also used to refer to a single viral particle that is released from the cell and is capable of infecting other cells of the same type.

Scientific opinions differ on whether viruses are a form of life, or organic structures that interact with living organisms. They have been described as “organisms at the edge of life”, since they resemble organisms in that they possess genes, evolve by natural selection, and reproduce by creating multiple copies of themselves through self-assembly. Although they have genes, they do not have a cellular structure, which is often seen as the basic unit of life. Viruses do not have their own metabolism, and require a host cell to make new products. They therefore cannot naturally reproduce outside a host cell—although bacterial species such as rickettsia and chlamydia are considered living organisms despite the same limitation. Accepted forms of life use cell division to reproduce, whereas viruses spontaneously assemble within cells. They differ from autonomous growth of crystals as they inherit genetic mutations while being subject to natural selection. Virus self-assembly within host cells has implications for the study of the origin of life, as it lends further credence to the hypothesis that life could have started as self-assembling organic molecules.

Viruses display a wide diversity of shapes and sizes, called ‘morphologies’. In general, viruses are much smaller than bacteria. Most viruses that have been studied have a diameter between 20 and 300 nanometres. Some filoviruses have a total length of up to 1400 nm; their diameters are only about 80 nm. Most viruses cannot be seen with an optical microscope, so scanning and transmission electron microscopes are used to visualise them. To increase the contrast between viruses and the background, electron-dense “stains” are used. These are solutions of salts of heavy metals, such as tungsten, that scatter the electrons from regions covered with the stain. When virions are coated with stain (positive staining),

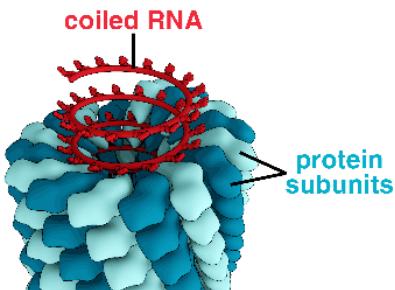


Figure 13.9: Structure of tobacco mosaic virus: RNA coiled in a helix of repeating protein sub-units¹²

fine detail is obscured. Negative staining overcomes this problem by staining the background only.

A complete virus particle, known as a virion, consists of nucleic acid surrounded by a protective coat of protein called a capsid. These are formed from identical protein subunits called capsomeres. Viruses can have a lipid “envelope” derived from the host cell membrane. The capsid is made from proteins encoded by the viral genome and its shape serves as the basis for morphological distinction. Virally-coded protein subunits will self-assemble to form a capsid, in general requiring the presence of the virus genome. Complex viruses code for proteins that assist in the construction of their capsid. Proteins associated with nucleic acid are known as nucleoproteins, and the association of viral capsid proteins with viral nucleic acid is called a nucleocapsid. The capsid and entire virus structure can be mechanically (physically) probed through atomic force microscopy. In general, there are four main morphological virus types:

Helical viruses are composed of a single type of capsomere stacked around a central axis to form a helical structure, which may have a central cavity, or tube. This arrangement results in rod-shaped or filamentous virions which can be short and highly rigid, or long and very flexible. The genetic material (typically single-stranded RNA, but ssDNA in some cases) is bound into the protein helix by interactions between the negatively charged nucleic acid and positive charges on the protein. Overall, the length of a helical capsid is related to the length of the nucleic acid contained within it, and the diameter is dependent on the size and arrangement of capsomeres. The well-studied tobacco mosaic virus is an example of a helical virus.

Most animal viruses are icosahedral or near-spherical with chiral icosahedral symmetry. A regular icosahedron is the optimum way of forming a closed shell from identical sub-units. The minimum number of identical capsomeres required for each triangular face is 3, which gives 60 for the icosahedron. Many viruses, such as rotavirus, have

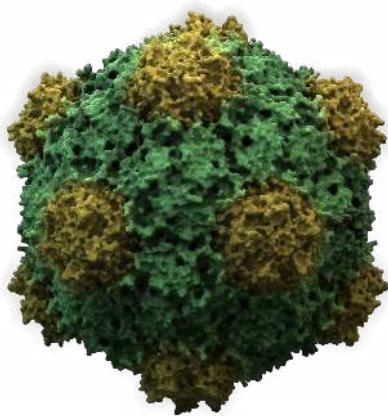


Figure 13.10: Structure of an icosahedral cowpea mosaic virus¹³

more than 60 capsomeres and appear spherical but they retain this symmetry. To achieve this, the capsomeres at the apices are surrounded by five other capsomeres and are called pentons. Capsomeres on the triangular faces are surrounded by six others and are called hexons. Hexons are in essence flat and pentons, which form the 12 vertices, are curved. The same protein may act as the subunit of both the pentamers and hexamers or they may be composed of different proteins.

Some species of virus envelop themselves in a modified form of one of the cell membranes, either the outer membrane surrounding an infected host cell or internal membranes such as nuclear membrane or endoplasmic reticulum, thus gaining an outer lipid bilayer known as a viral envelope. This membrane is studded with proteins coded for by the viral genome and host genome; the lipid membrane itself and any carbohydrates present originate entirely from the host. The influenza virus and HIV use this strategy. Most enveloped viruses are dependent on the envelope for their infectivity.

These viruses possess a capsid that is neither purely helical nor purely icosahedral, and that may possess extra structures such as protein tails or a complex outer wall. Some bacteriophages, such as Enterobacteria phage T4, have a complex structure consisting of an icosahedral head bound to a helical tail, which may have a hexagonal base plate with protruding protein tail fibres. This tail structure acts like a molecular syringe, attaching to the bacterial host and then injecting the viral genome into the cell. The poxviruses are large, complex viruses that have an unusual morphology. The viral genome is associated with proteins within a central disc structure known as a nucleoid. The nucleoid is surrounded by a membrane and two lateral bodies of unknown function. The virus has an outer envelope with a thick layer of protein studded over its surface. The whole

virion is slightly pleomorphic, ranging from ovoid to brick-shaped.

Mimivirus is one of the largest characterised viruses, with a capsid diameter of 400 nm. Protein filaments measuring 100 nm project from the surface. The capsid appears hexagonal under an electron microscope, therefore the capsid is probably icosahedral. In 2011, researchers discovered the largest then known virus in samples of water collected from the ocean floor off the coast of Las Cruces, Chile. Provisionally named Megavirus chilensis, it can be seen with a basic optical microscope. In 2013, the Pandoravirus genus was discovered in Chile and Australia, and has genomes about twice as large as Megavirus and Mimivirus. All giant viruses have dsDNA genomes and they are classified into several families: Mimiviridae, Pithoviridae, Pandoraviridae, Phycodnaviridae, and the Mollivirus genus.

Some viruses that infect Archaea have complex structures unrelated to any other form of virus, with a wide variety of unusual shapes, ranging from spindle-shaped structures to viruses that resemble hooked rods, teardrops or even bottles. Other archaeal viruses resemble the tailed bacteriophages, and can have multiple tail structures.

An enormous variety of genomic structures can be seen among viral species; as a group, they contain more structural genomic diversity than plants, animals, archaea, or bacteria. There are millions of different types of viruses, although fewer than 7,000 types have been described in detail. As of September 2015, the NCBI Virus genome database has more than 75,000 complete genome sequences, but there are doubtlessly many more to be discovered.

A virus has either a DNA or an RNA genome and is called a DNA virus or an RNA virus, respectively. The vast majority of viruses have RNA genomes. Plant viruses tend to have single-stranded RNA genomes and bacteriophages tend to have double-stranded DNA genomes.

Viral genomes are circular, as in the polyomaviruses, or linear, as in the adenoviruses. The type of nucleic acid is irrelevant to the shape of the genome. Among RNA viruses and certain DNA viruses, the genome is often divided up into separate parts, in which case it is called segmented. For RNA viruses, each segment often codes for only one protein and they are usually found together in one capsid. All segments are not required to be in the same virion for the virus to be infectious, as demonstrated by brome mosaic virus and several other plant viruses.

A viral genome, irrespective of nucleic acid type, is almost always either single-stranded or double-stranded. Single-stranded genomes consist of an unpaired nucleic acid, analogous to one-half of a ladder split down the mid-

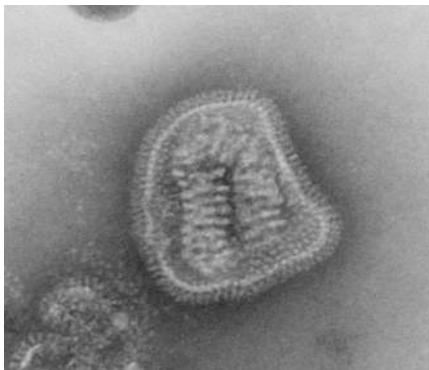


Figure 13.11: This negative-stained transmission electron micrograph (TEM) depicts the ultrastructural details of an influenza virus particle, or “virion”. A member of the taxonomic family Orthomyxoviridae, the influenza virus is an enveloped, single-stranded RNA virus. Eight helical capsids are surrounded by the viral envelope. Particle diameter: 80–120 nm¹⁴

Double-stranded genomes consist of two complementary paired nucleic acids, analogous to a ladder. The virus particles of some virus families, such as those belonging to the Hepadnaviridae, contain a genome that is partially double-stranded and partially single-stranded.

For most viruses with RNA genomes and some with single-stranded DNA genomes, the single strands are said to be either positive-sense (called the ‘plus-strand’) or negative-sense (called the ‘minus-strand’), depending on if they are complementary to the viral messenger RNA (mRNA). Positive-sense viral RNA is in the same sense as viral mRNA and thus at least a part of it can be immediately translated by the host cell. Negative-sense viral RNA is complementary to mRNA and thus must be converted to positive-sense RNA by an RNA-dependent RNA polymerase before translation. DNA nomenclature for viruses with single-sense genomic ssDNA is similar to RNA nomenclature, in that positive-strand viral ssDNA is identical in sequence to the viral mRNA and is thus a coding strand, while negative-strand viral ssDNA is complementary to the viral mRNA and is thus a template strand. Several types of ssDNA and ssRNA viruses have genomes that are ambisense in that transcription can occur off both strands in a double-stranded replicative intermediate. Examples include geminiviruses, which are ssDNA plant viruses and arenaviruses, which are ssRNA viruses of animals.

Genome size varies greatly between species. The smallest—the ssDNA circoviruses, family Circoviridae—code for only two proteins and have a genome size of only two kilobases; the largest—the pandoraviruses—have genome sizes of around two megabases which code for

about 2500 proteins. Virus genes rarely have introns and often are arranged in the genome so that they overlap.

In general, RNA viruses have smaller genome sizes than DNA viruses because of a higher error-rate when replicating, and have a maximum upper size limit. Beyond this, errors when replicating render the virus useless or uncompetitive. To compensate, RNA viruses often have segmented genomes—the genome is split into smaller molecules—thus reducing the chance that an error in a single-component genome will incapacitate the entire genome. In contrast, DNA viruses generally have larger genomes because of the high fidelity of their replication enzymes. Single-strand DNA viruses are an exception to this rule, as mutation rates for these genomes can approach the extreme of the ssRNA virus case.

Viruses undergo genetic change by several mechanisms. These include a process called antigenic drift where individual bases in the DNA or RNA mutate to other bases. Most of these point mutations are “silent”—they do not change the protein that the gene encodes—but others can confer evolutionary advantages such as resistance to antiviral drugs. Antigenic shift occurs when there is a major change in the genome of the virus. This can be a result of recombination or reassortment. When this happens with influenza viruses, pandemics might result. RNA viruses often exist as quasispecies or swarms of viruses of the same species but with slightly different genome nucleoside sequences. Such quasispecies are a prime target for natural selection.

Segmented genomes confer evolutionary advantages; different strains of a virus with a segmented genome can shuffle and combine genes and produce progeny viruses (or offspring) that have unique characteristics. This is called reassortment or ‘viral sex’.

Genetic recombination is the process by which a strand of DNA is broken and then joined to the end of a different DNA molecule. This can occur when viruses infect cells simultaneously and studies of viral evolution have shown that recombination has been rampant in the species studied. Recombination is common to both RNA and DNA viruses.

Viral populations do not grow through cell division, because they are acellular. Instead, they use the machinery and metabolism of a host cell to produce multiple copies of themselves, and they assemble in the cell. When infected, the host cell is forced to rapidly produce thousands of identical copies of the original virus.

Their life cycle differs greatly between species, but there are six basic stages in their life cycle:

Attachment is a specific binding between viral capsid

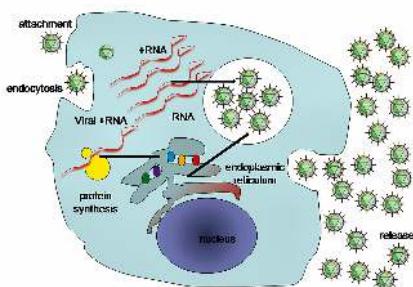


Figure 13.12: A typical virus replication cycle.¹⁵

proteins and specific receptors on the host cellular surface. This specificity determines the host range and type of host cell of a virus. For example, HIV infects a limited range of human leucocytes. This is because its surface protein, gp120, specifically interacts with the CD4 molecule—a chemokine receptor—which is most commonly found on the surface of CD4+ T-Cells. This mechanism has evolved to favour those viruses that infect only cells in which they are capable of replication. Attachment to the receptor can induce the viral envelope protein to undergo changes that result in the fusion of viral and cellular membranes, or changes of non-enveloped virus surface proteins that allow the virus to enter.

Penetration follows attachment: Virions enter the host cell through receptor-mediated endocytosis or membrane fusion in a process often known as viral entry. The infection of plant and fungal cells is different from that of animal cells. Plants have a rigid cell wall made of cellulose, and fungi one of chitin, so most viruses can get inside these cells only after trauma to the cell wall. Nearly all plant viruses (such as tobacco mosaic virus) can also move directly from cell to cell, in the form of single-stranded nucleoprotein complexes, through pores called plasmodesmata. Bacteria, like plants, have strong cell walls that a virus must breach to infect the cell. Given that bacterial cell walls are much thinner than plant cell walls due to their much smaller size, some viruses have evolved mechanisms that inject their genome into the bacterial cell across the cell wall, while the viral capsid remains outside.

Uncoating is a process in which the viral capsid is removed: This may be by degradation by viral enzymes or host enzymes or by simple dissociation; the end-result is the releasing of the viral genomic nucleic acid.

Replication of viruses involves primarily multiplication of the genome. Replication involves synthesis of viral messenger RNA (mRNA) from “early” genes (with exceptions for positive sense RNA viruses), viral protein synthesis, possible assembly of viral proteins, then viral genome replication mediated by early or regulatory

protein expression. This may be followed, for complex viruses with larger genomes, by one or more further rounds of mRNA synthesis: “late” gene expression is, in general, of structural or virion proteins.

Assembly – Following the structure-mediated self-assembly of the virus particles, some modification of the proteins often occurs. In viruses such as HIV, this modification (sometimes called maturation) occurs after the virus has been released from the host cell.

Release – Viruses can be released from the host cell by lysis, a process that kills the cell by bursting its membrane and cell wall if present: this is a feature of many bacterial and some animal viruses. Some viruses undergo a lytic cycle where the viral genome is incorporated by genetic recombination into a specific place in the host's chromosome. The viral genome is then known as a “provirus” or, in the case of bacteriophages a “prophage”. Whenever the host divides, the viral genome is also replicated. The viral genome is mostly silent within the host. At some point, the provirus or prophage may give rise to active virus, which may lyse the host cells. Enveloped viruses (e.g., HIV) typically are released from the host cell by budding. During this process the virus acquires its envelope, which is a modified piece of the host's plasma or other, internal membrane.

The genetic material within virus particles, and the method by which the material is replicated, varies considerably between different types of viruses.

The genome replication of most DNA viruses takes place in the cell's nucleus. If the cell has the appropriate receptor on its surface, these viruses enter the cell either by direct fusion with the cell membrane (e.g., herpesviruses) or—more usually—by receptor-mediated endocytosis. Most DNA viruses are entirely dependent on the host cell's DNA and RNA synthesising machinery, and RNA processing machinery. Viruses with larger genomes may encode much of this machinery themselves. In eukaryotes the viral genome must cross the cell's nuclear membrane to access this machinery, while in bacteria it need only enter the cell.

Replication of RNA viruses usually takes place in the cytoplasm. RNA viruses can be placed into four different groups depending on their modes of replication. The polarity (whether or not it can be used directly by ribosomes to make proteins) of single-stranded RNA viruses largely determines the replicative mechanism; the other major criterion is whether the genetic material is single-stranded or double-stranded. All RNA viruses use their own RNA replicate enzymes to create copies of their genomes.

Reverse transcribing viruses have ssRNA (Retroviridae, Metaviridae, Pseudoviridae) or dsDNA (Caulimoviridae,

and Hepadnaviridae) in their particles. Reverse transcribing viruses with RNA genomes (retroviruses) use a DNA intermediate to replicate, whereas those with DNA genomes (pararetroviruses) use an RNA intermediate during genome replication. Both types use a reverse transcriptase, or RNA-dependent DNA polymerase enzyme, to carry out the nucleic acid conversion. Retroviruses integrate the DNA produced by reverse transcription into the host genome as a provirus as a part of the replication process; pararetroviruses do not, although integrated genome copies of especially plant pararetroviruses can give rise to infectious virus. They are susceptible to antiviral drugs that inhibit the reverse transcriptase enzyme, e.g. zidovudine and lamivudine. An example of the first type is HIV, which is a retrovirus. Examples of the second type are the Hepadnaviridae, which includes Hepatitis B virus.

The range of structural and biochemical effects that viruses have on the host cell is extensive. These are called 'cytopathic effects'. Most virus infections eventually result in the death of the host cell. The causes of death include cell lysis, alterations to the cell's surface membrane and apoptosis. Often cell death is caused by cessation of its normal activities because of suppression by virus-specific proteins, not all of which are components of the virus particle. The distinction between cytopathic and harmless is gradual. Some viruses, such as Epstein-Barr virus, can cause cells to proliferate without causing malignancy, while others, such as papillomaviruses, are established causes of cancer.

Some viruses cause no apparent changes to the infected cell. Cells in which the virus is latent and inactive show few signs of infection and often function normally. This causes persistent infections and the virus is often dormant for many months or years. This is often the case with herpes viruses.

Viruses are by far the most abundant biological entities on Earth and they outnumber all the others put together. They infect all types of cellular life including animals, plants, bacteria and fungi. Different types of viruses can infect only a limited range of hosts and many are species-specific. Some, such as smallpox virus for example, can infect only one species—in this case humans, and are said to have a narrow host range. Other viruses, such as rabies virus, can infect different species of mammals and are said to have a broad range. The viruses that infect plants are harmless to animals, and most viruses that infect other animals are harmless to humans. The host range of some bacteriophages is limited to a single strain of bacteria and they can be used to trace the source of outbreaks of infections by a method called phage typing. The complete set of viruses in an organism

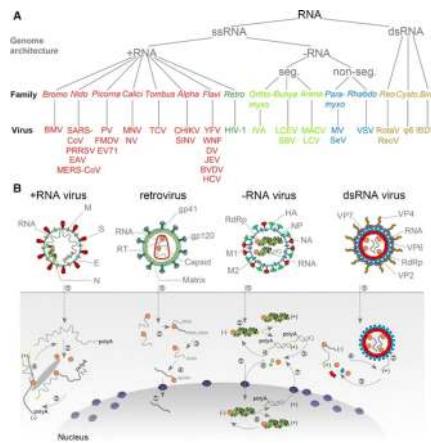


Figure 13.13: Taxonomy and replication strategies of RNA viruses.¹⁶ A) Simplified taxonomy of the genome architecture of the RNA viruses described in this review. See main text for used abbreviations. B) (+RNA virus)—as exemplified here with a CoV-like virion—releases a single-stranded RNA genome into the cytoplasm (1). (2) Translation of the 5'-terminal open-reading frame of the genome produces the viral replicase. (3) This multi-enzyme complex includes RdRp activity (orange) and associates with intracellular membranes before –RNA synthesis commences. Newly synthesised –RNAs are subsequently used to produce new +RNAs (4), which are typically capped (yellow) and polyadenylated (polyA). (Retrovirus) HIV-1 genomes are packaged as ssRNA in virions. When the ssRNA is released (1) a cDNA copy is synthesised by the RT (2). The RNA is next degraded by the intrinsic RNase H activity in the RT (3) and the single stranded cDNA converted to dsDNA (4). The dsDNA is imported in the nucleus (5) for integration into the host's genetic material. (–RNA virus) (1) As illustrated here with an IAV-like particle, infection with an –RNA virus releases a viral RNA genome that is associated with a viral polymerase (orange) and nucleoprotein (green). (2) In the case of non-segmented –RNA viruses, these complexes support transcription to produce viral mRNAs or cRNAs. (3) Viral mRNAs are next translated and new viral proteins complex with cRNAs to synthesise new vRNAs. (5) The vRNA-containing complexes of some segmented –RNA viruses are imported into the nucleus of the host cell, where (6) the RdRp produces mRNAs or cRNAs. (7) mRNAs are transported to the cytoplasm, while cRNAs are bound by new viral proteins to form cRNPs for –RNA synthesis. (dsRNA virus) Fully duplexed RNA genomes lack cap and polyA elements. (1) The RdRp (orange), therefore, transcribes the viral genome inside the capsid of the virion (blue and red), so viral mRNAs can be (2) released into the cytoplasm as illustrated here with a rotavirus-like virion. In the cytoplasm the mRNA is translated (3) or replicated by newly synthesised viral RdRps (4)]

or habitat is called the virome; for example, all human viruses constitute the human virome.

13.3.1 Role in human disease

Examples of common human diseases caused by viruses include the common cold, influenza, chickenpox, and cold sores. Many serious diseases such as rabies, Ebola virus disease, AIDS (HIV), avian influenza, and SARS are caused by viruses. The relative ability of viruses to cause disease is described in terms of virulence. Other diseases are under investigation to discover if they have a virus as the causative agent, such as the possible connection between human herpesvirus 6 (HHV6) and neurological diseases such as multiple sclerosis and chronic fatigue syndrome. There is controversy over whether the bornavirus, previously thought to cause neurological diseases in horses, could be responsible for psychiatric illnesses in humans.

Viruses have different mechanisms by which they produce disease in an organism, which depends largely on the viral species. Mechanisms at the cellular level primarily include cell lysis, the breaking open and subsequent death of the cell. In multicellular organisms, if enough cells die, the whole organism will start to suffer the effects. Although viruses cause disruption of healthy homeostasis, resulting in disease, they may exist relatively harmlessly within an organism. An example would include the ability of the herpes simplex virus, which causes cold sores, to remain in a dormant state within the human body. This is called latency and is a characteristic of the herpes viruses, including Epstein-Barr virus, which causes glandular fever, and varicella zoster virus, which causes chickenpox and shingles. Most people have been infected with at least one of these types of herpes virus. These latent viruses might sometimes be beneficial, as the presence of the virus can increase immunity against bacterial pathogens, such as Yersinia pestis.

Some viruses can cause lifelong or chronic infections, where the viruses continue to replicate in the body despite the host's defence mechanisms. This is common in hepatitis B virus and hepatitis C virus infections. People chronically infected are known as carriers, as they serve as reservoirs of infectious virus. In populations with a high proportion of carriers, the disease is said to be endemic.

Viral epidemiology is the branch of medical science that deals with the transmission and control of virus infections in humans. Transmission of viruses can be vertical, which means from mother to child, or horizontal, which means from person to person. Examples of vertical transmission include hepatitis B virus and HIV, where the baby is born already infected with the virus. Another, more rare, example is the varicella zoster virus, which, although causing relatively mild infections in children and adults, can be

fatal to the foetus and newborn baby.

Horizontal transmission is the most common mechanism of spread of viruses in populations. Horizontal transmission can occur when body fluids are exchanged during sexual activity, by exchange of saliva or when contaminated food or water is ingested. It can also occur when aerosols containing viruses are inhaled or by insect vectors such as when infected mosquitoes penetrate the skin of a host. Most types of viruses are restricted to just one or two of these mechanisms and they are referred to as "respiratory viruses" or "enteric viruses" and so forth. The rate or speed of transmission of viral infections depends on factors that include population density, the number of susceptible individuals, (i.e., those not immune), the quality of healthcare and the weather.

Epidemiology is used to break the chain of infection in populations during outbreaks of viral diseases. Control measures are used that are based on knowledge of how the virus is transmitted. It is important to find the source, or sources, of the outbreak and to identify the virus. Once the virus has been identified, the chain of transmission can sometimes be broken by vaccines. When vaccines are not available, sanitation and disinfection can be effective. Often, infected people are isolated from the rest of the community, and those that have been exposed to the virus are placed in quarantine. To control the outbreak of foot-and-mouth disease in cattle in Britain in 2001, thousands of cattle were slaughtered. Most viral infections of humans and other animals have incubation periods during which the infection causes no signs or symptoms. Incubation periods for viral diseases range from a few days to weeks, but are known for most infections. Somewhat overlapping, but mainly following the incubation period, there is a period of communicability—a time when an infected individual or animal is contagious and can infect another person or animal. This, too, is known for many viral infections, and knowledge of the length of both periods is important in the control of outbreaks. When outbreaks cause an unusually high proportion of cases in a population, community, or region, they are called epidemics. If outbreaks spread worldwide, they are called pandemics.

A pandemic is a worldwide epidemic. The 1918 flu pandemic, which lasted until 1919, was a category 5 influenza pandemic caused by an unusually severe and deadly influenza A virus. The victims were often healthy young adults, in contrast to most influenza outbreaks, which predominantly affect juvenile, elderly, or otherwise-weakened patients. Older estimates say it killed 40–50 million people, while more recent research suggests that it may have killed as many as 100 million people, or 5% of the world's population in 1918.

Although viral pandemics are rare events, HIV—which

evolved from viruses found in monkeys and chimpanzees—has been pandemic since at least the 1980s. During the 20th century there were four pandemics caused by influenza virus and those that occurred in 1918, 1957 and 1968 were severe. Most researchers believe that HIV originated in sub-Saharan Africa during the 20th century; it is now a pandemic, with an estimated 37.9 million people now living with the disease worldwide. There were about 770,000 deaths from AIDS in 2018. The Joint United Nations Programme on HIV/AIDS (UNAIDS) and the World Health Organization (WHO) estimate that AIDS has killed more than 25 million people since it was first recognised on 5 June 1981, making it one of the most destructive epidemics in recorded history. In 2007 there were 2.7 million new HIV infections and 2 million HIV-related deaths.

Several highly lethal viral pathogens are members of the Filoviridae. Filoviruses are filament-like viruses that cause viral hemorrhagic fever, and include ebolaviruses and marburgviruses. Marburg virus, first discovered in 1967, attracted widespread press attention in April 2005 for an outbreak in Angola. Ebola virus disease has also caused intermittent outbreaks with high mortality rates since 1976 when it was first identified. The worst and most recent one is the 2013–2016 West Africa epidemic.

With the exception of smallpox, most pandemics are caused by newly evolved viruses. These “emergent” viruses are usually mutants of less harmful viruses that have circulated previously either in humans or other animals.

Severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) are caused by new types of coronaviruses. Other coronaviruses are known to cause mild infections in humans, so the virulence and rapid spread of SARS infections—that by July 2003 had caused around 8,000 cases and 800 deaths—was unexpected and most countries were not prepared.

A related coronavirus emerged in Wuhan, China in November 2019 and spread rapidly around the world. Thought to have originated in bats and subsequently named severe acute respiratory syndrome coronavirus 2, infections with the virus caused a pandemic in 2020. Unprecedented restrictions in peacetime have been placed on international travel, and curfews imposed in several major cities worldwide.

Viruses are an established cause of cancer in humans and other species. Viral cancers occur only in a minority of infected persons (or animals). Cancer viruses come from a range of virus families, including both RNA and DNA viruses, and so there is no single type of “oncovirus” (an obsolete term originally used for acutely transforming retroviruses). The development of cancer is determined

by a variety of factors such as host immunity and mutations in the host. Viruses accepted to cause human cancers include some genotypes of human papillomavirus, hepatitis B virus, hepatitis C virus, Epstein–Barr virus, Kaposi’s sarcoma-associated herpesvirus and human T-lymphotropic virus. The most recently discovered human cancer virus is a polyomavirus (Merkel cell polyomavirus) that causes most cases of a rare form of skin cancer called Merkel cell carcinoma. Hepatitis viruses can develop into a chronic viral infection that leads to liver cancer. Infection by human T-lymphotropic virus can lead to tropical spastic paraparesis and adult T-cell leukaemia. Human papillomaviruses are an established cause of cancers of cervix, skin, anus, and penis. Within the Herpesviridae, Kaposi’s sarcoma-associated herpesvirus causes Kaposi’s sarcoma and body-cavity lymphoma, and Epstein–Barr virus causes Burkitt’s lymphoma, Hodgkin’s lymphoma, B lymphoproliferative disorder, and nasopharyngeal carcinoma. Merkel cell polyomavirus closely related to SV40 and mouse polyomaviruses that have been used as animal models for cancer viruses for over 50 years.

The body’s first line of defence against viruses is the innate immune system. This comprises cells and other mechanisms that defend the host from infection in a non-specific manner. This means that the cells of the innate system recognise, and respond to, pathogens in a generic way, but, unlike the adaptive immune system, it does not confer long-lasting or protective immunity to the host.

RNA interference is an important innate defence against viruses. Many viruses have a replication strategy that involves double-stranded RNA (dsRNA). When such a virus infects a cell, it releases its RNA molecule or molecules, which immediately bind to a protein complex called a dicer that cuts the RNA into smaller pieces. A biochemical pathway—the RISC complex—is activated, which ensures cell survival by degrading the viral mRNA. Rotaviruses have evolved to avoid this defence mechanism by not uncoating fully inside the cell, and releasing newly produced mRNA through pores in the particle’s inner capsid. Their genomic dsRNA remains protected inside the core of the virion.

When the adaptive immune system of a vertebrate encounters a virus, it produces specific antibodies that bind to the virus and often render it non-infectious. This is called humoral immunity. Two types of antibodies are important. The first, called IgM, is highly effective at neutralising viruses but is produced by the cells of the immune system only for a few weeks. The second, called IgG, is produced indefinitely. The presence of IgM in the blood of the host is used to test for acute infection, whereas IgG indicates an infection sometime in the past. IgG antibody is measured when tests for immunity are carried out.

Antibodies can continue to be an effective defence mechanism even after viruses have managed to gain entry to the host cell. A protein that is in cells, called TRIM21, can attach to the antibodies on the surface of the virus particle. This primes the subsequent destruction of the virus by the enzymes of the cell's proteosome system.

A second defence of vertebrates against viruses is called cell-mediated immunity and involves immune cells known as T cells. The body's cells constantly display short fragments of their proteins on the cell's surface, and, if a T cell recognises a suspicious viral fragment there, the host cell is destroyed by 'killer T' cells and the virus-specific T-cells proliferate. Cells such as the macrophage are specialists at this antigen presentation. The production of interferon is an important host defence mechanism. This is a hormone produced by the body when viruses are present. Its role in immunity is complex; it eventually stops the viruses from reproducing by killing the infected cell and its close neighbours.

Not all virus infections produce a protective immune response in this way. HIV evades the immune system by constantly changing the amino acid sequence of the proteins on the surface of the virion. This is known as "escape mutation" as the viral epitopes escape recognition by the host immune response. These persistent viruses evade immune control by sequestration, blockade of antigen presentation, cytokine resistance, evasion of natural killer cell activities, escape from apoptosis, and antigenic shift. Other viruses, called 'neurotropic viruses', are disseminated by neural spread where the immune system may be unable to reach them.

Because viruses use vital metabolic pathways within host cells to replicate, they are difficult to eliminate without using drugs that cause toxic effects to host cells in general. The most effective medical approaches to viral diseases are vaccinations to provide immunity to infection, and antiviral drugs that selectively interfere with viral replication.

Vaccination is a cheap and effective way of preventing infections by viruses. Vaccines were used to prevent viral infections long before the discovery of the actual viruses. Their use has resulted in a dramatic decline in morbidity (illness) and mortality (death) associated with viral infections such as polio, measles, mumps and rubella. Smallpox infections have been eradicated. Vaccines are available to prevent over thirteen viral infections of humans, and more are used to prevent viral infections of animals. Vaccines can consist of live-attenuated or killed viruses, or viral proteins (antigens). Live vaccines contain weakened forms of the virus, which do not cause the disease but, nonetheless, confer immunity. Such viruses are called attenuated. Live vaccines can be dangerous when given to people with a weak immunity (who are

described as immunocompromised), because in these people, the weakened virus can cause the original disease. Biotechnology and genetic engineering techniques are used to produce subunit vaccines. These vaccines use only the capsid proteins of the virus. Hepatitis B vaccine is an example of this type of vaccine. Subunit vaccines are safe for immunocompromised patients because they cannot cause the disease. The yellow fever virus vaccine, a live-attenuated strain called 17D, is probably the safest and most effective vaccine ever generated.

Antiviral drugs are often nucleoside analogues (fake DNA building-blocks), which viruses mistakenly incorporate into their genomes during replication. The life-cycle of the virus is then halted because the newly synthesised DNA is inactive. This is because these analogues lack the hydroxyl groups, which, along with phosphorus atoms, link together to form the strong "backbone" of the DNA molecule. This is called DNA chain termination. Examples of nucleoside analogues are aciclovir for Herpes simplex virus infections and lamivudine for HIV and hepatitis B virus infections. Aciclovir is one of the oldest and most frequently prescribed antiviral drugs. Other antiviral drugs in use target different stages of the viral life cycle. HIV is dependent on a proteolytic enzyme called the HIV-1 protease for it to become fully infectious. There is a large class of drugs called protease inhibitors that inactivate this enzyme.

Hepatitis C is caused by an RNA virus. In 80% of people infected, the disease is chronic, and without treatment, they are infected for the remainder of their lives. There is now an effective treatment that uses the nucleoside analogue drug ribavirin combined with interferon. The treatment of chronic carriers of the hepatitis B virus by using a similar strategy using lamivudine has been developed.

13.3.2 Animal viruses

Viruses are important pathogens of livestock. Diseases such as foot-and-mouth disease and bluetongue are caused by viruses. Companion animals such as cats, dogs, and horses, if not vaccinated, are susceptible to serious viral infections. Canine parvovirus is caused by a small DNA virus and infections are often fatal in pups. Like all invertebrates, the honey bee is susceptible to many viral infections. Most viruses co-exist harmlessly in their host and cause no signs or symptoms of disease.

13.3.3 Plant viruses

There are many types of plant virus, but often they cause only a loss of yield, and it is not economically viable to try to control them. Plant viruses are often spread from plant to plant by organisms, known as vectors. These are usu-

ally insects, but some fungi, nematode worms, and single-celled organisms have been shown to be vectors. When control of plant virus infections is considered economical, for perennial fruits, for example, efforts are concentrated on killing the vectors and removing alternate hosts such as weeds. Plant viruses cannot infect humans and other animals because they can reproduce only in living plant cells.

Originally from Peru, the potato has become a staple crop worldwide. The potato virus Y causes disease in potatoes and related species including tomatoes and peppers. In the 1980s, this virus acquired economical importance when it proved difficult to control in seed potato crops. Transmitted by aphids, this virus can reduce crop yields by up to 80 per cent, causing significant losses to potato yields.

Plants have elaborate and effective defence mechanisms against viruses. One of the most effective is the presence of so-called resistance (R) genes. Each R gene confers resistance to a particular virus by triggering localised areas of cell death around the infected cell, which can often be seen with the unaided eye as large spots. This stops the infection from spreading. RNA interference is also an effective defence in plants. When they are infected, plants often produce natural disinfectants that kill viruses, such as salicylic acid, nitric oxide, and reactive oxygen molecules.

Plant virus particles or virus-like particles (VLPs) have applications in both biotechnology and nanotechnology. The capsids of most plant viruses are simple and robust structures and can be produced in large quantities either by the infection of plants or by expression in a variety of heterologous systems. Plant virus particles can be modified genetically and chemically to encapsulate foreign material and can be incorporated into supramolecular structures for use in biotechnology.

13.3.4 Bacterial viruses

A bacteriophage, also known informally as a phage, is a virus that infects and replicates within bacteria and archaea. The term was derived from "bacteria" and the Greek φάγειν (phagein), meaning "to devour". Bacteriophages are composed of proteins that encapsulate a DNA or RNA genome, and may have structures that are either simple or elaborate. Their genomes may encode as few as four genes and as many as hundreds of genes.

These viruses infect specific bacteria by binding to surface receptor molecules and then entering the cell. Within a short amount of time, in some cases just minutes, bacterial polymerase starts translating viral mRNA into protein. These proteins go on to become either new virions within the cell, helper proteins, which help assembly of new viri-

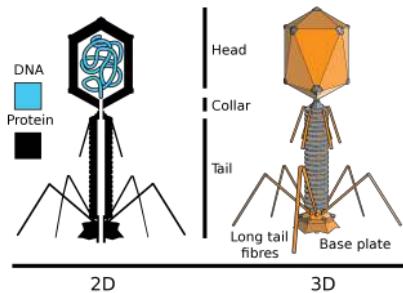


Figure 13.14: Structural overview of T2 phage¹⁷

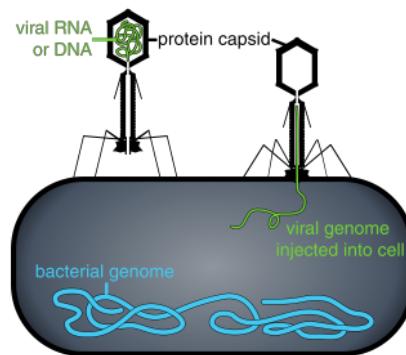


Figure 13.15: Diagram showing bacteriophages infecting a bacterial cell (not to scale; bacteriophages are much smaller than bacteria)¹⁸

ons, or proteins involved in cell lysis. Viral enzymes aid in the breakdown of the cell membrane, and, in the case of the T4 phage, in just over twenty minutes after injection over three hundred phages could be released.

The major way bacteria defend themselves from bacteriophages is by producing enzymes that destroy foreign DNA. These enzymes, called restriction endonucleases, cut up the viral DNA that bacteriophages inject into bacterial cells. Bacteria also contain a system that uses CRISPR sequences to retain fragments of the genomes of viruses that the bacteria have come into contact with in the past, which allows them to block the virus's replication through a form of RNA interference. This genetic system provides bacteria with acquired immunity to infection.

Bacteriophages are among the most common and diverse entities in the biosphere. Bacteriophages are ubiquitous viruses, found wherever bacteria exist. It is estimated there are more than 10^{31} bacteriophages on the planet, more than every other organism on Earth, including bacteria, combined. One of the densest natural sources for phages and other viruses is seawater, where up to 9×10^8 virions per millilitre have been found in microbial mats at the surface, and up to 70% of marine bacteria may be in-

fected by phages.

Phages have been used since the late 20th century as an alternative to antibiotics in the former Soviet Union and Central Europe, as well as in France. They are seen as a possible therapy against multi-drug-resistant strains of many bacteria (see phage therapy). On the other hand, phages of Inoviridae have been shown to complicate biofilms involved in pneumonia and cystic fibrosis and to shelter the bacteria from drugs meant to eradicate disease, thus promoting persistent infection.

Chapter 14

Protists

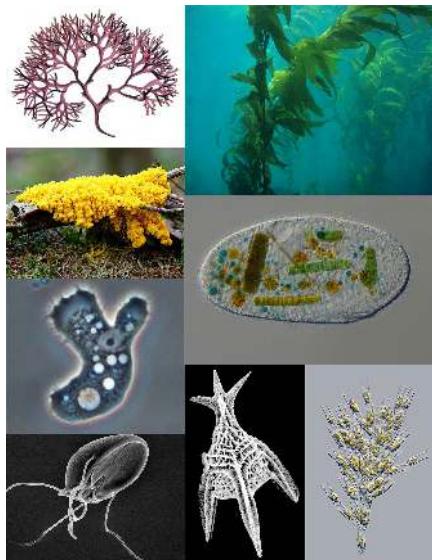


Figure 14.1: A sampling of protists:² red algae (*Chondrus crispus*); brown algae (Giant Kelp); ciliate (*Frontonia*); golden algae (*Dinobryon*); Foraminifera (*Radiolaria*); parasitic flagellate (*Giardia muris*); pathogenic amoeba (*Acanthamoeba*); amoebozoan slime mold (*Fuligo septica*)

A protist¹ is any eukaryotic organism (one with cells containing a nucleus) that is not an animal, plant, or fungus. While it is likely that protists share a common ancestor (the last eukaryotic common ancestor), the exclusion of other eukaryotes means that protists do not form a natural group, or clade. So some protists may be more closely related to animals, plants, or fungi than they are to other protists; however, like algae, invertebrates, or protozoans, the grouping is used for convenience.

In antiquity, the two lineages of animals and plants were recognized. They were given the taxonomic rank of Kingdom by the Swedish botanist Carl Linnaeus³. Though he included the fungi with plants with some reservations, it was later realized that they are quite distinct and warrant

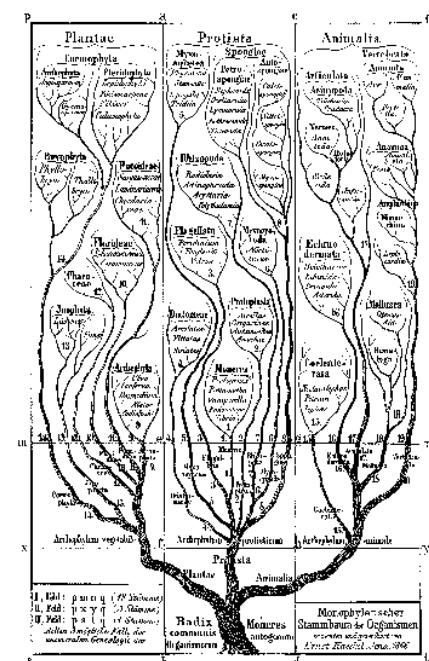


Figure 14.2: “Monophyletischer Stammbaum der Organismen” from Generelle Morphologie der Organismen (1866) with the three branches Plantae, Protista, Animalia.⁵

a separate kingdom, the composition of which was not entirely clear until the 1980s. The various single-cell eukaryotes were originally placed with plants or animals when they became known. In 1818, the German biologist Georg A. Goldfuss coined the word *protozoa* to refer to organisms such as ciliates, and this group was expanded until it encompassed all single-celled eukaryotes, and given their own kingdom, the Protista, by Ernst Haeckel⁴ in 1866.

The eukaryotes thus came to be composed of four kingdoms:

- Kingdom Protista
- Kingdom Plantae

¹<https://en.wikipedia.org/wiki/Protist>

³https://en.wikipedia.org/wiki/Carl_Linnaeus

⁴https://en.wikipedia.org/wiki/Ernst_Haeckel

- Kingdom Fungi
- Kingdom Animalia

The protists were understood to be “primitive forms”, and thus an evolutionary grade, united by their primitive unicellular nature. The disentanglement of the deep splits in the tree of life only really started with DNA sequencing, leading to a system of domains rather than kingdoms as top level rank being put forward by Carl Woese, uniting all the eukaryote kingdoms under the eukaryote domain. At the same time, work on the protist tree intensified, and is still actively going on today. Several alternative classifications have been forwarded, though there is no consensus in the field.

A revised classification in 2012 recognizes five supergroups of eukaryotes as shown in Table 14.1.

Table 14.1: A recent (2012) classification of protists.

Supergroup Name	Organisms
Archaeplastida (or Primoplantae)	Land plants, green algae, red algae, and glaucophytes
SAR supergroup	Stramenopiles (brown algae, diatoms, etcetera), Alveolata, and Rhizaria (Foraminifera, Radiolaria, and various other amoeboid protozoa)
Excavata	Various flagellate protozoa
Amoebozoa	Most lobose amoeboids and slime molds
Opisthokonta	Animals, fungi, choanoflagellates, etcetera

The classification of a separate kingdom to the animals and plants was first proposed by John Hogg in 1860 as the kingdom Protoctista; in 1866 Ernst Haeckel also proposed a third kingdom Protista as “the kingdom of primitive forms”. Originally these also included prokaryotes, but with time these would be removed to a fourth kingdom Monera. In the popular five-kingdom scheme proposed by Robert Whittaker in 1969, the Protista was defined as eukaryotic “organisms which are unicellular or unicellular-colonial and which form no tissues”, and the fifth kingdom Fungi was established. In the five-kingdom system of Lynn Margulis, the term protist is reserved for microscopic organisms, while the more inclusive kingdom Protoctista (or protocists) included certain large multicellular eukaryotes, such as kelp, red algae and slime molds. Others use the term protist interchangeably with Margulis’s protocist, to encompass both single-celled and multicellular eukaryotes, including those that form specialized tissues but do

not fit into any of the other traditional kingdoms.

Besides their relatively simple levels of organization, protists do not necessarily have much in common. When used, the term “protists” is now considered to mean a paraphyletic assemblage of similar–appearing but diverse taxa (biological groups); these taxa do not have an exclusive common ancestor beyond being composed of eukaryotes, and have different life cycles, trophic levels, modes of locomotion and cellular structures. Examples of protists include: amoebas (including nucleariids and Foraminifera); choanoflagellates; ciliates; diatoms; dinoflagellates; Giardia; Plasmodium (which causes malaria); oomycetes (including Phytophthora, the cause of the Great Famine of Ireland); and slime molds. These examples are unicellular, although oomycetes can form filaments, and slime molds can aggregate.

In cladistic systems (classifications based on common ancestry), there are no equivalents to the taxa Protista or Protoctista, as both terms refer to a paraphyletic group that spans the entire eukaryotic tree of life. In cladistic classification, the contents of Protista are mostly distributed among various supergroups: examples include the SAR supergroup (of stramenopiles or heterokonts, alveolates, and Rhizaria); Archaeplastida (or Plantae sensu lato); Excavata (which is mostly unicellular flagellates); and Opisthokonta (which commonly includes unicellular flagellates, but also animals and fungi). “Protista”, “Protoctista”, and “Protozoa” are therefore considered obsolete. However, the term “protist” continues to be used informally as a catch-all term for eukaryotic organisms that aren’t within other traditional kingdoms. For example, the word “protist pathogen” may be used to denote any disease-causing organism that is not plant, animal, fungal, prokaryotic, viral, or subviral.

The term protista was first used by Ernst Haeckel in 1866. Protists were traditionally subdivided into several groups based on similarities to the “higher” kingdoms such as:

- **Protozoa:** unicellular “animal-like” (heterotrophic, and sometimes parasitic) organisms that are further sub-divided based on characteristics such as motility, such as the (flagellated) Flagellata, the (ciliated) Ciliophora, the (phagocytic) amoeba, and the (spore-forming) Sporozoa.
- **Protophyta:** plant-like” (autotrophic) organisms that are composed mostly of unicellular algae. The dinoflagellates, diatoms and Euglena-like flagellates are photosynthetic protists.
- **Molds:** “Mold” generally refer to fungi; but slime molds and water molds are fungus-like” (saprophytic) protists, although some are pathogens.

Some protists, sometimes called ambiregnal protists,

have been considered to be both protozoa and algae or fungi (e.g., slime molds and flagellated algae). Conflicts, such as these – for example the dual-classification of Euglenids and Dinobryons, which are mixotrophic – is an example of why the kingdom Protista was adopted.

These traditional subdivisions, largely based on superficial commonalities, have been replaced by classifications based on phylogenetics (evolutionary relatedness among organisms). Molecular analyses in modern taxonomy have been used to redistribute former members of this group into diverse and sometimes distantly related phyla. For instance, the water molds are now considered to be closely related to photosynthetic organisms such as Brown algae and Diatoms, the slime molds are grouped mainly under Amoebozoa, and the Amoebozoa itself includes only a subset of the “Amoeba” group, and significant number of erstwhile “Amoeboid” genera are distributed among Rhizarians and other Phyla.

However, the older terms are still used as informal names to describe the morphology and ecology of various protists. For example, the term protozoa is used to refer to heterotrophic species of protists that do not form filaments.

Systematists today do not treat Protista as a formal taxon, but the term “protist” is still commonly used for convenience in two ways. The most popular contemporary definition is a phylogenetic one, that identifies a paraphyletic group: a protist is any eukaryote that is not an animal, (land) plant, or (true) fungus; this definition excludes many unicellular groups, like the Microsporidia (fungi), many Chytridiomycetes (fungi), and yeasts (fungi), and also a non-unicellular group included in Protista in the past, the Myxozoa (animal).

The taxonomy of protists is still changing. Newer classifications attempt to present monophyletic groups based on morphological (especially ultrastructural), biochemical (chemotaxonomy) and DNA sequence (molecular research) information. Because the protists as a whole are paraphyletic, new systems often split up or abandon the kingdom, instead treating the protist groups as separate lines of eukaryotes.

Table 14.2: Nutritional types in protist metabolism

Nutritional Type	Source of Energy	Source of Carbon	Examples
Photoautotrophs	Sunlight	Organic compounds or carbon fixation	Most algae
Chemoheterotrophs	Organic compounds	Organic compounds	Apicomplexa, Trypanosomes or Amoebae

14.1 Protozoa

Protozoa⁶ (also protozoan, plural protozoans) is an informal term for a group of single-celled eukaryotes, either free-living or parasitic, which feed on organic matter such as other microorganisms or organic tissues and debris. Historically, protozoans were regarded as “one-celled animals”, because they often possess animal-like behaviours, such as motility and predation, and lack a cell wall, as found in plants and many algae. Although the traditional practice of grouping protozoa with animals is no longer considered valid, the term continues to be used in a loose way to describe single-celled protists (that is, eukaryotes that aren’t animals, plants, or fungi) that feed by heterotrophy.

In some systems of biological classification, Protozoa remains a high-level taxonomic group. When first introduced by Georg Goldfuss in 1818, Protozoa was erected as a class within the animals, and its etymology is literally “first animals”. In later classification schemes it was elevated to a variety of higher ranks, including phylum, subkingdom and kingdom, and sometimes included within Protocista or Protista. With the advent of techniques such as molecular phylogenetics, it was realized that Protozoa did not represent a natural group; but while it is not an accepted taxon in cladistic analyses, some systematists continue to use it as a formal taxon.

In a series of classifications proposed by Thomas Cavalier-Smith and his collaborators since 1981, Protozoa has been ranked as a kingdom. The seven-kingdom scheme presented by Ruggiero et al. in 2015, places eight phyla under Kingdom Protozoa: Euglenozoa, Amoebozoa, Metamonada, Choanozoa sensu Cavalier-Smith, Loukozoa, Percolozoa, Microsporidia and Sulcozoa. Notably, this kingdom excludes several major groups of organisms traditionally placed among the protozoa, including the ciliates, dinoflagellates, foraminifera, and the parasitic apicomplexans, all of which are classified under Kingdom Chromista. Kingdom Protozoa, as defined in this scheme, does not form a natural group or clade, but a paraphyletic group or evolutionary grade, within which the members of Fungi, Animalia and Chromista are thought to have evolved.

The word “protozoa” (singular protozoon or protozoan) was coined in 1818 by zoologist Georg August Goldfuss, as the Greek equivalent of the German Urthiere, meaning “primitive, or original animals” (ur- ‘proto-’ + Thier ‘animal’). Goldfuss created Protozoa as a class containing what he believed to be the simplest animals. Originally, the group included not only single-celled microorganisms but also some “lower” multicellular animals,

such as rotifers, corals, sponges, jellyfish, bryozoa and polychaete worms. The term Protozoa is formed from the Greek words πρῶτος (prôtos), meaning “first”, and ζῷα (zôa), plural of ζῶον (zôn), meaning “animal”. The use of Protozoa as a formal taxon has been discouraged by some researchers, mainly because the term implies kinship with animals (Metazoa) and promotes an arbitrary separation of “animal-like” from “plant-like” organisms.

In 1848, as a result of advancements in cell theory pioneered by Theodor Schwann and Matthias Schleiden, the anatomist and zoologist C. T. von Siebold proposed that the bodies of protozoans such as ciliates and amoebae consisted of single cells, similar to those from which the multicellular tissues of plants and animals were constructed. Von Siebold redefined Protozoa to include only such unicellular forms, to the exclusion of all metazoa (animals). At the same time, he raised the group to the level of a phylum containing two broad classes of microorganisms: Infusoria (mostly ciliates and flagellated algae), and Rhizopoda (amoeboid organisms). The definition of Protozoa as a phylum or sub-kingdom composed of “unicellular animals” was adopted by the zoologist Otto Bütschli—celebrated at his centenary as the “architect of protozoology”—and the term came into wide use.

As a phylum under Animalia, the Protozoa were firmly rooted in the old “two-kingdom” classification of life, according to which all living beings were classified as either animals or plants. As long as this scheme remained dominant, the protozoa were understood to be animals and studied in departments of Zoology, while photosynthetic microorganisms and microscopic fungi—the so-called Protophyta—were assigned to the Plants, and studied in departments of Botany.

Criticism of this system began in the latter half of the 19th century, with the realization that many organisms met the criteria for inclusion among both plants and animals. For example, the algae Euglena and Dinobryon have chloroplasts for photosynthesis, but can also feed on organic matter and are motile. In 1860, John Hogg argued against the use of “protozoa”, on the grounds that “naturalists are divided in opinion—and probably some will ever continue so—whether many of these organisms, or living beings, are animals or plants.” As an alternative, he proposed a new kingdom called Primigenum, consisting of both the protozoa and unicellular algae (protophyta), which he combined together under the name “Protocista”. In Hogg’s conception, the animal and plant kingdoms were likened to two great “pyramids” blending at their bases in the Kingdom Primigenum.

Six years later, Ernst Haeckel also proposed a third kingdom of life, which he named Protista. At first, Haeckel included a few multicellular organisms in this kingdom, but

⁶<https://en.wikipedia.org/wiki/Protozoa>

in later work he restricted the Protista to single-celled organisms, or simple colonies whose individual cells are not differentiated into different kinds of tissues.

Despite these proposals, Protozoa emerged as the preferred taxonomic placement for heterotrophic microorganisms such as amoebae and ciliates, and remained so for more than a century. In the course of the 20th century, however, the old “two kingdom” system began to weaken, with the growing awareness that fungi did not belong among the plants, and that most of the unicellular protozoa were no more closely related to the animals than they were to the plants. By mid-century, some biologists, such as Herbert Copeland, Robert H. Whittaker and Lynn Margulis, advocated the revival of Haeckel’s Protista or Hogg’s Protoctista as a kingdom-level eukaryotic group, alongside Plants, Animals and Fungi. A variety of multi-kingdom systems were proposed, and Kingdoms Protista and Protoctista became well established in biology texts and curricula.

While many taxonomists have abandoned Protozoa as a high-level group, Thomas Cavalier-Smith has retained it as a kingdom in the various classifications he has proposed. As of 2015, Cavalier-Smith’s Protozoa excludes several major groups of organisms traditionally placed among the protozoa, including the ciliates, dinoflagellates and foraminifera (all members of the SAR supergroup). In its current form, his kingdom Protozoa is a paraphyletic group which includes a common ancestor and most of its descendants, but excludes two important clades that branch within it: the animals and fungi.

Since the protozoa, as traditionally defined, can no longer be regarded as “primitive animals” the terms “protists”, “Protista” or “Protoctista” are sometimes preferred. In 2005, members of the Society of Protozoologists voted to change its name to the International Society of Protistologists. Free-living protozoans are common and often abundant in fresh, brackish and salt water, as well as other moist environments, such as soils and mosses. Some species thrive in extreme environments such as hot springs and hypersaline lakes and lagoons. All protozoa require a moist habitat; however, some can survive for long periods of time in dry environments, by forming resting cysts which enable them to remain dormant until conditions improve.

Parasitic and symbiotic protozoa live on or within other organisms, including vertebrates and invertebrates, as well as plants and other single-celled organisms. Some are harmless or beneficial to their host organisms; others may be significant causes of diseases, such as babesia, malaria and toxoplasmosis.

Association between protozoan symbionts and their

host organisms can be mutually beneficial. Flagellated protozoans such as Trichonympha and Pyrsonympha inhabit the guts of termites, where they enable their insect host to digest wood by helping to break down complex sugars into smaller, more easily digested molecules. A wide range of protozoans live commensally in the rumens of ruminant animals, such as cattle and sheep. These include flagellates, such as Trichomonas, and ciliated protozoa, such as Isotricha and Entodinium. The ciliate subclass Astomatia is composed entirely of mouthless symbionts adapted for life in the guts of annelid worms.

All protozoans are heterotrophic, deriving nutrients from other organisms, either by ingesting them whole or consuming their organic remains and waste-products. Some protozoans take in food by phagocytosis, engulfing organic particles with pseudopodia (as amoebae do), or taking in food through a specialized mouth-like aperture called a cytostome. Others take in food by osmotrophy, absorbing dissolved nutrients through their cell membranes.

Parasitic protozoans use a wide variety of feeding strategies, and some may change methods of feeding in different phases of their life cycle. For instance, the malaria parasite Plasmodium feeds by pinocytosis during its immature trophozoite stage of life (ring phase), but develops a dedicated feeding organelle (cytostome) as it matures within a host’s red blood cell.

Protozoa may also live as mixotrophs, supplementing a heterotrophic diet with some form of autotrophy. Some protozoa form close associations with symbiotic photosynthetic algae, which live and grow within the membranes of the larger cell and provide nutrients to the host. Others practice kleptoplasty, stealing chloroplasts from prey organisms and maintaining them within their own cell bodies as they continue to produce nutrients through photosynthesis. The ciliate Mesodinium rubrum retains functioning plastids from the cryptophyte algae on which it feeds, using them to nourish themselves by autotrophy. These, in turn, may be passed along to dinoflagellates of the genus Dinophysis, which prey on Mesodinium rubrum but keep the enslaved plastids for themselves. Within Dinophysis, these plastids can continue to function for months.

Organisms traditionally classified as protozoa are abundant in aqueous environments and soil, occupying a range of trophic levels. The group includes flagellates (which move with the help of whip-like structures called flagella), ciliates (which move by using hair-like structures called cilia) and amoebae (which move by the use of foot-like structures called pseudopodia). Some protozoa are sessile, and do not move at all.

Unlike plants, fungi and most types of algae, proto-

zoans do not typically have a rigid cell wall, but are usually enveloped by elastic structures of membranes that permit movement of the cell. In some protozoans, such as the ciliates and euglenozoans, the cell is supported by a composite membranous envelope called the “pellicle”. The pellicle gives some shape to the cell, especially during locomotion. Pellicles of protozoan organisms vary from flexible and elastic to fairly rigid. In ciliates and Apicomplexa, the pellicle is supported by closely packed vesicles called alveoli. In euglenids, it is formed from protein strips arranged spirally along the length of the body. Familiar examples of protists with a pellicle are the euglenoids and the ciliate *Paramecium*. In some protozoa, the pellicle hosts epibiotic bacteria that adhere to the surface by their fimbriae (attachment pili).

Some protozoa have two-phase life cycles, alternating between proliferative stages (e.g., trophozoites) and dormant cysts. As cysts, protozoa can survive harsh conditions, such as exposure to extreme temperatures or harmful chemicals, or long periods without access to nutrients, water, or oxygen for periods of time. Being a cyst enables parasitic species to survive outside of a host, and allows their transmission from one host to another. When protozoa are in the form of trophozoites (Greek *tropho* = to nourish), they actively feed. The conversion of a trophozoite to cyst form is known as encystation, while the process of transforming back into a trophozoite is known as excystation.

Protozoans reproduce asexually by binary fission or multiple fission. Many protozoan species also exchange genetic material by sexual means (typically, through conjugation), but this is generally decoupled from the process of reproduction, and does not immediately result in increased population.

Although meiotic sex is widespread among present day eukaryotes, it has, until recently, been unclear whether or not eukaryotes were sexual early in their evolution. Due to recent advances in gene detection and other techniques, evidence has been found for some form of meiotic sex in an increasing number of protozoans of ancient lineage that diverged early in eukaryotic evolution. (See eukaryote reproduction.) Thus, such findings suggest that meiotic sex arose early in eukaryotic evolution. Examples of protozoan meiotic sexuality are described in the articles *Amoebozoa*, *Giardia lamblia*, *Leishmania*, *Plasmodium falciparum* biology, *Paramecium*, *Toxoplasma gondii*, *Trichomonas vaginalis* and *Trypanosoma brucei*.

Historically, the Protozoa were classified as “unicellular animals”, as distinct from the Protophyta, single-celled photosynthetic organisms (algae) which were considered primitive plants. Both groups were commonly given the rank of phylum, under the kingdom Protista. In older sys-

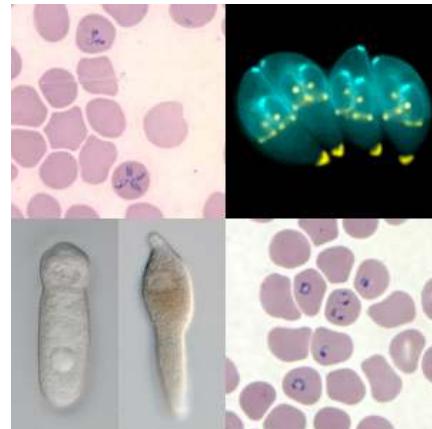


Figure 14.3: Composite image⁷ of various Apicomplexan parasites showing *Babesia microti* in red blood cells (top left), *Toxoplasma gondii* (top right), a septate eugregarine (bottom left), *Lankesteria cystodytae* (bottom middle), and *Plasmodium falciparum* in red blood cells (bottom right).

tems of classification, the phylum Protozoa was commonly divided into several sub-groups, reflecting the means of locomotion. Classification schemes differed, but throughout much of the 20th century the major groups of Protozoa included:

- Flagellates, or Mastigophora (motile cells equipped with whiplike organelles of locomotion, e.g., *Giardia lamblia*)
- Amoebea or Sarcodina (cells that move by extending pseudopodia or lamellipodia, e.g., *Entamoeba histolytica*)
- Sporozoans, or Sporozoa (parasitic, spore-producing cells, whose adult form lacks organs of motility, e.g., *Plasmodium knowlesi*)
 - Apicomplexa (now in Alveolata)
 - Microsporidia (now in Fungi)
 - Ascetosporea (now in Rhizaria)
 - Myxosporidia (now in Cnidaria)
- Ciliates, or Ciliophora (cells equipped with large numbers of short hairlike organs of locomotion, e.g. *Balantidium coli*)

With the emergence of molecular phylogenetics and tools enabling researchers to directly compare the DNA of different organisms, it became evident that, of the main sub-groups of Protozoa, only the ciliates (Ciliophora) formed a natural group, or monophyletic clade (that is, a distinct lineage of organisms sharing common ancestry). The other classes or subphyla of Protozoa were all polyphyletic groups composed of organisms that, despite similarities of appearance or way of life, were not necessarily closely related to one another. In the system of eukaryote classification currently endorsed by

the International Society of Protistologists, members of the old phylum Protozoa have been distributed among a variety of supergroups.

A number of protozoan pathogens are human parasites, causing diseases such as malaria (by *Plasmodium*), amoebiasis, giardiasis, toxoplasmosis, cryptosporidiosis, trichomoniasis, Chagas disease, leishmaniasis, African trypanosomiasis (sleeping sickness), *Acanthamoeba* keratitis, and primary amoebic meningoencephalitis (naegleriasis).

The protozoan *Ophryocystis elektroscirra* is a parasite of butterfly larvae, passed from female to caterpillar. Severely infected individuals are weak, unable to expand their wings, or unable to eclose, and have shortened lifespans, but parasite levels vary in populations. Infection creates a culling effect, whereby infected migrating animals are less likely to complete the migration. This results in populations with lower parasite loads at the end of the migration. This is not the case in laboratory or commercial rearing, where after a few generations, all individuals can be infected.

14.2 Flagellates

In older classifications, flagellated protozoa were grouped in Flagellata (= Mastigophora), sometimes divided in Phytoflagellata (= Phytomastigina, mostly autotrophic) and Zooflagellata (= Zoomastigina, heterotrophic). They were sometimes grouped with Sarcodina (ameboids) in the group Sarcomastigophora.

Excavata is a major supergroup of unicellular organisms eukaryotes that are classified based on their flagellar structures, and they are considered to be the most basal Flagellate lineage. It contains a variety of free-living and symbiotic forms, and also includes some important parasites of humans, including *Giardia* and *Trichomonas*. Except for Euglenozoa, they are all non-photosynthetic.

Euglena is a genus of single cell flagellate eukaryotes. It is the best known and most widely studied member of the class Euglenoidea, a diverse group containing some 54 genera and at least 800 species. Species of *Euglena* are found in freshwater and salt water. They are often abundant in quiet inland waters where they may bloom in numbers sufficient to color the surface of ponds and ditches green (*E. viridis*) or red (*E. sanguinea*).

The species *Euglena gracilis* has been used extensively in the laboratory as a model organism.

Most species of *Euglena* have photosynthesizing chloroplasts within the body of the cell, which enable them to feed by autotrophy, like plants. However, they can also take nourishment heterotrophically, like ani-



Figure 14.4: *Euglena*.

mals. Since *Euglena* have features of both animals and plants, early taxonomists, working within the Linnaean two-kingdom system of biological classification, found them difficult to classify. It was the question of where to put such "unclassifiable" creatures that prompted Ernst Haeckel to add a third living kingdom (a fourth kingdom *in toto*) to the Animale, Vegetable (and Lapideum meaning Mineral) of Linnaeus: the Kingdom Protista.

When feeding as a heterotroph, *Euglena* takes in nutrients by osmotrophy, and can survive without light on a diet of organic matter, such as beef extract, peptone, acetate, ethanol or carbohydrates. When there is sufficient sunlight for it to feed by phototrophy, it uses chloroplasts containing the pigments chlorophyll a and chlorophyll b to produce sugars by photosynthesis. *Euglena*'s chloroplasts are surrounded by three membranes, while those of plants and the green algae (among which earlier taxonomists often placed *Euglena*) have only two membranes. This fact has been taken as morphological evidence that *Euglena*'s chloroplasts evolved from a eukaryotic green alga. Thus, the similarities between *Euglena* and plants would have arisen not because of kinship but because of a secondary endosymbiosis. Molecular phylogenetic analysis has lent support to this hypothesis, and it is now generally accepted.

Euglena chloroplasts contain pyrenoids, used in the synthesis of paramylon, a form of starch energy storage enabling *Euglena* to survive periods of light deprivation. The presence of pyrenoids is used as an identifying feature of the genus, separating it from other euglenoids, such as *Lepocinclis* and *Phacus*.

Euglena have two flagella rooted in basal bodies located in a small reservoir at the front of the cell. Typically, one flagellum is very short, and does not protrude from the cell, while the other is long enough to be seen with light microscopy. In some species, such as *Euglena mutabilis*, both flagella are "non-emergent"—entirely confined to the interior of the cell's reservoir—and consequently cannot be seen in the light microscope. In species that possess a long,

emergent flagellum, it may be used to help the organism swim. The surface of the flagellum is coated with about 30,000 extremely fine filaments called mastigonemes.

Like other euglenoids, Euglena possess a red eyespot, an organelle composed of carotenoid pigment granules. The red spot itself is not thought to be photosensitive. Rather, it filters the sunlight that falls on a light-detecting structure at the base of the flagellum (a swelling, known as the paraflagellar body), allowing only certain wavelengths of light to reach it. As the cell rotates with respect to the light source, the eyespot partially blocks the source, permitting the Euglena to find the light and move toward it (a process known as phototaxis).

Euglena lacks a cell wall. Instead, it has a pellicle made up of a protein layer supported by a substructure of microtubules, arranged in strips spiraling around the cell. The action of these pellicle strips sliding over one another, known as metaboly, gives Euglena its exceptional flexibility and contractility. The mechanism of this euglenoid movement is not understood, but its molecular basis may be similar to that of amoeboid movement.

In low moisture conditions, or when food is scarce, Euglena forms a protective wall around itself and lies dormant as a resting cyst until environmental

Euglena reproduce asexually through binary fission, a form of cell division. Reproduction begins with the mitosis of the cell nucleus, followed by the division of the cell itself. Euglena divide longitudinally, beginning at the front end of the cell, with the duplication of flagellar processes, gullet and stigma. Presently, a cleavage forms in the anterior, and a V-shaped bifurcation gradually moves toward the posterior, until the two halves are entirely separated

14.3 Ciliates

The ciliates are a group of protists characterized by the presence of hair-like organelles called cilia, which are identical in structure to eukaryotic flagella, but are in general shorter and present in much larger numbers, with a different undulating pattern than flagella. Cilia occur in all members of the group (although the peculiar Suctoria only have them for part of their life-cycle) and are variously used in swimming, crawling, attachment, feeding, and sensation.

Ciliates are an important group of protists, common almost anywhere there is water — in lakes, ponds, oceans, rivers, and soils. About 4,500 unique free-living species have been described, and the potential number of extant species is estimated at 27,000–40,000. Included in this number are many ectosymbiotic and endosymbiotic species, as well as some obligate and opportunistic parasites. Ciliate species range in size from as little

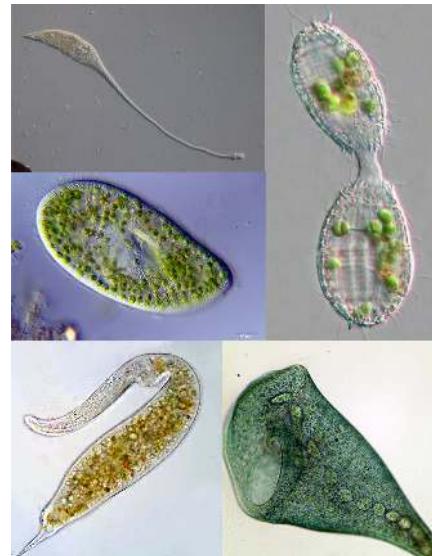


Figure 14.5: Composite image⁸ of various ciliates: *Lacrymaria olor*, *Paramecium bursaria*, *Coleps*, *Dileptus*, *Stentor coeruleus*.

as 10 µm in some colpodeans to as much as 4 mm in length in some geleids, and include some of the most morphologically complex protozoans.

Unlike most other eukaryotes, ciliates have two different sorts of nuclei: a tiny, diploid micronucleus (the “generative nucleus,” which carries the germline of the cell), and a large, polyploid macronucleus (the “vegetative nucleus,” which takes care of general cell regulation, expressing the phenotype of the organism). The latter is generated from the micronucleus by amplification of the genome and heavy editing. The micronucleus passes its genetic material to offspring, but does not express its genes. The macronucleus provides the small nuclear RNA for vegetative growth.

Division of the macronucleus occurs by amitosis, and the segregation of the chromosomes occurs by a process whose mechanism is unknown. This process is not perfect, and after about 200 generations the cell shows signs of aging. Periodically the macronuclei must be regenerated from the micronuclei. In most, this occurs during conjugation. Here two cells line up, the micronuclei undergo meiosis, some of the haploid daughters are exchanged and then fuse to form new micronuclei and macronuclei.

Food vacuoles are formed through phagocytosis and typically follow a particular path through the cell as their contents are digested and broken down by lysosomes so the substances the vacuole contains are then small enough to diffuse through the membrane of the food vacuole into the cell. Anything left in the food vacuole by the time it reaches the cytoproct (anal pore) is discharged by exocytosis. Most ciliates also have one or more prominent con-

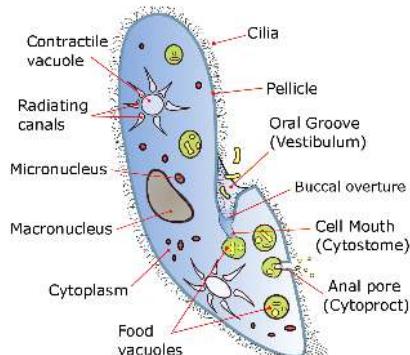


Figure 14.6: Diagram of the ciliate Paramecium⁹

tractile vacuoles, which collect water and expel it from the cell to maintain osmotic pressure, or in some function to maintain ionic balance. In some genera, such as *Paramecium*, these have a distinctive star shape, with each point being a collecting tube.

Most ciliates are heterotrophs, feeding on smaller organisms, such as bacteria and algae, and detritus swept into the oral groove (mouth) by modified oral cilia. This usually includes a series of membranelles to the left of the mouth and a paroral membrane to its right, both of which arise from polykinetids, groups of many cilia together with associated structures. The food is moved by the cilia through the mouth pore into the gullet, which forms food vacuoles.

Feeding techniques vary considerably, however. Some ciliates are mouthless and feed by absorption (osmotrophy), while others are predatory and feed on other protozoa and in particular on other ciliates. Some ciliates parasitize animals, although only one species, *Balantidium coli*, is known to cause disease in humans.

Ciliates reproduce asexually, by various kinds of fission. During fission, the micronucleus undergoes mitosis and the macronucleus elongates and undergoes amitosis (except among the Karyorelictan ciliates, whose macronuclei do not divide). The cell then divides in two, and each new cell obtains a copy of the micronucleus and the macronucleus.

Typically, the cell is divided transversally, with the anterior half of the ciliate (the proter) forming one new organism, and the posterior half (the opisthe) forming another. However, other types of fission occur in some ciliate groups. These include budding (the emergence of small ciliated offspring, or "swarmers", from the body of a mature parent); strobilation (multiple divisions along the cell body, producing a chain of new organisms); and palintomy (multiple fissions, usually within a cyst).

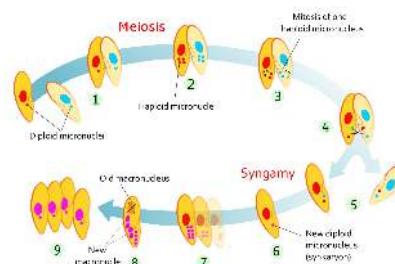


Figure 14.7: Stages of conjugation in *Paramecium caudatum*¹⁰

Fission may occur spontaneously, as part of the vegetative cell cycle. Alternatively, it may proceed as a result of self-fertilization (autogamy), or it may follow conjugation, a sexual phenomenon in which ciliates of compatible mating types exchange genetic material. While conjugation is sometimes described as a form of reproduction, it is not directly connected with reproductive processes, and does not directly result in an increase in the number of individual ciliates or their progeny.

Ciliate conjugation is a sexual phenomenon that results in genetic recombination and nuclear reorganization within the cell. During conjugation, two ciliates of a compatible mating type form a bridge between their cytoplasms. The micronuclei undergo meiosis, the macronuclei disappear, and haploid micronuclei are exchanged over the bridge. In some ciliates (peritrichs, chonotrichs and some suuctarians), conjugating cells become permanently fused, and one conjugant is absorbed by the other. In most ciliate groups, however, the cells separate after conjugation, and both form new macronuclei from their micronuclei. Conjugation and autogamy are always followed by fission.

In many ciliates, such as *Paramecium*, conjugating partners (gamonts) are similar or indistinguishable in size and shape. This is referred to as "isogamontic" conjugation. In some groups, partners are different in size and shape. This is referred to as "anisogamontic" conjugation. In sessile peritrichs, for instance, one sexual partner (the microconjugant) is small and mobile, while the other (macroconjugant) is large and sessile.

In *Paramecium caudatum*, the stages of conjugation are as follows (see diagram at right):

1. Compatible mating strains meet and partly fuse
2. The micronuclei undergo meiosis, producing four haploid micronuclei per cell.
3. Three of these micronuclei disintegrate. The fourth undergoes mitosis.
4. The two cells exchange a micronucleus.

5. The cells then separate.
6. The micronuclei in each cell fuse, forming a diploid micronucleus.
7. Mitosis occurs three times, giving rise to eight micronuclei.
8. Four of the new micronuclei transform into macronuclei, and the old macronucleus disintegrates.
9. Binary fission occurs twice, yielding four identical daughter cells.

Ciliates contain two types of nuclei: somatic “macronucleus” and the germline “micronucleus”. Only the DNA in the micronucleus is passed on during sexual reproduction (conjugation). On the other hand, only the DNA in the macronucleus is actively expressed and results in the phenotype of the organism. Macronuclear DNA is derived from micronuclear DNA by amazingly extensive DNA rearrangement and amplification.

The macronucleus begins as a copy of the micronucleus. The micronuclear chromosomes are fragmented into many smaller pieces and amplified to give many copies. The resulting macronuclear chromosomes often contain only a single gene. In *Tetrahymena*, the micronucleus has 10 chromosomes (five per haploid genome), while the macronucleus has over 20,000 chromosomes.

In addition, the micronuclear genes are interrupted by numerous “internal eliminated sequences” (IESs). During development of the macronucleus, IESs are deleted and the remaining gene segments, macronuclear destined sequences (MDSs), are spliced together to give the operational gene. *Tetrahymena* has about 6,000 IESs and about 15% of micronuclear DNA is eliminated during this process. The process is guided by small RNAs and epigenetic chromatin marks.

In spirotrich ciliates (such as *Oxytricha*), the process is even more complex due to “gene scrambling”: the MDSs in the micronucleus are often in different order and orientation from that in the macronuclear gene, and so in addition to deletion, DNA inversion and translocation are required for “unscrambling”. This process is guided by long RNAs derived from the parental macronucleus. More than 95% of micronuclear DNA is eliminated during spirotrich macronuclear development.

14.4 Amoebozoa

Amoebozoa is a major taxonomic group containing about 2,400 described species of amoeboid protists, often possessing blunt, fingerlike, lobose pseudopods and tubular mitochondrial cristae. In most classification schemes, Amoebozoa is ranked as a phylum within either the kingdom Protista or the kingdom Protozoa. In the

classification favored by the International Society of Protistologists, it is retained as an unranked “super-group” within Eukaryota. Molecular genetic analysis supports Amoebozoa as a monophyletic clade. Most phylogenetic trees identify it as the sister group to Opisthokonta, another major clade which contains both fungi and animals as well as some 300 species of unicellular protists. Amoebozoa and Opisthokonta are sometimes grouped together in a high-level taxon, variously named Unikonta, Amorphea or Opimoda.

Amoebozoa includes many of the best-known amoeboid organisms, such as *Chaos*, *Entamoeba*, *Pelomyxa* and the genus *Amoeba* itself. Species of Amoebozoa may be either shelled (testate), or naked, and cells may possess flagella. Free-living species are common in both salt and freshwater as well as soil, moss and leaf litter. Some live as parasites or symbionts of other organisms, and some are known to cause disease in humans and other organisms.

While the majority of amoeboid species are unicellular, the group also includes several varieties of slime molds, which have a macroscopic, multicellular stage of life during which individual amoeboid cells aggregate to produce spores.

Amoebozoa vary greatly in size. Some are only 10–20 µm in diameter, while others are among the largest protozoa. The well-known species *Amoeba proteus*, which may reach 800 µm in length, is often studied in schools and laboratories as a representative cell or model organism, partly because of its convenient size. Multinucleate amoebae like *Chaos* and *Pelomyxa* may be several millimeters in length, and some multicellular amoeboida, such as the “dog vomit” slime mold *Fuligo septica*, can cover an area of several square meters.

Amoebozoa is a large and diverse group, but certain features are common to many of its members. The amoeboid cell is typically divided into a granular central mass, called endoplasm, and a clear outer layer, called ectoplasm. During locomotion, the endoplasm flows forwards and the ectoplasm runs backwards along the outside of the cell. In motion, many amoeboids have a clearly defined anterior and posterior and may assume a “monopodial” form, with the entire cell functioning as a single pseudopod. Large pseudopods may produce numerous clear projections called subpseudopodia (or determinate pseudopodia), which are extended to a certain length and then retracted, either for the purpose of locomotion or food intake. A cell may also form multiple indeterminate pseudopodia, through which the entire contents of the cell flow in the direction of locomotion. These are more or less tubular and are mostly filled with granular endoplasm. The cell mass flows into a leading pseudopod, and the others ultimately retract, unless the

organism changes direction.

While most amoebozoans are “naked,” like the familiar *Amoeba* and *Chaos*, or covered with a loose coat of minute scales, like *Cochliopodium* and *Korotnevella*, members of the order Arcellinida form rigid shells, or tests, equipped with a single aperture through which the pseudopods emerge. Arcellinid tests may be secreted from organic materials, as in *Arcella*, or built up from collected particles cemented together, as in *Diffugia*.

In all amoebozoans, the primary mode of nutrition is phagocytosis, in which the cell surrounds potential food particles with its pseudopods, sealing them into vacuoles within which they may be digested and absorbed. Some amoebozoans have a posterior bulb called a uroid, which may serve to accumulate waste, periodically detaching from the rest of the cell.[citation needed] When food is scarce, most species can form cysts, which may be carried aerially and introduce them to new environments.[citation needed] In slime moulds, these structures are called spores, and form on stalked structures called fruiting bodies or sporangia.

The majority of Amoebozoa lack flagella and more generally do not form microtubule-supported structures except during mitosis. However, flagella do occur among the Archamoebae, and many slime moulds produce biflagellate gametes[citation needed]. The flagellum is generally anchored by a cone of microtubules, suggesting a close relationship to the opisthokonts.[citation needed] The mitochondria in amoebozoan cells characteristically have branching tubular cristae. However, among the Archamoebae, which are adapted to anoxic or microaerophilic habitats, mitochondria have been lost.

Amoebiasis, also known as amebiasis or entamoebiasis, is an infection caused by any of the amoebozoans of the Entamoeba group. Symptoms are most common upon infection by *Entamoeba histolytica*. Amoebiasis can present with no, mild, or severe symptoms. Symptoms may include abdominal pain, mild diarrhoea, bloody diarrhea or severe colitis with tissue death and perforation. This last complication may cause peritonitis. People affected may develop anemia due to loss of blood.

Invasion of the intestinal lining causes amoebic bloody diarrhea or amoebic colitis. If the parasite reaches the bloodstream it can spread through the body, most frequently ending up in the liver where it causes amoebic liver abscesses. Liver abscesses can occur without previous diarrhea. Cysts of *Entamoeba* can survive for up to a month in soil or for up to 45 minutes under fingernails. It is important to differentiate between amoebiasis and bacterial colitis. The preferred diagnostic method is through faecal examination under microscope,

but requires a skilled microscopist and may not be reliable when excluding infection. This method however may not be able to separate between specific types. Increased white blood cell count is present in severe cases, but not in mild ones. The most accurate test is for antibodies in the blood, but it may remain positive following treatment.

Prevention of amoebiasis is by separating food and water from faeces and by proper sanitation measures. There is no vaccine. There are two treatment options depending on the location of the infection. Amoebiasis in tissues is treated with either metronidazole, tinidazole, nitazoxanide, dehydroemetine or chloroquine, while luminal infection is treated with diloxanide furoate or iodoquinoline. For treatment to be effective against all stages of the amoeba may require a combination of medications. Infections without symptoms do not require treatment but infected individuals can spread the parasite to others and treatment can be considered. Treatment of other *Entamoeba* infections apart from *E. histolytica* is not needed.

Amoebiasis is present all over the world. About 480 million people are infected with what appears to be *E. histolytica* and these result in the death of between 40,000–110,000 people every year. Most infections are now ascribed to *E. dispar*. *E. dispar* is more common in certain areas and symptomatic cases may be fewer than previously reported. The first case of amoebiasis was documented in 1875 and in 1891 the disease was described in detail, resulting in the terms amoebic dysentery and amoebic liver abscess. Further evidence from the Philippines in 1913 found that upon ingesting cysts of *E. histolytica* volunteers developed the disease. It has been known since 1897 that at least one non-disease-causing species of *Entamoeba* existed (*Entamoeba coli*), but it was first formally recognized by the WHO in 1997 that *E. histolytica* was two species, despite this having first been proposed in 1925. In addition to the now-recognized *E. dispar* evidence shows there are at least two other species of *Entamoeba* that look the same in humans – *E. moshkovskii* and *Entamoeba bangladeshi*. The reason these species haven't been differentiated until recently is because of the reliance on appearance.

14.5 Algae

The Archaeplastida (or kingdom Plantae sensu lato) are a major group of autotrophic eukaryotes, comprising the red algae (Rhodophyta), the green algae, and the land plants, together with a small group of freshwater unicellular algae called glaucophytes. The Archaeplastida have chloroplasts that are surrounded by two membranes, suggesting that they were acquired directly from endosymbiotic cyanobacteria. All other groups besides the amoeboid *Paulinella*

chromatophora, have chloroplasts surrounded by three or four membranes, suggesting they were acquired secondarily from red or green algae. Unlike red and green algae, glaucophytes have never been involved in secondary endosymbiosis events.

The cells of the Archaeplastida typically lack centrioles and have mitochondria with flat cristae. They usually have a cell wall that contain cellulose, and food is stored in the form of starch. However, these characteristics are also shared with other eukaryotes. The main evidence that the Archaeplastida form a monophyletic group comes from genetic studies, which indicate their plastids probably had a single origin. This evidence is disputed. Based on the evidence to date, it is not possible to confirm or refute alternative evolutionary scenarios to a single primary endosymbiosis. Photosynthetic organisms with plastids of different origin (such as brown algae) do not belong to the Archaeplastida.

The archaeplastidans fall into two main evolutionary lines. The red algae are pigmented with chlorophyll a and phycobiliproteins, like most cyanobacteria, and accumulate starch outside the chloroplasts. The green algae and land plants – together known as Viridiplantae (Latin for “green plants”) or Chloroplastida – are pigmented with chlorophylls a and b, but lack phycobiliproteins, and starch is accumulated inside the chloroplasts. The glaucophytes have typical cyanobacterial pigments, and are unusual in retaining a cell wall within their plastids (called cyanelles).

Algae¹¹ is an informal term for a large and diverse group of photosynthetic eukaryotic organisms. It is a polyphyletic grouping, including species from multiple distinct clades. Included organisms range from unicellular microalgae, such as Chlorella and the diatoms, to multicellular forms, such as the giant kelp, a large brown alga which may grow up to 50 m in length. Most are aquatic and autotrophic and lack many of the distinct cell and tissue types, such as stomata, xylem and phloem, which are found in land plants. The largest and most complex marine algae are called seaweeds, while the most complex freshwater forms are the Charophyta, a division of green algae which includes, for example, Spirogyra and stoneworts.

No definition of algae is generally accepted. One definition is that algae “have chlorophyll as their primary photosynthetic pigment and lack a sterile covering of cells around their reproductive cells”. Although cyanobacteria are often referred to as “blue-green algae”, most authorities exclude all prokaryotes from the definition of algae.



Figure 14.8: A variety of microscopic unicellular and colonial freshwater algae¹²

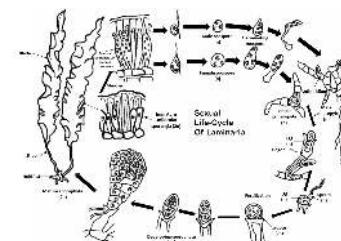


Figure 14.9: The sexual life cycle of Laminaria, a representative of some 30 different species of brown algae that are commonly called “kelp”.¹³

Algae constitute a polyphyletic group since they do not include a common ancestor, and although their plastids seem to have a single origin, from cyanobacteria, they were acquired in different ways. Green algae are examples of algae that have primary chloroplasts derived from endosymbiotic cyanobacteria. Diatoms and brown algae are examples of algae with secondary chloroplasts derived from an endosymbiotic red alga.

Algae exhibit a wide range of reproductive strategies, from simple asexual cell division to complex forms of sexual reproduction.

Algae lack the various structures that characterize land plants, such as the phyllids (leaf-like structures) of bryophytes, rhizoids in nonvascular plants, and the roots, leaves, and other organs found in tracheophytes (vascular plants). Most are phototrophic, although some are mixotrophic, deriving energy both from photosynthesis and uptake of organic carbon either by osmotrophy, myzotrophy, or phagotrophy. Some unicellular species of green algae, many golden algae, euglenids, dinoflagellates, and other algae have become heterotrophs (also called colorless or apochlorotic algae), sometimes parasitic, relying entirely on external energy sources and

¹¹<https://en.wikipedia.org/wiki/Algae>

have limited or no photosynthetic apparatus. Some other heterotrophic organisms, such as the apicomplexans, are also derived from cells whose ancestors possessed plastids, but are not traditionally considered as algae. Algae have photosynthetic machinery ultimately derived from cyanobacteria that produce oxygen as a by-product of photosynthesis, unlike other photosynthetic bacteria such as purple and green sulfur bacteria. Fossilized filamentous algae from the Vindhya basin have been dated back to 1.6 to 1.7 billion years ago.

The first land plants probably evolved from shallow freshwater charophyte algae much like Chara almost 500 million years ago. These probably had an isomorphic alternation of generations and were probably filamentous. Fossils of isolated land plant spores suggest land plants may have been around as long as 475 million years ago.

Most of the simpler algae are unicellular flagellates or amoeboids, but colonial and nonmotile forms have developed independently among several of the groups.

In three lines of algae, even higher levels of organization have been reached, with full tissue differentiation. These are the brown algae,—some of which may reach 50 m in length (kelps)—the red algae, and the green algae. The most complex forms are found among the charophyte algae (see Charales and Charophyta), in a lineage that eventually led to the higher land plants. The innovation that defines these nonalgal plants is the presence of female reproductive organs with protective cell layers that protect the zygote and developing embryo. Hence, the land plants are referred to as the Embryophytes.

Rhodophyta, Chlorophyta, and Heterokontophyta, the three main algal divisions, have lifecycles which show considerable variation and complexity. In general, an asexual phase exists where the seaweed's cells are diploid, a sexual phase where the cells are haploid, followed by fusion of the male and female gametes. Asexual reproduction permits efficient population increases, but less variation is possible. Commonly, in sexual reproduction of unicellular and colonial algae, two specialized, sexually compatible, haploid gametes make physical contact and fuse to form a zygote. To ensure a successful mating, the development and release of gametes is highly synchronized and regulated; pheromones may play a key role in these processes. Meiosis has been shown to occur in many different species of algae.

For example, *Fucus* is a genus of brown algae found in the intertidal zones of rocky seashores almost throughout the world. Species of *Fucus* are recorded almost worldwide. They are dominant on the shores of the British Isles, the northeastern coast of North America and California. These algae have a relatively simple life cycle and produce only

one type of thallus which grows to a maximum size of 2 m. Fertile cavities, the conceptacles, containing the reproductive cells are immersed in the receptacles near the ends of the branches. After meiosis oogonia and antheridia, the female and male reproductive organs, produce egg cells and sperm respectively that are released into the sea where fertilisation takes place. The resulting zygote develops directly into the diploid plant. This contrasts with the life cycle of the flowering plant, where the egg cells and sperm are produced by a haploid multicellular generation, albeit very strongly reduced, and the egg cells are fertilised within the ovules of the parent plant and then released as seeds.

Algae are prominent in bodies of water, common in terrestrial environments, and are found in unusual environments, such as on snow and ice. Seaweeds grow mostly in shallow marine waters, under 100 m (330 ft) deep; however, some such as *Navicula pennata* have been recorded to a depth of 360 m (1,180 ft). A type of algae, *Ancylonema nordenskioeldii*, was found in Greenland in areas known as the 'Dark Zone', which caused an increase in the rate of melting ice sheet. Same algae was found in the Italian Alps, after pink ice appeared on parts of the Presena glacier.

The various sorts of algae play significant roles in aquatic ecology. Microscopic forms that live suspended in the water column (phytoplankton) provide the food base for most marine food chains. In very high densities (algal blooms), these algae may discolor the water and outcompete, poison, or asphyxiate other life forms.

14.6 Slime Molds

Slime mold or slime mould is an informal name given to several kinds of unrelated eukaryotic organisms that can live freely as single cells, but can aggregate together to form multicellular reproductive structures. Slime molds were formerly classified as fungi but are no longer considered part of that kingdom. Although not forming a single monophyletic clade, they are grouped within the paraphyletic group referred to as kingdom Protista.

More than 900 species of slime mold occur globally. Their common name refers to part of some of these organisms' life cycles where they can appear as gelatinous "slime". This is mostly seen with the Myxogastria, which are the only macroscopic slime molds. Most slime molds are smaller than a few centimeters, but some species may reach sizes up to several square meters and masses up to 20 kilograms.

Many slime molds, mainly the "cellular" slime molds, do not spend most of their time in this state. When food is abundant, these slime molds exist as single-celled organisms. When food is in short supply, many of these single-



Figure 14.10: *Fuligo septica*, a slime mold¹⁴ *Fuligo septica* is a species of plasmodial slime mold, and a member of the Myxomycetes class. It is commonly known as the scrambled egg slime, or flowers of tan because of its peculiar yellowish, bile-colored appearance. Also known as the dog vomit slime mold, it is common with a worldwide distribution, and it is often found on bark mulch in urban areas after heavy rain or excessive watering. Their spores are produced on or in aerial sporangia and are spread by wind.

celled organisms will congregate and start moving as a single body. In this state they are sensitive to airborne chemicals and can detect food sources. They can readily change the shape and function of parts, and may form stalks that produce fruiting bodies, releasing countless spores, light enough to be carried on the wind or hitch a ride on passing animals.

They feed on microorganisms that live in any type of dead plant material. They contribute to the decomposition of dead vegetation, and feed on bacteria, yeasts, and fungi. For this reason, slime molds are usually found in soil, lawns, and on the forest floor, commonly on deciduous logs. In tropical areas they are also common on inflorescences and fruits, and in aerial situations (e.g., in the canopy of trees). In urban areas, they are found on mulch or in the leaf mold in rain gutters, and also grow in air conditioners, especially when the drain is blocked.

Slime molds begin life as amoeba-like cells. These unicellular amoebae are commonly haploid and feed on bacteria. These amoebae can mate if they encounter the correct mating type and form zygotes that then grow into plasmodia. These contain many nuclei without cell membranes between them, and can grow to meters in size. The species *Fuligo septica* is often seen as a slimy yellow network in and on rotting logs. The amoebae and the plasmodia engulf microorganisms. The plasmodium grows into an interconnected network of protoplasmic strands.

Within each protoplasmic strand, the cytoplasmic contents rapidly stream. If one strand is carefully watched for about 50 seconds, the cytoplasm can be seen to slow, stop, and then reverse direction. The streaming proto-



Figure 14.11: *Dictyostelium* Fruiting Body¹⁵ *Dictyostelium* is a genus of single- and multi-celled eukaryotic, phagotrophic bacterivores. Though they are Protista and in no way fungal, they traditionally are known as "slime molds". They are present in most terrestrial ecosystems as a normal and often abundant component of the soil microflora, and play an important role in the maintenance of balanced bacterial populations in soils.

plasm within a plasmodial strand can reach speeds of up to 1.35 mm per second, which is the fastest rate recorded for any microorganism. Migration of the plasmodium is accomplished when more protoplasm streams to advancing areas and protoplasm is withdrawn from rear areas. When the food supply wanes, the plasmodium will migrate to the surface of its substrate and transform into rigid fruiting bodies. The fruiting bodies or sporangia are what are commonly seen. They superficially look like fungi or molds but are not related to the true fungi. These sporangia will then release spores which hatch into amoebae to begin the life cycle again.

14.6.1 Reproduction of Protists

Some protists reproduce sexually using gametes, while others reproduce asexually.

Some species, for example *Plasmodium falciparum*, have extremely complex life cycles that involve multiple forms of the organism, some of which reproduce sexually and others asexually. However, it is unclear how frequently sexual reproduction causes genetic exchange between different strains of *Plasmodium* in nature and



Figure 14.12: Dictyostelium colony in process of aggregation¹⁶



Figure 14.13: Pseudoplasmodium or “slug” of a Dictyostelium¹⁷

most populations of parasitic protists may be clonal lines that rarely exchange genes with other members of their species.

Eukaryotes emerged in evolution more than 1.5 billion years ago. The earliest eukaryotes were likely protists. Although sexual reproduction is widespread among extant eukaryotes, it seemed unlikely until recently, that sex could be a primordial and fundamental characteristic of eukaryotes. A principal reason for this view was that sex appeared to be lacking in certain pathogenic protists whose ancestors branched off early from the eukaryotic family tree. However, several of these protists are now known to be capable of, or to recently have had the capability for, meiosis and hence sexual reproduction. For example, the common intestinal parasite *Giardia lamblia* was once considered to be a descendant of a protist lineage that predated the emergence of meiosis and sex. However, *G. lamblia* was recently found to have a core set of genes that function in meiosis and that are widely present among sexual eukaryotes. These results suggested that *G. lamblia* is capable of meiosis and thus sexual reproduction. Furthermore, direct evidence for meiotic recombination, indicative of sex, was also found in *G. lamblia*.

The pathogenic parasitic protists of the genus Leish-

mania have been shown to be capable of a sexual cycle in the invertebrate vector, likened to the meiosis undertaken in the trypanosomes.

Protists generally reproduce asexually under favorable environmental conditions, but tend to reproduce sexually under stressful conditions, such as starvation or heat shock. Oxidative stress, which is associated with the production of reactive oxygen species leading to DNA damage, also appears to be an important factor in the induction of sex in protists.

Some commonly found Protist pathogens such as *Toxoplasma gondii* are capable of infecting and undergoing asexual reproduction in a wide variety of animals – which act as secondary or intermediate host – but can undergo sexual reproduction only in the primary or definitive host (for example: felids such as domestic cats in this case).

Free-living Protists occupy almost any environment that contains liquid water. Many protists, such as algae, are photosynthetic and are vital primary producers in ecosystems, particularly in the ocean as part of the plankton. Protists make up a large portion of the biomass in both marine and terrestrial environments.

Other protists include pathogenic species, such as the kinetoplastid *Trypanosoma brucei*, which causes sleeping sickness, and species of the apicomplexan *Plasmodium*, which cause malaria.

Some protists are significant parasites of animals (e.g.; five species of the parasitic genus *Plasmodium* cause malaria in humans and many others cause similar diseases in other vertebrates), plants (the oomycete *Phytophthora infestans* causes late blight in potatoes) or even of other protists. Protist pathogens share many metabolic pathways with their eukaryotic hosts. This makes therapeutic target development extremely difficult – a drug that harms a protist parasite is also likely to harm its animal/plant host. A more thorough understanding of protist biology may allow these diseases to be treated more efficiently. For example, the apicoplast (a nonphotosynthetic chloroplast but essential to carry out important functions other than photosynthesis) present in apicomplexans provides an attractive target for treating diseases caused by dangerous pathogens such as plasmodium.

Chapter 15

The Theory of Evolution

Evolution¹ is change in the heritable characteristics of biological populations over successive generations. These characteristics are the expressions of genes that are passed on from parent to offspring during reproduction. Different characteristics tend to exist within any given population as a result of mutation, genetic recombination and other sources of genetic variation. Evolution occurs when evolutionary processes such as natural selection (including sexual selection) and genetic drift act on this variation, resulting in certain characteristics becoming more common or rare within a population. It is this process of evolution that has given rise to biodiversity at every level of biological organisation, including the levels of species, individual organisms and molecules.

Repeated formation of new species (speciation), change within species (anagenesis), and loss of species (extinction) throughout the evolutionary history of life on Earth are demonstrated by shared sets of morphological and biochemical traits, including shared DNA sequences. These shared traits are more similar among species that share a more recent common ancestor, and can be used to reconstruct a biological “tree of life” based on evolutionary relationships (phylogenetics), using both existing species and fossils. The fossil record includes a progression from early biogenic graphite, to microbial mat fossils, to fossilized multicellular organisms. Existing patterns of biodiversity have been shaped both by speciation and by extinction.

Charles Darwin² developed his theory of “natural selection” from 1838 onwards and was writing up his “big book” on the subject when Alfred Russel Wallace³ sent him a version of virtually the same theory in 1858. Their separate papers were presented together at an 1858 meeting of the Linnaean Society of London. At the end of 1859, Darwin’s book “On the Origin of Species” explained natural selection in detail and in a way, that led to an increasingly wide acceptance of Darwin’s concepts of evolution at the expense

of alternative theories.

According to Ernst Mayr⁴, Darwin’s theory actually consists of a number of different theories that can be best understood when they are clearly distinguished from each other. Mayr distinguished five independent theories:

1. The non-constancy of species (the basic theory of evolution)
2. The descent of all organisms from common ancestors (branching evolution)
3. The gradualness of evolution (no saltations, no discontinuities)
4. The multiplication of species (the origin of diversity)
5. Natural selection

The first and second theories were widely accepted by biologists rather quickly following Darwin’s publication. The other three theories were not widely accepted until the arrival of the so-called modern synthesis in the 20th century (see below).

Evolution by natural selection⁵ is a process demonstrated by the observation that more offspring are produced than can possibly survive, along with three facts about populations: 1) traits vary among individuals with respect to morphology, physiology, and behavior (phenotypic variation), 2) different traits confer different rates of survival and reproduction (differential fitness), and 3) traits can be passed from generation to generation (heritability of fitness). Thus, in successive generations members of a population are replaced by progeny of parents better adapted to survive and reproduce in the biophysical environment in which natural selection takes place.

The four most widely recognized evolutionary processes are natural selection (including sexual selection), genetic drift⁶, mutation⁷ and gene migration⁸ due to

¹<https://en.wikipedia.org/wiki/Evolution>

²https://en.wikipedia.org/wiki/Charles_Darwin

³https://en.wikipedia.org/wiki/Alfred_Russel_Wallace

⁴https://en.wikipedia.org/wiki/Ernst_Mayr

⁵https://en.wikipedia.org/wiki/Natural_selection

⁶https://en.wikipedia.org/wiki/Genetic_drift

⁷<https://en.wikipedia.org/wiki/Mutation>

⁸https://en.wikipedia.org/wiki/Gene_flow

genetic admixture. Natural selection and genetic drift sort variation; mutation and gene migration create variation.

The mechanisms of reproductive heritability and the origin of new traits remained a mystery. Towards this end, Darwin developed his provisional theory of pangenesis. In 1865, Gregor Mendel⁹ reported that traits were inherited in a predictable manner through the independent assortment and segregation of elements (later known as genes). Mendel's laws of inheritance eventually supplanted most of Darwin's pangenesis theory. August Weismann¹⁰ made the important distinction between germ cells that give rise to gametes (such as sperm and egg cells) and the somatic cells of the body, demonstrating that heredity passes through the germ line only. Hugo de Vries¹¹ connected Darwin's pangenesis theory to Weismann's germ/soma cell distinction and proposed that Darwin's pangenes were concentrated in the cell nucleus and when expressed they could move into the cytoplasm to change the cells structure. de Vries was also one of the researchers who made Mendel's work well-known, believing that Mendelian traits corresponded to the transfer of heritable variations along the germline. de Vries developed a mutation theory to explain how new variants originate. This led to a temporary rift between those who accepted Darwinian evolution and biometricalists who allied with de Vries. In the 1930s, pioneers in the field of population genetics, such as Ronald Fisher¹², Sewall Wright¹³ and J. B. S. Haldane¹⁴ set the foundations of evolution onto a robust statistical philosophy. The false contradiction between Darwin's theory, genetic mutations, and Mendelian inheritance was thus reconciled.

In the 1920s and 1930s the so-called modern synthesis¹⁵ connected natural selection and population genetics, based on Mendelian inheritance, into a unified theory that applied generally to any branch of biology. The modern synthesis explained patterns observed across species in populations, through fossil transitions in paleontology, and complex cellular mechanisms in developmental biology. The publication of the structure of DNA¹⁶ by James Watson¹⁷ and Francis Crick¹⁸ in 1953 demonstrated a physical mechanism for inheritance. Molecular biology improved our understanding of the relationship between genotype and phenotype. Advancements were also made in phylogenetic systematics, mapping the transition of

traits into a comparative and testable framework through the publication and use of evolutionary trees. In 1973, evolutionary biologist Theodosius Dobzhansky¹⁹ penned that "nothing in biology makes sense except in the light of evolution," because it has brought to light the relations of what first seemed disjointed facts in natural history into a coherent explanatory body of knowledge that describes and predicts many observable facts about life on this planet.

All life on Earth shares a common ancestor known as the last universal common ancestor²⁰ (LUCA), which lived approximately 3.5–3.8 billion years ago. The fossil record includes a progression from early biogenic graphite, to microbial mat fossils, to fossilised multicellular organisms. Existing patterns of biodiversity have been shaped by repeated formations of new species (speciation), changes within species (anagenesis) and loss of species (extinction) throughout the evolutionary history of life on Earth. Morphological and biochemical traits are more similar among species that share a more recent common ancestor, and can be used to reconstruct phylogenetic trees.

In terms of practical application, an understanding of evolution has been instrumental to developments in numerous scientific and industrial fields, including agriculture, human and veterinary medicine, and the life sciences in general. Discoveries in evolutionary biology have made a significant impact not just in the traditional branches of biology but also in other academic disciplines, including biological anthropology, and evolutionary psychology.

15.1 History of Evolutionary Thought

The proposal that one type of organism could descend from another type goes back to some of the first pre-Socratic Greek philosophers, such as Anaximander and Empedocles. Such proposals survived into Roman times.

In contrast to these materialistic views, Aristotelianism considered all natural things as actualisations of fixed natural possibilities, known as forms. This was part of a medieval teleological understanding of nature in which all things have an intended role to play in a divine cosmic order. Variations of this idea became the standard understanding of the Middle Ages and were integrated into Christian learning, but Aristotle did not demand that real types of organisms always correspond one-for-one with exact metaphysical forms and specifically gave examples of how new types of living things could come to be.

In the 17th century, the new method of modern science rejected the Aristotelian approach. It sought explanations

⁹https://en.wikipedia.org/wiki/Gregor_Mendel

¹⁰https://en.wikipedia.org/wiki/August_Weismann

¹¹https://en.wikipedia.org/wiki/Hugo_de_Vries

¹²https://en.wikipedia.org/wiki/Ronald_Fisher

¹³https://en.wikipedia.org/wiki/Sewall_Wright

¹⁴https://en.wikipedia.org/wiki/J._B._S._Haldane

¹⁵[https://en.wikipedia.org/wiki/Modern_synthesis_\(20th_century\)](https://en.wikipedia.org/wiki/Modern_synthesis_(20th_century))

¹⁶<https://en.wikipedia.org/wiki/DNA>

¹⁷https://en.wikipedia.org/wiki/James_Watson

¹⁸https://en.wikipedia.org/wiki/Francis_Crick

¹⁹https://en.wikipedia.org/wiki/Theodosius_Dobzhansky

²⁰https://en.wikipedia.org/wiki/Last_universal_common_ancestor

of natural phenomena in terms of physical laws that were the same for all visible things and that did not require the existence of any fixed natural categories or divine cosmic order. However, this new approach was slow to take root in the biological sciences, the last bastion of the concept of fixed natural types. John Ray applied one of the previously more general terms for fixed natural types, "species", to plant and animal types, but he strictly identified each type of living thing as a species and proposed that each species could be defined by the features that perpetuated themselves generation after generation. The biological classification introduced by Carl Linnaeus in 1735 explicitly recognised the hierarchical nature of species relationships, but still viewed species as fixed according to a divine plan.

Other naturalists of this time speculated on the evolutionary change of species over time according to natural laws. In 1751, Pierre Louis Maupertuis wrote of natural modifications occurring during reproduction and accumulating over many generations to produce new species. Georges-Louis Leclerc, Comte de Buffon suggested that species could degenerate into different organisms, and Erasmus Darwin proposed that all warm-blooded animals could have descended from a single microorganism (or "filament"). The first full-fledged evolutionary scheme was Jean-Baptiste Lamarck's "transmutation" theory of 1809, which envisaged spontaneous generation continually producing simple forms of life that developed greater complexity in parallel lineages with an inherent progressive tendency, and postulated that on a local level, these lineages adapted to the environment by inheriting changes caused by their use or disuse in parents. (The latter process was later called Lamarckism.) These ideas were condemned by established naturalists as speculation lacking empirical support. In particular, Georges Cuvier insisted that species were unrelated and fixed, their similarities reflecting divine design for functional needs. In the meantime, Ray's ideas of benevolent design had been developed by William Paley into the Natural Theology or Evidences of the Existence and Attributes of the Deity (1802), which proposed complex adaptations as evidence of divine design and which was admired by Charles Darwin.

The crucial break from the concept of constant typological classes or types in biology came with the theory of evolution through natural selection, which was formulated by Charles Darwin in terms of variable populations. Darwin used the expression "descent with modification" rather than "evolution". Partly influenced by An Essay on the Principle of Population (1798) by Thomas Robert Malthus, Darwin noted that population growth would lead to a "struggle for existence" in which favourable variations prevailed as others perished. In each generation, many offspring fail to survive to an age of reproduction because of limited

resources. This could explain the diversity of plants and animals from a common ancestry through the working of natural laws in the same way for all types of organism. Darwin developed his theory of "natural selection" from 1838 onwards and was writing up his "big book" on the subject when Alfred Russel Wallace sent him a version of virtually the same theory in 1858. Their separate papers were presented together at an 1858 meeting of the Linnean Society of London. At the end of 1859, Darwin's publication of his "abstract" as On the Origin of Species explained natural selection in detail and in a way that led to an increasingly wide acceptance of Darwin's concepts of evolution at the expense of alternative theories. Thomas Henry Huxley applied Darwin's ideas to humans, using paleontology and comparative anatomy to provide strong evidence that humans and apes shared a common ancestry. Some were disturbed by this since it implied that humans did not have a special place in the universe.

The mechanisms of reproductive heritability and the origin of new traits remained a mystery. Towards this end, Darwin developed his provisional theory of pangenesis. In 1865, Gregor Mendel reported that traits were inherited in a predictable manner through the independent assortment and segregation of elements (later known as genes). Mendel's laws of inheritance eventually supplanted most of Darwin's pangenesis theory. August Weismann made the important distinction between germ cells that give rise to gametes (such as sperm and egg cells) and the somatic cells of the body, demonstrating that heredity passes through the germ line only. Hugo de Vries connected Darwin's pangenesis theory to Weismann's germ/soma cell distinction and proposed that Darwin's pangenesis were concentrated in the cell nucleus and when expressed they could move into the cytoplasm to change the cell's structure. De Vries was also one of the researchers who made Mendel's work well known, believing that Mendelian traits corresponded to the transfer of heritable variations along the germline. To explain how new variants originate, de Vries developed a mutation theory that led to a temporary rift between those who accepted Darwinian evolution and biometricalians who allied with de Vries. In the 1930s, pioneers in the field of population genetics, such as Ronald Fisher, Sewall Wright and J. B. S. Haldane set the foundations of evolution onto a robust statistical philosophy. The false contradiction between Darwin's theory, genetic mutations, and Mendelian inheritance was thus reconciled.

15.2 The Modern Synthesis

In the 1920s and 1930s the so-called modern synthesis connected natural selection and population genetics, based on Mendelian inheritance, into a unified theory that

applied generally to any branch of biology. The modern synthesis explained patterns observed across species in populations, through fossil transitions in palaeontology, and complex cellular mechanisms in developmental biology. The publication of the structure of DNA by James Watson and Francis Crick with contribution of Rosalind Franklin in 1953 demonstrated a physical mechanism for inheritance. Molecular biology improved understanding of the relationship between genotype and phenotype. Advancements were also made in phylogenetic systematics, mapping the transition of traits into a comparative and testable framework through the publication and use of evolutionary trees. In 1973, evolutionary biologist Theodosius Dobzhansky penned that “nothing in biology makes sense except in the light of evolution,” because it has brought to light the relations of what first seemed disjointed facts in natural history into a coherent explanatory body of knowledge that describes and predicts many observable facts about life on this planet.

Since then, the modern synthesis has been further extended to explain biological phenomena across the full and integrative scale of the biological hierarchy, from genes to species. One extension, known as evolutionary developmental biology and informally called “evo-devo,” emphasises how changes between generations (evolution) acts on patterns of change within individual organisms (development). Since the beginning of the 21st century and in light of discoveries made in recent decades, some biologists have argued for an extended evolutionary synthesis, which would account for the effects of non-genetic inheritance modes, such as epigenetics, parental effects, ecological inheritance and cultural inheritance, and evolvability.

15.3 Mechanisms of Evolution

From a neo-Darwinian perspective, evolution occurs when there are changes in the frequencies of alleles within a population of interbreeding organisms, for example, the allele for black colour in a population of moths becoming more common. Mechanisms that can lead to changes in allele frequencies include natural selection, genetic drift, gene flow and mutation bias.

Evolution in organisms occurs through changes in heritable traits—the inherited characteristics of an organism. In humans, for example, eye colour is an inherited characteristic and an individual might inherit the “brown-eye trait” from one of their parents. Inherited traits are controlled by genes and the complete set of genes within an organism’s genome (genetic material) is called its genotype.

The complete set of observable traits that make up the structure and behaviour of an organism is called its phe-

notype. These traits come from the interaction of its genotype with the environment. As a result, many aspects of an organism’s phenotype are not inherited. For example, suntanned skin comes from the interaction between a person’s genotype and sunlight; thus, suntans are not passed on to people’s children. However, some people tan more easily than others, due to differences in genotypic variation; a striking example are people with the inherited trait of albinism, who do not tan at all and are very sensitive to sunburn.

Heritable traits are passed from one generation to the next via DNA, a molecule that encodes genetic information. DNA is a long biopolymer composed of four types of bases. The sequence of bases along a particular DNA molecule specify the genetic information, in a manner similar to a sequence of letters spelling out a sentence. Before a cell divides, the DNA is copied, so that each of the resulting two cells will inherit the DNA sequence. Portions of a DNA molecule that specify a single functional unit are called genes; different genes have different sequences of bases. Within cells, the long strands of DNA form condensed structures called chromosomes. The specific location of a DNA sequence within a chromosome is known as a locus. If the DNA sequence at a locus varies between individuals, the different forms of this sequence are called alleles. DNA sequences can change through mutations, producing new alleles. If a mutation occurs within a gene, the new allele may affect the trait that the gene controls, altering the phenotype of the organism. However, while this simple correspondence between an allele and a trait works in some cases, most traits are more complex and are controlled by quantitative trait loci (multiple interacting genes).

Recent findings have confirmed important examples of heritable changes that cannot be explained by changes to the sequence of nucleotides in the DNA. These phenomena are classed as epigenetic inheritance systems. DNA methylation marking chromatin, self-sustaining metabolic loops, gene silencing by RNA interference and the three-dimensional conformation of proteins (such as prions) are areas where epigenetic inheritance systems have been discovered at the organismic level. Developmental biologists suggest that complex interactions in genetic networks and communication among cells can lead to heritable variations that may underlay some of the mechanics in developmental plasticity and canalisation. Heritability may also occur at even larger scales. For example, ecological inheritance through the process of niche construction is defined by the regular and repeated activities of organisms in their environment. This generates a legacy of effects that modify and feed back into the selection regime of subsequent generations. Descendants inherit genes plus environmental characteristics generated by the ecological actions of ancestors. Other examples of heritability in evolution that

are not under the direct control of genes include the inheritance of cultural traits and symbiogenesis.

An individual organism's phenotype results from both its genotype and the influence from the environment it has lived in. A substantial part of the phenotypic variation in a population is caused by genotypic variation. The modern evolutionary synthesis defines evolution as the change over time in this genetic variation. The frequency of one particular allele will become more or less prevalent relative to other forms of that gene. Variation disappears when a new allele reaches the point of fixation—when it either disappears from the population or replaces the ancestral allele entirely.

Natural selection will only cause evolution if there is enough genetic variation in a population. Before the discovery of Mendelian genetics, one common hypothesis was blending inheritance. But with blending inheritance, genetic variance would be rapidly lost, making evolution by natural selection implausible. The Hardy-Weinberg principle provides the solution to how variation is maintained in a population with Mendelian inheritance. The frequencies of alleles (variations in a gene) will remain constant in the absence of selection, mutation, migration and genetic drift.

Variation comes from mutations in the genome, reshuffling of genes through sexual reproduction and migration between populations (gene flow). Despite the constant introduction of new variation through mutation and gene flow, most of the genome of a species is identical in all individuals of that species. However, even relatively small differences in genotype can lead to dramatic differences in phenotype: for example, chimpanzees and humans differ in only about 5% of their genomes.

Mutations are changes in the DNA sequence of a cell's genome. When mutations occur, they may alter the product of a gene, or prevent the gene from functioning, or have no effect. Based on studies in the fly *Drosophila melanogaster*, it has been suggested that if a mutation changes a protein produced by a gene, this will probably be harmful, with about 70% of these mutations having damaging effects, and the remainder being either neutral or weakly beneficial.

Mutations can involve large sections of a chromosome becoming duplicated (usually by genetic recombination), which can introduce extra copies of a gene into a genome. Extra copies of genes are a major source of the raw material needed for new genes to evolve. This is important because most new genes evolve within gene families from pre-existing genes that share common ancestors. For example, the human eye uses four genes to make structures that sense light: three for colour vision and one for night

vision; all four are descended from a single ancestral gene.

New genes can be generated from an ancestral gene when a duplicate copy mutates and acquires a new function. This process is easier once a gene has been duplicated because it increases the redundancy of the system; one gene in the pair can acquire a new function while the other copy continues to perform its original function. Other types of mutations can even generate entirely new genes from previously noncoding DNA.

The generation of new genes can also involve small parts of several genes being duplicated, with these fragments then recombining to form new combinations with new functions. When new genes are assembled from shuffling pre-existing parts, domains act as modules with simple independent functions, which can be mixed together to produce new combinations with new and complex functions. For example, polyketide synthases are large enzymes that make antibiotics; they contain up to one hundred independent domains that each catalyse one step in the overall process, like a step in an assembly line.

In asexual organisms, genes are inherited together, or linked, as they cannot mix with genes of other organisms during reproduction. In contrast, the offspring of sexual organisms contain random mixtures of their parents' chromosomes that are produced through independent assortment. In a related process called homologous recombination, sexual organisms exchange DNA between two matching chromosomes. Recombination and reassortment do not alter allele frequencies, but instead change which alleles are associated with each other, producing offspring with new combinations of alleles. Sex usually increases genetic variation and may increase the rate of evolution.

The two-fold cost of sex was first described by John Maynard Smith. The first cost is that in sexually dimorphic species only one of the two sexes can bear young. (This cost does not apply to hermaphroditic species, like most plants and many invertebrates.) The second cost is that any individual who reproduces sexually can only pass on 50% of its genes to any individual offspring, with even less passed on as each new generation passes. Yet sexual reproduction is the more common means of reproduction among eukaryotes and multicellular organisms. The Red Queen hypothesis has been used to explain the significance of sexual reproduction as a means to enable continual evolution and adaptation in response to coevolution with other species in an ever-changing environment.

Gene flow is the exchange of genes between populations and between species. It can therefore be a source of variation that is new to a population or to a species. Gene flow can be caused by the movement of individuals between separate populations of organisms, as might

be caused by the movement of mice between inland and coastal populations, or the movement of pollen between heavy-metal-tolerant and heavy-metal-sensitive populations of grasses.

Gene transfer between species includes the formation of hybrid organisms and horizontal gene transfer. Horizontal gene transfer is the transfer of genetic material from one organism to another organism that is not its offspring; this is most common among bacteria. In medicine, this contributes to the spread of antibiotic resistance, as when one bacteria acquires resistance genes it can rapidly transfer them to other species. Horizontal transfer of genes from bacteria to eukaryotes such as the yeast *Saccharomyces cerevisiae* and the adzuki bean weevil *Callosobruchus chinensis* has occurred. An example of larger-scale transfers are the eukaryotic bdelloid rotifers, which have received a range of genes from bacteria, fungi and plants. Viruses can also carry DNA between organisms, allowing transfer of genes even across biological domains.

Large-scale gene transfer has also occurred between the ancestors of eukaryotic cells and bacteria, during the acquisition of chloroplasts and mitochondria. It is possible that eukaryotes themselves originated from horizontal gene transfers between bacteria and archaea.

15.3.1 Natural selection

Evolution by means of natural selection is the process by which traits that enhance survival and reproduction become more common in successive generations of a population. It has often been called a “self-evident” mechanism because it necessarily follows from three simple facts:

Variation exists within populations of organisms with respect to morphology, physiology, and behaviour (phenotypic variation). Different traits confer different rates of survival and reproduction (differential fitness). These traits can be passed from generation to generation (heritability of fitness). More offspring are produced than can possibly survive, and these conditions produce competition between organisms for survival and reproduction. Consequently, organisms with traits that give them an advantage over their competitors are more likely to pass on their traits to the next generation than those with traits that do not confer an advantage. This teleonomy is the quality whereby the process of natural selection creates and preserves traits that are seemingly fitted for the functional roles they perform. Consequences of selection include nonrandom mating and genetic hitchhiking.

The central concept of natural selection is the evolutionary fitness of an organism. Fitness is measured by an

organism's ability to survive and reproduce, which determines the size of its genetic contribution to the next generation. However, fitness is not the same as the total number of offspring: instead fitness is indicated by the proportion of subsequent generations that carry an organism's genes. For example, if an organism could survive well and reproduce rapidly, but its offspring were all too small and weak to survive, this organism would make little genetic contribution to future generations and would thus have low fitness.

If an allele increases fitness more than the other alleles of that gene, then with each generation this allele will become more common within the population. These traits are said to be “selected for.” Examples of traits that can increase fitness are enhanced survival and increased fertility. Conversely, the lower fitness caused by having a less beneficial or deleterious allele results in this allele becoming rarer—they are “selected against.” Importantly, the fitness of an allele is not a fixed characteristic; if the environment changes, previously neutral or harmful traits may become beneficial and previously beneficial traits become harmful. However, even if the direction of selection does reverse in this way, traits that were lost in the past may not re-evolve in an identical form (see Dollo’s law). However, a re-activation of dormant genes, as long as they have not been eliminated from the genome and were only suppressed perhaps for hundreds of generations, can lead to the re-occurrence of traits thought to be lost like hindlegs in dolphins, teeth in chickens, wings in wingless stick insects, tails and additional nipples in humans etc. “Throwbacks” such as these are known as atavisms.

These charts depict the different types of genetic selection. On each graph, the x-axis variable is the type of phenotypic trait and the y-axis variable is the number of organisms. Group A is the original population and Group B is the population after selection. · Graph 1 shows directional selection, in which a single extreme phenotype is favoured. · Graph 2 depicts stabilizing selection, where the intermediate phenotype is favoured over the extreme traits. · Graph 3 shows disruptive selection, in which the extreme phenotypes are favoured over the intermediate. Natural selection within a population for a trait that can vary across a range of values, such as height, can be categorised into three different types. The first is directional selection, which is a shift in the average value of a trait over time—for example, organisms slowly getting taller. Secondly, disruptive selection is selection for extreme trait values and often results in two different values becoming most common, with selection against the average value. This would be when either short or tall organisms had an advantage, but not those of medium height. Finally, in stabilising selection there is selection against extreme trait values on both ends, which causes a decrease in variance around the average value and

less diversity. This would, for example, cause organisms to eventually have a similar height.

A special case of natural selection is sexual selection, which is selection for any trait that increases mating success by increasing the attractiveness of an organism to potential mates. Traits that evolved through sexual selection are particularly prominent among males of several animal species. Although sexually favoured, traits such as cumbersome antlers, mating calls, large body size and bright colours often attract predation, which compromises the survival of individual males. This survival disadvantage is balanced by higher reproductive success in males that show these hard-to-fake, sexually selected traits.

Natural selection most generally makes nature the measure against which individuals and individual traits, are more or less likely to survive. "Nature" in this sense refers to an ecosystem, that is, a system in which organisms interact with every other element, physical as well as biological, in their local environment. Eugene Odum, a founder of ecology, defined an ecosystem as: "Any unit that includes all of the organisms...in a given area interacting with the physical environment so that a flow of energy leads to clearly defined trophic structure, biotic diversity, and material cycles (i.e., exchange of materials between living and nonliving parts) within the system...." Each population within an ecosystem occupies a distinct niche, or position, with distinct relationships to other parts of the system. These relationships involve the life history of the organism, its position in the food chain and its geographic range. This broad understanding of nature enables scientists to delineate specific forces which, together, comprise natural selection.

Natural selection can act at different levels of organisation, such as genes, cells, individual organisms, groups of organisms and species. Selection can act at multiple levels simultaneously. An example of selection occurring below the level of the individual organism are genes called transposons, which can replicate and spread throughout a genome. Selection at a level above the individual, such as group selection, may allow the evolution of cooperation, as discussed below.

15.3.2 Genetic Hitchhiking

Recombination allows alleles on the same strand of DNA to become separated. However, the rate of recombination is low (approximately two events per chromosome per generation). As a result, genes close together on a chromosome may not always be shuffled away from each other and genes that are close together tend to be inherited together, a phenomenon known as linkage. This tendency is measured by finding how often two alleles occur together on a single chromosome compared to expectations, which

is called their linkage disequilibrium. A set of alleles that is usually inherited in a group is called a haplotype. This can be important when one allele in a particular haplotype is strongly beneficial: natural selection can drive a selective sweep that will also cause the other alleles in the haplotype to become more common in the population; this effect is called genetic hitchhiking or genetic draft. Genetic draft caused by the fact that some neutral genes are genetically linked to others that are under selection can be partially captured by an appropriate effective population size.

15.3.3 Genetic Drift

Simulation of genetic drift of 20 unlinked alleles in populations of 10 (top) and 100 (bottom). Drift to fixation is more rapid in the smaller population. Genetic drift is the random fluctuations of allele frequencies within a population from one generation to the next. When selective forces are absent or relatively weak, allele frequencies are equally likely to drift upward or downward at each successive generation because the alleles are subject to sampling error. This drift halts when an allele eventually becomes fixed, either by disappearing from the population or replacing the other alleles entirely. Genetic drift may therefore eliminate some alleles from a population due to chance alone. Even in the absence of selective forces, genetic drift can cause two separate populations that began with the same genetic structure to drift apart into two divergent populations with different sets of alleles.

The neutral theory of molecular evolution proposed that most evolutionary changes are the result of the fixation of neutral mutations by genetic drift. Hence, in this model, most genetic changes in a population are the result of constant mutation pressure and genetic drift. This form of the neutral theory is now largely abandoned, since it does not seem to fit the genetic variation seen in nature. However, a more recent and better-supported version of this model is the nearly neutral theory, where a mutation that would be effectively neutral in a small population is not necessarily neutral in a large population. Other alternative theories propose that genetic drift is dwarfed by other stochastic forces in evolution, such as genetic hitchhiking, also known as genetic draft.

The time for a neutral allele to become fixed by genetic drift depends on population size, with fixation occurring more rapidly in smaller populations. The number of individuals in a population is not critical, but instead a measure known as the effective population size. The effective population is usually smaller than the total population since it takes into account factors such as the level of inbreeding and the stage of the lifecycle in which the population is the smallest. The effective population size may not be the same for every gene in the same population.

It is usually difficult to measure the relative importance of selection and neutral processes, including drift. The comparative importance of adaptive and non-adaptive forces in driving evolutionary change is an area of current research.

15.3.4 Gene Flow

Gene flow involves the exchange of genes between populations and between species. The presence or absence of gene flow fundamentally changes the course of evolution. Due to the complexity of organisms, any two completely isolated populations will eventually evolve genetic incompatibilities through neutral processes, as in the Bateson-Dobzhansky-Muller model, even if both populations remain essentially identical in terms of their adaptation to the environment.

If genetic differentiation between populations develops, gene flow between populations can introduce traits or alleles which are disadvantageous in the local population and this may lead to organisms within these populations evolving mechanisms that prevent mating with genetically distant populations, eventually resulting in the appearance of new species. Thus, exchange of genetic information between individuals is fundamentally important for the development of the Biological Species Concept (BSC).

During the development of the modern synthesis, Sewall Wright developed his shifting balance theory, which regarded gene flow between partially isolated populations as an important aspect of adaptive evolution. However, recently there has been substantial criticism of the importance of the shifting balance theory.

15.3.5 A Classic Example: Evolution of The Peppered Moth

The evolution of the peppered moth is an evolutionary instance of directional colour change in the moth population as a consequence of air pollution during the Industrial Revolution. The frequency of dark-coloured moths increased at that time, an example of industrial melanism. Later, when pollution was reduced, the light-coloured form again predominated. Industrial melanism in the peppered moth was an early test of Charles Darwin's natural selection in action, and remains as a classic example in the teaching of evolution. In 1978 Sewall Wright described it as "the clearest case in which a conspicuous evolutionary process has actually been observed."

The dark-coloured or melanic form of the peppered moth (var. carbonaria) was not known before 1811. After field collection in 1848 from Manchester, an industrial city in England, the frequency of the variety was found to have

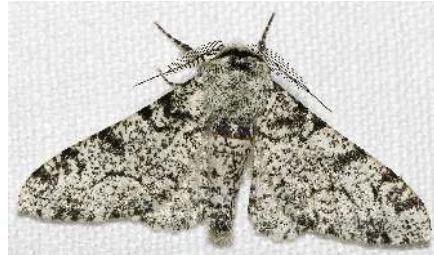


Figure 15.1: *Biston betularia* f. *typica*, the white-bodied peppered moth.²¹



Figure 15.2: *Biston betularia* f. *carbonaria*, the black-bodied peppered moth.²²

increased drastically. By the end of the 19th century it almost completely outnumbered the original light-coloured type (var. *typica*), with a record of 98% in 1895. The evolutionary importance of the moth was only speculated upon during Darwin's lifetime. It was 14 years after Darwin's death, in 1896, that J.W. Tutt presented it as a case of natural selection. Due to this, the idea widely spread, and more people believed in Darwin's theory.

Bernard Kettlewell was the first to investigate the evolutionary mechanism behind peppered moth adaptation, between 1953 and 1956. He found that a light-coloured body was an effective camouflage in a clean environment, such as in Dorset, while the dark colour was beneficial in a polluted environment like in Birmingham. This selective survival was due to birds which easily caught dark moths on clean trees, and white moths on trees darkened with soot. The story, supported by Kettlewell's experiment, became the canonical example of Darwinian evolution and evidence for natural selection used in standard textbooks.

However, failure to replicate the experiment and criticism of Kettlewell's methods by Theodore David Sargent in the late 1960s led to general skepticism. When Judith Hooper's *Of Moths and Men* was published in 2002, Kettlewell's story was more sternly attacked, accused of fraud, and became widely disregarded. The criticism became a major argument for creationists. Michael Majerus was the principal defender. His seven-year experiment beginning in 2001, the most elaborate of its kind in population biology, the results of which were published posthumously in 2012, vindicated Kettlewell's work in great detail. This



Figure 15.3: Typica and carbonaria morphs on the same tree.²³ The light-coloured typica (below the bark's scar) is nearly invisible on this pollution-free tree, camouflaging it from predators.

restored peppered moth evolution as “the most direct evidence”, and “one of the clearest and most easily understood examples of Darwinian evolution in action”.

Before the Industrial Revolution, the black peppered moth was rare. The first black specimen (of unknown origin) was kept in the University of Oxford in 1811. The first live specimen was caught by R.S. Edleston in Manchester, England in 1848, but he reported this only 16 years later in 1864 in the journal Entomologist. Edleston notes that by 1864 it was the more common type of moth in his garden in Manchester. The light-bodied moths were able to blend in with the light-coloured lichens and tree bark, and the less common black moth was more likely to be eaten by birds. As a result of the common light-coloured lichens and English trees, therefore, the light-coloured moths were much more effective at hiding from predators, and the frequency of the dark allele was about 0.01%.

During the early decades of the Industrial Revolution in England, the countryside between London and Manchester became blanketed with soot from the new coal-burning factories. Many of the light-bodied lichens died from sulphur dioxide emissions, and the trees became darkened. This led to an increase in bird predation for light-coloured moths, as they no longer blended in as well in their polluted ecosystem: indeed, their bodies now dramatically contrasted with the colour of the bark. Dark-coloured moths, on the other hand, were camouflaged very well by the blackened trees. The population of dark-coloured moth rapidly increased. By the mid-19th century, the number of dark-coloured moths had risen noticeably, and by 1895, the percentage of dark-coloured moths in Manchester was reported at 98%, a dramatic change (of almost 100%) from the original frequency. This effect of industrialization in body colour led to the coining of the term “industrial melanism”.

The implication that industrial melanism could be evi-

dence supporting Charles Darwin’s theory of natural selection was noticed during his lifetime. Albert Brydges Farn (1841–1921), a British entomologist, wrote to Darwin on 18 November 1878 to discuss his observation of colour variations in the Annulet moth (then *Gnophos obscurata*, now *Charissa obscurata*). He noted the existence of dark moths in peat in the New Forest, brown moths on clay and red soil in Herefordshire, and white moths on chalk cliffs in Lewes, then suggested this variation was an example of “survival of the fittest”. He told Darwin that he had found dark moths on a chalk slope where the foliage had been blackened by smoke from lime kilns, and he had also heard that white moths had become less common at Lewes after lime kilns had been in operation for a few years. Darwin does not seem to have responded to this information, possibly because he thought natural selection would be a much slower process. A scientific explanation of moth colouration was only published in 1896, 14 years after Darwin’s death, when J.W. Tutt explicitly linked peppered moth melanism to natural selection.

15.4 Outcomes of Evolution

Evolution influences every aspect of the form and behaviour of organisms. Most prominent are the specific behavioural and physical adaptations that are the outcome of natural selection. These adaptations increase fitness by aiding activities such as finding food, avoiding predators or attracting mates. Organisms can also respond to selection by cooperating with each other, usually by aiding their relatives or engaging in mutually beneficial symbiosis. In the longer term, evolution produces new species through splitting ancestral populations of organisms into new groups that cannot or will not interbreed.

These outcomes of evolution are distinguished based on time scale as macroevolution versus microevolution. Macroevolution refers to evolution that occurs at or above the level of species, in particular speciation and extinction; whereas microevolution refers to smaller evolutionary changes within a species or population, in particular shifts in allele frequency and adaptation. In general, macroevolution is regarded as the outcome of long periods of microevolution. Thus, the distinction between micro- and macroevolution is not a fundamental one—the difference is simply the time involved. However, in macroevolution, the traits of the entire species may be important. For instance, a large amount of variation among individuals allows a species to rapidly adapt to new habitats, lessening the chance of it going extinct, while a wide geographic range increases the chance of speciation, by making it more likely that part of the population will become isolated. In this sense, microevolution and macroevolution might involve selection at

different levels—with microevolution acting on genes and organisms, versus macroevolutionary processes such as species selection acting on entire species and affecting their rates of speciation and extinction.

A common misconception is that evolution has goals, long-term plans, or an innate tendency for “progress”, as expressed in beliefs such as orthogenesis and evolutionism; realistically however, evolution has no long-term goal and does not necessarily produce greater complexity. Although complex species have evolved, they occur as a side effect of the overall number of organisms increasing and simple forms of life still remain more common in the biosphere. For example, the overwhelming majority of species are microscopic prokaryotes, which form about half the world’s biomass despite their small size, and constitute the vast majority of Earth’s biodiversity. Simple organisms have therefore been the dominant form of life on Earth throughout its history and continue to be the main form of life up to the present day, with complex life only appearing more diverse because it is more noticeable. Indeed, the evolution of microorganisms is particularly important to modern evolutionary research, since their rapid reproduction allows the study of experimental evolution and the observation of evolution and adaptation in real time.

15.4.1 Adaptation

Adaptation is the process that makes organisms better suited to their habitat. Also, the term adaptation may refer to a trait that is important for an organism’s survival. For example, the adaptation of horses’ teeth to the grinding of grass. By using the term adaptation for the evolutionary process and adaptive trait for the product (the bodily part or function), the two senses of the word may be distinguished. Adaptations are produced by natural selection. The following definitions are due to Theodosius Dobzhansky:

Adaptation is the evolutionary process whereby an organism becomes better able to live in its habitat or habitats. Adaptedness is the state of being adapted: the degree to which an organism is able to live and reproduce in a given set of habitats. An adaptive trait is an aspect of the developmental pattern of the organism which enables or enhances the probability of that organism surviving and reproducing. Adaptation may cause either the gain of a new feature, or the loss of an ancestral feature. An example that shows both types of change is bacterial adaptation to antibiotic selection, with genetic changes causing antibiotic resistance by both modifying the target of the drug, or increasing the activity of transporters that pump the drug out of the cell. Other striking examples are the bacteria *Escherichia coli* evolving the ability to use citric acid as a

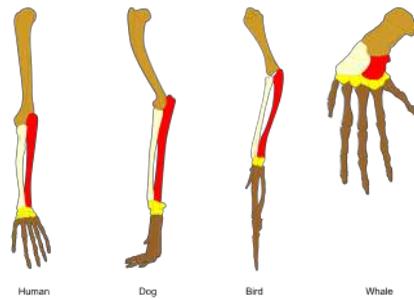


Figure 15.4: Homologous bones in the limbs of tetrapods. The bones of these animals have the same basic structure, but have been adapted for specific uses.²⁴

nutrient in a long-term laboratory experiment, *Flavobacterium* evolving a novel enzyme that allows these bacteria to grow on the by-products of nylon manufacturing, and the soil bacterium *Sphingobium* evolving an entirely new metabolic pathway that degrades the synthetic pesticide pentachlorophenol. An interesting but still controversial idea is that some adaptations might increase the ability of organisms to generate genetic diversity and adapt by natural selection (increasing organisms’ evolvability).

Adaptation occurs through the gradual modification of existing structures. Consequently, structures with similar internal organisation may have different functions in related organisms. This is the result of a single ancestral structure being adapted to function in different ways. The bones within bat wings, for example, are very similar to those in mice feet and primate hands, due to the descent of all these structures from a common mammalian ancestor. However, since all living organisms are related to some extent, even organs that appear to have little or no structural similarity, such as arthropod, squid and vertebrate eyes, or the limbs and wings of arthropods and vertebrates, can depend on a common set of homologous genes that control their assembly and function; this is called deep homology.

During evolution, some structures may lose their original function and become vestigial structures. Such structures may have little or no function in a current species, yet have a clear function in ancestral species, or other closely related species. Examples include pseudogenes, the non-functional remains of eyes in blind cave-dwelling fish, wings in flightless birds, the presence of hip bones in whales and snakes, and sexual traits in organisms that reproduce via asexual reproduction. Examples of vestigial structures in humans include wisdom teeth, the coccyx, the vermiform appendix, and other behavioural vestiges such as goose bumps and primitive reflexes.

However, many traits that appear to be simple adapta-

tions are in fact exaptations: structures originally adapted for one function, but which coincidentally became somewhat useful for some other function in the process. One example is the African lizard *Holaspis guentheri*, which developed an extremely flat head for hiding in crevices, as can be seen by looking at its near relatives. However, in this species, the head has become so flattened that it assists in gliding from tree to tree—an exaptation. Within cells, molecular machines such as the bacterial flagella and protein sorting machinery evolved by the recruitment of several pre-existing proteins that previously had different functions. Another example is the recruitment of enzymes from glycolysis and xenobiotic metabolism to serve as structural proteins called crystallins within the lenses of organisms' eyes.

An area of current investigation in evolutionary developmental biology is the developmental basis of adaptations and exaptations. This research addresses the origin and evolution of embryonic development and how modifications of development and developmental processes produce novel features. These studies have shown that evolution can alter development to produce new structures, such as embryonic bone structures that develop into the jaw in other animals instead forming part of the middle ear in mammals. It is also possible for structures that have been lost in evolution to reappear due to changes in developmental genes, such as a mutation in chickens causing embryos to grow teeth similar to those of crocodiles. It is now becoming clear that most alterations in the form of organisms are due to changes in a small set of conserved genes.

15.4.2 Coevolution

Interactions between organisms can produce both conflict and cooperation. When the interaction is between pairs of species, such as a pathogen and a host, or a predator and its prey, these species can develop matched sets of adaptations. Here, the evolution of one species causes adaptations in a second species. These changes in the second species then, in turn, cause new adaptations in the first species. This cycle of selection and response is called co-evolution. An example is the production of tetrodotoxin in the rough-skinned newt and the evolution of tetrodotoxin resistance in its predator, the common garter snake. In this predator-prey pair, an evolutionary arms race has produced high levels of toxin in the newt and correspondingly high levels of toxin resistance in the snake.

Chapter 16

Taxonomy, Binomial Nomenclature And Systematics

Taxonomy is the identification, naming and classification of organisms. Binomial nomenclature is a formal system of naming species of living things by giving each a name composed of two parts. Systematics is the branch of science that deals with unique properties of species and groups to recognise, describe name and arrange the diverse organisms according to an organised plan. In biology, phylogenetics (Greek: φυλή, φῦλον – phylé, phylon = tribe, clan, race + γενετικός – genetikós = origin, source, birth) is a part of systematics that addresses the inference of the evolutionary history and relationships among or within groups of organisms (e.g. species, or more inclusive taxa). Classifications are now usually based on phylogenetic data. The degree to which classification depends on inferred evolutionary history differs depending on the school of taxonomy: phenetics ignores phylogenetic speculation altogether, trying to represent the similarity between organisms instead; cladistics (phylogenetic systematics) tries to reflect phylogeny in its classifications by only recognizing groups based on shared, derived characters (synapomorphies); evolutionary taxonomy tries to take into account both the branching pattern and “degree of difference” to find a compromise between them.

Biological classification is a critical component of the taxonomic process. As a result, it informs the user as to what the relatives of the taxon are hypothesized to be. Biological classification uses taxonomic ranks, including among others (in order from most inclusive to least inclusive): Domain, Kingdom, Phylum, Class, Order, Family, Genus, and Species

The “definition” of a taxon is encapsulated by its description or its diagnosis or by both combined. There are no set rules governing the definition of taxa, but the naming and publication of new taxa is governed by sets of rules. In zoology, the nomenclature for the more commonly used ranks (superfamily to subspecies), is regulated by the Inter-

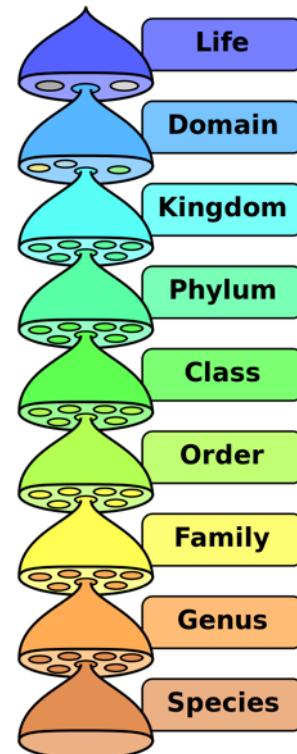


Figure 16.1: Biological classification.¹

national Code of Zoological Nomenclature (ICZN Code). In the fields of botany, phycology, and mycology, the naming of taxa is governed by the International Code of Nomenclature for algae, fungi, and plants (ICN).

The initial description of a taxon involves five main requirements:

1. The taxon must be given a name based on the 26 letters of the Latin alphabet (a binomial for new species, or uninomial for other ranks).
2. The name must be unique (i.e. not a homonym).
3. The description must be based on at least one name-bearing type specimen.
4. It should include statements about appropriate attributes either to describe (define) the taxon or to differentiate it from other taxa (the diagnosis, ICZN Code, Article 13.1.1, ICN, Article 38). Both codes deliberately separate defining the content of a taxon (its circumscription) from defining its name.
5. These first four requirements must be published in a work that is obtainable in numerous identical copies, as a permanent scientific record.

However, often much more information is included, like the geographic range of the taxon, ecological notes, chemistry, behavior, etc. How researchers arrive at their taxa varies: depending on the available data, and resources, methods vary from simple quantitative or qualitative comparisons of striking features, to elaborate computer analyses of large amounts of DNA sequence data.

An "authority" may be placed after a scientific name. The authority is the name of the scientist or scientists who first validly published the name. For example, in 1758 Linnaeus gave the Asian elephant the scientific name *Elephas maximus*, so the name is sometimes written as "Elephas maximus Linnaeus, 1758". The names of authors are frequently abbreviated: the abbreviation L., for Linnaeus, is commonly used. In botany, there is, in fact, a regulated list of standard abbreviations (see list of botanists by author abbreviation). The system for assigning authorities differs slightly between botany and zoology. However, it is standard that if a species' name or placement has been changed since the original description, the original authority's name is placed in parentheses.

16.1 The Concept of Species in Biology

In biology, a species is the basic unit of classification and a taxonomic rank of an organism, as well as a unit of biodiversity. A species is often defined as the largest group of organisms in which any two individuals of the appro-

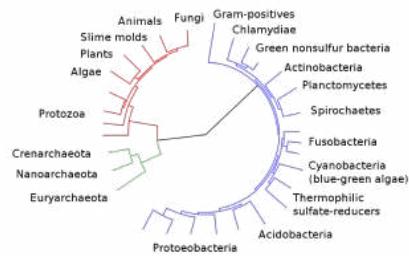


Figure 16.2: The tree of life.²

priate sexes or mating types can produce fertile offspring, typically by sexual reproduction. Other ways of defining species include their karyotype, DNA sequence, morphology, behaviour or ecological niche. In addition, paleontologists use the concept of the chronospecies since fossil reproduction cannot be examined.

Species were seen from the time of Aristotle until the 18th century as fixed categories that could be arranged in a hierarchy, the great chain of being. In the 19th century, biologists grasped that species could evolve given sufficient time. Charles Darwin's 1859 book *On the Origin of Species* explained how species could arise by natural selection. That understanding was greatly extended in the 20th century through genetics and population ecology. Genetic variability arises from mutations and recombination, while organisms themselves are mobile, leading to geographical isolation and genetic drift with varying selection pressures. Genes can sometimes be exchanged between species by horizontal gene transfer; new species can arise rapidly through hybridisation and polyploidy; and species may become extinct for a variety of reasons.

Biologists and taxonomists have made many attempts to define species, beginning from morphology and moving towards genetics. Early taxonomists such as Linnaeus had no option but to describe what they saw: this was later formalised as the typological or morphological species concept. Ernst Mayr³ proposed the widely used Biological Species Concept of reproductive isolation in 1942. Later biologists have tried to refine Mayr's definition. Many of the concepts are quite similar or overlap, so they are not easy to count: the biologist R. L. Mayden recorded about 24 concepts, and the philosopher of science John Wilkins counted 26. Wilkins further grouped the species concepts into seven basic kinds of concepts: (1) agamospecies for asexual organisms (2) biospecies for reproductively isolated sexual organisms (3) ecospecies based on ecological niches (4) evolutionary species based on lineage (5) genetic species based on gene pool (6) morphospecies based on form or phenotype and (7) taxonomic species, a

³https://en.wikipedia.org/wiki/Ernst_Mayr

species as determined by a taxonomist.

A typological species is a group of organisms in which individuals conform to certain fixed properties (a type), so that even pre-literate people often recognise the same taxon as do modern taxonomists. The clusters of variations or phenotypes within specimens (such as longer or shorter tails) would differentiate the species. This method was used as a “classical” method of determining species, such as with Linnaeus early in evolutionary theory. However, different phenotypes are not necessarily different species (e.g. a four-winged *Drosophila* born to a two-winged mother is not a different species). Species named in this manner are called morphospecies.

A species is given a taxonomic name when a type specimen is described formally, in a publication that assigns it a unique scientific name. The description typically provides means for identifying the new species, differentiating it from other previously described and related or confusable species and provides a validly published name (in botany) or an available name (in zoology) when the paper is accepted for publication. The type material is usually held in a permanent repository, often the research collection of a major museum or university, that allows independent verification and the means to compare specimens. Describers of new species are asked to choose names that, in the words of the International Code of Zoological Nomenclature, are “appropriate, compact, euphonious, memorable, and do not cause offence”.

The naming of a particular species, including which genus (and higher taxa) it is placed in, is a hypothesis about the evolutionary relationships and distinguishability of that group of organisms. As further information comes to hand, the hypothesis may be confirmed or refuted. Sometimes, especially in the past when communication was more difficult, taxonomists working in isolation have given two distinct names to individual organisms later identified as the same species. When two named species are discovered to be of the same species, the older species name is given priority and usually retained, and the newer name considered as a junior synonym, a process called synonymisation. Dividing a taxon into multiple, often new, taxa is called splitting. Taxonomists are often referred to as “lumpers” or “splitters” by their colleagues, depending on their personal approach to recognising differences or commonalities between organisms.

It is difficult to define a species in a way that applies to all organisms. The debate about species delimitation is called the species problem. The problem was recognized even in 1859, when Darwin wrote in *On the Origin of Species*:

No one definition has satisfied all naturalists;



Figure 16.3: The butterfly genus *Heliconius* contains many similar species.⁴

yet every naturalist knows vaguely what he means when he speaks of a species. Generally the term includes the unknown element of a distinct act of creation.

The evolutionary process by which biological populations evolve to become distinct or reproductively isolated as species is called speciation. Charles Darwin was the first to describe the role of natural selection in speciation in his 1859 book *The Origin of Species*. Speciation depends on a measure of reproductive isolation, a reduced gene flow. This occurs most easily in allopatric speciation, where populations are separated geographically and can diverge gradually as mutations accumulate. Reproductive isolation is threatened by hybridisation, but this can be selected against once a pair of populations have incompatible alleles of the same gene, as described in the Bateson-Dobzhansky-Muller model. A different mechanism, phyletic speciation, involves one lineage gradually changing over time into a new and distinct form, without increasing the number of resultant species.

16.2 Binomial Nomenclature

The commonly used names for kinds of organisms are often ambiguous: “cat” could mean the domestic cat or the cat family. Another problem with common names is that they often vary from place to place, so that puma, cougar, catamount, panther, painter and mountain lion all mean *Puma concolor* in various parts of America, while “panther” may also mean the jaguar (*Panthera onca*) of Latin America or the leopard (*Panthera pardus*) of Africa and Asia. In contrast, the scientific names of species are chosen to be unique and universal; they are in two parts used together: the genus as in *Puma*, and the specific epithet as in *concolor*.

Binomial nomenclature ("two-term naming system"), also called binominal nomenclature ("two-name naming system") or binary nomenclature, is a formal system of naming species of living things by giving each a name composed of two parts, both of which use Latin grammatical forms, although they can be based on words from other languages. Such a name is called a binomial name (which may be shortened to just "binomial"), a binomen, binominal name or a scientific name; more informally it is also called a Latin name.

The first part of the name – the generic name – identifies the genus to which the species belongs, while the second part – the specific name or specific epithet – identifies the species within the genus. For example, humans belong to the genus *Homo* and within this genus to the species *Homo sapiens*. *Tyrannosaurus rex* is probably the most widely known binomial. The formal introduction of this system of naming species is credited to Carl Linnaeus, effectively beginning with his work *Species Plantarum* in 1753. But Gaspard Bauhin, in as early as 1622, had introduced in his book *Pinax theatri botanici* (English, Illustrated exposition of plants) many names of genera that were later adopted by Linnaeus.

The application of binomial nomenclature is now governed by various internationally agreed codes of rules, of which the two most important are the International Code of Zoological Nomenclature (ICZN) for animals and the International Code of Nomenclature for algae, fungi, and plants (ICNfp). Although the general principles underlying binomial nomenclature are common to these two codes, there are some differences, both in the terminology they use and in their precise rules.

In modern usage, the first letter of the first part of the name, the genus, is always capitalized in writing, while that of the second part is not, even when derived from a proper noun such as the name of a person or place. Similarly, both parts are italicized when a binomial name occurs in normal text (or underlined in handwriting). Thus the binomial name of the annual phlox (named after botanist Thomas Drummond) is now written as *Phlox drummondii*.

In scientific works, the authority for a binomial name is usually given, at least when it is first mentioned, and the date of publication may be specified.

- In zoology

- "*Patella vulgata* Linnaeus, 1758". The name "Linnaeus" tells the reader who it was that first published a description and name for this species of limpet; 1758 is the date of the publication in which the original description can be found (in this case the 10th edition of the book *Systema Naturae*).

- "*Passer domesticus* (Linnaeus, 1758)". The original name given by Linnaeus was *Fringilla domestica*; the parentheses indicate that the species is now considered to belong in a different genus. The ICZN does not require that the name of the person who changed the genus be given, nor the date on which the change was made, although nomenclatorial catalogs usually include such information.

- In botany

- "*Amaranthus retroflexus* L." – "L." is the standard abbreviation used in botany for "Linnaeus".
- "*Hyacinthoides italicica* (L.) Rothm." – Linnaeus first named this bluebell species *Scilla italicica*; Rothmaler transferred it to the genus *Hyacinthoides*; the ICNfp does not require that the dates of either publication be specified.

Prior to the adoption of the modern binomial system of naming species, a scientific name consisted of a generic name combined with a specific name that was from one to several words long. Together they formed a system of polynomial nomenclature. These names had two separate functions. First, to designate or label the species, and second, to be a diagnosis or description; however these two goals were eventually found to be incompatible. In a simple genus, containing only two species, it was easy to tell them apart with a one-word genus and a one-word specific name; but as more species were discovered, the names necessarily became longer and unwieldy, for instance, *Plantago foliis ovato-lanceolatus pubescensibus, spica cylindrica, scapo tereti* ("plantain with pubescent ovate-lanceolate leaves, a cylindric spike and a terete scape"), which we know today as *Plantago media*.

Such "polynomial names" may sometimes look like binomials, but are significantly different. For example, Gerard's herbal (as amended by Johnson) describes various kinds of spiderwort: "The first is called *Phalangium ramosum*, Branched Spiderwort; the second, *Phalangium non ramosum*, Unbranched Spiderwort. The other ... is aptly termed *Phalangium Ephemorum Virginianum*, Soon-Fading Spiderwort of Virginia". The Latin phrases are short descriptions, rather than identifying labels.

The Bauhins, in particular Caspar Bauhin (1560–1624), took some important steps towards the binomial system, by pruning the Latin descriptions, in many cases to two words. The adoption by biologists of a system of strictly binomial nomenclature is due to Swedish botanist and physician Carl Linnaeus (1707–1778). It was in Linnaeus's 1753 *Species Plantarum* that he began consistently using a one-word "trivial name" (*nomen triviale*) after a generic name (genus name) in a system of binomial nomenclature. Trivial

names had already appeared in his *Critica Botanica* (1737) and *Philosophia Botanica* (1751). This trivial name is what is now known as a specific epithet (ICNafp) or specific name (ICZN). The Bauhins' genus names were retained in many of these, but the descriptive part was reduced to a single word.

Linnaeus's trivial names introduced an important new idea, namely that the function of a name could simply be to give a species a unique label. This meant that the name no longer need be descriptive; for example both parts could be derived from the names of people. Thus Gerard's *Phalangium ephemerum virginianum* became *Tradescantia virginiana*, where the genus name honoured John Tradescant the Younger,[note 1] an English botanist and gardener. A bird in the parrot family was named *Psittacus alexandri*, meaning "Alexander's parrot", after Alexander the Great, whose armies introduced eastern parakeets to Greece. Linnaeus's trivial names were much easier to remember and use than the parallel polynomial names and eventually replaced them.

The value of the binomial nomenclature system derives primarily from its economy, its widespread use, and the uniqueness and stability of names it generally favors:

- **Economy.** Compared to the polynomial system which it replaced, a binomial name is shorter and easier to remember. It corresponds to the widespread system of family name plus given name(s) used to name people in many cultures.
- **Widespread use.** The binomial system of nomenclature is governed by international codes and is used by biologists worldwide. A few binomials have also entered common speech, such as *Homo sapiens*, *E. coli*, *Boa constrictor*, and *Tyrannosaurus rex*.
- **Uniqueness.** Provided that taxonomists agree as to the limits of a species, it can have only one name that is correct under the appropriate nomenclature code, generally the earliest published if two or more names are accidentally assigned to a species. However, establishing that two names actually refer to the same species and then determining which has priority can be difficult, particularly if the species was named by biologists from different countries. Therefore, a species may have more than one regularly used name; all but one of these names are "synonyms".
- **Stability.** Although stability is far from absolute, the procedures associated with establishing binomial names, such as the principle of priority, tend to favor stability. For example, when species are transferred between genera (as not uncommonly happens as a result of new knowledge), if possible the second part of the binomial is kept the same.

Thus there is disagreement among botanists as to whether the genera *Chionodoxa* and *Scilla* are sufficiently different for them to be kept separate. Those who keep them separate give the plant commonly grown in gardens in Europe the name *Chionodoxa siehei*; those who do not give it the name *Scilla siehei*. The *siehei* element is constant. Similarly if what were previously thought to be two distinct species are demoted to a lower rank, such as subspecies, where possible the second part of the binomial name is retained as the third part of the new name. Thus the Tenerife robin may be treated as a different species from the European robin, in which case its name is *Erithacus superbus*, or as only a subspecies, in which case its name is *Erithacus rubecula superbus*. The *superbus* element of the name is constant.

Nomenclature (including binomial nomenclature) is not the same as classification, although the two are related. Classification is the ordering of items into groups based on similarities or differences; in biological classification, species are one of the kinds of item to be classified. In principle, the names given to species could be completely independent of their classification. This is not the case for binomial names, since the first part of a binomial is the name of the genus into which the species is placed. Above the rank of genus, binomial nomenclature and classification are partly independent; for example, a species retains its binomial name if it is moved from one family to another or from one order to another, unless it better fits a different genus in the same or different family, or it is split from its old genus and placed in a newly created genus. The independence is only partial since the names of families and other higher taxa are usually based on genera.

Taxonomy includes both nomenclature and classification. Its first stages (sometimes called "alpha taxonomy") are concerned with finding, describing and naming species of living or fossil organisms. Binomial nomenclature is thus an important part of taxonomy as it is the system by which species are named. Taxonomists are also concerned with classification, including its principles, procedures and rules.

A complete binomial name is always treated grammatically as if it were a phrase in the Latin language (hence the common use of the term "Latin name" for a binomial name). However, the two parts of a binomial name can each be derived from a number of sources, of which Latin is only one. These include:

- Latin, either classical or medieval. Thus, both parts of the binomial name *Homo sapiens* are Latin words, meaning "wise" (*sapiens*) "human/man" (*Homo*).
- Classical Greek. The genus *Rhododendron*

was named by Linnaeus from the Greek word ρόδόδενδρον, itself derived from *rhodon*, “rose”, and *dendron*, “tree”. Greek words are often converted to a Latinized form. Thus *coca* (the plant from which cocaine is obtained) has the name *Erythroxylum coca*. *Erythroxylum* is derived from the Greek words *erythros*, red, and *xylon*, wood. The Greek neuter ending *-ov* (*-on*) is often converted to the Latin neuter ending *-um*.

- Other languages. The second part of the name *Erythroxylum coca* is derived from *kuka*, the name of the plant in Aymara and Quechua. Since many dinosaur fossils were found in Mongolia, their names often use Mongolian words, e.g. *Tarchia* from *tarkhi*, meaning “brain”, or *Saichania* meaning “beautiful one”.
- Names of people (often naturalists or biologists). The name *Magnolia campbellii* commemorates two people: Pierre Magnol, a French botanist, and Archibald Campbell, a doctor in British India.
- Names of places. The lone star tick, *Amblyomma americanum*, is widespread in the United States.
- Other sources. Some binomial names have been constructed from taxonomic anagrams or other reorderings of existing names. Thus the name of the genus *Muilla* is derived by reversing the name *Allium*. Names may also be derived from jokes or puns. For example, Ratcliffe described a number of species of rhinoceros beetle, including *Cyclocephala nodanoth-erwon*.

The first part of the name, which identifies the genus, must be a word which can be treated as a Latin singular noun in the nominative case. It must be unique within each kingdom, but can be repeated between kingdoms. Thus *Huia recurvata* is an extinct species of plant, found as fossils in Yunnan, China, whereas *Huia masonii* is a species of frog found in Java, Indonesia.

The second part of the name, which identifies the species within the genus, is also treated grammatically as a Latin word. It can have one of a number of forms:

- The second part of a binomial may be an adjective. The adjective must agree with the genus name in gender. Latin has three genders, masculine, feminine and neuter, shown by varying endings to nouns and adjectives. The house sparrow has the binomial name *Passer domesticus*. Here *domesticus* (“domestic”) simply means “associated with the house”. The sacred bamboo is *Nandina domestica* rather than *Nandina domesticus*, since *Nandina* is feminine whereas *Passer* is masculine. The tropical fruit langsat is a product of the plant *Lansium parasiticum*, since *Lansium* is neuter.

Some common endings for Latin adjectives in the three genders (masculine, feminine, neuter) are *-us*, *-a*, *-um* (as in the previous example of *domesticus*); *-is*, *-is*, *-e* (e.g. *tristis*, meaning “sad”); and *-or*, *-or*, *-us* (e.g. *minor*, meaning “smaller”). For further information, see Latin declension: Adjectives.

- The second part of a binomial may be a noun in the nominative case. An example is the binomial name of the lion, which is *Panthera leo*. Grammatically the noun is said to be in apposition to the genus name and the two nouns do not have to agree in gender; in this case, *Panthera* is feminine and *leo* is masculine. The second part of a binomial may be a noun in the genitive (possessive) case. The genitive case is constructed in a number of ways in Latin, depending on the declension of the noun. Common endings for masculine and neuter nouns are *-ii* or *-i* in the singular and *-orum* in the plural, and for feminine nouns *-ae* in the singular and *-arum* in the plural. The noun may be part of a person’s name, often the surname, as in the Tibetan antelope (*Pantholops hodgsonii*), the shrub *Magnolia hodgsonii*, or the olive-backed pipit (*Anthus hodgsoni*). The meaning is “of the person named”, so that *Magnolia hodgsonii* means “Hodgson’s magnolia”. The *-ii* or *-i* endings show that in each case Hodgson was a man (not the same one); had Hodgson been a woman, *hodgsonae* would have been used. The person commemorated in the binomial name is not usually (if ever) the person who created the name; for example *Anthus hodgsoni* was named by Charles Wallace Richmonde, in honour of Hodgson. Rather than a person, the noun may be related to a place, as with *Latimeria chalumnae*, meaning “of the Chalumna River”. Another use of genitive nouns is in, for example, the name of the bacterium *Escherichia coli*, where *coli* means “of the colon”. This formation is common in parasites, as in *Xenos vesparum*, where *vesparum* means “of the wasps”, since *Xenos vesparum* is a parasite of wasps.

Whereas the first part of a binomial name must be unique within a kingdom, the second part is quite commonly used in two or more genera (as is shown by examples of *hodgsonii* above). The full binomial name must be unique within a kingdom

16.3 Taxonomy And Systematics

Taxonomy⁵ (from Ancient Greek *taxis*, meaning ‘arrangement’, and *nomia*, meaning ‘method’) is the science of defining and naming groups of biological organisms on the

⁵[https://en.wikipedia.org/wiki/Taxonomy_\(biology\)](https://en.wikipedia.org/wiki/Taxonomy_(biology))

basis of shared characteristics. Organisms are grouped together into taxa (singular: taxon) and these groups are given a taxonomic rank; groups of a given rank can be aggregated to form a super-group of higher rank, thus creating a taxonomic hierarchy. The principal ranks in modern use are domain, kingdom, phylum (division is sometimes used in botany in place of phylum), class, order, family, genus and species. The Swedish botanist Carl Linnaeus⁶ (1707–1778) is regarded as the father of taxonomy, as he developed a system known as Linnaean taxonomy for categorization of organisms and binomial nomenclature for naming organisms. With the advent of such fields of study as phylogenetics, cladistics, and systematics, the Linnaean system has progressed to a system of modern biological classification based on the evolutionary relationships between organisms, both living and extinct.

Systematics is the branch of science that deals with unique properties of species and groups to recognise, describe name and arrange the diverse organisms according to an organised plan. The word systematics is derived from Latin word *systema* which means systematic arrangement of organisms. Linneous used ‘*systema naturae*’ as the title of his book.

The term “*taxonomy*” was coined by Augustin Pyramus de Candolle while the term “*systematic*” was coined by Carl Linnaeus the father of taxonomy.

Biological systematics is the study of the diversification of living forms, both past and present, and the relationships among living things through time. Relationships are visualized as evolutionary trees (synonyms: cladograms, phylogenetic trees, phylogenies). Phylogenies have two components: branching order (showing group relationships) and branch length (showing amount of evolution). Phylogenetic trees of species and higher taxa are used to study the evolution of traits (e.g., anatomical or molecular characteristics) and the distribution of organisms (biogeography). Systematics, in other words, is used to understand the evolutionary history of life on Earth.

Biological systematics classifies species by using three specific branches. Numerical systematics, or biometry, uses biological statistics to identify and classify animals. Biochemical systematics classifies and identifies animals based on the analysis of the material that makes up the living part of a cell—such as the nucleus, organelles, and cytoplasm. Experimental systematics identifies and classifies animals based on the evolutionary units that comprise a species, as well as their importance in evolution itself. Factors such as mutations, genetic divergence, and hybridization all are considered evolutionary units.

⁶https://en.wikipedia.org/wiki/Carl_Linnaeus

With the specific branches, researchers are able to determine the applications and uses for modern-day systematics. These applications include:

- Studying the diversity of organisms and the differentiation between extinct and living creatures. Biologists study the well-understood relationships by making many different diagrams and “trees” (cladograms, phylogenetic trees, phylogenies, etc.).
- Including the scientific names of organisms, species descriptions and overviews, taxonomic orders, and classifications of evolutionary and organism histories.
- Explaining the biodiversity of the planet and its organisms. The systematic study is that of conservation.
- Manipulating and controlling the natural world. This includes the practice of ‘biological control’, the intentional introduction of natural predators and disease.

Linnaeus ushered in a new era of taxonomy. With his major works *Systema Naturae* 1st Edition in 1735, *Species Plantarum* in 1753, and *Systema Naturae* 10th Edition, he revolutionized modern taxonomy. His works implemented a standardized binomial naming system for animal and plant species, which proved to be an elegant solution to a chaotic and disorganized taxonomic literature. He not only introduced the standard of class, order, genus, and species, but also made it possible to identify plants and animals from his book, by using the smaller parts of the flower. Thus, the Linnaean system was born, and is still used in essentially the same way today as it was in the 18th century. Currently, plant and animal taxonomists regard Linnaeus’ work as the “starting point” for valid names (at 1753 and 1758 respectively). Names published before these dates are referred to as “pre-Linnaean”, and not considered valid (with the exception of spiders published in *Svenska Spindlar*). Even taxonomic names published by Linnaeus himself before these dates are considered pre-Linnaean.

Whereas Linnaeus classified for ease of identification, the idea of the Linnaean taxonomy as translating into a sort of dendrogram of the Animal- and Plant Kingdoms was formulated toward the end of the 18th century, well before *On the Origin of Species* was published. Among early works exploring the idea of a transmutation of species were Erasmus Darwin’s⁷ 1796 *Zoönomia* and Jean-Baptiste Lamarck’s⁸ *Philosophie Zoologique* of 1809. The idea was popularized in the Anglophone world by the speculative but widely read *Vestiges of the Natural History of Creation*⁹, published anonymously by Robert

⁷https://en.wikipedia.org/wiki/Erasmus_Darwin

⁸https://en.wikipedia.org/wiki/Jean-Baptiste_Lamarck

⁹https://en.wikipedia.org/wiki/Vestiges_of_the_Natural_History_of_Creation

Chambers¹⁰ in 1844.

With Darwin's theory, a general acceptance quickly appeared that a classification should reflect the Darwinian principle of common descent. Tree of life representations became popular in scientific works, with known fossil groups incorporated. One of the first modern groups tied to fossil ancestors was birds. Using the then newly discovered fossils of Archaeopteryx and Hesperornis, Thomas Henry Huxley pronounced that they had evolved from dinosaurs, a group formally named by Richard Owen¹¹ in 1842. The resulting description, that of dinosaurs "giving rise to" or being "the ancestors of" birds, is the essential hallmark of evolutionary taxonomic thinking. As more and more fossil groups were found and recognized in the late 19th and early 20th centuries, paleontologists worked to understand the history of animals through the ages by linking together known groups. With the modern evolutionary synthesis of the early 1940s, an essentially modern understanding of the evolution of the major groups was in place.

The cladistic method has emerged since the 1960s. In 1958, Julian Huxley¹² used the term clade. Later, in 1960, Cain and Harrison introduced the term cladistic. The salient feature is arranging taxa in a hierarchical evolutionary tree, ignoring ranks. A taxon is called monophyletic, if it includes all the descendants of an ancestral form. Groups that have descendant groups removed from them (e.g. dinosaurs, with birds as offspring group) are termed paraphyletic, while groups representing more than one branch from the tree of life are called polyphyletic. The International Code of Phylogenetic Nomenclature or PhyloCode is intended to regulate the formal naming of clades. Linnaean ranks will be optional under the PhyloCode, which is intended to coexist with the current, rank-based codes.

16.4 Kingdoms And Domains

Well before Linnaeus, plants and animals were considered separate Kingdoms. Linnaeus used this as the top rank, dividing the physical world into the plant, animal and mineral kingdoms. As advances in microscopy made classification of microorganisms possible, the number of kingdoms increased, five and six-kingdom systems being the most common.

When Carl Linnaeus introduced the rank-based system of nomenclature into biology in 1735, the highest rank was

given the name "kingdom" and was followed by four other main or principal ranks: class, order, genus and species. Later two further main ranks were introduced, making the sequence kingdom, phylum or division, class, order, family, genus and species. In 1990, the rank of domain was introduced above kingdom.

Prefixes can be added so subkingdom (subregnum) and infrakingdom (also known as infraregnum) are the two ranks immediately below kingdom. Superkingdom may be considered as an equivalent of domain or empire or as an independent rank between kingdom and domain or subdomain. In some classification systems the additional rank branch (Latin: ramus) can be inserted between subkingdom and infrakingdom, e.g., Protostomia and Deuterostomia in the classification of Cavalier-Smith.

Some recent classifications based on modern cladistics have explicitly abandoned the term "kingdom", noting that the traditional kingdoms are not monophyletic, i.e., do not consist of all the descendants of a common ancestor.

By tradition, the binomial names of species are usually typeset in italics; for example, *Homo sapiens*. Generally, the binomial should be printed in a font style different from that used in the normal text; for example, "Several more *Homo sapiens* fossils were discovered." When handwritten, a binomial name should be underlined; for example, Homo sapiens.

The first part of the binomial, the genus name, is always written with an initial capital letter. In current usage, the second part is never written with an initial capital. Older sources, particularly botanical works published before the 1950s, use a different convention. If the second part of the name is derived from a proper noun, e.g. the name of a person or place, a capital letter was used. Thus the modern form *Berberis darwinii* was written as *Berberis Darwinii*. A capital was also used when the name is formed by two nouns in apposition, e.g. *Panthera Leo* or *Centaurea Cyanus*.[note 3]

When used with a common name, the scientific name often follows in parentheses, although this varies with publication. For example, "The house sparrow (*Passer domesticus*) is decreasing in Europe."

The binomial name should generally be written in full. The exception to this is when several species from the same genus are being listed or discussed in the same paper or report, or the same species is mentioned repeatedly; in which case the genus is written in full when it is first used, but may then be abbreviated to an initial (and a period/full stop). For example, a list of members of the genus *Canis* might be written as "*Canis lupus*, *C. aureus*, *C. simensis*". In rare cases, this abbreviated form has spread to more general use; for example, the

Creation

¹⁰[https://en.wikipedia.org/wiki/Robert_Chambers_\(publisher,_born_1802\)](https://en.wikipedia.org/wiki/Robert_Chambers_(publisher,_born_1802))

¹¹https://en.wikipedia.org/wiki/Richard_Owen

¹²https://en.wikipedia.org/wiki/Julian_Huxley

bacterium *Escherichia coli* is often referred to as just *E. coli*, and *Tyrannosaurus rex* is perhaps even better known simply as *T. rex*, these two both often appearing in this form in popular writing even where the full genus name has not already been given.

16.4.1 The Three-Domain System

The three-domain system is a biological classification first proposed in 1977 by Carl Woese's¹³ that divides cellular life forms into Archaea¹⁴, Bacteria¹⁵ and Eukaryota¹⁶. The key difference from earlier classifications is the splitting of archaea and bacteria, previously grouped into the single kingdom Bacteria (a kingdom also sometimes called Monera).

The three-domain system adds a level of classification (the domains) "above" the kingdoms present in the previously used five- or six-kingdom systems. This classification system recognizes the fundamental divide between the two prokaryotic groups, insofar as Archaea appear to be more closely related to Eukaryotes than they are to other prokaryotes – bacteria-like organisms with no cell nucleus. The current system sorts the previously known kingdoms into these three domains: Archaea, Bacteria, and Eukarya.

Woese argued, on the basis of differences in 16S rRNA genes, that bacteria, archaea, and eukaryotes each arose separately from an ancestor with poorly developed genetic machinery, often called a progenote. To reflect these primary lines of descent, he treated each as a domain, divided into several different kingdoms. Originally his split of the prokaryotes was into Eubacteria (now Bacteria) and Archaeabacteria (now Archaea). Woese initially used the term "kingdom" to refer to the three primary phylogenetic groupings, and this nomenclature was widely used until the term "domain" was adopted in 1990.

Acceptance of the validity of Woese's phylogenetically valid classification was a slow process. Prominent biologists including Salvador Luria and Ernst Mayr objected to his division of the prokaryotes. Not all criticism of him was restricted to the scientific level. A decade of labor-intensive oligonucleotide cataloging left him with a reputation as "a crank," and Woese would go on to be dubbed as "Microbiology's Scarred Revolutionary" by a news article printed in the journal Science. The growing amount of supporting data led the scientific community to accept the Archaea by the mid-1980s. Today, few scientists cling to the idea of a unified Prokarya.

¹³https://en.wikipedia.org/wiki/Carl_Woese

¹⁴<https://en.wikipedia.org/wiki/Archaea>

¹⁵<https://en.wikipedia.org/wiki/Bacteria>

¹⁶<https://en.wikipedia.org/wiki/Eukaryote>

16.5 Cladistics

Cladistics¹⁷ (from Greek κλάδος, kládos, "branch") is an approach to biological classification in which organisms are categorized based on shared derived characteristics that can be traced to a group's most recent common ancestor and are not present in more distant ancestors. Therefore, members of a group are assumed to share a common history and are considered to be closely related.

The original methods used in cladistic analysis and the school of taxonomy derived from the work of the German entomologist Willi Hennig¹⁸, who referred to it as phylogenetic systematics (also the title of his 1966 book); the terms "cladistics" and "clade" were popularized by other researchers. Cladistics in the original sense refers to a particular set of methods used in phylogenetic analysis, although it is now sometimes used to refer to the whole field.

The following terms, coined by Hennig, are used to identify shared or distinct character states among groups of organisms:

- A **plesiomorphy** ("close form") or ancestral state is a character state that a taxon has retained from its ancestors. When two or more taxa that are not nested within each other share a plesiomorphy, it is a symplesiomorphy (from syn-, "together"). Symplesiomorphies do not mean that the taxa that exhibit that character state are necessarily closely related. For example, Reptilia is traditionally characterized by (among other things) being cold-blooded (i.e., not maintaining a constant high body temperature), whereas birds are warm-blooded. Since cold-bloodedness is a plesiomorphy, inherited from the common ancestor of traditional reptiles and birds, and thus a symplesiomorphy of turtles, snakes and crocodiles (among others), it does not mean that turtles, snakes and crocodiles form a clade that excludes the birds.
- An **apomorphy** ("separate form") or derived state is an innovation. It can thus be used to diagnose a clade – or even to help define a clade name in phylogenetic nomenclature. Features that are derived in individual taxa (a single species or a group that is represented by a single terminal in a given phylogenetic analysis) are called autapomorphies (from auto-, "self"). Autapomorphies express nothing about relationships among groups; clades are identified (or defined) by synapomorphies (from syn-, "together"). For example, the possession of digits that are homologous with those of *Homo sapiens* is a synapomorphy within the vertebrates.

¹⁷<https://en.wikipedia.org/wiki/Cladistics>

¹⁸https://en.wikipedia.org/wiki/Willi_Hennig

The tetrapods can be singled out as consisting of the first vertebrate with such digits homologous to those of *Homo sapiens* together with all descendants of this vertebrate (an apomorphy-based phylogenetic definition). Importantly, snakes and other tetrapods that do not have digits are nonetheless tetrapods: other characters, such as amniotic eggs and diapsid skulls, indicate that they descended from ancestors that possessed digits which are homologous with ours.

- A **character state** is homoplastic or “an instance of homoplasy” if it is shared by two or more organisms but is absent from their common ancestor or from a later ancestor in the lineage leading to one of the organisms. It is therefore inferred to have evolved by convergence or reversal. Both mammals and birds are able to maintain a high constant body temperature (i.e., they are warm-blooded). However, the accepted cladogram explaining their significant features indicates that their common ancestor is in a group lacking this character state, so the state must have evolved independently in the two clades. Warm-bloodedness is separately a synapomorphy of mammals (or a larger clade) and of birds (or a larger clade), but it is not a synapomorphy of any group including both these clades. Hennig's Auxiliary Principle states that shared character states should be considered evidence of grouping unless they are contradicted by the weight of other evidence; thus, homoplasy of some feature among members of a group may only be inferred after a phylogenetic hypothesis for that group has been established.

The terms plesiomorphy and apomorphy are relative; their application depends on the position of a group within a tree. For example, when trying to decide whether the tetrapods form a clade, an important question is whether having four limbs is a synapomorphy of the earliest taxa to be included within Tetrapoda: did all the earliest members of the Tetrapoda inherit four limbs from a common ancestor, whereas all other vertebrates did not, or at least not homologously? By contrast, for a group within the tetrapods, such as birds, having four limbs is a plesiomorphy. Using these two terms allows a greater precision in the discussion of homology, in particular allowing clear expression of the hierarchical relationships among different homologous features.

It can be difficult to decide whether a character state is in fact the same and thus can be classified as a synapomorphy, which may identify a monophyletic group, or whether it only appears to be the same and is thus a homoplasy, which cannot identify such a group. There is a danger of

circular reasoning: assumptions about the shape of a phylogenetic tree are used to justify decisions about character states, which are then used as evidence for the shape of the tree. Phylogenetics uses various forms of parsimony to decide such questions; the conclusions reached often depend on the dataset and the methods. Such is the nature of empirical science, and for this reason, most cladists refer to their cladograms as hypotheses of relationship. Cladograms that are supported by a large number and variety of different kinds of characters are viewed as more robust than those based on more limited evidence.

From the late-20th century onwards, cladistics superseded phenetics. Phenetics was an attempt to determine the relationships of organisms through a measure of overall similarity, making no distinction between plesiomorphies (shared ancestral traits) and apomorphies (derived traits).

A monophyletic¹⁹ group is a group of organisms that forms a clade, which consists of all the descendants of a common ancestor. Monophyletic groups are typically characterized by shared derived characteristics (synapomorphies), which distinguish organisms in the clade from other organisms. The arrangement of the members of a monophyletic group is called a monophly.

A group is paraphyletic²⁰ if it consists of the group's last common ancestor and all descendants of that ancestor excluding a few—typically only one or two—monophyletic subgroups. The group is said to be paraphyletic with respect to the excluded subgroups. The arrangement of the members of a paraphyletic group is called a paraphly.

A polyphyletic²¹ (Greek for “of many races”) group is a set of organisms, or other evolving elements, that have been grouped together but do not share an immediate common ancestor. The term is often applied to groups that share characteristics that appear to be similar but have not been inherited from common ancestors; these characteristics are known as homoplasies, and the development and phenomenon of homoplasies is known as convergent evolution. The arrangement of the members of a polyphyletic group is called a polyphly.

A cladogram (from Greek *clados* “branch” and *gramma* “character”) is a diagram used in cladistics to show relations among organisms. A cladogram is not, however, an evolutionary tree because it does not show how ancestors are related to descendants, nor does it show how much they have changed; nevertheless, many evolutionary trees can be inferred from a single cladogram. A cladogram uses lines that branch off in different directions ending at a clade, a

¹⁹<https://en.wikipedia.org/wiki/Monophly>

²⁰<https://en.wikipedia.org/wiki/Paraphly>

²¹<https://en.wikipedia.org/wiki/Polyphly>

group of organisms with a last common ancestor. There are many shapes of cladograms but they all have lines that branch off from other lines. The lines can be traced back to where they branch off. These branching off points represent a hypothetical ancestor (not an actual entity) which can be inferred to exhibit the traits shared among the terminal taxa above it. This hypothetical ancestor might then provide clues about the order of evolution of various features, adaptation, and other evolutionary narratives about ancestors. Although traditionally such cladograms were generated largely on the basis of morphological characters, DNA and RNA sequencing data and computational phylogenetics are now very commonly used in the generation of cladograms, either on their own or in combination with morphology.

The characteristics used to create a cladogram can be roughly categorized as either morphological (synapsid skull, warm blooded, notochord, unicellular, etc.) or molecular (DNA, RNA, or other genetic information). Prior to the advent of DNA sequencing, cladistic analysis primarily used morphological data. Behavioral data (for animals) may also be used.

As DNA sequencing has become cheaper and easier, molecular systematics has become a more and more popular way to infer phylogenetic hypotheses. Using a parsimony criterion is only one of several methods to infer a phylogeny from molecular data. Approaches such as maximum likelihood, which incorporate explicit models of sequence evolution, are non-Hennigian ways to evaluate sequence data. Another powerful method of reconstructing phylogenies is the use of genomic retrotransposon markers, which are thought to be less prone to the problem of reversion that plagues sequence data. They are also generally assumed to have a low incidence of homoplasies because it was once thought that their integration into the genome was entirely random; this seems at least sometimes not to be the case, however.

Researchers must decide which character states are “ancestral” (plesiomorphies) and which are derived (synapomorphies), because only synapomorphic character states provide evidence of grouping. This determination is usually done by comparison to the character states of one or more outgroups. States shared between the outgroup and some members of the in-group are symplesiomorphies; states that are present only in a subset of the in-group are synapomorphies. Note that character states unique to a single terminal (autapomorphies) do not provide evidence of grouping. The choice of an outgroup is a crucial step in cladistic analysis because different outgroups can produce trees with profoundly different topologies.

A homoplasy is a character state that is shared by two

or more taxa due to some cause other than common ancestry. The two main types of homoplasy are convergence (evolution of the “same” character in at least two distinct lineages) and reversion (the return to an ancestral character state). Characters that are obviously homoplastic, such as white fur in different lineages of Arctic mammals, should not be included as a character in a phylogenetic analysis as they do not contribute anything to our understanding of relationships. However, homoplasy is often not evident from inspection of the character itself (as in DNA sequence, for example), and is then detected by its incongruence (unparsimonious distribution) on a most-parsimonious cladogram. Note that characters that are homoplastic may still contain phylogenetic signal.

A well-known example of homoplasy due to convergent evolution would be the character, “presence of wings”. Although the wings of birds, bats, and insects serve the same function, each evolved independently, as can be seen by their anatomy. If a bird, bat, and a winged insect were scored for the character, “presence of wings”, a homoplasy would be introduced into the dataset, and this could potentially confound the analysis, possibly resulting in a false hypothesis of relationships. Of course, the only reason a homoplasy is recognizable in the first place is because there are other characters that imply a pattern of relationships that reveal its homoplastic distribution.

The cladogram shown in Figure 16.4 represents the current universally accepted hypothesis that all primates, including strepsirrhines like the lemurs and lorises, had a common ancestor all of whose descendants were primates, and so form a clade; the name Primates is therefore recognized for this clade. Within the primates, all anthropoids (monkeys, apes and humans) are hypothesized to have had a common ancestor all of whose descendants were anthropoids, so they form the clade called Anthropoidea. The “prosimians”, on the other hand, form a paraphyletic taxon. The name Prosimii is not used in phylogenetic nomenclature, which names only clades; the “prosimians” are instead divided between the clades Strepsirrhini and Haplorrhini, where the latter contains Tarsiiformes and Anthropoidea.

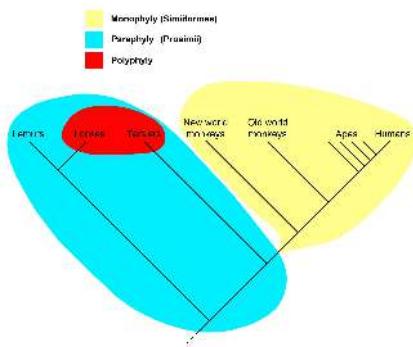


Figure 16.4: Cladogram of the primates²² showing a monophyletic taxon (a clade: the simians or Anthropoidea, in yellow), a paraphyletic taxon (the prosimians, in blue, including the red patch), and a polyphyletic taxon (the nocturnal primates – the lorises and the tarsiers – in red)]

Chapter 17

Fungi



Figure 17.1: A collage of five fungi (clockwise from top left):¹ a mushroom with a flat, red top with white-spots, and a white stem growing on the ground; a red cup-shaped fungus growing on wood; a stack of green and white moldy bread slices on a plate; a microscopic, spherical grey semi-transparent cell, with a smaller spherical cell beside it; a microscopic view of an elongated cellular structure shaped like a microphone, attached to the larger end is a number of smaller roughly circular elements that collectively form a mass around sites.

Fungi (singular: fungus) are heterotrophic uni- or multicellular eukaryotic organisms with external digestion and include microorganisms such as yeasts and molds, as well as the more familiar mushrooms. These organisms are classified as a kingdom, which is separate from the other eukaryotic life kingdoms of plants and animals.

A characteristic that places fungi in a different kingdom from plants, bacteria, and some protists is chitin in their cell walls. Similar to animals, fungi are heterotrophs; they acquire their food by absorbing dissolved molecules, typically by secreting digestive enzymes into their environment. Fungi do not photosynthesize. Growth is their means of mobility, except for spores (a few of which are flagellated), which may travel through the air or water. Fungi are the principal decomposers in ecological systems. These and other differences place fungi in a single group

of related organisms, named the Eumycota (true fungi or Eumycetes), which share a common ancestor (from a monophyletic group), an interpretation that is also strongly supported by molecular phylogenetics. This fungal group is distinct from the structurally similar myxomycetes (slime molds) and oomycetes (water molds). The discipline of biology devoted to the study of fungi is known as mycology (from the Greek μύκης mykes, mushroom). In the past, mycology was regarded as a branch of botany, although it is now known fungi are genetically more closely related to animals than to plants.

Abundant worldwide, most fungi are inconspicuous because of the small size of their structures, and their cryptic lifestyles in soil or on dead matter. Fungi include symbionts of plants, animals, or other fungi and also parasites. They may become noticeable when fruiting, either as mushrooms or as molds. Fungi perform an essential role in the decomposition of organic matter and have fundamental roles in nutrient cycling and exchange in the environment. They have long been used as a direct source of human food, in the form of mushrooms and truffles; as a leavening agent for bread; and in the fermentation of various food products, such as wine, beer, and soy sauce. Since the 1940s, fungi have been used for the production of antibiotics, and, more recently, various enzymes produced by fungi are used industrially and in detergents. Fungi are also used as biological pesticides to control weeds, plant diseases and insect pests. Many species produce bioactive compounds called mycotoxins, such as alkaloids and polyketides, that are toxic to animals including humans. The fruiting structures of a few species contain psychotropic compounds and are consumed recreationally or in traditional spiritual ceremonies. Fungi can break down manufactured materials and buildings, and become significant pathogens of humans and other animals. Losses of crops due to fungal diseases (e.g., rice blast disease) or food spoilage can have a large impact on human food supplies and local economies.

The fungus kingdom encompasses an enormous diversity of taxa with varied ecologies, life cycle strate-

gies, and morphologies ranging from unicellular aquatic chytrids to large mushrooms. However, little is known of the true biodiversity of Kingdom Fungi, which has been estimated at 2.2 million to 3.8 million species. Of these, only about 120,000 have been described, with over 8,000 species known to be detrimental to plants and at least 300 that can be pathogenic to humans. Ever since the pioneering 18th and 19th century taxonomical works of Carl Linnaeus, Christian Hendrik Persoon, and Elias Magnus Fries, fungi have been classified according to their morphology (e.g., characteristics such as spore color or microscopic features) or physiology. Advances in molecular genetics have opened the way for DNA analysis to be incorporated into taxonomy, which has sometimes challenged the historical groupings based on morphology and other traits. Phylogenetic studies published in the last decade have helped reshape the classification within Kingdom Fungi, which is divided into one subkingdom, seven phyla, and ten subphyla.

The English word fungus is directly adopted from the Latin *fungus* (mushroom), used in the writings of Horace and Pliny. This in turn is derived from the Greek word *sphongos* (σφόγγος “sponge”), which refers to the macroscopic structures and morphology of mushrooms and molds; the root is also used in other languages, such as the German *Schwamm* (“sponge”) and *Schimmel* (“mold”).

The word mycology is derived from the Greek *mykes* (μύκης “mushroom”) and *logos* (λόγος “discourse”). It denotes the scientific study of fungi. The Latin adjectival form of “mycology” (*mycologicæ*) appeared as early as 1796 in a book on the subject by Christiaan Hendrik Persoon. The word appeared in English as early as 1824 in a book by Robert Kaye Greville. In 1836 the English naturalist Miles Joseph Berkeley’s publication *The English Flora of Sir James Edward Smith*, Vol. 5. also refers to mycology as the study of fungi.

A group of all the fungi present in a particular area or geographic region is known as mycobiota (plural noun, no singular), e.g., “the mycobiota of Ireland”.

Before the introduction of molecular methods for phylogenetic analysis, taxonomists considered fungi to be members of the plant kingdom because of similarities in lifestyle: both fungi and plants are mainly immobile, and have similarities in general morphology and growth habitat. Like plants, fungi often grow in soil and, in the case of mushrooms, form conspicuous fruit bodies, which sometimes resemble plants such as mosses. The fungi are now considered a separate kingdom, distinct from both plants and animals, from which they appear to have diverged around one billion years ago (around the start of the Neoproterozoic Era). Some morphological, biochemical, and genetic features are shared with other

organisms, while others are unique to the fungi, clearly separating them from the other kingdoms:

Shared features include:

- With other eukaryotes: Fungal cells contain membrane-bound nuclei with chromosomes that contain DNA with noncoding regions called introns and coding regions called exons. Fungi have membrane-bound cytoplasmic organelles such as mitochondria, sterol-containing membranes, and ribosomes of the 80S type. They have a characteristic range of soluble carbohydrates and storage compounds, including sugar alcohols (e.g., mannitol), disaccharides, (e.g., trehalose), and polysaccharides (e.g., glycogen, which is also found in animals).
- With animals: Fungi lack chloroplasts and are heterotrophic organisms and so require preformed organic compounds as energy sources.
- With plants: Fungi have a cell wall and vacuoles. They reproduce by both sexual and asexual means, and like basal plant groups (such as ferns and mosses) produce spores. Similar to mosses and algae, fungi typically have haploid nuclei.
- With euglenoids and bacteria: Higher fungi, euglenoids, and some bacteria produce the amino acid L-lysine in specific biosynthesis steps, called the α -aminoacid pathway.
- The cells of most fungi grow as tubular, elongated, and thread-like (filamentous) structures called hyphae, which may contain multiple nuclei and extend by growing at their tips. Each tip contains a set of aggregated vesicles—cellular structures consisting of proteins, lipids, and other organic molecules—called the Spitzenkörper. Both fungi and oomycetes grow as filamentous hyphal cells. In contrast, similar-looking organisms, such as filamentous green algae, grow by repeated cell division within a chain of cells. There are also single-celled fungi (yeasts) that do not form hyphae, and some fungi have both hyphal and yeast forms.

Unique features of fungi:

- Some species grow as unicellular yeasts that reproduce by budding or fission. Dimorphic fungi can switch between a yeast phase and a hyphal phase in response to environmental conditions.
- The fungal cell wall is composed of glucans and chitin; while glucans are also found in plants and chitin in the exoskeleton of arthropods, fungi are the only organisms that combine these two structural molecules in their cell wall. Unlike those of plants and oomycetes, fungal cell walls do not contain cellulose.

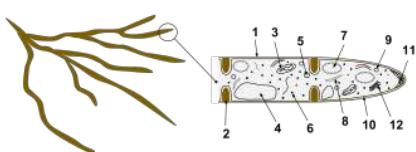


Figure 17.2: Fungal hyphae cells² 1. Hyphal wall 2. Septum 3. Mitochondrion 4. Vacuole 5. Ergosterol crystal 6. Ribosome 7. Nucleus 8. Endoplasmic reticulum 9. Lipid body 10. Plasma membrane 11. Spitzenkörper 12. Golgi apparatus

Most fungi lack an efficient system for the long-distance transport of water and nutrients, such as the xylem and phloem in many plants. To overcome this limitation, some fungi, such as *Armillaria*, form rhizomorphs, which resemble and perform functions similar to the roots of plants. As eukaryotes, fungi possess a biosynthetic pathway for producing terpenes that uses mevalonic acid and pyrophosphate as chemical building blocks. Plants and some other organisms have an additional terpene biosynthesis pathway in their chloroplasts, a structure fungi and animals do not have. Fungi produce several secondary metabolites that are similar or identical in structure to those made by plants. Many of the plant and fungal enzymes that make these compounds differ from each other in sequence and other characteristics, which indicates separate origins and convergent evolution of these enzymes in the fungi and plants.

Fungi have a worldwide distribution, and grow in a wide range of habitats, including extreme environments such as deserts or areas with high salt concentrations or ionizing radiation, as well as in deep sea sediments. Some can survive the intense UV and cosmic radiation encountered during space travel. Most grow in terrestrial environments, though several species live partly or solely in aquatic habitats, such as the chytrid fungus *Batrachochytrium dendrobatidis*, a parasite that has been responsible for a worldwide decline in amphibian populations. This organism spends part of its life cycle as a motile zoospore, enabling it to propel itself through water and enter its amphibian host. Other examples of aquatic fungi include those living in hydrothermal areas of the ocean.

Around 120,000 species of fungi have been described by taxonomists, but the global biodiversity of the fungus kingdom is not fully understood. A 2017 estimate suggests there may be between 2.2 and 3.8 million species. In mycology, species have historically been distinguished by a variety of methods and concepts. Classification

based on morphological characteristics, such as the size and shape of spores or fruiting structures, has traditionally dominated fungal taxonomy. Species may also be distinguished by their biochemical and physiological characteristics, such as their ability to metabolize certain biochemicals, or their reaction to chemical tests. The biological species concept discriminates species based on their ability to mate. The application of molecular tools, such as DNA sequencing and phylogenetic analysis, to study diversity has greatly enhanced the resolution and added robustness to estimates of genetic diversity within various taxonomic groups.

17.1 Morphology of Fungi

Most fungi grow as hyphae, which are cylindrical, thread-like structures 2–10 µm in diameter and up to several centimeters in length. Hyphae grow at their tips (apices); new hyphae are typically formed by emergence of new tips along existing hyphae by a process called branching, or occasionally growing hyphal tips fork, giving rise to two parallel-growing hyphae. Hyphae also sometimes fuse when they come into contact, a process called hyphal fusion (or anastomosis). These growth processes lead to the development of a mycelium, an interconnected network of hyphae. Hyphae can be either septate or coenocytic. Septate hyphae are divided into compartments separated by cross walls (internal cell walls, called septa, that are formed at right angles to the cell wall giving the hypha its shape), with each compartment containing one or more nuclei; coenocytic hyphae are not compartmentalized. Septa have pores that allow cytoplasm, organelles, and sometimes nuclei to pass through; an example is the dolipore septum in fungi of the phylum Basidiomycota. Coenocytic hyphae are in essence multinucleate supercells.

Many species have developed specialized hyphal structures for nutrient uptake from living hosts; examples include haustoria in plant-parasitic species of most fungal phyla, and arbuscules of several mycorrhizal fungi, which penetrate into the host cells to consume nutrients.

Although fungi are opistokonts—a grouping of evolutionarily related organisms broadly characterized by a single posterior flagellum—all phyla except for the chytrids have lost their posterior flagella. Fungi are unusual among the eukaryotes in having a cell wall that, in addition to glucans (e.g., β-1,3-glucan) and other typical components, also contains the biopolymer chitin.

Fungal mycelia can become visible to the naked eye, for example, on various surfaces and substrates, such as damp walls and spoiled food, where they are commonly called molds. Mycelia grown on solid agar

media in laboratory petri dishes are usually referred to as colonies. These colonies can exhibit growth shapes and colors (due to spores or pigmentation) that can be used as diagnostic features in the identification of species or groups. Some individual fungal colonies can reach extraordinary dimensions and ages as in the case of a clonal colony of *Armillaria solidipes*, which extends over an area of more than 900 ha (3.5 square miles), with an estimated age of nearly 9,000 years.

The apothecium—a specialized structure important in sexual reproduction in the ascomycetes—is a cup-shaped fruit body that is often macroscopic and holds the hymenium, a layer of tissue containing the spore-bearing cells. The fruit bodies of the basidiomycetes (basidiocarps) and some ascomycetes can sometimes grow very large, and many are well known as mushrooms.

17.2 Growth And Physiology of Fungi

The growth of fungi as hyphae on or in solid substrates or as single cells in aquatic environments is adapted for the efficient extraction of nutrients, because these growth forms have high surface area to volume ratios. Hyphae are specifically adapted for growth on solid surfaces, and to invade substrates and tissues. They can exert large penetrative mechanical forces; for example, many plant pathogens, including *Magnaporthe grisea*, form a structure called an appressorium that evolved to puncture plant tissues. The pressure generated by the appressorium, directed against the plant epidermis, can exceed 8 megapascals (1,200 psi). The filamentous fungus *Paecilomyces lilacinus* uses a similar structure to penetrate the eggs of nematodes.

The mechanical pressure exerted by the appressorium is generated from physiological processes that increase intracellular turgor by producing osmolytes such as glycerol. Adaptations such as these are complemented by hydrolytic enzymes secreted into the environment to digest large organic molecules—such as polysaccharides, proteins, and lipids—into smaller molecules that may then be absorbed as nutrients. The vast majority of filamentous fungi grow in a polar fashion (extending in one direction) by elongation at the tip (apex) of the hypha. Other forms of fungal growth include intercalary extension (longitudinal expansion of hyphal compartments that are below the apex) as in the case of some endophytic fungi, or growth by volume expansion during the development of mushroom stipes and other large organs. Growth of fungi as multicellular structures consisting of somatic and reproductive cells—a feature independently evolved in animals and plants—has several functions, including the development of fruit bodies for dissemination of sexual spores (see above) and biofilms for substrate colonization and intercellular communication.

The fungi are traditionally considered heterotrophs, organisms that rely solely on carbon fixed by other organisms for metabolism. Fungi have evolved a high degree of metabolic versatility that allows them to use a diverse range of organic substrates for growth, including simple compounds such as nitrate, ammonia, acetate, or ethanol. In some species the pigment melanin may play a role in extracting energy from ionizing radiation, such as gamma radiation. This form of “radiotrophic” growth has been described for only a few species, the effects on growth rates are small, and the underlying biophysical and biochemical processes are not well known. This process might bear similarity to CO₂fixation via visible light, but instead uses ionizing radiation as a source of energy.

17.3 Reproduction of Fungi

Fungal reproduction is complex, reflecting the differences in lifestyles and genetic makeup within this diverse kingdom of organisms. It is estimated that a third of all fungi reproduce using more than one method of propagation; for example, reproduction may occur in two well-differentiated stages within the life cycle of a species, the teleomorph and the anamorph. Environmental conditions trigger genetically determined developmental states that lead to the creation of specialized structures for sexual or asexual reproduction. These structures aid reproduction by efficiently dispersing spores or spore-containing propagules.

17.3.1 Asexual reproduction

Asexual reproduction occurs via vegetative spores (conidia) or through mycelial fragmentation. Mycelial fragmentation occurs when a fungal mycelium separates into pieces, and each component grows into a separate mycelium. Mycelial fragmentation and vegetative spores maintain clonal populations adapted to a specific niche, and allow more rapid dispersal than sexual reproduction. The “Fungi imperfecti” (fungi lacking the perfect or sexual stage) or Deuteromycota comprise all the species that lack an observable sexual cycle. Deuteromycota is not an accepted taxonomic clade, and is now taken to mean simply fungi that lack a known sexual stage.

17.3.2 Sexual reproduction

Sexual reproduction with meiosis has been directly observed in all fungal phyla except Glomeromycota (genetic analysis suggests meiosis in Glomeromycota as well). It differs in many aspects from sexual reproduction in animals or plants. Differences also exist between

fungal groups and can be used to discriminate species by morphological differences in sexual structures and reproductive strategies. Mating experiments between fungal isolates may identify species on the basis of biological species concepts. The major fungal groupings have initially been delineated based on the morphology of their sexual structures and spores; for example, the spore-containing structures, asci and basidia, can be used in the identification of ascomycetes and basidiomycetes, respectively. Fungi employ two mating systems: heterothallic species allow mating only between individuals of opposite mating type, whereas homothallic species can mate, and sexually reproduce, with any other individual or itself.

Most fungi have both a haploid and a diploid stage in their life cycles. In sexually reproducing fungi, compatible individuals may combine by fusing their hyphae together into an interconnected network; this process, anastomosis, is required for the initiation of the sexual cycle. Many ascomycetes and basidiomycetes go through a dikaryotic stage, in which the nuclei inherited from the two parents do not combine immediately after cell fusion, but remain separate in the hyphal cells (see heterokaryosis).

In ascomycetes, dikaryotic hyphae of the hymenium (the spore-bearing tissue layer) form a characteristic hook at the hyphal septum. During cell division, formation of the hook ensures proper distribution of the newly divided nuclei into the apical and basal hyphal compartments. An ascus (plural asci) is then formed, in which karyogamy (nuclear fusion) occurs. Asci are embedded in an ascocarp, or fruiting body. Karyogamy in the asci is followed immediately by meiosis and the production of ascospores. After dispersal, the ascospores may germinate and form a new haploid mycelium.

Sexual reproduction in basidiomycetes is similar to that of the ascomycetes. Compatible haploid hyphae fuse to produce a dikaryotic mycelium. However, the dikaryotic phase is more extensive in the basidiomycetes, often also present in the vegetatively growing mycelium. A specialized anatomical structure, called a clamp connection, is formed at each hyphal septum. As with the structurally similar hook in the ascomycetes, the clamp connection in the basidiomycetes is required for controlled transfer of nuclei during cell division, to maintain the dikaryotic stage with two genetically different nuclei in each hyphal compartment. A basidiocarp is formed in which club-like structures known as basidia generate haploid basidiospores after karyogamy and meiosis. The most commonly known basidiocarps are mushrooms, but they may also take other forms (see Morphology section).

In fungi formerly classified as Zygomycota, haploid hyphae of two individuals fuse, forming a gametangium,

a specialized cell structure that becomes a fertile gamete-producing cell. The gametangium develops into a zygosporangium, a thick-walled spore formed by the union of gametes. When the zygosporangium germinates, it undergoes meiosis, generating new haploid hyphae, which may then form asexual sporangiospores. These sporangiospores allow the fungus to rapidly disperse and germinate into new genetically identical haploid fungal mycelia.

17.3.3 Spore dispersal

Both asexual and sexual spores or sporangiospores are often actively dispersed by forcible ejection from their reproductive structures. This ejection ensures exit of the spores from the reproductive structures as well as traveling through the air over long distances.

Specialized mechanical and physiological mechanisms, as well as spore surface structures (such as hydrophobins), enable efficient spore ejection. For example, the structure of the spore-bearing cells in some ascomycete species is such that the buildup of substances affecting cell volume and fluid balance enables the explosive discharge of spores into the air. The forcible discharge of single spores termed ballistospores involves formation of a small drop of water (Buller's drop), which upon contact with the spore leads to its projectile release with an initial acceleration of more than 10,000 g; the net result is that the spore is ejected 0.01–0.02 cm, sufficient distance for it to fall through the gills or pores into the air below. Other fungi, like the puffballs, rely on alternative mechanisms for spore release, such as external mechanical forces. The hydnoid fungi (tooth fungi) produce spores on pendant, tooth-like or spine-like projections. The bird's nest fungi use the force of falling water drops to liberate the spores from cup-shaped fruiting bodies. Another strategy is seen in the stinkhorns, a group of fungi with lively colors and putrid odor that attract insects to disperse their spores.

The most common means of spore dispersal is by wind – species using this form of dispersal often produce dry or hydrophobic spores which do not absorb water and are readily scattered by raindrops, for example. Most of the researched species of fungus are transported by wind.

17.4 Evolution of Fungi

In contrast to plants and animals, the early fossil record of the fungi is meager. The earliest fossils possessing features typical of fungi date to the Paleoproterozoic era, some 2,400 million years ago (Ma); these multicellular benthic organisms had filamentous structures capable of anastomosis. Other studies (2009) estimate the arrival of fungal organisms at about 760–1060 Ma on the basis of comparisons of the rate of evolution in closely related groups.

For much of the Paleozoic Era (542–251 Ma), the fungi appear to have been aquatic and consisted of organisms similar to the extant chytrids in having flagellum-bearing spores. The evolutionary adaptation from an aquatic to a terrestrial lifestyle necessitated a diversification of ecological strategies for obtaining nutrients, including parasitism, saprobism, and the development of mutualistic relationships such as mycorrhiza and lichenization. Recent (2009) studies suggest that the ancestral ecological state of the Ascomycota was saprobism, and that independent lichenization events have occurred multiple times.

In May 2019, scientists reported the discovery of a fossilized fungus, named *Ourasphaira giraldae*, in the Canadian Arctic, that may have grown on land a billion years ago, well before plants were living on land. Earlier, it had been presumed that the fungi colonized the land during the Cambrian (542–488.3 Ma), also long before land plants. Fossilized hyphae and spores recovered from the Ordovician of Wisconsin (460 Ma) resemble modern-day Glomerales, and existed at a time when the land flora likely consisted of only non-vascular bryophyte-like plants. Prototaxites, which was probably a fungus or lichen, would have been the tallest organism of the late Silurian and early Devonian. Fungal fossils do not become common and uncontroversial until the early Devonian (416–359.2 Ma), when they occur abundantly in the Rhynie chert, mostly as Zygomycota and Chytridiomycota. At about this same time, approximately 400 Ma, the Ascomycota and Basidiomycota diverged, and all modern classes of fungi were present by the Late Carboniferous (Pennsylvanian, 318.1–299 Ma).

Lichen-like fossils have been found in the Doushantuo Formation in southern China dating back to 635–551 Ma. Lichens formed a component of the early terrestrial ecosystems, and the estimated age of the oldest terrestrial lichen fossil is 400 Ma; this date corresponds to the age of the oldest known sporocarp fossil, a Paleopyrenomyces species found in the Rhynie Chert. The oldest fossil with microscopic features resembling modern-day basidiomycetes is Palaeoancistrus, found permineralized with a fern from the Pennsylvanian. Rare in the fossil record are the Homobasidiomycetes (a taxon roughly equivalent to the mushroom-producing species of the Agaricomycetes). Two amber-preserved specimens provide evidence that the earliest known mushroom-forming fungi (the extinct species Archaeomarasmius leggerti) appeared during the late Cretaceous, 90 Ma.

Some time after the Permian-Triassic extinction event (251.4 Ma), a fungal spike (originally thought to be an extraordinary abundance of fungal spores in sediments) formed, suggesting that fungi were the dominant life form at this time, representing nearly 100% of the available fossil record for this period. However, the relative proportion

of fungal spores relative to spores formed by algal species is difficult to assess, the spike did not appear worldwide, and in many places it did not fall on the Permian-Triassic boundary.

65 million years ago, immediately after the Cretaceous-Paleogene extinction event that famously killed off most dinosaurs, there is a dramatic increase in evidence of fungi, apparently the death of most plant and animal species leading to a huge fungal bloom like “a massive compost heap”.

Although commonly included in botany curricula and textbooks, fungi are more closely related to animals than to plants and are placed with the animals in the monophyletic group of opisthokonts. Analyses using molecular phylogenetics support a monophyletic origin of fungi. The taxonomy of fungi is in a state of constant flux, especially due to recent research based on DNA comparisons. These current phylogenetic analyses often overturn classifications based on older and sometimes less discriminative methods based on morphological features and biological species concepts obtained from experimental matings.

There is no unique generally accepted system at the higher taxonomic levels and there are frequent name changes at every level, from species upwards. Efforts among researchers are now underway to establish and encourage usage of a unified and more consistent nomenclature. Fungal species can also have multiple scientific names depending on their life cycle and mode (sexual or asexual) of reproduction. Web sites such as Index Fungorum and ITIS list current names of fungal species (with cross-references to older synonyms).

The 2007 classification of Kingdom Fungi is the result of a large-scale collaborative research effort involving dozens of mycologists and other scientists working on fungal taxonomy. It recognizes seven phyla, two of which—the Ascomycota and the Basidiomycota—are contained within a branch representing subkingdom Dikarya, the most species rich and familiar group, including all the mushrooms, most food-spoilage molds, most plant pathogenic fungi, and the beer, wine, and bread yeasts.

The major phyla (sometimes called divisions) of fungi have been classified mainly on the basis of characteristics of their sexual reproductive structures.

Phylogenetic analysis has demonstrated that the Microsporidia, unicellular parasites of animals and protists, are fairly recent and highly derived endobiotic fungi (living within the tissue of another species). One 2006 study concludes that the Microsporidia are a sister group to the true fungi; that is, they are each other's closest evolutionary relative. Hibbett and colleagues suggest that this analysis does not clash with their classification of the Fungi, and although the Microsporidia are elevated to phylum status, it

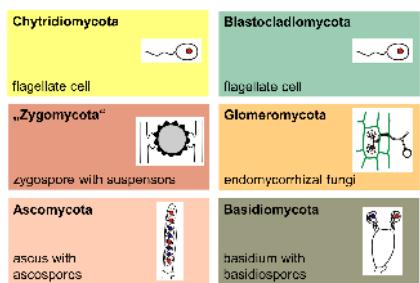


Figure 17.3: Six commonly recognized groups of fungi.³ Zygomycota, or zygote fungi, is a former division or phylum of the kingdom Fungi. The members are now part of two recently proposed phyla.

is acknowledged that further analysis is required to clarify evolutionary relationships within this group.

The Chytridiomycota are commonly known as chytrids. These fungi are distributed worldwide. Chytrids and their close relatives Neocallimastigomycota and Blastocladiomycota (below) are the only fungi with active motility, producing zoospores that are capable of active movement through aqueous phases with a single flagellum, leading early taxonomists to classify them as protists. Molecular phylogenies, inferred from rRNA sequences in ribosomes, suggest that the Chytrids are a basal group divergent from the other fungal phyla, consisting of four major clades with suggestive evidence for paraphyly or possibly polyphyly.

The Blastocladiomycota were previously considered a taxonomic clade within the Chytridiomycota. Recent molecular data and ultrastructural characteristics, however, place the Blastocladiomycota as a sister clade to the Zygomycota, Glomeromycota, and Dikarya (Ascomycota and Basidiomycota). The blastocladiomycetes are saprotrophs, feeding on decomposing organic matter, and they are parasites of all eukaryotic groups. Unlike their close relatives, the chytrids, most of which exhibit zygotic meiosis, the blastocladiomycetes undergo sporic meiosis.

The Neocallimastigomycota were earlier placed in the phylum Chytridiomycota. Members of this small phylum are anaerobic organisms, living in the digestive system of larger herbivorous mammals and in other terrestrial and aquatic environments enriched in cellulose (e.g., domestic waste landfill sites). They lack mitochondria but contain hydrogenosomes of mitochondrial origin. As in the related chytrids, neocallimastigomycetes form zoospores that are posteriorly uniflagellate or polyflagellate.

Members of the Glomeromycota form arbuscular mycorrhizae, a form of mutualist symbiosis wherein fungal hyphae invade plant root cells and both species

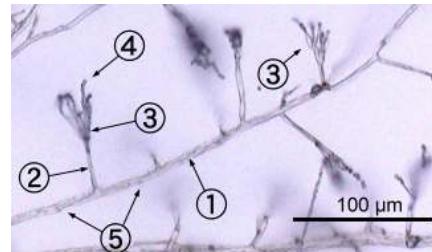


Figure 17.4: An environmental isolate of *Penicillium* hypha conidiophore phialide conidia septa⁴ 1. hypha 2. conidiophore 3. phialide 4. conidia 5. septa

benefit from the resulting increased supply of nutrients. All known Glomeromycota species reproduce asexually. The symbiotic association between the Glomeromycota and plants is ancient, with evidence dating to 400 million years ago. Formerly part of the Zygomycota (commonly known as 'sugar' and 'pin' molds), the Glomeromycota were elevated to phylum status in 2001 and now replace the older phylum Zygomycota. Fungi that were placed in the Zygomycota are now being reassigned to the Glomeromycota, or the subphyla incertae sedis Mucoromycotina, Kickxellomycotina, the Zoopagomycotina and the Entomophthoromycotina. Some well-known examples of fungi formerly in the Zygomycota include black bread mold (*Rhizopus stolonifer*), and *Pilobolus* species, capable of ejecting spores several meters through the air. Medically relevant genera include *Mucor*, *Rhizomucor*, and *Rhizopus*.

The Ascomycota, commonly known as sac fungi or ascomycetes, constitute the largest taxonomic group within the Eumycota. These fungi form meiotic spores called ascospores, which are enclosed in a special sac-like structure called an ascus. This phylum includes morels, a few mushrooms and truffles, unicellular yeasts (e.g., of the genera *Saccharomyces*, *Kluyveromyces*, *Pichia*, and *Candida*), and many filamentous fungi living as saprotrophs, parasites, and mutualistic symbionts (e.g. lichens). Prominent and important genera of filamentous ascomycetes include *Aspergillus*, *Penicillium*, *Fusarium*, and *Claviceps*. Many ascomycete species have only been observed undergoing asexual reproduction (called anamorphic species), but analysis of molecular data has often been able to identify their closest teleomorphs in the Ascomycota. Because the products of meiosis are retained within the sac-like ascus, ascomycetes have been used for elucidating principles of genetics and heredity (e.g., *Neurospora crassa*).

Members of the Basidiomycota, commonly known as the club fungi or basidiomycetes, produce meiospores called basidiospores on club-like stalks called basidia. Most common mushrooms belong to this group, as well

as rust and smut fungi, which are major pathogens of grains. Other important basidiomycetes include the maize pathogen *Ustilago maydis*, human commensal species of the genus *Malassezia*, and the opportunistic human pathogen, *Cryptococcus neoformans*.

17.5 Fungus-like organisms

Because of similarities in morphology and lifestyle, the slime molds, water molds (oomycetes) and hyphochytrids (both Stramenopiles) were formerly classified in the kingdom Fungi.

Unlike true fungi, the cell walls of oomycetes contain cellulose and lack chitin. Hyphochytrids have both chitin and cellulose. Slime molds lack a cell wall during the assimilative phase (except labyrinthulids, which have a wall of scales), and ingest nutrients by ingestion (phagocytosis, except labyrinthulids) rather than absorption (osmotrophy, as fungi, labyrinthulids, oomycetes and hyphochytrids). Neither water molds nor slime molds are closely related to the true fungi, and, therefore, taxonomists no longer group them in the kingdom Fungi. Nonetheless, studies of the oomycetes and myxomycetes are still often included in mycology textbooks and primary research literature.

17.6 Ecology of Fungi

Although often inconspicuous, fungi occur in every environment on Earth and play very important roles in most ecosystems. Along with bacteria, fungi are the major decomposers in most terrestrial (and some aquatic) ecosystems, and therefore play a critical role in biogeochemical cycles and in many food webs. As decomposers, they play an essential role in nutrient cycling, especially as saprotrophs and symbionts, degrading organic matter to inorganic molecules, which can then re-enter anabolic metabolic pathways in plants or other organisms.

17.6.1 Symbiosis

Many fungi have important symbiotic relationships with organisms from most if not all kingdoms. These interactions can be mutualistic or antagonistic in nature, or in the case of commensal fungi are of no apparent benefit or detriment to the host.

Mycorrhizal symbiosis between plants and fungi is one of the most well-known plant-fungus associations and is of significant importance for plant growth and persistence in many ecosystems; over 90% of all plant species engage in mycorrhizal relationships with fungi and are dependent upon this relationship for survival.

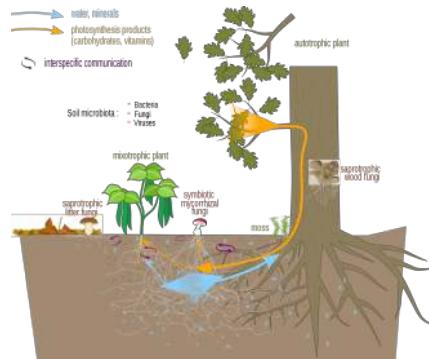


Figure 17.5: Nutrient exchanges and communication between a mycorrhizal fungus and plants.⁵

The mycorrhizal symbiosis is ancient, dating to at least 400 million years ago. It often increases the plant's uptake of inorganic compounds, such as nitrate and phosphate from soils having low concentrations of these key plant nutrients. The fungal partners may also mediate plant-to-plant transfer of carbohydrates and other nutrients. Such mycorrhizal communities are called "common mycorrhizal networks". A special case of mycorrhiza is myco-heterotrophy, whereby the plant parasitizes the fungus, obtaining all of its nutrients from its fungal symbiont. Some fungal species inhabit the tissues inside roots, stems, and leaves, in which case they are called endophytes. Similar to mycorrhiza, endophytic colonization by fungi may benefit both symbionts; for example, endophytes of grasses impart to their host increased resistance to herbivores and other environmental stresses and receive food and shelter from the plant in return.

Lichens are a symbiotic relationship between fungi and photosynthetic algae or cyanobacteria. The photosynthetic partner in the relationship is referred to in lichen terminology as a "photobiont".

The fungal part of the relationship is composed mostly of various species of ascomycetes and a few basidiomycetes.

Lichens occur in every ecosystem on all continents, play a key role in soil formation and the initiation of biological succession, and are prominent in some extreme environments, including polar, alpine, and semiarid desert regions. They are able to grow on inhospitable surfaces, including bare soil, rocks, tree bark, wood, shells, barnacles and leaves. As in mycorrhizas, the photobiont provides sugars and other carbohydrates via photosynthesis to the fungus, while the fungus provides minerals and water to the photobiont. The functions of both symbiotic organisms are so closely intertwined that they function almost as a single organism; in most cases the resulting organism



Figure 17.6: Fruticose (light green) and crustose (yellow) lichens growing on a tree.



Figure 17.8: Cup-shaped apothecia and thallus of a foliose lichen.

differs greatly from the individual components. Lichenization is a common mode of nutrition for fungi; around 20% of fungi—between 17,500 and 20,000 described species—are lichenized. Characteristics common to most lichens include obtaining organic carbon by photosynthesis, slow growth, small size, long life, long-lasting (seasonal) vegetative reproductive structures, mineral nutrition obtained largely from airborne sources, and greater tolerance of desiccation than most other photosynthetic organisms in the same habitat.

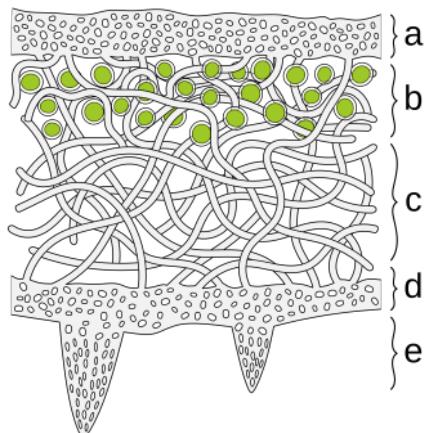


Figure 17.7: Schematic cross section of foliose lichen:⁶ a) The cortex is the outer layer of tightly woven fungus filaments (hyphae) b) This photobiont layer has photosynthesizing green algae c) Loosely packed hyphae in the medulla d) A tightly woven lower cortex e) Anchoring hyphae called rhizines where the fungus attaches to the substrate.

Many insects also engage in mutualistic relationships with fungi. Several groups of ants cultivate fungi in the order Agaricales as their primary food source, while ambrosia beetles cultivate various species of fungi in the bark of trees that they infest. Likewise, females of several wood wasp species (genus *Sirex*) inject their eggs together with spores of the wood-rotting fungus *Amylostereum areolatum* into the sapwood of pine trees; the growth of the fungus provides ideal nutritional conditions for the development of the wasp larvae. At least one species of stingless bee has a relationship with a fungus in the genus *Monascus*, where the larvae consume and depend on fungus transferred from old to new nests. Termites on the African savannah are also known to cultivate fungi, and yeasts of the genera *Candida* and *Lachancea* inhabit the gut of a wide range of insects, including neuropterans, beetles, and cockroaches; it is not known whether these fungi benefit their hosts. Fungi ingrowing dead wood are essential for xylophagous insects (e.g. woodboring beetles). They deliver nutrients needed by xylophages to nutritionally scarce dead wood. Thanks to this nutritional enrichment the larvae of woodboring insect is able to grow and develop to adulthood. The larvae of many families of fungicolous flies, particularly those within the superfamily Sciaroidea such as the Mycetophilidae and some Keroplatidae feed on fungal fruiting bodies and sterile mycorrhizae.

17.7 Fungi as Pathogens And Parasites

Many fungi are parasites on plants, animals (including humans), and other fungi. Serious pathogens of many cultivated plants causing extensive damage and losses to agriculture and forestry include the rice blast fungus *Magnaporthe oryzae*, tree pathogens such as *Ophiostoma ulmi* and *Ophiostoma novo-ulmi* causing Dutch elm disease and *Cryphonectria parasitica* responsible for chestnut blight, and plant pathogens in the genera *Fusarium*, *Ustilago*, *Alternaria*, and *Cochliobolus*. Some carnivorous fungi, like *Paecilomyces lilacinus*, are predators of nematodes, which they capture using an array of specialized structures such as constricting rings or adhesive nets. Many fungi that are plant pathogens, such as *Magnaporthe oryzae*, can switch from being biotrophic (parasitic on living plants) to being necrotrophic (feeding on the dead tissues of plants they have killed). This same principle is applied to fungi-feeding parasites, including *Asterotremella albida*, which feeds on the fruit bodies of other fungi both while they are living and after they are dead.

Some fungi can cause serious diseases in humans, several of which may be fatal if untreated. These include aspergillosis, candidiasis, coccidioidomycosis, cryptococcosis, histoplasmosis, mycetomas, and paracoccidioidomycosis. Furthermore, persons with immuno-deficiencies are particularly susceptible to disease by genera such as *Aspergillus*, *Candida*, *Cryptococcus*, *Histoplasma*, and *Pneumocystis*. Other fungi can attack eyes, nails, hair, and especially skin, the so-called dermatophytic and keratinophilic fungi, and cause local infections such as ringworm and athlete's foot. Fungal spores are also a cause of allergies, and fungi from different taxonomic groups can evoke allergic reactions.

The organisms which parasitize fungi are known as mycoparasitic organisms. Certain species of the genus *Pythium*, which are oomycetes, have potential as biocontrol agents against certain fungi. Fungi can also act as mycoparasites or antagonists of other fungi, such as *Hypomyces chrysospermus*, which grows on bolete mushrooms. Fungi can also become the target of infection by mycoviruses.

17.8 Mycotoxins

Many fungi produce biologically active compounds, several of which are toxic to animals or plants and are therefore called mycotoxins. Of particular relevance to humans are mycotoxins produced by molds causing food spoilage, and poisonous mushrooms (see above). Particularly infamous are the lethal amatoxins in some *Amanita* mushrooms, and ergot alkaloids, which have a long history of causing se-

rious epidemics of ergotism (St Anthony's Fire) in people consuming rye or related cereals contaminated with sclerotia of the ergot fungus, *Claviceps purpurea*. Other notable mycotoxins include the aflatoxins, which are insidious liver toxins and highly carcinogenic metabolites produced by certain *Aspergillus* species often growing in or on grains and nuts consumed by humans, ochratoxins, patulin, and trichothecenes (e.g., T-2 mycotoxin) and fumonisins, which have significant impact on human food supplies or animal livestock.

Mycotoxins are secondary metabolites (or natural products), and research has established the existence of biochemical pathways solely for the purpose of producing mycotoxins and other natural products in fungi. Mycotoxins may provide fitness benefits in terms of physiological adaptation, competition with other microbes and fungi, and protection from consumption (fungivory). Many fungal secondary metabolites (or derivatives) are used medically, as described under Human Use below.

The human use of fungi for food preparation or preservation and other purposes is extensive and has a long history. Mushroom farming and mushroom gathering are large industries in many countries. The study of the historical uses and sociological impact of fungi is known as ethnomycology. Because of the capacity of this group to produce an enormous range of natural products with antimicrobial or other biological activities, many species have long been used or are being developed for industrial production of antibiotics, vitamins, and anti-cancer and cholesterol-lowering drugs. More recently, methods have been developed for genetic engineering of fungi, enabling metabolic engineering of fungal species. For example, genetic modification of yeast species—which are easy to grow at fast rates in large fermentation vessels—has opened up ways of pharmaceutical production that are potentially more efficient than production by the original source organisms.

Many species produce metabolites that are major sources of pharmacologically active drugs. Particularly important are the antibiotics, including the penicillins, a structurally related group of β -lactam antibiotics that are synthesized from small peptides. Although naturally occurring penicillins such as penicillin G (produced by *Penicillium chrysogenum*) have a relatively narrow spectrum of biological activity, a wide range of other penicillins can be produced by chemical modification of the natural penicillins. Modern penicillins are semisynthetic compounds, obtained initially from fermentation cultures, but then structurally altered for specific desirable properties. Other antibiotics produced by fungi include: ciclosporin, commonly used as an immunosuppressant during transplant surgery; and fusidic acid,

used to help control infection from methicillin-resistant *Staphylococcus aureus* bacteria. Widespread use of antibiotics for the treatment of bacterial diseases, such as tuberculosis, syphilis, leprosy, and others began in the early 20th century and continues to date. In nature, antibiotics of fungal or bacterial origin appear to play a dual role: at high concentrations they act as chemical defense against competition with other microorganisms in species-rich environments, such as the rhizosphere, and at low concentrations as quorum-sensing molecules for intra- or interspecies signaling. Other drugs produced by fungi include griseofulvin isolated from *Penicillium griseofulvum*, used to treat fungal infections, and statins (HMG-CoA reductase inhibitors), used to inhibit cholesterol synthesis. Examples of statins found in fungi include mevastatin from *Penicillium citrinum* and lovastatin from *Aspergillus terreus* and the oyster mushroom. Fungi produce compounds that inhibit viruses and cancer cells. Specific metabolites, such as polysaccharide-K, ergotamine, and β -lactam antibiotics, are routinely used in clinical medicine. The shiitake mushroom is a source of lentinan, a clinical drug approved for use in cancer treatments in several countries, including Japan. In Europe and Japan, polysaccharide-K (brand name Krestin), a chemical derived from *Trametes versicolor*, is an approved adjuvant for cancer therapy.

Certain mushrooms enjoy usage as therapeutics in folk medicines, such as Traditional Chinese medicine. Notable medicinal mushrooms with a well-documented history of use include *Agaricus subrufescens*, *Ganoderma lucidum*, *Psilocybe* and *Ophiocordyceps sinensis*.

Baker's yeast or *Saccharomyces cerevisiae*, a unicellular fungus, is used to make bread and other wheat-based products, such as pizza dough and dumplings. Yeast species of the genus *Saccharomyces* are also used to produce alcoholic beverages through fermentation. Shoyu koji mold (*Aspergillus oryzae*) is an essential ingredient in brewing Shoyu (soy sauce) and sake, and the preparation of miso, while *Rhizopus* species are used for making tempeh. Several of these fungi are domesticated species that were bred or selected according to their capacity to ferment food without producing harmful mycotoxins (see below), which are produced by very closely related *Aspergilli*. Quorn, a meat substitute, is made from *Fusarium venenatum*.

Edible mushrooms include commercially raised and wild-harvested fungi. *Agaricus bisporus*, sold as button mushrooms when small or Portobello mushrooms when larger, is the most widely cultivated species in the West, used in salads, soups, and many other dishes. Many Asian fungi are commercially grown and have increased in popularity in the West. They are often available fresh in

grocery stores and markets, including straw mushrooms (*Volvariella volvacea*), oyster mushrooms (*Pleurotus ostreatus*), shiitakes (*Lentinula edodes*), and enokitake (*Flammulina spp.*).

Many other mushroom species are harvested from the wild for personal consumption or commercial sale. Milk mushrooms, morels, chanterelles, truffles, black trumpets, and porcini mushrooms (*Boletus edulis*) (also known as king boletes) demand a high price on the market. They are often used in gourmet dishes.

Certain types of cheeses require inoculation of milk curds with fungal species that impart a unique flavor and texture to the cheese. Examples include the blue color in cheeses such as Stilton or Roquefort, which are made by inoculation with *Penicillium roqueforti*. Molds used in cheese production are non-toxic and are thus safe for human consumption; however, mycotoxins (e.g., aflatoxins, roquefortine C, patulin, or others) may accumulate because of growth of other fungi during cheese ripening or storage.

Many mushroom species are poisonous to humans and cause a range of reactions including slight digestive problems, allergic reactions, hallucinations, severe organ failure, and death. Genera with mushrooms containing deadly toxins include *Conocybe*, *Galerina*, *Lepiota*, and, the most infamous, *Amanita*. The latter genus includes the destroying angel (*A. virosa*) and the death cap (*A. phalloides*), the most common cause of deadly mushroom poisoning. The false morel (*Gyromitra esculenta*) is occasionally considered a delicacy when cooked, yet can be highly toxic when eaten raw. *Tricholoma equestre* was considered edible until it was implicated in serious poisonings causing rhabdomyolysis. Fly agaric mushrooms (*Amanita muscaria*) also cause occasional non-fatal poisonings, mostly as a result of ingestion for its hallucinogenic properties. Historically, fly agaric was used by different peoples in Europe and Asia and its present usage for religious or shamanic purposes is reported from some ethnic groups such as the Koryak people of northeastern Siberia.

As it is difficult to accurately identify a safe mushroom without proper training and knowledge, it is often advised to assume that a wild mushroom is poisonous and not to consume it.

In agriculture, fungi may be useful if they actively compete for nutrients and space with pathogenic microorganisms such as bacteria or other fungi via the competitive exclusion principle, or if they are parasites of these pathogens. For example, certain species may be used to eliminate or suppress the growth of harmful plant pathogens, such as insects, mites, weeds, nematodes, and other fungi that cause diseases of important crop

plants. This has generated strong interest in practical applications that use these fungi in the biological control of these agricultural pests. Entomopathogenic fungi can be used as biopesticides, as they actively kill insects. Examples that have been used as biological insecticides are *Beauveria bassiana*, *Metarhizium* spp., *Hirsutella* spp., *Paecilomyces (Isaria)* spp., and *Lecanicillium lecanii*. Endophytic fungi of grasses of the genus *Neotyphodium*, such as *N. coenophialum*, produce alkaloids that are toxic to a range of invertebrate and vertebrate herbivores. These alkaloids protect grass plants from herbivory, but several endophyte alkaloids can poison grazing animals, such as cattle and sheep. Infecting cultivars of pasture or forage grasses with *Neotyphodium* endophytes is one approach being used in grass breeding programs; the fungal strains are selected for producing only alkaloids that increase resistance to herbivores such as insects, while being non-toxic to livestock.

Certain fungi, in particular white-rot fungi, can degrade insecticides, herbicides, pentachlorophenol, creosote, coal tars, and heavy fuels and turn them into carbon dioxide, water, and basic elements. Fungi have been shown to biomimicrize uranium oxides, suggesting they may have application in the bioremediation of radioactively polluted sites.

Several pivotal discoveries in biology were made by researchers using fungi as model organisms, that is, fungi that grow and sexually reproduce rapidly in the laboratory. For example, the one gene-one enzyme hypothesis was formulated by scientists using the bread mold *Neurospora crassa* to test their biochemical theories. Other important model fungi are *Aspergillus nidulans* and the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, each of which with a long history of use to investigate issues in eukaryotic cell biology and genetics, such as cell cycle regulation, chromatin structure, and gene regulation. Other fungal models have more recently emerged that address specific biological questions relevant to medicine, plant pathology, and industrial uses; examples include *Candida albicans*, a dimorphic, opportunistic human pathogen, *Magnaporthe grisea*, a plant pathogen, and *Pichia pastoris*, a yeast widely used for eukaryotic protein production.

Chapter 18

Plants



Figure 18.1: The diversity of plants.¹

Plants are multicellular organisms, predominantly photosynthetic eukaryotes of the kingdom Plantae. Historically, plants were treated as one of two kingdoms including all living things that were not animals, and all algae and fungi were treated as plants. However, all current definitions of Plantae exclude the fungi and some algae, as well as the prokaryotes (the archaea and bacteria). By one definition, plants form the clade Viridiplantae (Latin name for “green plants”), a group that includes the flowering plants, conifers and other gymnosperms, ferns and their allies, hornworts, liverworts, mosses and the green algae, but excludes the red and brown algae.

Green plants obtain most of their energy from sunlight via photosynthesis by primary chloroplasts that are derived from endosymbiosis with cyanobacteria. Their chloroplasts contain chlorophylls a and b, which gives them their green color. Some plants are parasitic or

mycotrophic and have lost the ability to produce normal amounts of chlorophyll or to photosynthesize. Plants are characterized by sexual reproduction and alternation of generations, although asexual reproduction is also common.

There are about 320,000 species of plants, of which the great majority, some 260–290 thousand, produce seeds. Green plants provide a substantial proportion of the world’s molecular oxygen, and are the basis of most of Earth’s ecosystems. Plants that produce grain, fruit and vegetables also form basic human foods and have been domesticated for millennia. Plants have many cultural and other uses, as ornaments, building materials, writing material and, in great variety, they have been the source of medicines and psychoactive drugs. The scientific study of plants is known as botany, a branch of biology.

All living things were traditionally placed into one of two groups, plants and animals. This classification may date from Aristotle (384 BC – 322 BC), who made the distinction between plants, which generally do not move, and animals, which often are mobile to catch their food. Much later, when Linnaeus (1707–1778) created the basis of the modern system of scientific classification, these two groups became the kingdoms Vegetabilia (later Metaphyta or Plantae) and Animalia (also called Metazoa). Since then, it has become clear that the plant kingdom as originally defined included several unrelated groups, and the fungi and several groups of algae were removed to new kingdoms. However, these organisms are still often considered plants, particularly in popular contexts.

The term “plant” generally implies the possession of the following traits: multicellularity, possession of cell walls containing cellulose, and the ability to carry out photosynthesis with primary chloroplasts.

When the name Plantae or plant is applied to a specific group of organisms or taxon, it usually refers to one of four concepts. From least to most inclusive, these four groupings are:

Table 18.1: Four commonly used concepts of plants.

Name	Scope	Distribution
Land plants (Embryophyta)	Plantae sensu strictissimo	Plants in the strictest sense include the liverworts, hornworts, mosses, and vascular plants, as well as fossil plants similar to these surviving groups.
Green plants (Viridiplantae or Viridiphyta or Chlorobionta or Chloroplastida)	Plantae sensu stricto	Plants in a strict sense include the green algae, and land plants that emerged within them, including stoneworts. The relationships between plant groups are still being worked out, and the names given to them vary considerably. The clade Viridiplantae encompasses a group of organisms that have cellulose in their cell walls, possess chlorophylls a and b and have plastids bound by only two membranes that are capable of photosynthesis and of storing starch. This clade is the main subject of this article.
Archaeplastida (Plastida or Primoplantae)	Plantae sensu lato	Plants in a broad sense comprise the green plants listed above plus the red algae (Rhodophyta) and the glaucophyte algae (Glaucophyta) that store Floridean starch outside the plastids, in the cytoplasm. This clade includes all of the organisms that eons ago acquired their primary chloroplasts directly by engulfing cyanobacteria.
Old definitions of plant (obsolete)	Plantae sensu amplio	Plants in the widest sense refers to older, obsolete classifications that placed diverse algae, fungi or bacteria in Plantae.

The evolution of plants has resulted in increasing levels of complexity, from the earliest algal mats, through bryophytes, lycopods, ferns to the complex gymnosperms and angiosperms of today. Plants in all of these groups continue to thrive, especially in the environments in which they evolved.

An algal scum formed on the land 1,200 million years ago, but it was not until the Ordovician Period, around 450 million years ago, that land plants appeared. However, new evidence from the study of carbon isotope ratios in Precambrian rocks has suggested that complex photosynthetic plants developed on the earth over 1000 m.y.a. For more than a century it has been assumed that the ancestors of land plants evolved in aquatic environments and then adapted to a life on land, an idea usually credited to botanist Frederick Orpen Bower in his 1908 book *The Origin of a Land Flora*. A recent alternative view, supported by genetic evidence, is that they evolved from terrestrial single-celled algae, and that even the common ancestor of red and green algae, and the unicellular freshwater algae glaucophytes, originated in a terrestrial environment in freshwater biofilms or microbial mats. Primitive land plants began to diversify in the late Silurian Period, around 420 million years ago, and the results of their diversification are displayed in remarkable detail in an early Devonian fossil assemblage from the Rhynie chert. This chert preserved early plants in cellular detail, petrified in volcanic springs. By the middle of the Devonian Period most of the features recognised in plants today are present, including roots, leaves and secondary wood, and by late Devonian times seeds had evolved. Late Devonian plants had thereby reached a degree of sophistication that allowed them to form forests of tall trees. Evolutionary innovation continued in the Carboniferous and later geological periods and is ongoing today. Most plant groups were relatively unscathed by the Permo-Triassic extinction event, although the structures of communities changed. This may have set the scene for the evolution of flowering plants in the Triassic (~200 million years ago), which exploded in the Cretaceous and Tertiary. The latest major group of plants to evolve were the grasses, which became important in the mid Tertiary, from around 40 million years ago. The grasses, as well as many other groups, evolved new mechanisms of metabolism to survive the low CO₂ and warm, dry conditions of the tropics over the last 10 million years.

18.1 Embryophytes

The plants that are likely most familiar to us are the multicellular land plants, called embryophytes. Embryophytes include the vascular plants, such as ferns, conifers and flowering plants. They also include the bryophytes, of which mosses and liverworts are the most common.

All of these plants have eukaryotic cells with cell walls composed of cellulose, and most obtain their energy through photosynthesis, using light, water and carbon dioxide to synthesize food. About three hundred plant species do not photosynthesize but are parasites on other species of photosynthetic plants. Embryophytes are distinguished from green algae, which represent a mode of photosynthetic life similar to the kind modern plants are believed to have evolved from, by having specialized reproductive organs protected by non-reproductive tissues.

Bryophytes first appeared during the early Paleozoic. They mainly live in habitats where moisture is available for significant periods, although some species, such as *Targionia*, are desiccation-tolerant. Most species of bryophytes remain small throughout their life-cycle. This involves an alternation between two generations: a haploid stage, called the gametophyte, and a diploid stage, called the sporophyte. In bryophytes, the sporophyte is always unbranched and remains nutritionally dependent on its parent gametophyte. The embryophytes have the ability to secrete a cuticle on their outer surface, a waxy layer that confers resistant to desiccation. In the mosses and hornworts a cuticle is usually only produced on the sporophyte. Stomata are absent from liverworts, but occur on the sporangia of mosses and hornworts, allowing gas exchange.

Vascular plants first appeared during the Silurian period, and by the Devonian had diversified and spread into many different terrestrial environments. They developed a number of adaptations that allowed them to spread into increasingly more arid places, notably the vascular tissues xylem and phloem, that transport water and food throughout the organism. Root systems capable of obtaining soil water and nutrients also evolved during the Devonian. In modern vascular plants, the sporophyte is typically large, branched, nutritionally independent and long-lived, but there is increasing evidence that Paleozoic gametophytes were just as complex as the sporophytes. The gametophytes of all vascular plant groups evolved to become reduced in size and prominence in the life cycle.

In seed plants, the microgametophyte is reduced from a multicellular free-living organism to a few cells in a pollen grain and the miniaturised megagametophyte remains inside the megasporangium, attached to and dependent on the parent plant. A megasporangium enclosed in a protective layer called an integument is known as an ovule. After fertilisation by means of sperm produced by pollen grains, an embryo sporophyte develops inside the ovule. The integument becomes a seed coat, and the ovule develops into a seed. Seed plants can survive and reproduce in extremely arid conditions, because they are

not dependent on free water for the movement of sperm, or the development of free living gametophytes.

The first seed plants, pteridosperms (seed ferns), now extinct, appeared in the Devonian and diversified through the Carboniferous. They were the ancestors of modern gymnosperms, of which four surviving groups are widespread today, particularly the conifers, which are dominant trees in several biomes. The name gymnosperm comes from the Greek composite word γυμνόσπερμος (γυμνός *gymnos*, “naked” and σπέρμα *sperma*, “seed”), as the ovules and subsequent seeds are not enclosed in a protective structure (carpels or fruit), but are borne naked, typically on cone scales.

Plant fossils include roots, wood, leaves, seeds, fruit, pollen, spores, phytoliths, and amber (the fossilized resin produced by some plants). Fossil land plants are recorded in terrestrial, lacustrine, fluvial and nearshore marine sediments. Pollen, spores and algae (dinoflagellates and acritarchs) are used for dating sedimentary rock sequences. The remains of fossil plants are not as common as fossil animals, although plant fossils are locally abundant in many regions worldwide.

The earliest fossils clearly assignable to Kingdom Plantae are fossil green algae from the Cambrian. These fossils resemble calcified multicellular members of the Dasycladales. Earlier Precambrian fossils are known that resemble single-cell green algae, but definitive identity with that group of algae is uncertain.

The earliest fossils attributed to green algae date from the Precambrian (ca. 1200 mya). The resistant outer walls of prasinophyte cysts (known as phycomata) are well preserved in fossil deposits of the Paleozoic (ca. 250–540 mya). A filamentous fossil (*Proterocladius*) from middle Neoproterozoic deposits (ca. 750 mya) has been attributed to the Cladophorales, while the oldest reliable records of the Bryopsidales, Dasycladales) and stoneworts are from the Paleozoic.

The oldest known fossils of embryophytes date from the Ordovician, though such fossils are fragmentary. By the Silurian, fossils of whole plants are preserved, including the simple vascular plant *Cooksonia* in mid-Silurian and the much larger and more complex lycophyte *Baragwanathia longifolia* in late Silurian. From the early Devonian Rhynie chert, detailed fossils of lycophytes and rhyniophytes have been found that show details of the individual cells within the plant organs and the symbiotic association of these plants with fungi of the order Glomales. The Devonian period also saw the evolution of leaves and roots, and the first modern tree, *Archaeopteris*. This tree with fern-like foliage and a trunk with conifer-like wood was heterosporous producing spores of two different sizes, an

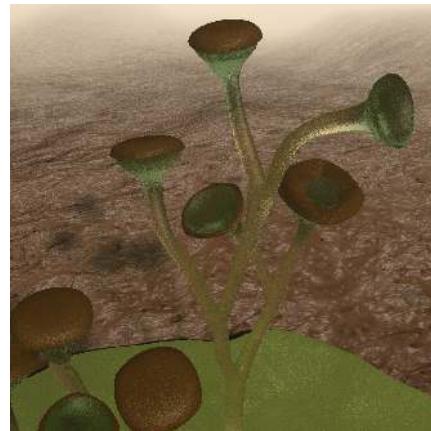


Figure 18.2: Reconstruction of *Cooksonia*, a vascular plant from the Silurian²

early step in the evolution of seeds.

The Coal measures are a major source of Paleozoic plant fossils, with many groups of plants in existence at this time. The spoil heaps of coal mines are the best places to collect; coal itself is the remains of fossilised plants, though structural detail of the plant fossils is rarely visible in coal. In the Fossil Grove at Victoria Park in Glasgow, Scotland, the stumps of *Lepidodendron* trees are found in their original growth positions.

The fossilized remains of conifer and angiosperm roots, stems and branches may be locally abundant in lake and inshore sedimentary rocks from the Mesozoic and Cenozoic eras. *Sequoia* and its allies, magnolia, oak, and palms are often found.

Petrified wood is common in some parts of the world, and is most frequently found in arid or desert areas where it is more readily exposed by erosion. Petrified wood is often heavily silicified (the organic material replaced by silicon dioxide), and the impregnated tissue is often preserved in fine detail. Such specimens may be cut and polished using lapidary equipment. Fossil forests of petrified wood have been found in all continents.

Fossils of seed ferns such as *Glossopteris* are widely distributed throughout several continents of the Southern Hemisphere, a fact that gave support to Alfred Wegener's early ideas regarding Continental drift theory.

18.2 Non-vascular Plants

Bryophytes are an informal group consisting of three divisions of non-vascular land plants (embryophytes): the liverworts, hornworts and mosses. They are characteristically limited in size and prefer moist habitats although they can



Figure 18.3: *Marchantia polymorpha*, with antheridial and archegonial stalks.³



Figure 18.4: Hornworts are bryophytes that are believed to be the closest living relatives of the vascular plants.⁴

survive in drier environments. The bryophytes consist of about 20,000 plant species. Bryophytes produce enclosed reproductive structures (gametangia and sporangia), but they do not produce flowers or seeds. They reproduce via spores. Bryophytes are usually considered to be a paraphyletic group and not a monophyletic group, although some studies have produced contrary results. Regardless of their status, the name is convenient and remains in use as an informal collective term. The term “bryophyte” comes from Greek βρύον, bryon “tree-moss, oyster-green” and φυτόν, phyton “plant”.



Figure 18.5: A moss showing both gametophytes (the low, leaf-like forms) and sporophytes (the tall, stalk-like forms).

The defining features of bryophytes are:

- Their life cycles are dominated by the gametophyte stage
- Their sporophytes are unbranched
- They do not have a true vascular tissue containing lignin (although some have specialized tissues for the transport of water)

Bryophytes exist in a wide variety of habitats. They can be found growing in a range of temperatures (cold arctics and in hot deserts), elevations (sea-level to alpine), and moisture (dry deserts to wet rainforests).

Bryophytes can grow where vascularized plants cannot because they do not depend on roots for an uptake of nutrients from soil. Bryophytes can survive on rocks and bare soil.

Traditionally, all living land plants without vascular tissues were classified in a single taxonomic group, often a division (or phylum). More recently, phylogenetic research has questioned whether the bryophytes form a monophyletic group and thus whether they should form a single taxon. Although a 2005 study supported the traditional view that the bryophytes form a monophyletic group, by 2010 a broad consensus had emerged among systematists that bryophytes as a whole are not a natural group (i.e., are paraphyletic), although each of the three extant (living) groups is monophyletic.

The three bryophyte clades (which may be treated as divisions) are the Marchantiophyta (liverworts), Bryophyta (mosses) and Anthocerotophyta (hornworts). The vascular plants or tracheophytes form a fourth, unranked clade of land plants called the “Polysporangiophyta”. In this analysis, hornworts are sister to vascular plants and liverworts are sister to all other land plants, including the hornworts and mosses. Phylogenetic studies continue to produce conflicting results. In particular those based on gene sequences suggest the bryophytes are paraphyletic, whereas those based on the amino acid translations of the same genes suggest they are monophyletic. A 2014 study concluded that composition biases were responsible for these differences and that the bryophytes are monophyletic. The issue remains unresolved.

18.3 Vascular Plants

Vascular plants (from Latin *vasculum*: duct), also known as tracheophytes (from the equivalent Greek term *trachea*), form a large group of plants (c. 308,312 accepted known species) that are defined as land plants that have lignified tissues (the xylem) for conducting water and minerals throughout the plant. They also have a specialized non-lignified tissue (the phloem) to conduct products of

photosynthesis. Vascular plants include the clubmosses, horsetails, ferns, gymnosperms (including conifers) and angiosperms (flowering plants). Scientific names for the group include Tracheophyta,:251 Tracheobionta and Equisetopsida sensu lato. Some early land plants (the rhyniophytes) had less developed vascular tissue; the term eutracheophyte has been used for all other vascular plants.

Botanists define vascular plants by three primary characteristics:

- Vascular plants have vascular tissues which distribute resources through the plant. Two kinds of vascular tissue occur in plants: xylem and phloem. Phloem and xylem are closely associated with one another and are typically located immediately adjacent to each other in the plant. The combination of one xylem and one phloem strand adjacent to each other is known as a vascular bundle. The evolution of vascular tissue in plants allowed them to evolve to larger sizes than non-vascular plants, which lack these specialized conducting tissues and are thereby restricted to relatively small sizes.
- In vascular plants, the principal generation phase is the sporophyte, which produces spores and is diploid (having two sets of chromosomes per cell). (By contrast, the principal generation phase in non-vascular plants is the gametophyte, which produces gametes and is haploid – with one set of chromosomes per cell.)
- Vascular plants have true roots, leaves, and stems, even if some groups have secondarily lost one or more of these traits.

Cavalier-Smith (1998) treated the Tracheophyta as a phylum or botanical division encompassing two of these characteristics defined by the Latin phrase “*facies diploida xylem et phloem instructa*” (diploid phase with xylem and phloem).:251

One possible mechanism for the presumed evolution from emphasis on haploid generation to emphasis on diploid generation is the greater efficiency in spore dispersal with more complex diploid structures. Elaboration of the spore stalk enabled the production of more spores and the development of the ability to release them higher and to broadcast them farther. Such developments may include more photosynthetic area for the spore-bearing structure, the ability to grow independent roots, woody structure for support, and more branching.

Water and nutrients in the form of inorganic solutes are drawn up from the soil by the roots and transported throughout the plant by the xylem. Organic compounds such as sucrose produced by photosynthesis in leaves are distributed by the phloem sieve tube elements.

The xylem consists of vessels in flowering plants and tracheids in other vascular plants, which are dead hard-walled hollow cells arranged to form files of tubes that function in water transport. A tracheid cell wall usually contains the polymer lignin. The phloem, however, consists of living cells called sieve-tube members. Between the sieve-tube members are sieve plates, which have pores to allow molecules to pass through. Sieve-tube members lack such organs as nuclei or ribosomes, but cells next to them, the companion cells, function to keep the sieve-tube members alive.

The most abundant compound in all plants, as in all cellular organisms, is water, which serves an important structural role and a vital role in plant metabolism. Transpiration is the main process of water movement within plant tissues. Water is constantly transpired from the plant through its stomata to the atmosphere and replaced by soil water taken up by the roots. The movement of water out of the leaf stomata creates a transpiration pull or tension in the water column in the xylem vessels or tracheids. The pull is the result of water surface tension within the cell walls of the mesophyll cells, from the surfaces of which evaporation takes place when the stomata are open. Hydrogen bonds exist between water molecules, causing them to line up; as the molecules at the top of the plant evaporate, each pulls the next one up to replace it, which in turn pulls on the next one in line. The draw of water upwards may be entirely passive and can be assisted by the movement of water into the roots via osmosis. Consequently, transpiration requires very little energy to be used by the plant. Transpiration assists the plant in absorbing nutrients from the soil as soluble salts.

Living root cells passively absorb water in the absence of transpiration pull via osmosis creating root pressure. It is possible for there to be no evapotranspiration and therefore no pull of water towards the shoots and leaves. This is usually due to high temperatures, high humidity, darkness or drought.

Xylem and phloem tissues are involved in the conduction processes within plants. Sugars are conducted throughout the plant in the phloem, water and other nutrients through the xylem. Conduction occurs from a source to a sink for each separate nutrient. Sugars are produced in the leaves (a source) by photosynthesis and transported to the growing shoots and roots (sinks) for use in growth, cellular respiration or storage. Minerals are absorbed in the roots (a source) and transported to the shoots to allow cell division and growth.

18.3.1 Ferns

A fern (Polypodiopsida or Polypodiophyta) is a member of a group of vascular plants (plants with xylem and phloem)

that reproduce via spores and have neither seeds nor flowers. They differ from mosses by being vascular, i.e., having specialized tissues that conduct water and nutrients and in having life cycles in which the sporophyte is the dominant phase. Ferns have complex leaves called megaphylls, that are more complex than the microphylls of clubmosses. Most ferns are leptosporangiate ferns. They produce coiled fiddleheads that uncoil and expand into fronds. The group includes about 10,560 known extant species. Ferns are defined here in the broad sense, being all of the Polypodiopsida, comprising both the leptosporangiate (Polypodiidae) and eusporangiate ferns, the latter group including horse-tails or scouring rushes, whisk ferns, marattioid ferns, and ophioglossoid ferns.

Ferns first appear in the fossil record about 360 million years ago in the middle Devonian period, but many of the current families and species did not appear until roughly 145 million years ago in the early Cretaceous, after flowering plants came to dominate many environments. The fern *Osmunda claytoniana* is a paramount example of evolutionary stasis; paleontological evidence indicates it has remained unchanged, even at the level of fossilized nuclei and chromosomes, for at least 180 million years.

Ferns are not of major economic importance, but some are used for food, medicine, as biofertilizer, as ornamental plants and for remediating contaminated soil. They have been the subject of research for their ability to remove some chemical pollutants from the atmosphere. Some fern species, such as bracken (*Pteridium aquilinum*) and water fern (*Azolla filiculoides*) are significant weeds worldwide. Some fern genera, such as *Azolla*, can fix nitrogen and make a significant input to the nitrogen nutrition of rice paddies. They also play certain roles in folklore.

Like the sporophytes of seed plants, those of ferns consist of stems, leaves and roots. Ferns differ from seed plants in reproducing by spores and from bryophytes in that, like seed plants, they are Polysporangiophytes, their sporophytes branching and producing many sporangia. Unlike bryophytes, fern sporophytes are free-living and only briefly dependent on the maternal gametophyte.

Stems: Fern stems are often referred to as rhizomes, even though they grow underground only in some of the species. Epiphytic species and many of the terrestrial ones have above-ground creeping stolons (e.g., Polypodiaceae), and many groups have above-ground erect semi-woody trunks (e.g., Cyatheaceae). These can reach up to 20 meters (66 ft) tall in a few species (e.g., *Cyathea brownii* on Norfolk Island and *Cyathea medullaris* in New Zealand).

Leaf: The green, photosynthetic part of the plant is technically a megaphyll and in ferns, it is often referred to as a frond. New leaves typically expand by the unrolling of

a tight spiral called a crozier or fiddlehead into fronds. This uncurling of the leaf is termed circinate vernation. Leaves are divided into two types a trophophyll and a sporophyll. A trophophyll frond is a vegetative leaf analogous to the typical green leaves of seed plants that does not produce spores, instead only producing sugars by photosynthesis. A sporophyll frond is a fertile leaf that produces spores borne in sporangia that are usually clustered to form sori. In most ferns, fertile leaves are morphologically very similar to the sterile ones, and they photosynthesize in the same way. In some groups, the fertile leaves are much narrower than the sterile leaves, and may even have no green tissue at all (e.g., Blechnaceae, Lomariopsidaceae). The anatomy of fern leaves can either be simple or highly divided. In tree ferns, the main stalk that connects the leaf to the stem (known as the stipe), often has multiple leaflets. The leafy structures that grow from the stipe are known as pinnae and are often again divided into smaller pinnules.

Roots: The underground non-photosynthetic structures that take up water and nutrients from soil. They are always fibrous and structurally are very similar to the roots of seed plants.

Like all other vascular plants, the diploid sporophyte is the dominant phase or generation in the life cycle. The gametophytes of ferns, however, are very different from those of seed plants. They are free-living and resemble liverworts, whereas those of seed plants develop within the spore wall and are dependent on the parent sporophyte for their nutrition. A fern gametophyte typically consists of:

Prothallus: A green, photosynthetic structure that is one cell thick, usually heart or kidney shaped, 3–10 mm long and 2–8 mm broad. The prothallus produces gametes by means of: **Antheridia:** Small spherical structures that produce flagellate sperm. **Archegonia:** A flask-shaped structure that produces a single egg at the bottom, reached by the sperm by swimming down the neck. **Rhizoids:** root-like structures (not true roots) that consist of single greatly elongated cells, that absorb water and mineral salts over the whole structure. Rhizoids anchor the prothallus to the soil.

Ferns are widespread in their distribution, with the greatest abundance in the tropics, and least in arctic areas. The greatest diversity occurs in tropical rainforests. New Zealand, for which the fern is a symbol, has about 230 species, distributed throughout the country.

The stereotypical image of ferns growing in moist shady woodland nooks is far from a complete picture of the habitats where ferns can be found growing. Fern species live in a wide variety of habitats, from remote mountain elevations, to dry desert rock faces, to bodies of water or in open fields. Ferns in general may be thought

of as largely being specialists in marginal habitats, often succeeding in places where various environmental factors limit the success of flowering plants. Some ferns are among the world's most serious weed species, including the bracken fern growing in the Scottish highlands, or the mosquito fern (*Azolla*) growing in tropical lakes, both species forming large aggressively spreading colonies. There are four particular types of habitats that ferns are found in: moist, shady forests; crevices in rock faces, especially when sheltered from the full sun; acid wetlands including bogs and swamps; and tropical trees, where many species are epiphytes (something like a quarter to a third of all fern species).

Especially the epiphytic ferns have turned out to be hosts of a huge diversity of invertebrates. It is assumed that bird's-nest ferns alone contain up to half the invertebrate biomass within a hectare of rainforest canopy.

Many ferns depend on associations with mycorrhizal fungi. Many ferns grow only within specific pH ranges; for instance, the climbing fern (*Lygodium palmatum*) of eastern North America will grow only in moist, intensely acid soils, while the bulblet bladder fern (*Cystopteris bulbifera*), with an overlapping range, is found only on limestone.

The spores are rich in lipids, protein and calories, so some vertebrates eat these. The European woodmouse (*Apodemus sylvaticus*) has been found to eat the spores of *Culcita macrocarpa* and the bullfinch (*Pyrrhula murina*) and the New Zealand lesser short-tailed bat (*Mystacinia tuberculata*) also eat fern spores.

18.4 Seed Plants

The spermatophytes, also known as phanerogams (taxon *Phanerogamae*) or phaeogams (taxon *Phaenogamae*), comprise those plants that produce seeds, hence the alternative name seed plants. They are a subset of the embryophytes or land plants. The term phanerogams or *phanerogamae* is derived from the Greek φανέρως, *phanerós* meaning "visible", in contrast to the cryptogamae from Greek κρυπτός *kryptós* = "hidden" together with the suffix γάμεω, *gameo*, "to marry". These terms distinguished those plants with hidden sexual organs (cryptogamae) from those with visible sexual organs (phanerogamae).

The extant spermatophytes form five divisions, the first four of which are traditionally grouped as gymnosperms, plants that have unenclosed, "naked seeds":

- Cycadophyta, the cycads, a subtropical and tropical group of plants,
- Ginkgophyta, which includes a single living species of tree in the genus *Ginkgo*,



Figure 18.6: Microimages of seeds of various plants.⁵ The first row: Poppy, Red pepper, Strawberry, Apple tree, Blackberry, Rice, Carum. Second row: Mustard, Eggplant, Physalis, grapes, raspberries, red rice, Patchouli. The third row: Figs, Lycium barbarum, Beets, Blueberries, Golden Kiwifruit, Rosehip, Basil. The fourth row: Pink pepper, Tomato, Radish, Carrot, Matthiola, Dill, Coriander. Fifth row: Black pepper, White cabbage, Napa cabbage, Seabuckthorn, Parsley, Dandelion, Capsella bursa-pastoris. The sixth row: Cauliflower, Radish, Kiwifruit, Grenadilla, Passion fruit, Melissa, Tagetes erecta.

- Pinophyta, the conifers, which are cone-bearing trees and shrubs,
- and Gnethophyta, the gnethophytes, various woody plants in the relict genera *Ephedra*, *Gnetum*, and *Welwitschia*.

The fifth extant division is the flowering plants, also known as angiosperms or magnoliophytes, the largest and most diverse group of spermatophytes. Angiosperms possess seeds enclosed in a fruit, unlike gymnosperms.

In addition to the taxa listed above, the fossil record contains evidence of many extinct taxa of seed plants. The so-called "seed ferns" (*Pteridospermae*) were one of the earliest successful groups of land plants, and forests dominated by seed ferns were prevalent in the late Paleozoic. *Glossopteris* was the most prominent tree genus in the ancient southern supercontinent of Gondwana during the Permian period. By the Triassic period, seed ferns had declined in ecological importance, and representatives of modern gymnosperm groups were abundant and dominant through the end of the Cretaceous, when angiosperms radiated.

18.5 Gymnosperms

The gymnosperms, also known as *Acrogymnospermae*, are a group of seed-producing plants that includes conifers, cycads, Ginkgo, and gnethophytes. The term



Figure 18.7: Various gymnosperms Left 1 *Welwitschia mirabilis* 2 *Cycas revoluta* 3 *Taxus baccata* 4 *Ginkgo biloba* RIGHT 1 *Cupressus sempervirens* 2 *Sequoiadendron giganteum* 3 *Dammara orientalis* 4 *Araucaria heterophylla*⁶

"gymnosperm" comes from the composite word in Greek: γυμνόσπερμος (γυμνός, *gymnos*, 'naked' and σπέρμα, *sperma*, 'seed'), literally meaning "naked seeds". The name is based on the unenclosed condition of their seeds (called ovules in their unfertilized state). The non-encased condition of their seeds contrasts with the seeds and ovules of flowering plants (angiosperms), which are enclosed within an ovary. Gymnosperm seeds develop either on the surface of scales or leaves, which are often modified to form cones, or solitary as in yew, *Torreya*, *Ginkgo*.

The gymnosperms and angiosperms together compose the spermatophytes or seed plants. The gymnosperms are divided into six phyla. Organisms that belong to the Cycadophyta, Ginkgophyta, Gnetaophyta, and Pinophyta (also known as Coniferophyta) phyla are still in existence while those in the Pteridospermales and Cordaitales phyla are now extinct.

By far the largest group of living gymnosperms are the conifers (pines, cypresses, and relatives), followed by cycads, gnetaophytes (*Gnetum*, *Ephedra* and *Welwitschia*), and *Ginkgo biloba* (a single living species).

Some genera have mycorrhiza, fungal associations with roots (*Pinus*), while in some others (*Cycas*) small specialised roots called coraloid roots are associated with nitrogen-fixing cyanobacteria.

There are over 1000 living species of gymnosperm. It is widely accepted that the gymnosperms originated in the late Carboniferous period, replacing the lycopid rainforests of the tropical region. This appears to have been the result of a whole genome duplication event around 319 million years ago. Early characteristics of seed plants were evident in fossil progymnosperms of the late Devonian period around 383 million years ago. It has been suggested that during the mid-Mesozoic era, pollination

of some extinct groups of gymnosperms was by extinct species of scorpionflies that had specialized proboscis for feeding on pollination drops. The scorpionflies likely engaged in pollination mutualisms with gymnosperms, long before the similar and independent coevolution of nectar-feeding insects on angiosperms. Evidence has also been found that mid-Mesozoic gymnosperms were pollinated by Kalligrammatid lacewings, a now-extinct genus with members which (in an example of convergent evolution) resembled the modern butterflies that arose far later.

Conifers are by far the most abundant extant group of gymnosperms with six to eight families, with a total of 65–70 genera and 600–630 species (696 accepted names). Conifers are woody plants and most are evergreens. The leaves of many conifers are long, thin and needle-like, other species, including most Cupressaceae and some Podocarpaceae, have flat, triangular scale-like leaves. *Agathis* in Araucariaceae and *Nageia* in Podocarpaceae have broad, flat strap-shaped leaves.

Cycads are the next most abundant group of gymnosperms, with two or three families, 11 genera, and approximately 338 species. A majority of cycads are native to tropical climates and are most abundantly found in regions near the equator. The other extant groups are the 95–100 species of Gnetales and one species of *Ginkgo*.

Gymnosperms, like all vascular plants, have a sporophyte-dominant life cycle, which means they spend most of their life cycle with diploid cells, while the gametophyte (gamete-bearing phase) is relatively short-lived. Two spore types, microspores and megaspores, are typically produced in pollen cones or ovulate cones, respectively. Gametophytes, as with all heterosporous plants, develop within the spore wall. Pollen grains (microgametophytes) mature from microspores, and ultimately produce sperm cells. Megagametophytes develop from megaspores and are retained within the ovule. Gymnosperms produce multiple archegonia, which produce the female gamete. During pollination, pollen grains are physically transferred between plants from the pollen cone to the ovule. Pollen is usually moved by wind or insects. Whole grains enter each ovule through a microscopic gap in the ovule coat (integument) called the micropyle. The pollen grains mature further inside the ovule and produce sperm cells. Two main modes of fertilization are found in gymnosperms. Cycads and *Ginkgo* have motile sperm that swim directly to the egg inside the ovule, whereas conifers and gnetaophytes have sperm with no flagella that are moved along a pollen tube to the egg. After syngamy (joining of the sperm and egg cell), the zygote develops into an embryo (young



Figure 18.8: Flowers of different families:⁷ St Bernards Lilly (*Anthericum liliago*): Liliaceae Bermuda Buttercup (*Oxalis pes-caprae*): Oxalidaceae Oleander (*Nerium oleander*): Apocynaceae Lantana (*Lantana camara*): Verbenaceae Scarlet Pimpernel (*Anagallis arvensis*): Primulaceae Verbascum (*Verbascum sinuatum*): Scrophulariaceae Common Mallow (*Malva sylvestris*): Malvaceae Spanish Oyster (*Scolymus hispanicus*): Asteraceae Stork's Bill (*Erodium malacoides*): Geraniaceae Bindweed (*Convolvulus arvensis*): Convolvulaceae Blue Gem (*Hebe x franciscana*): Plantaginaceae Calla Lily (*Zantedeschia aethiopica*): Araceae

sporophyte). More than one embryo is usually initiated in each gymnosperm seed. The mature seed comprises the embryo and the remains of the female gametophyte, which serves as a food supply, and the seed coat.

18.6 Angiosperms

The flowering plants, also known as Angiospermae, or Magnoliophyta, are the most diverse group of land plants, with 64 orders, 416 families, approximately 13,000 known genera and 300,000 known species. Like gymnosperms, angiosperms are seed-producing plants. They are distinguished from gymnosperms by characteristics including flowers, endosperm within the seeds, and the production of fruits that contain the seeds. Etymologically, angiosperm means a plant that produces seeds within an enclosure; in other words, a fruiting plant. The term comes from the Greek words *angeion* ("case" or "casing") and *sperma* ("seed").

The ancestors of flowering plants diverged from gymnosperms in the Triassic Period, 245 to 202 million years ago (mya), and the first flowering plants are known from

~140 mya. They diversified extensively during the Early Cretaceous, became widespread by 120 mya, and replaced conifers as the dominant trees from 100 to 60 mya.

Angiosperms differ from other seed plants in several ways, described in the table below. These distinguishing characteristics taken together have made the angiosperms the most diverse and numerous land plants and the most commercially important group to humans.

Table 18.2: Distinctive features of angiosperms.

Feature	Description
Flowering organs	Flowers, the reproductive organs of flowering plants, are the most remarkable feature distinguishing them from the other seed plants. Flowers provided angiosperms with the means to have a more species-specific breeding system, and hence a way to evolve more readily into different species without the risk of crossing back with related species. Faster speciation enabled the Angiosperms to adapt to a wider range of ecological niches. This has allowed flowering plants to largely dominate terrestrial ecosystems. [citation needed]
Stamens with two pairs of pollen sacs	Stamens are much lighter than the corresponding organs of gymnosperms and have contributed to the diversification of angiosperms through time with adaptations to specialized pollination syndromes, such as particular pollinators. Stamens have also become modified through time to prevent self-fertilization, which has permitted further diversification, allowing angiosperms eventually to fill more niches.
Reduced male parts, three cells	The male gametophyte in angiosperms is significantly reduced in size compared to those of gymnosperm seed plants. The smaller size of the pollen reduces the amount of time between pollination — the pollen grain reaching the female plant — and fertilization. In gymnosperms, fertilization can occur up to a year after pollination, whereas in angiosperms, fertilization begins very soon after pollination. The shorter amount of time between pollination and fertilization allows angiosperms to produce seeds earlier after pollination than gymnosperms, providing angiosperms a distinct evolutionary advantage.
Closed carpel enclosing the ovules (carpel or carpels and accessory parts may become the fruit)	The closed carpel of angiosperms also allows adaptations to specialized pollination syndromes and controls. This helps to prevent self-fertilization, thereby maintaining increased diversity. Once the ovary is fertilized, the carpel and some surrounding tissues develop into a fruit. This fruit often serves as an attractant to seed-dispersing animals. The resulting cooperative relationship presents another advantage to angiosperms in the process of dispersal.
Reduced female gametophyte, seven cells with eight nuclei	The reduced female gametophyte, like the reduced male gametophyte, may be an adaptation allowing for more rapid seed set, eventually leading to such flowering plant adaptations as annual herbaceous life-cycles, allowing the flowering plants to fill even more niches.
Endosperm	In general, endosperm formation begins after fertilization and before the first division of the zygote. Endosperm is a highly nutritive tissue that can provide food for the developing embryo, the cotyledons, and sometimes the seedling when it first appears.

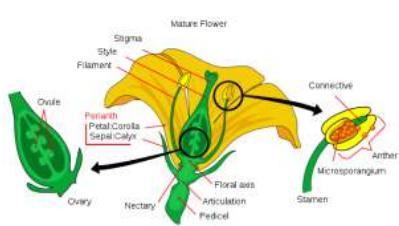


Figure 18.9: Anatomy of the flower.⁸

Angiosperm stems are made up of seven layers as shown on the right. The amount and complexity of tissue-formation in flowering plants exceeds that of gymnosperms. The vascular bundles of the stem are arranged such that the xylem and phloem form concentric rings.

In the dicotyledons, the bundles in the very young stem are arranged in an open ring, separating a central pith from an outer cortex. In each bundle, separating the xylem and phloem, is a layer of meristem or active formative tissue known as cambium. By the formation of a layer of cambium between the bundles (interfascicular cambium), a complete ring is formed, and a regular periodical increase in thickness results from the development of xylem on the inside and phloem on the outside. The soft phloem becomes crushed, but the hard wood persists and forms the bulk of the stem and branches of the woody perennial. Owing to differences in the character of the elements produced at the beginning and end of the season, the wood is marked out in transverse section into concentric rings, one for each season of growth, called annual rings.

Among the monocotyledons, the bundles are more numerous in the young stem and are scattered through the ground tissue. They contain no cambium and once formed the stem increases in diameter only in exceptional cases.

The characteristic feature of angiosperms is the flower. Flowers show remarkable variation in form and elaboration, and provide the most trustworthy external characteristics for establishing relationships among angiosperm species. The function of the flower is to ensure fertilization of the ovule and development of fruit containing seeds. The floral apparatus may arise terminally on a shoot or from the axil of a leaf (where the petiole attaches to the stem). Occasionally, as in violets, a flower arises singly in the axil of an ordinary foliage-leaf. More typically, the flower-bearing portion of the plant is sharply distinguished from the foliage-bearing or vegetative portion, and forms a more or less elaborate branch-system called an inflorescence.

There are two kinds of reproductive cells produced by flowers. Microspores, which will divide to become pollen grains, are the "male" cells and are borne in the stamens (or

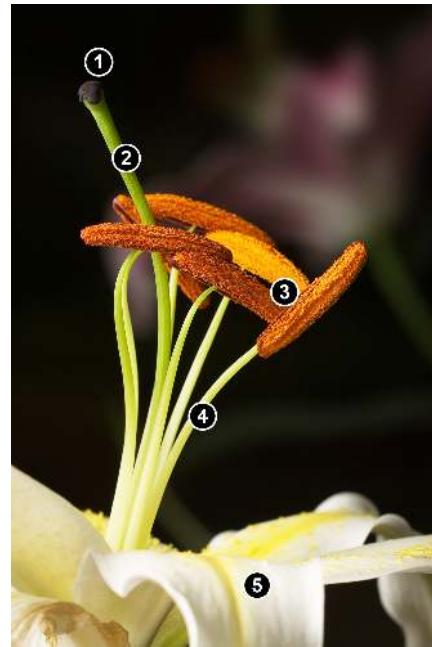


Figure 18.10: Reproductive parts of Easter Lily (*Lilium longiflorum*).⁹ 1. Stigma, 2. Style, 3. Stamens, 4. Filament, 5. Petal

microsporophylls). The "female" cells called megasporangia, which will divide to become the egg cell (megagametogenesis), are contained in the ovule and enclosed in the carpel (or megasporophyll).

The flower may consist only of these parts, as in willow, where each flower comprises only a few stamens or two carpels. Usually, other structures are present and serve to protect the sporophylls and to form an envelope attractive to pollinators. The individual members of these surrounding structures are known as sepals and petals (or tepals in flowers such as Magnolia where sepals and petals are not distinguishable from each other). The outer series (calyx of sepals) is usually green and leaf-like, and functions to protect the rest of the flower, especially the bud. The inner series (corolla of petals) is, in general, white or brightly colored, and is more delicate in structure. It functions to attract insect or bird pollinators. Attraction is effected by color, scent, and nectar, which may be secreted in some part of the flower. The characteristics that attract pollinators account for the popularity of flowers and flowering plants among humans.

While the majority of flowers are perfect or hermaphrodite (having both pollen and ovule producing parts in the same flower structure), flowering plants have developed numerous morphological and physiological mechanisms to reduce or prevent self-fertilization. Heteromorphic flowers have short carpels and long stamens,

or vice versa, so animal pollinators cannot easily transfer pollen to the pistil (receptive part of the carpel). Homomorphic flowers may employ a biochemical (physiological) mechanism called self-incompatibility to discriminate between self and non-self pollen grains. In other species, the male and female parts are morphologically separated, developing on different flowers.

The botanical term “Angiosperm”, from the Ancient Greek ἄγγεῖον, *angeíon* (bottle, vessel) and σπέρμα, *sperma* (seed), was coined in the form *Angiospermae* by Paul Hermann in 1690, as the name of one of his primary divisions of the plant kingdom. This included flowering plants possessing seeds enclosed in capsules, distinguished from his *Gymnospermae*, or flowering plants with achenial or schizo-carpic fruits, the whole fruit or each of its pieces being here regarded as a seed and naked. The term and its antonym were maintained by Carl Linnaeus with the same sense, but with restricted application, in the names of the orders of his class *Didynamia*. Its use with any approach to its modern scope became possible only after 1827, when Robert Brown established the existence of truly naked ovules in the Cycadeae and Coniferae, and applied to them the name *Gymnosperms*. From that time onward, as long as these *Gymnosperms* were, as was usual, reckoned as dicotyledonous flowering plants, the term *Angiosperm* was used antithetically by botanical writers, with varying scope, as a group-name for other dicotyledonous plants.

In 1851, Hofmeister discovered the changes occurring in the embryo-sac of flowering plants, and determined the correct relationships of these to the *Cryptogamia*. This fixed the position of *Gymnosperms* as a class distinct from Dicotyledons, and the term *Angiosperm* then gradually came to be accepted as the suitable designation for the whole of the flowering plants other than *Gymnosperms*, including the classes of Dicotyledons and Monocotyledons. This is the sense in which the term is used today.

In most taxonomies, the flowering plants are treated as a coherent group. The most popular descriptive name has been *Angiospermae* (*Angiosperms*), with *Anthophyta* (“flowering plants”) a second choice. These names are not linked to any rank. The Wettstein system and the Engler system use the name *Angiospermae*, at the assigned rank of subdivision. The Reveal system treated flowering plants as subdivision *Magnoliophytina*, but later split it to *Magnoliopsida*, *Liliopsida*, and *Rosopsida*. The Takhtajan system and Cronquist system treat this group at the rank of division, leading to the name *Magnoliophyta* (from the family name *Magnoliaceae*). The Dahlgren system and Thorne system (1992) treat this group at the rank of class, leading to the name *Magnoliopsida*. The APG system of 1998, and the later 2003 and 2009 revisions, treat the flower-

ing plants as a clade called angiosperms without a formal botanical name. A formal classification was published alongside the 2009 revision in which the flowering plants form the Subclass *Magnoliidae*.

The internal classification of this group has undergone considerable revision. The Cronquist system, proposed by Arthur Cronquist in 1968 and published in its full form in 1981, is still widely used but is no longer believed to accurately reflect phylogeny. A consensus about how the flowering plants should be arranged has recently begun to emerge through the work of the Angiosperm Phylogeny Group (APG), which published an influential reclassification of the angiosperms in 1998. Updates incorporating more recent research were published as the APG II system in 2003, the APG III system in 2009, and the APG IV system in 2016.

Traditionally, the flowering plants are divided into two groups,

- Dicotyledoneae or Magnoliopsida
- Monocotyledoneae or Liliopsida

which in the Cronquist system are called Magnoliopsida (at the rank of class, formed from the family name *Magnoliaceae*) and Liliopsida (at the rank of class, formed from the family name *Liliaceae*). Other descriptive names allowed by Article 16 of the ICBN include Dicotyledones or Dicotyledoneae, and Monocotyledones or Monocotyledoneae, which have a long history of use. In English a member of either group may be called a dicotyledon (plural dicotyledons) and monocotyledon (plural monocotyledons), or abbreviated, as dicot (plural dicots) and monocot (plural monocots). These names derive from the observation that the dicots most often have two cotyledons, or embryonic leaves, within each seed. The monocots usually have only one, but the rule is not absolute either way. From a broad diagnostic point of view, the number of cotyledons is neither a particularly handy, nor a reliable character.

Recent studies, as by the APG, show that the monocots form a monophyletic group (clade) but that the dicots do not (they are paraphyletic). Nevertheless, the majority of dicot species do form a monophyletic group, called the eudicots or tricolpates. Of the remaining dicot species, most belong to a third major clade known as the magnoliids, containing about 9,000 species. The rest include a paraphyletic grouping of early branching taxa known collectively as the basal angiosperms, plus the families Ceratophyllaceae and Chloranthaceae.

Fossilized spores suggest that land plants (embryophytes) have existed for at least 475 million years. Early land plants reproduced sexually with flagellated, swimming sperm, like the green algae from which they evolved. An adaptation to terrestrialization was the



Figure 18.11: A Bee orchid¹⁰ has evolved over many generations to better mimic a female bee to attract male bees as pollinators.

development of upright meiosporangia for dispersal by spores to new habitats. This feature is lacking in the descendants of their nearest algal relatives, the Charophycean green algae. A later terrestrial adaptation took place with retention of the delicate, avascular sexual stage, the gametophyte, within the tissues of the vascular sporophyte. This occurred by spore germination within sporangia rather than spore release, as in non-seed plants. A current example of how this might have happened can be seen in the precocious spore germination in *Selaginella*, the spike-moss. The result for the ancestors of angiosperms was enclosing them in a case, the seed.

The apparently sudden appearance of nearly modern flowers in the fossil record initially posed such a problem for the theory of evolution that Charles Darwin called it an “abominable mystery”. However, the fossil record has considerably grown since the time of Darwin, and recently discovered angiosperm fossils such as *Archae fructus*, along with further discoveries of fossil gymnosperms, suggest how angiosperm characteristics may have been acquired in a series of steps. Several groups of extinct gymnosperms, in particular seed ferns, have been proposed as the ancestors of flowering plants, but there is no continuous fossil evidence showing exactly how flowers evolved. Some older fossils, such as the upper Triassic *Sanmiguelia*, have been suggested.

The first seed bearing plants, like the ginkgo, and conifers (such as pines and firs), did not produce flowers. The pollen grains (male gametophytes) of Ginkgo and cycads produce a pair of flagellated, mobile sperm cells that “swim” down the developing pollen tube to the female

and her eggs.

Oleanane, a secondary metabolite produced by many flowering plants, has been found in Permian deposits of that age together with fossils of gigantopterids. Gigantopterids are a group of extinct seed plants that share many morphological traits with flowering plants, although they are not known to have been flowering plants themselves.

Based on current evidence, some propose that the ancestors of the angiosperms diverged from an unknown group of gymnosperms in the Triassic period (245–202 million years ago). Fossil angiosperm-like pollen from the Middle Triassic (247.2–242.0 Ma) suggests an older date for their origin. A close relationship between angiosperms and gnetophytes, proposed on the basis of morphological evidence, has more recently been disputed on the basis of molecular evidence that suggest gnetophytes are instead more closely related to other gymnosperms.

The fossil plant species *Nanjinganthus dendrostyla* from Early Jurassic China seems to share many exclusively angiosperm features, such as a thickened receptacle with ovules, and thus might represent a crown-group or a stem-group angiosperm. However, the interpretation of the structures in this fossils are highly contested.

The evolution of seed plants and later angiosperms appears to be the result of two distinct rounds of whole genome duplication events. These occurred at 319 million years ago and 192 million years ago. Another possible whole genome duplication event at 160 million years ago perhaps created the ancestral line that led to all modern flowering plants. That event was studied by sequencing the genome of an ancient flowering plant, *Amborella trichopoda*, and directly addresses Darwin’s “abominable mystery”.

One study has suggested that the early-middle Jurassic plant *Schmeissneria*, traditionally considered a type of ginkgo, may be the earliest known angiosperm, or at least a close relative.

It has been proposed that the swift rise of angiosperms to dominance was facilitated by a reduction in their genome size. During the early Cretaceous period, only angiosperms underwent rapid genome downsizing, while genome sizes of ferns and gymnosperms remained unchanged. Smaller genomes—and smaller nuclei—allow for faster rates of cell division and smaller cells. Thus, species with smaller genomes can pack more, smaller cells—in particular veins and stomata—into a given leaf volume. Genome downsizing therefore facilitated higher rates of leaf gas exchange (transpiration and photosynthesis) and faster rates of growth. This would have countered some of the negative physiological effects

of genome duplications, facilitated increased uptake of carbon dioxide despite concurrent declines in atmospheric CO₂ concentrations, and allowed the flowering plants to outcompete other land plants.

The earliest known macrofossil confidently identified as an angiosperm, *Archaefructus liaoningensis*, is dated to about 125 million years BP (the Cretaceous period), whereas pollen considered to be of angiosperm origin takes the fossil record back to about 130 million years BP, with *Montsechia* representing the earliest flower at that time. In 2018, scientists reported that the earliest flowers began about 180 million years ago, 50 million years earlier than thought earlier. Nonetheless, circumstantial chemical evidence has been found for the existence of angiosperms as early as 250 million years ago.

In 2013 flowers encased in amber were found and dated 100 million years before present. The amber had frozen the act of sexual reproduction in the process of taking place. Microscopic images showed tubes growing out of pollen and penetrating the flower's stigma. The pollen was sticky, suggesting it was carried by insects. In August 2017, scientists presented a detailed description and 3D model image of what the first flower possibly looked like, and presented the hypothesis that it may have lived about 140 million years ago. A Bayesian analysis of 52 angiosperm taxa suggested that the crown group of angiosperms evolved between 178 million years ago and 198 million years ago.

Recent DNA analysis based on molecular systematics showed that *Amborella trichopoda*, found on the Pacific island of New Caledonia, belongs to a sister group of the other flowering plants, and morphological studies suggest that it has features that may have been characteristic of the earliest flowering plants. The orders Amborellales, Nymphaeales, and Austrobaileyales diverged as separate lineages from the remaining angiosperm clade at a very early stage in flowering plant evolution.

The great angiosperm radiation, when a great diversity of angiosperms appears in the fossil record, occurred in the mid-Cretaceous (approximately 100 million years ago). However, a study in 2007 estimated that the division of the five most recent (the genus *Ceratophyllum*, the family Chloranthaceae, the eudicots, the magnoliids, and the monocots) of the eight main groups occurred around 140 million years ago. It is generally assumed that the function of flowers, from the start, was to involve mobile animals in their reproduction processes. That is, pollen can be scattered even if the flower is not brightly colored or oddly shaped in a way that attracts animals; however, by expending the energy required to create such traits, angiosperms can enlist the aid of animals and, thus, reproduce more efficiently.

Island genetics provides one proposed explanation for the sudden, fully developed appearance of flowering plants. Island genetics is believed to be a common source of speciation in general, especially when it comes to radical adaptations that seem to have required inferior transitional forms. Flowering plants may have evolved in an isolated setting like an island or island chain, where the plants bearing them were able to develop a highly specialized relationship with some specific animal (a wasp, for example). Such a relationship, with a hypothetical wasp carrying pollen from one plant to another much the way fig wasps do today, could result in the development of a high degree of specialization in both the plant(s) and their partners. Note that the wasp example is not incidental; bees, which, it is postulated, evolved specifically due to mutualistic plant relationships, are descended from wasps.

Animals are also involved in the distribution of seeds. Fruit, which is formed by the enlargement of flower parts, is frequently a seed-dispersal tool that attracts animals to eat or otherwise disturb it, incidentally scattering the seeds it contains (see frugivory). Although many such mutualistic relationships remain too fragile to survive competition and to spread widely, flowering proved to be an unusually effective means of reproduction, spreading (whatever its origin) to become the dominant form of land plant life.

Flower ontogeny uses a combination of genes normally responsible for forming new shoots. The most primitive flowers probably had a variable number of flower parts, often separate from (but in contact with) each other. The flowers tended to grow in a spiral pattern, to be bisexual (in plants, this means both male and female parts on the same flower), and to be dominated by the ovary (female part). As flowers evolved, some variations developed parts fused together, with a much more specific number and design, and with either specific sexes per flower or plant or at least "ovary-inferior". Flower evolution continues to the present day; modern flowers have been so profoundly influenced by humans that some of them cannot be pollinated in nature. Many modern domesticated flower species were formerly simple weeds, which sprouted only when the ground was disturbed. Some of them tended to grow with human crops, perhaps already having symbiotic companion plant relationships with them, and the prettiest did not get plucked because of their beauty, developing a dependence upon and special adaptation to human affection.

A few paleontologists have also proposed that flowering plants, or angiosperms, might have evolved due to interactions with dinosaurs. One of the idea's strongest proponents is Robert T. Bakker. He proposes that herbivorous dinosaurs, with their eating habits, provided a selective pressure on plants, for which adaptations either succeeded in deterring or coping with predation by herbivores.

By the late Cretaceous, angiosperms appear to have dominated environments formerly occupied by ferns and cycadophytes, but large canopy-forming trees replaced conifers as the dominant trees only close to the end of the Cretaceous 66 million years ago or even later, at the beginning of the Tertiary. The radiation of herbaceous angiosperms occurred much later. Yet, many fossil plants recognizable as belonging to modern families (including beech, oak, maple, and magnolia) had already appeared by the late Cretaceous. Flowering plants appeared in Australia about 126 million years ago. This also pushed the age of ancient Australian vertebrates, in what was then a south polar continent, to 126–110 million years old.

18.6.1 Fertilization and Embryogenesis

Double fertilization refers to a process in which two sperm cells fertilize cells in the ovule. This process begins when a pollen grain adheres to the stigma of the pistil (female reproductive structure), germinates, and grows a long pollen tube. While this pollen tube is growing, a haploid generative cell travels down the tube behind the tube nucleus. The generative cell divides by mitosis to produce two haploid (n) sperm cells. As the pollen tube grows, it makes its way from the stigma, down the style and into the ovary. Here the pollen tube reaches the micropyle of the ovule and digests its way into one of the synergids, releasing its contents (which include the sperm cells). The synergid that the cells were released into degenerates and one sperm makes its way to fertilize the egg cell, producing a diploid ($2n$) zygote. The second sperm cell fuses with both central cell nuclei, producing a triploid ($3n$) cell. As the zygote develops into an embryo, the triploid cell develops into the endosperm, which serves as the embryo's food supply. The ovary will now develop into a fruit and the ovule will develop into a seed.

As the development of embryo and endosperm proceeds within the embryo sac, the sac wall enlarges and combines with the nucellus (which is likewise enlarging) and the integument to form the seed coat. The ovary wall develops to form the fruit or pericarp, whose form is closely associated with type of seed dispersal system.

Frequently, the influence of fertilization is felt beyond the ovary, and other parts of the flower take part in the formation of the fruit, e.g., the floral receptacle in the apple, strawberry, and others.

The character of the seed coat bears a definite relation to that of the fruit. They protect the embryo and aid in dissemination; they may also directly promote germination. Among plants with indehiscent fruits, in general, the fruit provides protection for the embryo and secures dissemination. In this case, the seed coat is only slightly developed. If the fruit is dehiscent and the seed is exposed, in general,

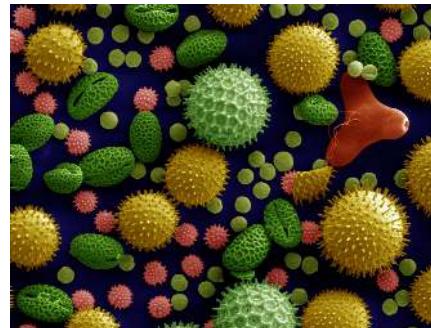


Figure 18.12: Scanning electron microscope image (500x magnification) of pollen grains from a variety of common plants:¹¹ sunflower (*Helianthus annuus*), morning glory (*Ipomoea purpurea*), prairie hollyhock (*Sidalcea malviflora*), oriental lily (*Lilium auratum*), evening primrose (*Oenothera fruticosa*), and castor bean (*Ricinus communis*). The image is magnified some $\times 500$, so the bean shaped grain in the bottom left corner is about $50 \mu\text{m}$ long. The colors are computer-generated.

the seed-coat is well developed, and must discharge the functions otherwise executed by the fruit.

Flowering plants generate gametes using a specialized cell division called meiosis. Meiosis takes place in the ovule (a structure within the ovary that is located within the pistil at the center of the flower) (see diagram labeled "Angiosperm lifecycle"). A diploid cell (megaspore mother cell) in the ovule undergoes meiosis (involving two successive cell divisions) to produce four cells (megaspores) with haploid nuclei. It is thought that the basal chromosome number in angiosperms is $n = 7$. One of these four cells (megaspore) then undergoes three successive mitotic divisions to produce an immature embryo sac (megagametophyte) with eight haploid nuclei. Next, these nuclei are segregated into separate cells by cytokinesis to produce 3 antipodal cells, 2 synergid cells and an egg cell. Two polar nuclei are left in the central cell of the embryo sac.

Pollen is also produced by meiosis in the male anther (microsporangium). During meiosis, a diploid microspore mother cell undergoes two successive meiotic divisions to produce 4 haploid cells (microspores or male gametes). Each of these microspores, after further mitoses, becomes a pollen grain (microgametophyte) containing two haploid generative (sperm) cells and a tube nucleus. When a pollen grain makes contact with the female stigma, the pollen grain forms a pollen tube that grows down the style into the ovary. In the act of fertilization, a male sperm nucleus fuses with the female egg nucleus to form a diploid zygote that can then develop into an embryo within the newly forming seed. Upon germination of the seed, a new plant can grow and mature.

The adaptive function of meiosis is currently a matter of debate. A key event during meiosis in a diploid cell is the pairing of homologous chromosomes and homologous recombination (the exchange of genetic information) between homologous chromosomes. This process promotes the production of increased genetic diversity among progeny and the recombinational repair of damages in the DNA to be passed on to progeny. To explain the adaptive function of meiosis in flowering plants, some authors emphasize diversity and others emphasize DNA repair.

18.7 Plant Anatomy

Plant anatomy or phytotomy is the general term for the study of the internal structure of plants. Originally it included plant morphology, the description of the physical form and external structure of plants, but since the mid-20th century plant anatomy has been considered a separate field referring only to internal plant structure. Plant anatomy is now frequently investigated at the cellular level, and often involves the sectioning of tissues and microscopy.

Some studies of plant anatomy use a systems approach, organized on the basis of the plant's activities, such as nutrient transport, flowering, pollination, embryogenesis or seed development. Others are more classically divided into the following structural categories:

- Flower anatomy, including study of the Calyx, Corolla, Androecium, and Gynoecium
- Leaf anatomy, including study of the Epidermis, stomata and Palisade cells
- Stem anatomy, including Stem structure and vascular tissues, buds and shoot apex
- Fruit/Seed anatomy, including structure of the Ovule, Seed, Pericarp and Accessory fruit
- Wood anatomy, including structure of the Bark, Cork, Xylem, Phloem, Vascular cambium, Heartwood and sapwood and branch collar
- Root anatomy, including structure of the Root, root tip, endodermis

Organs of plants can be divided into vegetative and reproductive. Vegetative plant organs include roots, stems, and leaves. The reproductive organs are variable. In flowering plants, they are represented by the flower, seed and fruit. In conifers, the organ that bears the reproductive structures is called a cone. In other divisions (phyla) of plants, the reproductive organs are called strobili, in Lycopodiophyta, or simply gametophores in mosses.

The vegetative organs are essential for maintaining the life of a plant. While there can be 11 organ systems in animals, there are far fewer in plants, where some perform

the vital functions, such as photosynthesis, while the reproductive organs are essential in reproduction. However, if there is asexual vegetative reproduction, the vegetative organs are those that create the new generation of plants (see clonal colony).

18.8 Plant Tissues

In plant anatomy, tissues are categorized broadly into three tissue systems: the epidermis, the ground tissue, and the vascular tissue.

- Epidermis – Cells forming the outer surface of the leaves and of the young plant body.
- Vascular tissue – The primary components of vascular tissue are the xylem and phloem. These transport fluids and nutrients internally.
- Ground tissue – Ground tissue is less differentiated than other tissues. Ground tissue manufactures nutrients by photosynthesis and stores reserve nutrients.

Plant tissues can also be divided differently into two types:

1. Meristematic tissues
2. Permanent tissues.

18.8.1 Meristematic Tissues

Meristematic tissue consists of actively dividing cells, and leads to increase in length and thickness of the plant. The primary growth of a plant occurs only in certain, specific regions, such as in the tips of stems or roots. It is in these regions that meristematic tissues are present. Cells in these tissues are roughly spherical or polyhedral, to rectangular in shape, and have thin cell walls. New cells produced by meristem are initially those of meristem itself, but as the new cells grow and mature, their characteristics slowly change and they become differentiated as components of the region of occurrence of meristematic tissues, being classified as:

- Apical meristem – It is present at the growing tips of stems and roots and increases the length of the stem and root. They form growing parts at the apices of roots and stems and are responsible for the increase in length, also called primary growth. This meristem is responsible for the linear growth of an organ.
- Lateral meristem – This meristem consists of cells which mainly divide in one plane and cause the organ to increase in diameter and growth. Lateral meristem usually occurs beneath the bark of the tree in the form of Cork Cambium and in vascular bundles of dicots in the form of vascular cambium. The activ-

ity of this cambium results in the formation of secondary growth.

- **Intercalary meristem** – This meristem is located in between permanent tissues. It is usually present at the base of the node, internode and on leaf base. They are responsible for growth in length of the plant and increasing the size of the internode. They result in branch formation and growth.

The cells of meristematic tissues are similar in structure and have thin and elastic primary cell wall made up of cellulose. They are compactly arranged without inter-cellular spaces between them. Each cell contains a dense cytoplasm and a prominent nucleus. The dense protoplasm of meristematic cells contains very few vacuoles. Normally the meristematic cells are oval, polygonal or rectangular in shape.

Meristematic tissue cells have a large nucleus with small or no vacuoles as they have no need to store anything, opposed to their function of multiplying and increasing the girth and length of the plant, and no intercellular spaces.

18.8.2 Permanent Tissues

Permanent tissues may be defined as a group of living or dead cells formed by meristematic tissue and have lost their ability to divide and have permanently placed at fixed positions in the plant body. Meristematic tissues that take up a specific role lose the ability to divide. This process of taking up a permanent shape, size and a function is called cellular differentiation. Cells of meristematic tissue differentiate to form different types of permanent tissues. There are 3 types of permanent tissues:

1. simple permanent tissues
2. complex permanent tissues
3. special or secretory tissues (glandular).

Simple Permanent tissues

A group of cells which are similar in origin; similar in structure and similar in function are called simple permanent tissue. They are of three types:

1. Parenchyma
2. Collenchyma
3. Sclerenchyma

18.8.3 Parenchyma

Parenchyma (para – ‘beside’; infusion – ‘tissue’) is the bulk of a substance. In plants, it consists of relatively unspecialized living cells with thin cell walls that are usually loosely packed so that intercellular spaces are found between cells of this tissue. These are generally

isodiametric, in shape. They contain small number of vacuoles or sometimes they even may not contain any vacuole. Even if they do so the vacuole is of much smaller size than of normal animal cells. This tissue provides support to plants and also stores food. Chlorenchyma is a special type of parenchyma that contains chlorophyll and performs photosynthesis. In aquatic plants, aerenchyma tissues, or large air cavities, give support to float on water by making them buoyant. Parenchyma cells called idioblasts have metabolic waste. Spindle shape fiber also contained into this cell to support them and known as prosenchyma, succulent parenchyma also noted. In xerophytes, parenchyma tissues store water.

18.8.4 Collenchyma

Collenchyma is Greek word where “Colla” means gum and “enchyma” means infusion. It is a living tissue of primary body like Parenchyma. Cells are thin-walled but possess thickening of cellulose, water and pectin substances (pectocellulose) at the corners where a number of cells join together. This tissue gives tensile strength to the plant and the cells are compactly arranged and have very little inter-cellular spaces. It occurs chiefly in hypodermis of stems and leaves. It is absent in monocots and in roots. Sometimes it contains chlorophyll which can help them photosynthesize.

Collenchymatous tissue acts as a supporting tissue in stems of young plants. It provides mechanical support, elasticity, and tensile strength to the plant body. It helps in manufacturing sugar and storing it as starch. It is present in the margin of leaves and resists tearing effect of the wind.

18.8.5 Sclerenchyma

Sclerenchyma is Greek word where “Sclero-” means hard and “enchyma” means infusion. This tissue consists of thick-walled, dead cells and protoplasm is negligible. These cells have hard and extremely thick secondary walls due to uniform distribution and high secretion of lignin and have a function of providing mechanical support. They do not have inter-molecular space between them. Lignin deposition is so thick that the cell walls become strong, rigid and impermeable to water which is also known as a stone cell or sclereids. These tissues are mainly of two types: sclerenchyma fiber and sclereids. Sclerenchyma fiber cells have a narrow lumen and are long, narrow and unicellular. Fibers are elongated cells that are strong and flexible, often used in ropes. Sclereids have extremely thick cell walls and are brittle, and are found in nutshells and legumes.

18.8.6 Epidermis

The entire surface of the plant consists of a single layer of cells called epidermis or surface tissue. The entire surface of the plant has this outer layer of the epidermis. Hence it is also called surface tissue. Most of the epidermal cells are relatively flat. The outer and lateral walls of the cell are often thicker than the inner walls. The cells form a continuous sheet without intercellular spaces. It protects all parts of the plant. The outer epidermis is coated with a waxy thick layer called cuticle which prevents loss of water. The epidermis also consists of stomata(singular:stoma) which helps in transpiration.

18.8.7 Complex Permanent Tissue

The complex tissue consists of more than one type of cells which work together as a unit. Complex tissues help in the transportation of organic material, water, and minerals up and down the plants. That is why it is also known as conducting and vascular tissue. The common types of complex permanent tissue are:

- Xylem or wood
- Phloem or bast.

Xylem and phloem together form vascular bundles.

18.8.8 Xylem

Xylem consists of:

- Xylem tracheids
- Xylem vessel
- Xylem fibres or Xylem sclerenchyma
- Xylem parenchyma

Xylem serves as a chief conducting tissue of vascular plants.

It is responsible for the conduction of water and mineral ions/salt. Xylem tissue is organised in a tube-like fashion along the main axes of stems and roots. It consists of a combination of parenchyma cells, fibers, vessels, tracheids, and ray cells. Longer tubes made up of individual cells are vessels tracheids, while vessel members are open at each end. Internally, there may be bars of wall material extending across the open space. These cells are joined end to end to form long tubes. Vessel members and tracheids are dead at maturity. Tracheids have thick secondary cell walls and are tapered at the ends. They do not have end openings such as the vessels. The tracheids end overlap with each other, with pairs of pits present. The pit pairs allow water to pass from cell to cell.

Though most conduction in xylem tissue is vertical, lateral conduction along the diameter of a stem is facilitated via rays. Rays are horizontal rows of long-living

parenchyma cells that arise out of the vascular cambium. In trees and other woody plants, rays radiate out from the center of stems and roots and appear like spokes on a wheel in cross section. Rays, unlike vessel members and tracheids, are alive at functional maturity.

18.8.9 Phloem

Phloem consists of:

- Sieve tube
- Companion cell
- Phloem fibre
- Phloem parenchyma.

Phloem is an equally important plant tissue as it also is part of the 'plumbing system' of a plant. Primarily, phloem carries dissolved food substances throughout the plant. This conduction system is composed of sieve-tube member and companion cells, that are without secondary walls. The parent cells of the vascular cambium produce both xylem and phloem. This usually also includes fibers, parenchyma and ray cells. Sieve tubes are formed from sieve-tube members laid end to end. The end walls, unlike vessel members in xylem, do not have openings. The end walls, however, are full of small pores where cytoplasm extends from cell to cell. These porous connections are called sieve plates. In spite of the fact that their cytoplasm is actively involved in the conduction of food materials, sieve-tube members do not have nuclei at maturity. It is the companion cells that are nestled between sieve-tube members that function in some manner bringing about the conduction of food. Sieve-tube members that are alive contain a polymer called callose, a carbohydrate polymer, forming the callus pad/callus, the colourless substance that covers the sieve plate. Callose stays in solution as long as the cell contents are under pressure. Phloem transports food and materials in plants upwards and downwards as required.

18.8.10 Plant cells

Plant cells are typically distinguished by their large water-filled central vacuole, chloroplasts, and rigid cell walls that are made up of cellulose, hemicellulose, and pectin. Cell division is also characterized by the development of a phragmoplast for the construction of a cell plate in the late stages of cytokinesis. Just as in animals, plant cells differentiate and develop into multiple cell types. Totipotent meristematic cells can differentiate into vascular, storage, protective (e.g. epidermal layer), or reproductive tissues, with more primitive plants lacking some tissue types.

18.9 Biochemistry of Plants

The chemical elements of which plants are constructed—principally carbon, oxygen, hydrogen, nitrogen, phosphorus, sulfur, etc.—are the same as for all other life forms animals, fungi, bacteria and even viruses. Only the details of the molecules into which they are assembled differs.

Despite this underlying similarity, plants produce a vast array of chemical compounds with unique properties which they use to cope with their environment. Pigments are used by plants to absorb or detect light, and are extracted by humans for use in dyes. Other plant products may be used for the manufacture of commercially important rubber or biofuel. Perhaps the most celebrated compounds from plants are those with pharmacological activity, such as salicylic acid from which aspirin is made, morphine, and digoxin. Drug companies spend billions of dollars each year researching plant compounds for potential medicinal benefits.

Plants require some nutrients, such as carbon and nitrogen, in large quantities to survive. Some nutrients are termed macronutrients, where the prefix macro- (large) refers to the quantity needed, not the size of the nutrient particles themselves. Other nutrients, called micronutrients, are required only in trace amounts for plants to remain healthy. Such micronutrients are usually absorbed as ions dissolved in water taken from the soil, though carnivorous plants acquire some of their micronutrients from captured prey.

The following tables list element nutrients essential to plants. Uses within plants are generalized.

18.9.1 Pigments

Among the most important molecules for plant function are the pigments. Plant pigments include a variety of different kinds of molecules, including porphyrins, carotenoids, and anthocyanins. All biological pigments selectively absorb certain wavelengths of light while reflecting others. The light that is absorbed may be used by the plant to power chemical reactions, while the reflected wavelengths of light determine the color the pigment appears to the eye.

Chlorophyll is the primary pigment in plants; it is a porphyrin that absorbs red and blue wavelengths of light while reflecting green. It is the presence and relative abundance of chlorophyll that gives plants their green color. All land plants and green algae possess two forms of this pigment: chlorophyll a and chlorophyll b. Kelps, diatoms, and other photosynthetic heterokonts contain chlorophyll c instead of b, red algae possess chlorophyll a. All chlorophylls serve as the primary means plants use to intercept light to fuel

photosynthesis.

Carotenoids are red, orange, or yellow tetraterpenoids. They function as accessory pigments in plants, helping to fuel photosynthesis by gathering wavelengths of light not readily absorbed by chlorophyll. The most familiar carotenoids are carotene (an orange pigment found in carrots), lutein (a yellow pigment found in fruits and vegetables), and lycopene (the red pigment responsible for the color of tomatoes). Carotenoids have been shown to act as antioxidants and to promote healthy eyesight in humans.

Anthocyanins (literally “flower blue”) are water-soluble flavonoid pigments that appear red to blue, according to pH. They occur in all tissues of higher plants, providing color in leaves, stems, roots, flowers, and fruits, though not always in sufficient quantities to be noticeable. Anthocyanins are most visible in the petals of flowers, where they may make up as much as 30% of the dry weight of the tissue. They are also responsible for the purple color seen on the underside of tropical shade plants such as *Tradescantia zebrina*. In these plants, the anthocyanin catches light that has passed through the leaf and reflects it back towards regions bearing chlorophyll, in order to maximize the use of available light

Betalains are red or yellow pigments. Like anthocyanins they are water-soluble, but unlike anthocyanins they are indole-derived compounds synthesized from tyrosine. This class of pigments is found only in the Caryophyllales (including cactus and amaranth), and never co-occur in plants with anthocyanins. Betalains are responsible for the deep red color of beets, and are used commercially as food-coloring agents. Plant physiologists are uncertain of the function that betalains have in plants which possess them, but there is some preliminary evidence that they may have fungicidal properties.

18.9.2 Plant Hormones

Plants produce hormones and other growth regulators which act to signal a physiological response in their tissues. They also produce compounds such as phytochrome that are sensitive to light and which serve to trigger growth or development in response to environmental signals.

Plant hormones, known as plant growth regulators (PGRs) or phytohormones, are chemicals that regulate a plant's growth. According to a standard animal definition, hormones are signal molecules produced at specific locations, that occur in very low concentrations, and cause altered processes in target cells at other locations. Unlike animals, plants lack specific hormone-producing tissues or organs. Plant hormones are often not transported to other parts of the plant and production is not limited to

specific locations.

Plant hormones are chemicals that in small amounts promote and influence the growth, development and differentiation of cells and tissues. Hormones are vital to plant growth; affecting processes in plants from flowering to seed development, dormancy, and germination. They regulate which tissues grow upwards and which grow downwards, leaf formation and stem growth, fruit development and ripening, as well as leaf abscission and even plant death.

The most important plant hormones are abscissic acid (ABA), auxins, ethylene, gibberellins, and cytokinins, though there are many other substances that serve to regulate plant physiology.

18.9.3 Photomorphogenesis

While most people know that light is important for photosynthesis in plants, few realize that plant sensitivity to light plays a role in the control of plant structural development (morphogenesis). The use of light to control structural development is called photomorphogenesis, and is dependent upon the presence of specialized photoreceptors, which are chemical pigments capable of absorbing specific wavelengths of light.

Plants use four kinds of photoreceptors: phytochrome, cryptochrome, a UV-B photoreceptor, and protochlorophyllide a. The first two of these, phytochrome and cryptochrome, are photoreceptor proteins, complex molecular structures formed by joining a protein with a light-sensitive pigment. Cryptochrome is also known as the UV-A photoreceptor, because it absorbs ultraviolet light in the long wave "A" region. The UV-B receptor is one or more compounds not yet identified with certainty, though some evidence suggests carotene or riboflavin as candidates. Protochlorophyllide a, as its name suggests, is a chemical precursor of chlorophyll.

The most studied of the photoreceptors in plants is phytochrome. It is sensitive to light in the red and far-red region of the visible spectrum. Many flowering plants use it to regulate the time of flowering based on the length of day and night (photoperiodism) and to set circadian rhythms. It also regulates other responses including the germination of seeds, elongation of seedlings, the size, shape and number of leaves, the synthesis of chlorophyll, and the straightening of the epicotyl or hypocotyl hook of dicot seedlings.

18.9.4 Photoperiodism

Many flowering plants use the pigment phytochrome to sense seasonal changes in day length, which they take as signals to flower. This sensitivity to day length is termed

photoperiodism. Broadly speaking, flowering plants can be classified as long day plants, short day plants, or day neutral plants, depending on their particular response to changes in day length. Long day plants require a certain minimum length of daylight to start flowering, so these plants flower in the spring or summer. Conversely, short day plants flower when the length of daylight falls below a certain critical level. Day neutral plants do not initiate flowering based on photoperiodism, though some may use temperature sensitivity (vernization) instead.

Although a short day plant cannot flower during the long days of summer, it is not actually the period of light exposure that limits flowering. Rather, a short day plant requires a minimal length of uninterrupted darkness in each 24-hour period (a short daylength) before floral development can begin. It has been determined experimentally that a short day plant (long night) does not flower if a flash of phytochrome activating light is used on the plant during the night.

Plants make use of the phytochrome system to sense day length or photoperiod. This fact is utilized by florists and greenhouse gardeners to control and even induce flowering out of season, such as the Poinsettia.

18.10 Plant Secondary Metabolism

Secondary metabolism produces a large number of specialized compounds (estimated 200,000) that do not aid in the growth and development of plants but are required for the plant to survive in its environment. Secondary metabolism is connected to primary metabolism by using building blocks and biosynthetic enzymes derived from primary metabolism. Primary metabolism governs all basic physiological processes that allow a plant to grow and set seeds, by translating the genetic code into proteins, carbohydrates, and amino acids. Specialized compounds from secondary metabolism are essential for communicating with other organisms in mutualistic (e.g. attraction of beneficial organisms such as pollinators) or antagonistic interactions (e.g. deterrent against herbivores and pathogens). They further assist in coping with abiotic stress such as increased UV-radiation. The broad functional spectrum of specialized metabolism is still not fully understood. In any case, a good balance between products of primary and secondary metabolism is best for a plant's optimal growth and development as well as for its effective coping with often changing environmental conditions. Well known specialized compounds include alkaloids, polyphenols including flavonoids, and terpenoids. Humans use quite a lot of these compounds, or the plants from which they originate, for culinary, medicinal and nutraceutical purposes.

18.11 Tropisms And Nastic Movements

Plants may respond both to directional and non-directional stimuli. A response to a directional stimulus, such as gravity or sunlight, is called a tropism. A response to a nondirectional stimulus, such as temperature or humidity, is a nastic movement.

Tropisms in plants are the result of differential cell growth, in which the cells on one side of the plant elongates more than those on the other side, causing the part to bend toward the side with less growth. Among the common tropisms seen in plants is phototropism, the bending of the plant toward a source of light. Phototropism allows the plant to maximize light exposure in plants which require additional light for photosynthesis, or to minimize it in plants subjected to intense light and heat. Geotropism allows the roots of a plant to determine the direction of gravity and grow downwards. Tropisms generally result from an interaction between the environment and production of one or more plant hormones.

Nastic movements results from differential cell growth (e.g. epinasty and hiponasty), or from changes in turgor pressure within plant tissues (e.g., nyctinasty), which may occur rapidly. A familiar example is thigmonasty (response to touch) in the Venus fly trap, a carnivorous plant. The traps consist of modified leaf blades which bear sensitive trigger hairs. When the hairs are touched by an insect or other animal, the leaf folds shut. This mechanism allows the plant to trap and digest small insects for additional nutrients. Although the trap is rapidly shut by changes in internal cell pressures, the leaf must grow slowly to reset for a second opportunity to trap insects.

18.12 Plant Reproduction

Plant reproduction is the production of new offspring in plants, which can be accomplished by sexual or asexual reproduction. Sexual reproduction produces offspring by the fusion of gametes, resulting in offspring genetically different from the parent or parents. Asexual reproduction produces new individuals without the fusion of gametes, genetically identical to the parent plants and each other, except when mutations occur.

18.12.1 The Life Cycle of Plants

Like all land plants (embryophytes), bryophytes have life cycles with alternation of generations. In each cycle, a haploid gametophyte, each of whose cells contains a fixed number of unpaired chromosomes, alternates with a

diploid sporophyte, whose cell contain two sets of paired chromosomes. Gametophytes produce haploid sperm and eggs which fuse to form diploid zygotes that grow into sporophytes. Sporophytes produce haploid spores by meiosis, that grow into gametophytes.

Bryophytes are gametophyte dominant, meaning that the more prominent, longer-lived plant is the haploid gametophyte. The diploid sporophytes appear only occasionally and remain attached to and nutritionally dependent on the gametophyte. In bryophytes, the sporophytes are always unbranched and produce a single sporangium (spore producing capsule), but each gametophyte can give rise to several sporophytes at once.

The sporophyte develops differently in the three groups. Both mosses and hornworts have a meristem zone where cell division occur. In hornworts, the meristem starts at the base where the foot ends, and the division of cells is pushing the sporophyte body upwards. In mosses, the meristem is located between the capsule and the top of the stalk (seta), and produce cells downward, elongating the stalk and elevates the capsule. In liverworts the meristem is absent and the elongation of the sporophyte is caused almost exclusively by cell expansion.

Liverworts, mosses and hornworts spend most of their lives as gametophytes. Gametangia (gamete-producing organs), archegonia and antheridia, are produced on the gametophytes, sometimes at the tips of shoots, in the axils of leaves or hidden under thalli. Some bryophytes, such as the liverwort *Marchantia*, create elaborate structures to bear the gametangia that are called gametangiophores. Sperm are flagellated and must swim from the antheridia that produce them to archegonia which may be on a different plant. Arthropods can assist in transfer of sperm.

Fertilized eggs become zygotes, which develop into sporophyte embryos inside the archegonia. Mature sporophytes remain attached to the gametophyte. They consist of a stalk called a seta and a single sporangium or capsule. Inside the sporangium, haploid spores are produced by meiosis. These are dispersed, most commonly by wind, and if they land in a suitable environment can develop into a new gametophyte. Thus bryophytes disperse by a combination of swimming sperm and spores, in a manner similar to lycophytes, ferns and other cryptogams.

The arrangement of antheridia and archegonia on an individual bryophyte plant is usually constant within a species, although in some species it may depend on environmental conditions. The main division is between species in which the antheridia and archegonia occur on the same plant and those in which they occur on different plants. The term monoicous may be used where

antheridia and archegonia occur on the same gametophyte and the term dioicous where they occur on different gametophytes.

In seed plants, "monoecious" is used where flowers with anthers (microsporangia) and flowers with ovules (megasporangia) occur on the same sporophyte and "dioecious" where they occur on different sporophytes. These terms occasionally may be used instead of "monoecious" and "dioecious" to describe bryophyte gametophytes. "Monoecious" and "monoecious" are both derived from the Greek for "one house", "dioecious" and "dioecious" from the Greek for two houses. The use of the "oicy" terminology refers to the gametophyte sexuality of bryophytes as distinct from the sporophyte sexuality of seed plants.

Monoecious plants are necessarily hermaphroditic, meaning that the same plant has both sexes. The exact arrangement of the antheridia and archegonia in monoecious plants varies. They may be borne on different shoots (autoicous or autoecious), on the same shoot but not together in a common structure (paroicous or paroecious), or together in a common "inflorescence" (synoicous or synoecious). Dioecious plants are unisexual, meaning that the same plant has only one sex. All four patterns (autoicous, paroicous, synoicous and dioecious) occur in species of the moss genus *Bryum*.

18.12.2 Asexual Reproduction

In asexual reproduction male and female gametes do not fuse, as they do in sexual reproduction. Asexual reproduction may occur through Binary Fission, budding, fragmentation, spore formation, Regeneration and vegetative propagation. Plants have two main types of asexual reproduction in which new plants are produced that are genetically identical clone of the parent individual. Vegetative reproduction involves a vegetative piece of the original plant (budding, tillering, etc.) and is distinguished from apomixis, which is a replacement of sexual reproduction, and in some cases involves seeds. Apomixis appears in many plant species and also in some non-plant organisms. For apomixis and similar processes in non-plant organisms, see parthenogenesis.

Natural vegetative reproduction is a process mostly found in herbaceous and woody perennial plants, and typically involves structural modifications of the stem or roots and in a few species leaves. Most plant species that employ vegetative reproduction do so as a means to perennialize the plants, allowing them to survive from one season to the next and often facilitating their expansion in size. A plant that persists in a location through vegetative reproduction of individuals constitutes a clonal colony; a single ramet, or apparent individual, of a clonal colony is genetically identical to all others in the same colony.

The distance that a plant can move during vegetative reproduction is limited, though some plants can produce ramets from branching rhizomes or stolons that cover a wide area, often in only a few growing seasons. In a sense, this process is not one of reproduction but one of survival and expansion of biomass of the individual. When an individual organism increases in size via cell multiplication and remains intact, the process is called vegetative growth. However, in vegetative reproduction, the new plants that result are new individuals in almost every respect except genetic. A major disadvantage to vegetative reproduction, is the transmission of pathogens from parent to offspring; it is uncommon for pathogens to be transmitted from the plant to its seeds (in sexual reproduction or in apomixis), though there are occasions when it occurs.

Seeds generated by apomixis are a means of asexual reproduction, involving the formation and dispersal of seeds that do not originate from the fertilization of the embryos. Hawkweed (*Hieracium*), dandelion (*Taraxacum*), some *Citrus* (*Citrus*) and Kentucky blue grass (*Poa pratensis*) all use this form of asexual reproduction. Pseudogamy occurs in some plants that have apomictic seeds, where pollination is often needed to initiate embryo growth, though the pollen contributes no genetic material to the developing offspring. Other forms of apomixis occur in plants also, including the generation of a plantlet in replacement of a seed or the generation of bulbils instead of flowers, where new cloned individuals are produced. Asexual reproduction is a type of reproduction where the offspring comes from one parent only, thus, inheriting the characteristics of the parent.

A rhizome is a modified underground stem serving as an organ of vegetative reproduction; the growing tips of the rhizome can separate as new plants, e.g., polypody, iris, couch grass and nettles.

Prostrate aerial stems, called runners or stolons, are important vegetative reproduction organs in some species, such as the strawberry, numerous grasses, and some ferns.

Adventitious buds form on roots near the ground surface, on damaged stems (as on the stumps of cut trees), or on old roots. These develop into above-ground stems and leaves. A form of budding called suckering is the reproduction or regeneration of a plant by shoots that arise from an existing root system. Species that characteristically produce suckers include Elm (*Ulmus*), Dandelion (*Taraxacum*), and many members of the Rose family such as *Rosa* and *Rubus*.

Plants like onion (*Allium cepa*), hyacinth (*Hyacinth*), narcissus (*Narcissus*) and tulips (*Tulipa*) reproduce by dividing their underground bulbs into more bulbs. Other plants like potatoes (*Solanum tuberosum*) and dahlia (*Dahlia*) reproduce by a similar method involving under-

ground tubers. Gladioli and crocuses (*Crocus*) reproduce in a similar way with corms.

The most common form of plant reproduction utilized by people is seeds, but a number of asexual methods are utilized which are usually enhancements of natural processes, including: cutting, grafting, budding, layering, division, sectioning of rhizomes, roots, tubers, bulbs, stolons, tillers, etc., and artificial propagation by laboratory tissue cloning. Asexual methods are most often used to propagate cultivars with individual desirable characteristics that do not come true from seed. Fruit tree propagation is frequently performed by budding or grafting desirable cultivars (clones), onto rootstocks that are also clones, propagated by stooling.

In horticulture, a “cutting” is a branch that has been cut off from a mother plant below an internode and then rooted, often with the help of a rooting liquid or powder containing hormones. When a full root has formed and leaves begin to sprout anew, the clone is a self-sufficient plant, genetically identical.

Examples include cuttings from the stems of blackberries (*Rubus occidentalis*), African violets (*Saintpaulia*), verbenas (*Verbena*) to produce new plants. A related use of cuttings is grafting, where a stem or bud is joined onto a different stem. Nurseries offer for sale trees with grafted stems that can produce four or more varieties of related fruits, including apples. The most common usage of grafting is the propagation of cultivars onto already rooted plants, sometimes the rootstock is used to dwarf the plants or protect them from root damaging pathogens.

Since vegetatively propagated plants are clones, they are important tools in plant research. When a clone is grown in various conditions, differences in growth can be ascribed to environmental effects instead of genetic differences.

18.12.3 Sexual Reproduction

Sexual reproduction involves two fundamental processes: meiosis, which rearranges the genes and reduces the number of chromosomes, and fertilization, which restores the chromosome to a complete diploid number. In between these two processes, different types of plants and algae vary, but many of them, including all land plants, undergo alternation of generations, with two different multicellular structures (phases), a gametophyte and a sporophyte. The evolutionary origin and adaptive significance of sexual reproduction are discussed in the pages “Evolution of sexual reproduction” and “Origin and function of meiosis.”

The gametophyte is the multicellular structure (plant) that is haploid, containing a single set of chromosomes in each cell. The gametophyte produces male or female

gametes (or both), by a process of cell division, called mitosis. In vascular plants with separate gametophytes, female gametophytes are known as mega gametophytes (mega=large, they produce the large egg cells) and the male gametophytes are called micro gametophytes (micro=small, they produce the small sperm cells).

The fusion of male and female gametes (fertilization) produces a diploid zygote, which develops by mitotic cell divisions into a multicellular sporophyte.

The mature sporophyte produces spores by meiosis, sometimes referred to as “reduction division” because the chromosome pairs are separated once again to form single sets.

In mosses and liverworts, the gametophyte is relatively large, and the sporophyte is a much smaller structure that is never separated from the gametophyte. In ferns, gymnosperms, and flowering plants (angiosperms), the gametophytes are relatively small and the sporophyte is much larger. In gymnosperms and flowering plants the megagametophyte is contained within the ovule (that may develop into a seed) and the micro gametophyte is contained within a pollen grain. It is known as fertilization.

Unlike animals, plants are immobile, and cannot seek out sexual partners for reproduction. In the evolution of early plants, abiotic means, including water and wind, transported sperm for reproduction. The first plants were aquatic, as described in the page “Evolutionary history of plants”, and released sperm freely into the water to be carried with the currents. Primitive land plants like liverworts and mosses had motile sperm that swam in a thin film of water or were splashed in water droplets from the male reproduction organs onto the female organs. As taller and more complex plants evolved, modifications in the alternation of generations evolved; in the Paleozoic era progymnosperms reproduced by using spores dispersed on the wind. The seed plants including seed ferns, conifers and cordaites, which were all gymnosperms, evolved 350 million years ago; they had pollen grains that contained the male gametes for protection of the sperm during the process of transfer from the male to female parts. It is believed that insects fed on the pollen, and plants thus evolved to use insects to actively carry pollen from one plant to the next. Seed producing plants, which include the angiosperms and the gymnosperms, have a heteromorphic alternation of generations with large sporophytes containing much-reduced gametophytes. Angiosperms have distinctive reproductive organs called flowers, with carpels, and the female gametophyte is greatly reduced to a female embryo sac, with as few as eight cells. The male gametophyte consists of the pollen grains. The sperm of seed plants are non-motile, except for two older groups of plants, the Cycadophyta and the

Ginkgophyta, which have flagellated sperm.

18.12.4 Reproduction of Flowering Plants

Flowering plants are the dominant plant form on land and they reproduce by sexual and asexual means. Often their most distinguishing feature is their reproductive organs, commonly called flowers. Sexual reproduction in flowering plants involves the production of male and female gametes, the transfer of the male gametes to the female ovules in a process called pollination. After pollination occurs, fertilization happens and the ovules grow into seeds within a fruit. After the seeds are ready to for dispersal, the fruit ripens and by various means, the seeds are freed from the fruit and after varying amounts of time and under specific conditions the seeds germinate and grow into the next generation.

The anther produces male gametophytes which are pollen grains, which attach to the stigma on top of a carpel, in which the female gametophytes (inside ovules) are located. After the pollen tube grows through the carpel's style, the sperm from the pollen grain migrates into the ovule to fertilize the egg cell and central cell within the female gametophyte in a process termed double fertilization. The resulting zygote develops into an embryo, while the triploid endosperm (one sperm cell plus a binucleate female cell) and female tissues of the ovule give rise to the surrounding tissues in the developing seed. The ovary, which produced the female gametophyte(s), then grows into a fruit, which surrounds the seed(s). Plants may either self-pollinate or cross-pollinate.

18.12.5 Pollination

In plants that use insects or other animals to move pollen from one flower to the next, plants have developed greatly modified flower parts to attract pollinators and to facilitate the movement of pollen from one flower to the insect and from the insect back to the next flower. Flowers of wind-pollinated plants tend to lack petals and or sepals; typically large amounts of pollen are produced and pollination often occurs early in the growing season before leaves can interfere with the dispersal of the pollen. Many trees and all grasses and sedges are wind-pollinated, as such they have no need for any flowers.

Plants have a number of different means to attract pollinators including color, scent, heat, nectar glands, edible pollen and flower shape. Along with modifications involving the above structures two other conditions play a very important role in the sexual reproduction of flowering plants, the first is the timing of flowering and the other is the size or number of flowers produced. Often plant species have a few large, very showy flowers while others

produce many small flowers, often flowers are collected together into large inflorescences to maximize their visual effect, becoming more noticeable to passing pollinators. Flowers are attraction strategies and sexual expressions are functional strategies used to produce the next generation of plants, with pollinators and plants having co-evolved, often to some extraordinary degrees, very often rendering mutual benefit.

The largest family of flowering plants is the orchids (Orchidaceae), estimated by some specialists to include up to 35,000 species, which often have highly specialized flowers that attract particular insects for pollination. The stamens are modified to produce pollen in clusters called pollinia, which become attached to insects that crawl into the flower. The flower shapes may force insects to pass by the pollen, which is "glued" to the insect. Some orchids are even more highly specialized, with flower shapes that mimic the shape of insects to attract them to attempt to 'mate' with the flowers, a few even have scents that mimic insect pheromones.

Another large group of flowering plants is the Asteraceae or sunflower family with close to 22,000 species, which also have highly modified inflorescences that are flowers collected together in heads composed of a composite of individual flowers called florets. Heads with florets of one sex, when the flowers are pistillate or functionally staminate or made up of all bisexual florets, are called homogamous and can include discoid and liguliflorous type heads. Some radiate heads may be homogamous too. Plants with heads that have florets of two or more sexual forms are called heterogamous and include radiate and disciform head forms, though some radiate heads may be heterogamous too.

18.13 The Fruit

A fruit¹² is the seed-bearing structure in angiosperms formed from the ovary after flowering.

Fruits are the means by which angiosperms disseminate seeds. Edible fruits, in particular, have propagated with the movements of humans and animals in a symbiotic relationship as a means for seed dispersal and nutrition; in fact, humans and many animals have become dependent on fruits as a source of food. Accordingly, fruits account for a substantial fraction of the world's agricultural output, and some (such as the apple and the pomegranate) have acquired extensive cultural and symbolic meanings.

In common language usage, "fruit" normally means the fleshy seed-associated structures of a plant that are sweet

¹²<https://en.wikipedia.org/wiki/Fruit>



Figure 18.13: Flowers and fruit (capsules) of the ground orchid, *Spathoglottis plicata*, illustrating an inferior ovary.¹³

or sour, and edible in the raw state, such as apples, bananas, grapes, lemons, oranges, and strawberries. On the other hand, in botanical usage, “fruit” includes many structures that are not commonly called “fruits”, such as bean pods, corn kernels, tomatoes, and wheat grains. The section of a fungus that produces spores is also called a fruiting body.

The outer, often edible layer, is the pericarp, formed from the ovary and surrounding the seeds, although in some species other tissues contribute to or form the edible portion. The pericarp may be described in three layers from outer to inner, the epicarp, mesocarp and endocarp.

Fruit that bears a prominent pointed terminal projection is said to be beaked.

A fruit results from maturation of one or more flowers, and the gynoecium of the flower(s) forms all or part of the fruit. Gynoecium (from Ancient Greek γυνή, gyne, meaning woman, and οἶκος, oikos, meaning house) is most commonly used as a collective term for the parts of a flower that produce ovules and ultimately develop into the fruit and seeds. The gynoecium is the innermost whorl of a flower; it consists of (one or more) pistils and is typically surrounded by the pollen-producing reproductive organs, the stamens, collectively called the androecium. The gynoecium is often referred to as the “female” portion of the flower, although rather than directly producing female gametes (i.e. egg cells), the gynoecium produces megasporangia, each of which develops into a female gametophyte which then produces egg cells.

Inside the ovary/ovaries are one or more ovules where the megagametophyte contains the egg cell. After double fertilization, these ovules will become seeds. The ovules are fertilized in a process that starts with pollination, which involves the movement of pollen from the stamens to the stigma of flowers. After pollination, a tube grows from the pollen through the stigma into the ovary to the ovule and two sperm are transferred from the pollen to the megagametophyte. Within the megagametophyte one of the two sperm unites with the egg, forming a zygote, and the sec-

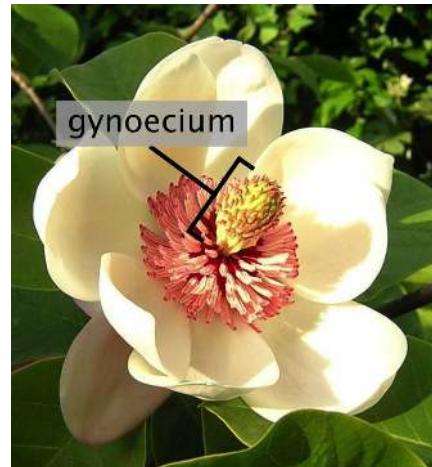


Figure 18.14: Flower of *Magnolia × wieseneri* showing the many pistils making up the gynoecium in the middle of the flower¹⁴



Figure 18.15: Cross-section through the ovary of *Narcissus* showing multiple carpels (the female reproductive part of the flower) fused along the placental line where the ovules form¹⁵

ond sperm enters the central cell forming the endosperm mother cell, which completes the double fertilization process. Later the zygote will give rise to the embryo of the seed, and the endosperm mother cell will give rise to endosperm, a nutritive tissue used by the embryo.

As the ovules develop into seeds, the ovary begins to ripen and the ovary wall, the pericarp, may become fleshy (as in berries or drupes), or form a hard outer covering (as in nuts). In some multiseeded fruits, the extent to which the flesh develops is proportional to the number of fertilized ovules. The pericarp is often differentiated into two or three distinct layers called the exocarp (outer layer, also called epicarp), mesocarp (middle layer), and endocarp (inner layer). In some fruits, especially simple fruits derived from an inferior ovary, other parts of the flower (such as the floral tube, including the petals, sepals, and stamens), fuse with the ovary and ripen with it. In other cases, the



Figure 18.16: The development sequence of a typical drupe, the nectarine (*Prunus persica*) over a 7.5 month period, from bud formation in early winter to fruit ripening in midsummer.¹⁶⁾ 1. Bud formation can be observed on new growth on the plant (early winter) 2. Flower buds clearly formed and leaves start to develop (early spring, \approx 3 months). 3. Flowers fully develop and are pollinated by wind or insects (early spring, \approx 3½ months). 4. If successfully pollinated, flowers die back and incipient fruit can be observed; leaves have quickly grown to provide tree with food and energy from photosynthesis (mid-spring, \approx 4 months). 5. Fruit is well developed and continues to grow (late spring, \approx 5½ months). 6. Fruit fully ripens to an edible form to encourage spreading of seed contained within by animals (midsummer, \approx 7½ months)

sepals, petals and/or stamens and style of the flower fall off. When such other floral parts are a significant part of the fruit, it is called an accessory fruit. Since other parts of the flower may contribute to the structure of the fruit, it is important to study flower structure to understand how a particular fruit forms.

There are three general modes of fruit development:

- Apocarpous fruits develop from a single flower having one or more separate carpels, and they are the simplest fruits.
- Syncarpous fruits develop from a single gynoecium having two or more carpels fused together.
- Multiple fruits form from many different flowers.

Plant scientists have grouped fruits into three main groups, simple fruits, aggregate fruits, and composite or multiple fruits. The groupings are not evolutionarily relevant, since many diverse plant taxa may be in the same group, but reflect how the flower organs are arranged and how the fruits develop.

18.13.1 Reproductin of Ferns

Ferns are vascular plants differing from lycophytes by having true leaves (megaphylls), which are often pinnate. They differ from seed plants (gymnosperms and angiosperms) in reproducing by means of spores and they lack flowers and seeds. Like all land plants, they have a life cycle referred to as alternation of generations, characterized by alternating diploid sporophytic and haploid gametophytic phases. The diploid sporophyte has $2n$ paired chromosomes, where n varies from species to species. The haploid gametophyte has n unpaired chromosomes, i.e. half the number of the sporophyte. The gametophyte of ferns is a free-living organism, whereas the gametophyte of the gymnosperms and angiosperms is dependent on the sporophyte.

The life cycle of a typical fern proceeds as follows:

1. A diploid sporophyte phase produces haploid spores by meiosis (a process of cell division which reduces the number of chromosomes by a half).
2. A spore grows into a free-living haploid gametophyte by mitosis (a process of cell division which maintains the number of chromosomes). The gametophyte typically consists of a photosynthetic prothallus.
3. The gametophyte produces gametes (often both sperm and eggs on the same prothallus) by mitosis.
4. A mobile, flagellate sperm fertilizes an egg that remains attached to the prothallus.
5. The fertilized egg is now a diploid zygote and grows by mitosis into a diploid sporophyte (the typical fern plant).

Ferns typically produce large diploids with stem, roots, and leaves; and on fertile leaves called sporangium, spores are produced. The spores are released and germinate to produce short, thin gametophytes that are typically heart-shaped, small and green in color. The gametophytes or thallus, produce both motile sperm in the antheridia and egg cells in separate archegonia. After rains or when dew deposits a film of water, the motile sperm are splashed away from the antheridia, which are normally produced on the top side of the thallus, and swim in the film of water to the antheridia where they fertilize the egg. To promote out crossing or cross-fertilization the sperm is released before the eggs are receptive of the sperm, making it more likely that the sperm will fertilize the eggs of the different thallus. A zygote is formed after fertilization, which grows into a new sporophytic plant. The condition of having separate sporophyte and gametophyte plants is called alternation of the generations. Other plants with similar reproductive means include the *Psilotum*, *Lycopodium*, *Selaginella* and *Equisetum*.

18.13.2 Reproductin of Bryophytes

The bryophytes, which include liverworts, hornworts and mosses, reproduce both sexually and vegetatively. The gametophyte is the most commonly known phase of the plant. All are small plants found growing in moist locations and like ferns, have motile sperm with flagella and need water to facilitate sexual reproduction. These plants start as a haploid spore that grows into the dominant form, which is a multicellular haploid body with leaf-like structures that photosynthesize. Haploid gametes are produced in antheridia and archegonia by mitosis. The sperm released from the antheridia respond to chemicals released by ripe archegonia and swim to them in a film of water and fertilize the egg cells, thus producing zygotes that are diploid. The zygote divides by mitotic division and grows into a sporophyte that is diploid. The multicellular diploid sporophyte produces structures called spore capsules. The spore capsules produce spores by meiosis, and when ripe, the capsules burst open and the spores are released. Bryophytes show considerable variation in their breeding structures and the above is a basic outline. In some species each gametophyte is one sex while other species produce both antheridia and archegonia on the same gametophyte which is thus hermaphrodite.

Chapter 19

Animals

Animals¹ (from the Latin *animalis*, meaning having breath, having soul or living being) are heterotrophic multicellular eukaryotic organisms with internal digestion that form the biological kingdom Animalia. Animals consume organic material, breathe oxygen, are able to move, can reproduce sexually, and grow from a hollow sphere of cells, the blastula, during embryonic development. Over 1.5 million living animal species have been described—of which around 1 million are insects—but it has been estimated there are over 7 million animal species in total. Animals range in length from 8.5 micrometres (0.00033 in) to 33.6 metres (110 ft). They have complex interactions with each other and their environments, forming intricate food webs. The kingdom Animalia includes humans but in colloquial use the term animal often refers only to non-human animals. The scientific study of animals is known as zoology.

Most living animal species are in Bilateria, a clade whose members have a bilaterally symmetric body plan. The Bilateria include the protostomes—in which many groups of invertebrates are found, such as nematodes, arthropods, and molluscs—and the deuterostomes, containing both the echinoderms as well as the chordates, the latter containing the vertebrates. Life forms interpreted as early animals were present in the Ediacaran biota of the late Precambrian. Many modern animal phyla became clearly established in the fossil record as marine species during the Cambrian explosion, which began around 542 million years ago. 6,331 groups of genes common to all living animals have been identified; these may have arisen from a single common ancestor that lived 650 million years ago.

Historically, Aristotle divided animals into those with blood and those without. Carl Linnaeus created the first hierarchical biological classification for animals in 1758 with his *Systema Naturae*, which Jean-Baptiste Lamarck expanded into 14 phyla by 1809. In 1874, Ernst Haeckel divided the animal kingdom into the multicellular Metazoa (synonymous for Animalia) and the Protozoa, single-celled



Figure 19.1: Diversity of animals.²

¹<https://en.wikipedia.org/wiki/Animal>

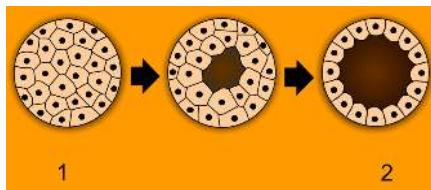


Figure 19.2: Animals are unique in having the ball of cells of the early embryo (1) develop into a hollow ball or blastula (2).³

organisms no longer considered animals. In modern times, the biological classification of animals relies on advanced techniques, such as molecular phylogenetics, which are effective at demonstrating the evolutionary relationships between animal taxa.

Humans make use of many other animal species, such as for food (including meat, milk, and eggs), for materials (such as leather and wool), and also as pets, and for transports, as working animals. Dogs have been used in hunting, while many terrestrial and aquatic animals were hunted for sports. Non-human animals have appeared in art from the earliest times and are featured in mythology and religion.

Animals have several characteristics that set them apart from other living things. Animals are eukaryotic and multicellular, unlike bacteria, which are prokaryotic, and unlike protists, which are eukaryotic but unicellular. Unlike plants and algae, which produce their own nutrients animals are heterotrophic, feeding on organic material and digesting it internally. With very few exceptions, animals respire aerobically. All animals are motile (able to spontaneously move their bodies) during at least part of their life cycle, but some animals, such as sponges, corals, mussels, and barnacles, later become sessile. The blastula is a stage in embryonic development that is unique to most animals, allowing cells to be differentiated into specialised tissues and organs.

All animals are composed of cells, surrounded by a characteristic extracellular matrix composed of collagen and elastic glycoproteins. During development, the animal extracellular matrix forms a relatively flexible framework upon which cells can move about and be reorganised, making the formation of complex structures possible. This may be calcified, forming structures such as shells, bones, and spicules. In contrast, the cells of other multicellular organisms (primarily algae, plants, and fungi) are held in place by cell walls, and so develop by progressive growth. Animal cells uniquely possess the cell junctions called tight junctions, gap junctions, and desmosomes.

With few exceptions—in particular, the sponges and placozoans—animal bodies are differentiated into tissues.

These include muscles, which enable locomotion, and nerve tissues, which transmit signals and coordinate the body. Typically, there is also an internal digestive chamber with either one opening (in Ctenophora, Cnidaria, and flatworms) or two openings (in most bilaterians).

Nearly all animals make use of some form of sexual reproduction. They produce haploid gametes by meiosis; the smaller, motile gametes are spermatozoa and the larger, non-motile gametes are ova. These fuse to form zygotes, which develop via mitosis into a hollow sphere, called a blastula. In sponges, blastula larvae swim to a new location, attach to the seabed, and develop into a new sponge. In most other groups, the blastula undergoes more complicated rearrangement. It first invaginates to form a gastrula with a digestive chamber and two separate germ layers, an external ectoderm and an internal endoderm. In most cases, a third germ layer, the mesoderm, also develops between them. These germ layers then differentiate to form tissues and organs.

Repeated instances of mating with a close relative during sexual reproduction generally leads to inbreeding depression within a population due to the increased prevalence of harmful recessive traits. Animals have evolved numerous mechanisms for avoiding close inbreeding.

Some animals are capable of asexual reproduction, which often results in a genetic clone of the parent. This may take place through fragmentation; budding, such as in Hydra and other cnidarians; or parthenogenesis, where fertile eggs are produced without mating, such as in aphids.

Animals are categorised into ecological groups depending on how they obtain or consume organic material, including carnivores, herbivores, omnivores, detritivores, and parasites. Interactions between animals form complex food webs. In carnivorous or omnivorous species, predation is a consumer-resource interaction where a predator feeds on another organism (called its prey). Selective pressures imposed on one another lead to an evolutionary arms race between predator and prey, resulting in various anti-predator adaptations. Almost all multicellular predators are animals. Some consumers use multiple methods; for example, in parasitoid wasps, the larvae feed on the hosts' living tissues, killing them in the process, but the adults primarily consume nectar from flowers. Other animals may have very specific feeding behaviours, such as hawksbill sea turtles primarily eating sponges.

Most animals rely on the biomass and energy produced by plants through photosynthesis. Herbivores eat plant material directly, while carnivores, and other animals on higher trophic levels typically acquire it indirectly by eat-



Figure 19.3: Tiktaalik, ≈ 375 Ma⁴

ing other animals. Animals oxidize carbohydrates, lipids, proteins, and other biomolecules to unlock the chemical energy of molecular oxygen, which allows the animal to grow and to sustain biological processes such as locomotion. Animals living close to hydrothermal vents and cold seeps on the dark sea floor consume organic matter of archaea and bacteria produced in these locations through chemosynthesis (by oxidizing inorganic compounds, such as hydrogen sulfide).

Animals originally evolved in the sea. Lineages of arthropods colonised land around the same time as land plants, probably between 510–471 million years ago during the Late Cambrian or Early Ordovician. Vertebrates such as the lobe-finned fish Tiktaalik started to move on to land in the late Devonian, about 375 million years ago. Animals occupy virtually all of earth's habitats and microhabitats, including salt water, hydrothermal vents, fresh water, hot springs, swamps, forests, pastures, deserts, air, and the interiors of animals, plants, fungi and rocks. Animals are however not particularly heat tolerant; very few of them can survive at constant temperatures above 50 °C (122 °F). Only very few species of animals (mostly nematodes) inhabit the most extreme cold deserts of continental Antarctica.

The blue whale (*Balaenoptera musculus*) is the largest animal that has ever lived, weighing up to at least 190 tonnes and measuring up to 33.6 metres (110 ft) long. The largest extant terrestrial animal is the African bush elephant (*Loxodonta africana*), weighing up to 12.25 tonnes and measuring up to 10.67 metres (35.0 ft) long. The largest terrestrial animals that ever lived were titanosaur sauropod dinosaurs such as Argentinosaurus, which may have weighed as much as 73 tonnes. Several animals are microscopic; some Myxozoa (obligate parasites within the Cnidaria) never grow larger than 20 µm, and one of the smallest species (*Myxobolus shekel*) is no more than 8.5 µm when fully grown.

The oldest animals are found in the Ediacaran biota, towards the end of the Precambrian, around 610 million years ago. It had long been doubtful whether these included animals, but the discovery of the animal lipid cholesterol in fossils of Dickinsonia establishes that these were indeed animals. Animals are thought to have originated under low-oxygen conditions, suggesting that they were capable of living entirely by anaerobic



Figure 19.4: Dickinsonia costata from the Ediacaran biota (c. 635–542 MYA) is one of the earliest animal species known.⁵



Figure 19.5: Anomalocaris canadensis⁶ is one of the many animal species that emerged in the Cambrian explosion, starting some 542 million years ago, and found in the fossil beds of the Burgess shale.

respiration, but as they became specialized for aerobic metabolism they became fully dependent on oxygen in their environments.

Many animal phyla first appear in the fossil record during the Cambrian explosion, starting about 542 million years ago, in beds such as the Burgess shale. Extant phyla in these rocks include molluscs, brachiopods, onychophorans, tardigrades, arthropods, echinoderms and hemichordates, along with numerous now-extinct forms such as the predatory Anomalocaris. The apparent suddenness of the event may however be an artefact of the fossil record, rather than showing that all these animals appeared simultaneously.

Some palaeontologists have suggested that animals appeared much earlier than the Cambrian explosion, possibly as early as 1 billion years ago. Trace fossils such as tracks and burrows found in the Tonian period may indicate the presence of triploblastic worm-like animals, roughly as large (about 5 mm wide) and complex as earthworms. However, similar tracks are produced today by the giant single-celled protist *Gromia sphaerica*, so the Tonian trace fossils may not indicate early animal

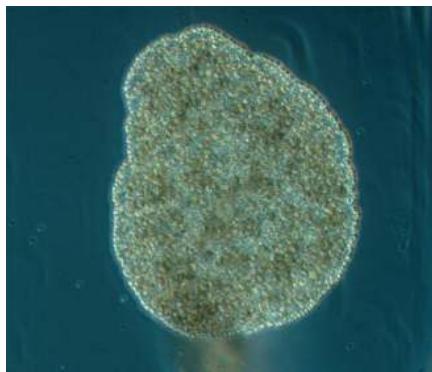


Figure 19.6: The Placozoa are a basal form of free-living (non-parasitic) multicellular organism. They are the simplest in structure of all animals. Three genera have been found: the classical *Trichoplax adhaerens*⁷, *Hoilungia hongkongensis*, and *Polyplacotoma mediterranea*.

evolution. Around the same time, the layered mats of microorganisms called stromatolites decreased in diversity, perhaps due to grazing by newly-evolved animals.

Animals are monophyletic, meaning they are derived from a common ancestor. Animals are sister to the Choanoflagellata, with which they form the Choanozoa. The most basal animals, the Porifera, Ctenophora, Cnidaria, and Placozoa, have body plans that lack bilateral symmetry. Their relationships are still disputed; the sister group to all other animals could be the Porifera or the Ctenophora, both of which lack hox genes, important in body plan development.

Table 19.1: Estimated numbers of described extant species for the animal groups with the largest numbers of species, along with their principal habitats (terrestrial, fresh water, and marine), and free-living or parasitic ways of life.

Phylum	No. of Species	Land	Sea	Freshwater	Free-living	Parasitic
Annelids	17,000	Yes (soil)	Yes	1,750	Yes	400
Arthropods	1,257,000	1,000,000(insects)	>40,000(Malacostraca)	94,000	Yes	>45,000[b]
Bryozoa	6,000	Yes	Yes	60-80	Yes	
Chordates	65,00045,000	23,000	13,000	18,0009,000	Yes	40(catfish)
Cnidaria	16,000	Yes	Yes (few)	Yes	Yes	>1,350(Myxozoa)
Echinoderms	7,500	7,500	7,500	Yes	Yes	
Molluscs	85,000107,000	35,000	60,000	5,00012,000	Yes	>5,600
Nematodes	25,000	Yes (soil)	4,000	2,000	11,000	14,000
Platyhelmini	29,500	Yes	Yes	1,300	Yes 3,000-6,500	>40,000
Rotifers	2,000	>400	2,000	2,000	Yes	4,000-25,000
Sponges	10,800	Yes	200-300	Yes	Yes	Yes

These genes are found in the Placozoa and the higher animals, the Bilateria. 6,331 groups of genes common to all living animals have been identified; these may have arisen from a single common ancestor that lived 650 million years ago in the Precambrian. 25 of these are novel core gene groups, found only in animals; of those, 8 are for essential components of the Wnt and TGF-beta signalling pathways which may have enabled animals to become multicellular by providing a pattern for the body's system of axes (in three dimensions), and another 7 are for transcription factors including homeodomain proteins involved in the control of development.

19.1 Non-bilaterian animals

Several animal phyla lack bilateral symmetry. Among these, the sponges (Porifera) probably diverged first, representing the oldest animal phylum. Sponges lack the complex organization found in most other animal phyla; their cells are differentiated, but in most cases not organised into distinct tissues. They typically feed by drawing in water through pores.

The Ctenophora (comb jellies) and Cnidaria (which includes jellyfish, sea anemones, and corals) are radially symmetric and have digestive chambers with a single opening, which serves as both mouth and anus. Animals in both phyla have distinct tissues, but these are not organised into organs. They are diploblastic, having only two main germ layers, ectoderm and endoderm. The tiny placozoans are similar, but they do not have a permanent digestive chamber.

19.1.1 Porifera

Sponges⁸, the members of the phylum Porifera (meaning “pore bearer”), are a basal Metazoa (animal) clade as a sister of the Diploblasts. They are multicellular organisms that have bodies full of pores and channels allowing water to circulate through them, consisting of jelly-like mesohyl sandwiched between two thin layers of cells.

Sponges have unspecialized cells that can transform into other types and that often migrate between the main cell layers and the mesohyl in the process. Sponges do not have nervous, digestive or circulatory systems. Instead, most rely on maintaining a constant water flow through their bodies to obtain food and oxygen and to remove wastes. Sponges were first to branch off the evolutionary tree from the common ancestor of all animals, making them the sister group of all other animals.

Sponges are similar to other animals in that they are multicellular, heterotrophic, lack cell walls and produce



Figure 19.7: Sponge biodiversity and morphotypes at the lip of a wall site in 60 feet (20 m) of water. Included are the yellow tube sponge, *Aplysina fistularis*, the purple vase sponge, *Niphates digitalis*, the red encrusting sponge, *Spirastrella coccinea* [nl], and the gray rope sponge, *Callyspongia* sp.⁹

sperm cells. Unlike other animals, they lack true tissues and organs. Some of them are radially symmetrical, but most are asymmetrical. The shapes of their bodies are adapted for maximal efficiency of water flow through the central cavity, where the water deposits nutrients and then leaves through a hole called the osculum. Many sponges have internal skeletons of spongin and/or spicules (skeletal-like fragments) of calcium carbonate or silicon dioxide. All sponges are sessile aquatic animals, meaning that they attach to an underwater surface and remain fixed in place (i.e., do not travel). Although there are freshwater species, the great majority are marine (salt-water) species, ranging in habitat from tidal zones to depths exceeding 8,800 m (5.5 mi).

Although most of the approximately 5,000–10,000 known species of sponges feed on bacteria and other microscopic food in the water, some host photosynthesizing microorganisms as endosymbionts, and these alliances often produce more food and oxygen than they consume. A few species of sponges that live in food-poor environments have evolved as carnivores that prey mainly on small crustaceans.

Most species use sexual reproduction, releasing sperm cells into the water to fertilize ova that in some species are released and in others are retained by the “mother.” The fertilized eggs develop into larvae, which swim off in search of places to settle. Sponges are known for regenerating from fragments that are broken off, although this only works if the fragments include the right types of cells. A few species reproduce by budding. When environmental conditions become less hospitable to the sponges, for example as temperatures drop, many freshwater species and a few marine ones produce gemmules, “survival pods” of unspecialized cells that remain dormant until conditions

⁸<https://en.wikipedia.org/wiki/Sponge>

improve; they then either form completely new sponges or recolonize the skeletons of their parents.

In most sponges, an internal gelatinous matrix called mesohyl functions as an endoskeleton, and it is the only skeleton in soft sponges that encrust such hard surfaces as rocks. More commonly, the mesohyl is stiffened by mineral spicules, by spongin fibers, or both. Demosponges use spongin; many species have silica spicules, whereas some species have calcium carbonate exoskeletons. Demosponges constitute about 90% of all known sponge species, including all freshwater ones, and they have the widest range of habitats. Calcareous sponges, which have calcium carbonate spicules and, in some species, calcium carbonate exoskeletons, are restricted to relatively shallow marine waters where production of calcium carbonate is easiest. The fragile glass sponges, with "scaffolding" of silica spicules, are restricted to polar regions and the ocean depths where predators are rare. Fossils of all of these types have been found in rocks dated from 580 million years ago.

The single-celled choanoflagellates resemble the choanocyte cells of sponges which are used to drive their water flow systems and capture most of their food. This along with phylogenetic studies of ribosomal molecules have been used as morphological evidence to suggest sponges are the sister group to the rest of animals.

The few species of demosponge that have entirely soft fibrous skeletons with no hard elements have been used by humans over thousands of years for several purposes, including as padding and as cleaning tools. By the 1950s, though, these had been overfished so heavily that the industry almost collapsed, and most sponge-like materials are now synthetic. Sponges and their microscopic endosymbionts are now being researched as possible sources of medicines for treating a wide range of diseases.

Even if a few sponges are able to produce mucus – which acts as a microbial barrier in all other animals – no sponge with the ability to secrete a functional mucus layer has been recorded. Without such a mucus layer their living tissue is covered by a layer of microbial symbionts, which can contribute up to 40–50% of the sponge wet mass.

Like cnidarians (jellyfish, etc.) and ctenophores (comb jellies), and unlike all other known metazoans, sponges' bodies consist of a non-living jelly-like mass (mesoglea) sandwiched between two main layers of cells. Cnidarians and ctenophores have simple nervous systems, and their cell layers are bound by internal connections and by being mounted on a basement membrane (thin fibrous mat, also known as "basal lamina"). Sponges have no nervous systems, their middle jelly-like layers have large and varied populations of cells, and some types of cells in their outer

layers may move into the middle layer and change their functions.

A sponge's body is hollow and is held in shape by the mesohyl, a jelly-like substance made mainly of collagen and reinforced by a dense network of fibers also made of collagen. The inner surface is covered with choanocytes, cells with cylindrical or conical collars surrounding one flagellum per choanocyte. The wave-like motion of the whip-like flagella drives water through the sponge's body. All sponges have ostia, channels leading to the interior through the mesohyl, and in most sponges these are controlled by tube-like porocytes that form closable inlet valves. Pinacocytes, plate-like cells, form a single-layered external skin over all other parts of the mesohyl that are not covered by choanocytes, and the pinacocytes also digest food particles that are too large to enter the ostia, while those at the base of the animal are responsible for anchoring it.

Other types of cell live and move within the mesohyl:

- Lophocytes are amoeba-like cells that move slowly through the mesohyl and secrete collagen fibres.
- Collencytes are another type of collagen-producing cell.
- Rhabdiferous cells secrete polysaccharides that also form part of the mesohyl.
- Oocytes and spermatocytes are reproductive cells.
- Sclerocytes secrete the mineralized spicules ("little spines") that form the skeletons of many sponges and in some species provide some defense against predators.
- In addition to or instead of sclerocytes, demosponges have spongocytes that secrete a form of collagen that polymerizes into spongin, a thick fibrous material that stiffens the mesohyl.
- Myocytes ("muscle cells") conduct signals and cause parts of the animal to contract.
- "Grey cells" act as sponges' equivalent of an immune system.
- Archaeocytes (or amoebocytes) are amoeba-like cells that are totipotent, in other words each is capable of transformation into any other type of cell. They also have important roles in feeding and in clearing debris that block the ostia.

Many larval sponges possess neuron-less eyes that are based on cryptochromes. They mediate phototoxic behavior.

Glass sponges present a distinctive variation on this basic plan. Their spicules, which are made of silica, form a scaffolding-like framework between whose rods the living tissue is suspended like a cobweb that contains most of the cell types. This tissue is a syncytium that in some

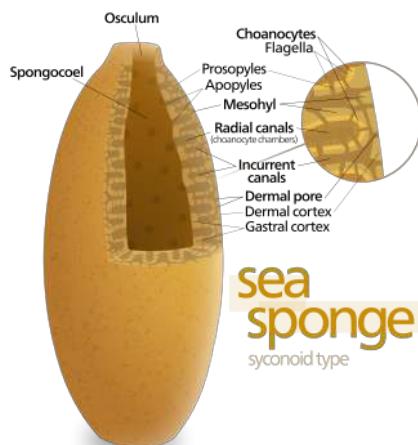


Figure 19.8: Diagram of the structure of a sponge.¹⁰

ways behaves like many cells that share a single external membrane, and in others like a single cell with multiple nuclei. The mesohyl is absent or minimal. The syncytium's cytoplasm, the soupy fluid that fills the interiors of cells, is organized into "rivers" that transport nuclei, organelles ("organs" within cells) and other substances. Instead of choanocytes, they have further syncytia, known as choanosyncytia, which form bell-shaped chambers where water enters via perforations. The insides of these chambers are lined with "collar bodies", each consisting of a collar and flagellum but without a nucleus of its own. The motion of the flagella sucks water through passages in the "cobweb" and expels it via the open ends of the bell-shaped chambers.

Some types of cells have a single nucleus and membrane each, but are connected to other single-nucleus cells and to the main syncytium by "bridges" made of cytoplasm. The sclerocytes that build spicules have multiple nuclei, and in glass sponge larvae they are connected to other tissues by cytoplasm bridges; such connections between sclerocytes have not so far been found in adults, but this may simply reflect the difficulty of investigating such small-scale features. The bridges are controlled by "plugged junctions" that apparently permit some substances to pass while blocking others. Most sponges work rather like chimneys: they take in water at the bottom and eject it from the osculum ("little mouth") at the top. Since ambient currents are faster at the top, the suction effect that they produce by Bernoulli's principle does some of the work for free. Sponges can control the water flow by various combinations of wholly or partially closing the osculum and ostia (the intake pores) and varying the beat of the flagella, and may shut it down if there is a lot of sand or silt in the water.

Although the layers of pinacocytes and choanocytes re-

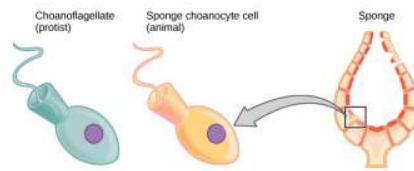


Figure 19.9: Cells of the protist choanoflagellate clade closely resemble sponge choanocyte cells. Beating of choanocyte flagella draws water through the sponge so that nutrients can be extracted and waste removed.¹¹

semble the epithelia of more complex animals, they are not bound tightly by cell-to-cell connections or a basal lamina (thin fibrous sheet underneath). The flexibility of these layers and re-modeling of the mesohyl by lophocytes allow the animals to adjust their shapes throughout their lives to take maximum advantage of local water currents.

The simplest body structure in sponges is a tube or vase shape known as "asconoid", but this severely limits the size of the animal. The body structure is characterized by a stalk-like spongocoel surrounded by a single layer of choanocytes. If it is simply scaled up, the ratio of its volume to surface area increases, because surface increases as the square of length or width while volume increases proportionally to the cube. The amount of tissue that needs food and oxygen is determined by the volume, but the pumping capacity that supplies food and oxygen depends on the area covered by choanocytes. Asconoid sponges seldom exceed 1 mm (0.039 in) in diameter.

Some sponges overcome this limitation by adopting the "syconoid" structure, in which the body wall is pleated. The inner pockets of the pleats are lined with choanocytes, which connect to the outer pockets of the pleats by ostia. This increase in the number of choanocytes and hence in pumping capacity enables syconoid sponges to grow up to a few centimeters in diameter.

The "leuconoid" pattern boosts pumping capacity further by filling the interior almost completely with mesohyl that contains a network of chambers lined with choanocytes and connected to each other and to the water intakes and outlet by tubes. Leuconid sponges grow to over 1 m (3.3 ft) in diameter, and the fact that growth in any direction increases the number of choanocyte chambers enables them to take a wider range of forms, for example "encrusting" sponges whose shapes follow those of the surfaces to which they attach. All freshwater and most shallow-water marine sponges have leuconid bodies. The networks of water passages in glass sponges are similar to the leuconid structure. In all three types of structure the cross-section area of the choanocyte-lined regions is

much greater than that of the intake and outlet channels. This makes the flow slower near the choanocytes and thus makes it easier for them to trap food particles. For example, in *Leuconia*, a small leuconoid sponge about 10 centimetres (3.9 in) tall and 1 centimetre (0.39 in) in diameter, water enters each of more than 80,000 intake canals at 6 cm per minute. However, because *Leuconia* has more than 2 million flagellated chambers whose combined diameter is much greater than that of the canals, water flow through chambers slows to 3.6 cm per hour, making it easy for choanocytes to capture food. All the water is expelled through a single osculum at about 8.5 cm per second, fast enough to carry waste products some distance away.

The mesohyl functions as an endoskeleton in most sponges, and is the only skeleton in soft sponges that encrust hard surfaces such as rocks. More commonly the mesohyl is stiffened by mineral spicules, by spongin fibers or both. Spicules, which are present in most but not all species, may be made of silica or calcium carbonate, and vary in shape from simple rods to three-dimensional "stars" with up to six rays. Spicules are produced by sclerocyte cells, and may be separate, connected by joints, or fused.

Some sponges also secrete exoskeletons that lie completely outside their organic components. For example, sclerosponges ("hard sponges") have massive calcium carbonate exoskeletons over which the organic matter forms a thin layer with choanocyte chambers in pits in the mineral. These exoskeletons are secreted by the pinacocytes that form the animals' skins.

Although adult sponges are fundamentally sessile animals, some marine and freshwater species can move across the sea bed at speeds of 1–4 mm (0.039–0.157 in) per day, as a result of amoeba-like movements of pinacocytes and other cells. A few species can contract their whole bodies, and many can close their oscula and ostia. Juveniles drift or swim freely, while adults are stationary.

Sponges do not have distinct circulatory, respiratory, digestive, and excretory systems – instead the water flow system supports all these functions. They filter food particles out of the water flowing through them. Particles larger than 50 micrometers cannot enter the ostia and pinacocytes consume them by phagocytosis (engulfing and internal digestion). Particles from 0.5 µm to 50 µm are trapped in the ostia, which taper from the outer to inner ends. These particles are consumed by pinacocytes or by archaeocytes which partially extrude themselves through the walls of the ostia. Bacteria-sized particles, below 0.5 micrometers, pass through the ostia and are caught and consumed by choanocytes. Since the smallest particles are

by far the most common, choanocytes typically capture 80% of a sponge's food supply. Archaeocytes transport food packaged in vesicles from cells that directly digest food to those that do not. At least one species of sponge has internal fibers that function as tracks for use by nutrient-carrying archaeocytes, and these tracks also move inert objects.

Sponges' cells absorb oxygen by diffusion from water into cells as water flows through body, into which carbon dioxide and other soluble waste products such as ammonia also diffuse. Archeocytes remove mineral particles that threaten to block the ostia, transport them through the mesohyl and generally dump them into the outgoing water current, although some species incorporate them into their skeletons.

Sponges have three asexual methods of reproduction: after fragmentation; by budding; and by producing gemmules. Fragments of sponges may be detached by currents or waves. They use the mobility of their pinacocytes and choanocytes and reshaping of the mesohyl to re-attach themselves to a suitable surface and then rebuild themselves as small but functional sponges over the course of several days. The same capabilities enable sponges that have been squeezed through a fine cloth to regenerate. A sponge fragment can only regenerate if it contains both collencytes to produce mesohyl and archeocytes to produce all the other cell types. A very few species reproduce by budding.

Gemmules are "survival pods" which a few marine sponges and many freshwater species produce by the thousands when dying and which some, mainly freshwater species, regularly produce in autumn. Spongocytes make gemmules by wrapping shells of spongin, often reinforced with spicules, round clusters of archeocytes that are full of nutrients. Freshwater gemmules may also include photosynthesizing symbionts. The gemmules then become dormant, and in this state can survive cold, drying out, lack of oxygen and extreme variations in salinity. Freshwater gemmules often do not revive until the temperature drops, stays cold for a few months and then reaches a near—"normal" level. When a gemmule germinates, the archeocytes round the outside of the cluster transform into pinacocytes, a membrane over a pore in the shell bursts, the cluster of cells slowly emerges, and most of the remaining archeocytes transform into other cell types needed to make a functioning sponge. Gemmules from the same species but different individuals can join forces to form one sponge. Some gemmules are retained within the parent sponge, and in spring it can be difficult to tell whether an old sponge has revived or been "recolonized" by its own gemmules.

Most sponges are hermaphrodites (function as both

sexes simultaneously), although sponges have no gonads (reproductive organs). Sperm are produced by choanocytes or entire choanocyte chambers that sink into the mesohyl and form spermatic cysts while eggs are formed by transformation of archeocytes, or of choanocytes in some species. Each egg generally acquires a yolk by consuming "nurse cells". During spawning, sperm burst out of their cysts and are expelled via the osculum. If they contact another sponge of the same species, the water flow carries them to choanocytes that engulf them but, instead of digesting them, metamorphose to an ameboid form and carry the sperm through the mesohyl to eggs, which in most cases engulf the carrier and its cargo.

A few species release fertilized eggs into the water, but most retain the eggs until they hatch. There are four types of larvae, but all are balls of cells with an outer layer of cells whose flagellae or cilia enable the larvae to move. After swimming for a few days the larvae sink and crawl until they find a place to settle. Most of the cells transform into archeocytes and then into the types appropriate for their locations in a miniature adult sponge.

Glass sponge embryos start by dividing into separate cells, but once 32 cells have formed they rapidly transform into larvae that externally are ovoid with a band of cilia round the middle that they use for movement, but internally have the typical glass sponge structure of spicules with a cobweb-like main syncitium draped around and between them and choanosyncitia with multiple collar bodies in the center. The larvae then leave their parents' bodies.

Sponges in temperate regions live for at most a few years, but some tropical species and perhaps some deep-ocean ones may live for 200 years or more. Some calcified demosponges grow by only 0.2 mm (0.0079 in) per year and, if that rate is constant, specimens 1 m (3.3 ft) wide must be about 5,000 years old. Some sponges start sexual reproduction when only a few weeks old, while others wait until they are several years old.

Adult sponges lack neurons or any other kind of nervous tissue. However, most species have the ability to perform movements that are coordinated all over their bodies, mainly contractions of the pinacocytes, squeezing the water channels and thus expelling excess sediment and other substances that may cause blockages. Some species can contract the osculum independently of the rest of the body. Sponges may also contract in order to reduce the area that is vulnerable to attack by predators. In cases where two sponges are fused, for example if there is a large but still unseparated bud, these contraction waves slowly become coordinated in both of the "Siamese twins". The coordinating mechanism is unknown, but may involve chemicals similar to neurotransmitters. However, glass sponges rapidly transmit electrical impulses through all parts of the syn-

cytium, and use this to halt the motion of their flagella if the incoming water contains toxins or excessive sediment. Myocytes are thought to be responsible for closing the osculum and for transmitting signals between different parts of the body.

Sponges contain genes very similar to those that contain the "recipe" for the post-synaptic density, an important signal-receiving structure in the neurons of all other animals. However, in sponges these genes are only activated in "flask cells" that appear only in larvae and may provide some sensory capability while the larvae are swimming. This raises questions about whether flask cells represent the predecessors of true neurons or are evidence that sponges' ancestors had true neurons but lost them as they adapted to a sessile lifestyle.

Sponges are worldwide in their distribution, living in a wide range of ocean habitats, from the polar regions to the tropics. Most live in quiet, clear waters, because sediment stirred up by waves or currents would block their pores, making it difficult for them to feed and breathe. The greatest numbers of sponges are usually found on firm surfaces such as rocks, but some sponges can attach themselves to soft sediment by means of a root-like base.

Sponges are more abundant but less diverse in temperate waters than in tropical waters, possibly because organisms that prey on sponges are more abundant in tropical waters. Glass sponges are the most common in polar waters and in the depths of temperate and tropical seas, as their very porous construction enables them to extract food from these resource-poor waters with the minimum of effort. Demosponges and calcareous sponges are abundant and diverse in shallower non-polar waters

Table 19.2: Comparison of 4 classes of sponges.

	Type of cells	Spicules	Spongin fibers	Massive exoskeleton	Body form
Calcarea	Single nucleus, single external membrane	Calcite May be individual or large masses	Never	Common. Made of calcite if present.	Asconoid, syconoid, leuconoid or solenoid
Hexactinellida	Mostly syncytia in all species	Silica May be individual or fused	Never	Never	Leuconoid
Demospongiae	Single nucleus, single external membrane	Silica	In many species	In some species. Made of aragonite if present.	Leuconoid
Homoscleromorpha	Single nucleus, single external membrane	Silica	In many species	Never	Sylleibid or leuconoid

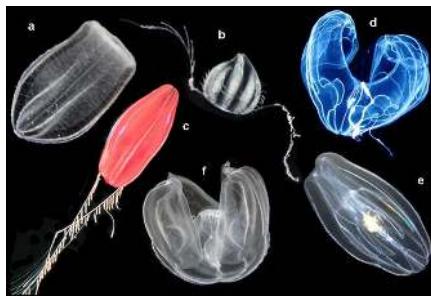


Figure 19.10: Pelagic (open ocean) ctenophores (a) *Beroe ovata*, (b) *Euplokamis* sp., (c) *Nepheloctena* sp., (d) *Bathocyroe fosteri*, (e) *Mnemiopsis leidyi*, and (f) *Ocyropsis* sp.¹³

19.1.2 Ctenophora

Ctenophora¹² (from Ancient Greek: κτείς, romanized: kteis, lit. ‘comb’ and φέρω, pherō, ‘to carry’; commonly known as comb jellies) comprise a phylum of invertebrate animals that live in marine waters worldwide. They are notable for the groups of cilia they use for swimming (commonly referred to as “combs”), and they are the largest animals to swim with the help of cilia. Depending on the species, adult ctenophores range from a few millimeters to 1.5 m (4 ft 11 in) in size. Only 100 to 150 species have been validated, and possibly another 25 have not been fully described and named. The textbook examples are cydippids with egg-shaped bodies and a pair of retractable tentacles fringed with tentilla (“little tentacles”) that are covered with colloblasts, sticky cells that capture prey. Their bodies consist of a mass of jelly, with a layer two cells thick on the outside, and another lining the internal cavity. The phylum has a wide range of body forms, including the egg-shaped cydippids with retractable tentacles that capture prey, the flat generally combless platyctenids, and the large-mouthed beroids, which prey on other ctenophores.

Almost all ctenophores function as predators, taking prey ranging from microscopic larvae and rotifers to the adults of small crustaceans; the exceptions are juveniles of two species, which live as parasites on the salps on which adults of their species feed.

Despite their soft, gelatinous bodies, fossils thought to represent ctenophores appear in lagerstätten dating as far back as the early Cambrian, about 525 million years ago. The position of the ctenophores in the “tree of life” has long been debated in molecular phylogenetics studies. Biologists proposed that ctenophores constitute the second-earliest branching animal lineage, with sponges being the sister-group to all other multicellular animals. Other biologists once believed that ctenophores were

emerging earlier than the sponges, which themselves appeared before the split between cnidarians and bilaterians. However reanalysis of the data showed that the computer algorithms used for analysis were misled by the presence of specific ctenophore genes that were markedly different from those of other species. Molecular phylogenetics studies indicate that the common ancestor of modern ctenophores was cydippid-like, descending from various cydippids after the Cretaceous–Paleogene extinction event 66 million years ago.

Among animal phyla, the Ctenophores are more complex than sponges, about as complex as cnidarians (jellyfish, sea anemones, etc.), and less complex than bilaterians (which include almost all other animals). Unlike sponges, both ctenophores and cnidarians have: cells bound by inter-cell connections and carpet-like basement membranes; muscles; nervous systems; and some have sensory organs. Ctenophores are distinguished from all other animals by having colloblasts, which are sticky and adhere to prey, although a few ctenophore species lack them.

Like sponges and cnidarians, ctenophores have two main layers of cells that sandwich a middle layer of jelly-like material, which is called the mesoglea in cnidarians and ctenophores; more complex animals have three main cell layers and no intermediate jelly-like layer. Hence ctenophores and cnidarians have traditionally been labelled diploblastic, along with sponges. Both ctenophores and cnidarians have a type of muscle that, in more complex animals, arises from the middle cell layer, and as a result some recent text books classify ctenophores as triploblastic, while others still regard them as diploblastic. The comb jellies have more than 80 different cell types, exceeding the numbers from other groups like placozoans, sponges, cnidarians, and some deep-branching bilaterians.

Ranging from about 1 millimeter (0.039 in) to 1.5 meters (4.9 ft) in size, ctenophores are the largest non-colonial animals that use cilia (“hairs”) as their main method of locomotion. Most species have eight strips, called comb rows, that run the length of their bodies and bear comb-like bands of cilia, called “ctenes”, stacked along the comb rows so that when the cilia beat, those of each comb touch the comb below.

The phylogenetic relationship of ctenophores to the rest of Metazoa is very important to our understanding of the early evolution of animals and the origin of multicellularity. It has been the focus of debate for many years. Ctenophores have been purported to be the sister lineage to the Bilateria, sister to the Cnidaria, sister to Cnidaria, Placozoa, and Bilateria, and sister to all other animals.

¹²<https://en.wikipedia.org/wiki/Ctenophora>

A series of studies that looked at the presence and absence of members of gene families and signalling pathways (e.g., homeoboxes, nuclear receptors, the Wnt signaling pathway, and sodium channels) showed evidence congruent with the latter two scenarios, that ctenophores are either sister to Cnidaria, Placozoa, and Bilateria or sister to all other animal phyla. Several more recent studies comparing complete sequenced genomes of ctenophores with other sequenced animal genomes have also supported ctenophores as the sister lineage to all other animals. This position would suggest that neural and muscle cell types either were lost in major animal lineages (e.g., Porifera and Placozoa) or evolved independently in the ctenophore lineage.

Other researchers have argued that the placement of Ctenophora as sister to all other animals is a statistical anomaly caused by the high rate of evolution in ctenophore genomes, and that Porifera (sponges) is the earliest-diverging animal taxon instead. As such, the Ctenophora appear to be a basal diploblast clade. In agreement with the latter point, the analysis of a very large sequence alignment at the metazoan taxonomic scale (1,719 proteins totalizing ca. 400,000 amino acid positions) showed that ctenophores emerge as the second-earliest branching animal lineage, and sponges are sister-group to all other multicellular animals. Also, research on mucin genes, which allow an animal to produce mucus, shows that sponges have never had them while all other animals, including comb jellies, appear to share genes with a common origin.

19.1.3 Cnidaria

Cnidaria¹⁴ is a phylum under kingdom Animalia containing over 11,000 species of aquatic animals found both in freshwater and marine environments: they are predominantly marine.

Their distinguishing feature is cnidocytes, specialized cells that they use mainly for capturing prey. Their bodies consist of mesoglea, a non-living jelly-like substance, sandwiched between two layers of epithelium that are mostly one cell thick.

They have two basic body forms: swimming medusae and sessile polyps, both of which are radially symmetrical with mouths surrounded by tentacles that bear cnidocytes. Both forms have a single orifice and body cavity that are used for digestion and respiration. Many cnidarian species produce colonies that are single organisms composed of medusa-like or polyp-like zooids, or both (hence they are trimorphic). Cnidarians' activities are coordinated by a decentralized nerve net and simple receptors. Several

free-swimming species of Cubozoa and Scyphozoa possess balance-sensing statocysts, and some have simple eyes. Not all cnidarians reproduce sexually, with many species having complex life cycles of asexual polyp stages and sexual medusae. Some, however, omit either the polyp or the medusa stage.

Cnidarians are classified into four main groups: the almost wholly sessile Anthozoa (sea anemones, corals, sea pens); swimming Scyphozoa (jellyfish); Cubozoa (box jellies); and Hydrozoa (a diverse group that includes all the freshwater cnidarians as well as many marine forms, and has both sessile members, such as Hydra, and colonial swimmers, such as the Portuguese Man o' War). Staurozoa have recently been recognised as a class in their own right rather than a sub-group of Scyphozoa, and the parasitic Myxozoa and Polypodiozoa were firmly recognized as cnidarians in 2007.

¹⁴<https://en.wikipedia.org/wiki/Cnidaria>

Table 19.3: Comparison of 4 classes of cnidaria.

	Hydrozoa	Scyphozoa	Cubozoa	Anthozoa	Myxozoa
Number of species	3,600	228	42	6,100	1300
Examples	Hydra, siphonophores	Jellyfish	Box jellies	Sea anemones, corals, sea pens	<i>Myxobolus</i> <i>cerebralis</i>
Cells found in mesoglea	No	Yes	Yes	Yes	
Nematocysts in exodermis	No	Yes	Yes	Yes	
Medusa phase in life cycle	In some species	Yes	Yes	No	
Number of medusae produced per polyp	Many	Many	One	(not applicable)	



Figure 19.11: Four examples of Cnidaria.¹⁵ A jellyfish *Chrysaora melanaster*, a gorgonian *Annella mollis*, a rocky coral *Acropora cervicornis*, and a sea anemone *Nemanthus annamensis*.

Most cnidarians prey on organisms ranging in size from plankton to animals several times larger than themselves, but many obtain much of their nutrition from dinoflagellates, and a few are parasites. Many are preyed on by other animals including starfish, sea slugs, fish, turtles, and even other cnidarians. Many scleractinian corals—which form the structural foundation for coral reefs—possess polyps that are filled with symbiotic photo-synthetic zooxanthellae. While reef-forming corals are almost entirely restricted to warm and shallow marine waters, other cnidarians can be found at great depths, in polar regions, and in freshwater.

Recent phylogenetic analyses support monophyly of cnidarians, as well as the position of cnidarians as the sister group of bilaterians. Fossil cnidarians have been found in rocks formed about 580 million years ago, and other fossils show that corals may have been present shortly before 490 million years ago and diversified a few million years later. However, molecular clock analysis of mitochondrial genes suggests a much older age for the crown group of cnidarians, estimated around 741 million years ago, almost 200 million years before the Cambrian period as well as any fossils.

Cnidarians form a phylum of animal that are more complex than sponges, about as complex as ctenophores (comb jellies), and less complex than bilaterians, which include almost all other animals. Both cnidarians and ctenophores are more complex than sponges as they have: cells bound by inter-cell connections and carpet-like basement membranes; muscles; nervous systems; and some have sensory organs. Cnidarians are distinguished from all other animals by having cnidocytes that fire harpoon like structures and are usually used mainly to capture prey. In some species, cnidocytes can also be used as anchors. Cnidarians are also distinguished by the fact that they have only one opening in their body for

ingestion and excretion i.e. they don't have a separate mouth and anus.

Like sponges and ctenophores, cnidarians have two main layers of cells that sandwich a middle layer of jelly-like material, which is called the mesoglea in cnidarians; more complex animals have three main cell layers and no intermediate jelly-like layer. Hence, cnidarians and ctenophores have traditionally been labelled diploblastic, along with sponges. However, both cnidarians and ctenophores have a type of muscle that, in more complex animals, arises from the middle cell layer. As a result, some recent text books classify ctenophores as triploblastic, and it has been suggested that cnidarians evolved from triploblastic ancestors.

Most adult cnidarians appear as either free-swimming medusae or sessile polyps, and many hydrozoans species are known to alternate between the two forms.

Both are radially symmetrical, like a wheel and a tube respectively. Since these animals have no heads, their ends are described as "oral" (nearest the mouth) and "aboral" (furthest from the mouth).

Most have fringes of tentacles equipped with cnidocytes around their edges, and medusae generally have an inner ring of tentacles around the mouth. Some hydroids may consist of colonies of zooids that serve different purposes, such as defense, reproduction and catching prey. The mesoglea of polyps is usually thin and often soft, but that of medusae is usually thick and springy, so that it returns to its original shape after muscles around the edge have contracted to squeeze water out, enabling medusae to swim by a sort of jet propulsion.

Cnidaria are diploblastic animals; in other words, they have two main cell layers, while more complex animals are triploblasts having three main layers. The two main cell layers of cnidarians form epithelia that are mostly one cell thick, and are attached to a fibrous basement membrane, which they secrete. They also secrete the jelly-like mesoglea that separates the layers. The layer that faces outwards, known as the ectoderm ("outside skin"), generally contains the following types of cells:

- Epitheliomuscular cells whose bodies form part of the epithelium but whose bases extend to form muscle fibers in parallel rows. The fibers of the outward-facing cell layer generally run at right angles to the fibers of the inward-facing one. In Anthozoa (anemones, corals, etc.) and Scyphozoa (jellyfish), the mesoglea also contains some muscle cells.
- Cnidocytes, the harpoon-like "nettle cells" that give the phylum Cnidaria its name. These appear between or sometimes on top of the muscle cells.

- Nerve cells. Sensory cells appear between or sometimes on top of the muscle cells, and communicate via synapses (gaps across which chemical signals flow) with motor nerve cells, which lie mostly between the bases of the muscle cells. Some form a simple nerve net.
- Interstitial cells, which are unspecialized and can replace lost or damaged cells by transforming into the appropriate types. These are found between the bases of muscle cells.

In addition to epitheliomuscular, nerve and interstitial cells, the inward-facing gastroderm ("stomach skin") contains gland cells that secrete digestive enzymes. In some species it also contains low concentrations of cnidocytes, which are used to subdue prey that is still struggling.

The mesoglea contains small numbers of amoeba-like cells, and muscle cells in some species. However, the number of middle-layer cells and types are much lower than in sponges.

Medusae swim by a form of jet propulsion: muscles, especially inside the rim of the bell, squeeze water out of the cavity inside the bell, and the springiness of the mesoglea powers the recovery stroke. Since the tissue layers are very thin, they provide too little power to swim against currents and just enough to control movement within currents.

Hydras and some sea anemones can move slowly over rocks and sea or stream beds by various means: creeping like snails, crawling like inchworms, or by somersaulting. A few can swim clumsily by wagging their bases.

Cnidarians are generally thought to have no brains or even central nervous systems. However, they do have integrative areas of neural tissue that could be considered some form of centralization. Most of their bodies are innervated by decentralized nerve nets that control their swimming musculature and connect with sensory structures, though each clade has slightly different structures. These sensory structures, usually called rhopalia, can generate signals in response to various types of stimuli such as light, pressure, and much more. Medusa usually have several of them around the margin of the bell that work together to control the motor nerve net, that directly innervates the swimming muscles. Most Cnidarians also have a parallel system. In scyphozoans, this takes the form of a diffuse nerve net, which has modulatory effects on the nervous system. As well as forming the "signal cables" between sensory neurons and motoneurons, intermediate neurons in the nerve net can also form ganglia that act as local coordination centers. Communication between nerve cells can occur by chemical synapses or gap junctions in hydrozoans, though gap junctions are not present in all groups. Cnidarians have many of the same

neurotransmitters as many animals, including chemicals such as glutamate, GABA, and acetylcholine.

This structure ensures that the musculature is excited rapidly and simultaneously, and can be directly stimulated from any point on the body, and it also is better able to recover after injury.

Medusae and complex swimming colonies such as siphonophores and chondrophores sense tilt and acceleration by means of statocysts, chambers lined with hairs which detect the movements of internal mineral grains called statoliths. If the body tilts in the wrong direction, the animal rights itself by increasing the strength of the swimming movements on the side that is too low. Most species have ocelli ("simple eyes"), which can detect sources of light. However, the agile box jellyfish are unique among Medusae because they possess four kinds of true eyes that have retinas, corneas and lenses. Although the eyes probably do not form images, Cubozoa can clearly distinguish the direction from which light is coming as well as negotiate around solid-colored objects.

Cnidarians feed in several ways: predation, absorbing dissolved organic chemicals, filtering food particles out of the water, obtaining nutrients from symbiotic algae within their cells, and parasitism. Most obtain the majority of their food from predation but some, including the corals *Hetoxenia* and *Leptogorgia*, depend almost completely on their endosymbionts and on absorbing dissolved nutrients. Cnidaria give their symbiotic algae carbon dioxide, some nutrients, a place in the sun and protection against predators.

Predatory species use their cnidocytes to poison or entangle prey, and those with venomous nematocysts may start digestion by injecting digestive enzymes. The "smell" of fluids from wounded prey makes the tentacles fold inwards and wipe the prey off into the mouth. In medusae the tentacles round the edge of the bell are often short and most of the prey capture is done by "oral arms", which are extensions of the edge of the mouth and are often frilled and sometimes branched to increase their surface area. Medusae often trap prey or suspended food particles by swimming upwards, spreading their tentacles and oral arms and then sinking. In species for which suspended food particles are important, the tentacles and oral arms often have rows of cilia whose beating creates currents that flow towards the mouth, and some produce nets of mucus to trap particles. Their digestion is both intra and extracellular.

Once the food is in the digestive cavity, gland cells in the gastroderm release enzymes that reduce the prey to slurry, usually within a few hours. This circulates through the digestive cavity and, in colonial cnidarians, through the

connecting tunnels, so that gastroderm cells can absorb the nutrients. Absorption may take a few hours, and digestion within the cells may take a few days. The circulation of nutrients is driven by water currents produced by cilia in the gastroderm or by muscular movements or both, so that nutrients reach all parts of the digestive cavity. Nutrients reach the outer cell layer by diffusion or, for animals or zooids such as medusae which have thick mesogleas, are transported by mobile cells in the mesoglea.

Indigestible remains of prey are expelled through the mouth. The main waste product of cells' internal processes is ammonia, which is removed by the external and internal water currents.

There are no respiratory organs, and both cell layers absorb oxygen from and expel carbon dioxide into the surrounding water. When the water in the digestive cavity becomes stale it must be replaced, and nutrients that have not been absorbed will be expelled with it. Some Anthozoa have ciliated grooves on their tentacles, allowing them to pump water out of and into the digestive cavity without opening the mouth. This improves respiration after feeding and allows these animals, which use the cavity as a hydrostatic skeleton, to control the water pressure in the cavity without expelling undigested food.

Cnidaria that carry photosynthetic symbionts may have the opposite problem, an excess of oxygen, which may prove toxic. The animals produce large quantities of antioxidants to neutralize the excess oxygen.

Cnidarian sexual reproduction often involves a complex life cycle with both polyp and medusa stages. For example, in Scyphozoa (jellyfish) and Cubozoa (box jellies) a larva swims until it finds a good site, and then becomes a polyp. This grows normally but then absorbs its tentacles and splits horizontally into a series of disks that become juvenile medusae, a process called strobilation. The juveniles swim off and slowly grow to maturity, while the polyp re-grows and may continue strobilating periodically. The adults have gonads in the gastroderm, and these release ova and sperm into the water in the breeding season.

This phenomenon of succession of differently organized generations (one asexually reproducing, sessile polyp, followed by a free-swimming medusa or a sessile polyp that reproduces sexually) is sometimes called "alternation of asexual and sexual phases" or "metagenesis", but should not be confused with the alternation of generations as found in plants.

Spawning is generally driven by environmental factors such as changes in the water temperature, and their release is triggered by lighting conditions such as sunrise, sunset or the phase of the moon. Many species of Cnidaria may spawn simultaneously in the same location, so that



Figure 19.12: Diversity of bilaterians.¹⁶

there are too many ova and sperm for predators to eat more than a tiny percentage — one famous example is the Great Barrier Reef, where at least 110 corals and a few non-cnidarian invertebrates produce enough gametes to turn the water cloudy. These mass spawnings may produce hybrids, some of which can settle and form polyps, but it is not known how long these can survive. In some species the ova release chemicals that attract sperm of the same species.

The fertilized eggs develop into larvae by dividing until there are enough cells to form a hollow sphere (blastula) and then a depression forms at one end (gastrulation) and eventually becomes the digestive cavity. However, in cnidarians the depression forms at the end further from the yolk (at the animal pole), while in bilaterians it forms at the other end (vegetal pole). The larvae, called planulae, swim or crawl by means of cilia. They are cigar-shaped but slightly broader at the "front" end, which is the aboral, vegetal-pole end and eventually attaches to a substrate if the species has a polyp stage.

Anthozoan larvae either have large yolks or are capable of feeding on plankton, and some already have endosymbiotic algae that help to feed them. Since the parents are immobile, these feeding capabilities extend the larvae's range and avoid overcrowding of sites. Scyphozoan and hydrozoan larvae have little yolk and most lack endosymbiotic algae, and therefore have to settle quickly and metamorphose into polyps. Instead, these species rely on their medusae to extend their ranges.

19.2 Bilaterian animals

The remaining animals, the great majority—comprising some 29 phyla and over a million species—form a clade, the Bilateria. The body is triploblastic, with three well-developed germ layers, and their tissues form distinct organs. The digestive chamber has two openings, a mouth and an anus, and there is an internal body cavity, a coelom or pseudocoelom. Animals with this bilaterally symmetric body plan and a tendency to move in one direction have a head end (anterior) and a tail end (posterior) as well as a back (dorsal) and a belly (ventral); therefore they also have a left side and a right side.

Having a front end means that this part of the body

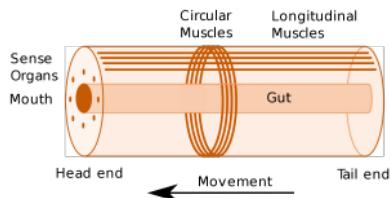


Figure 19.13: Idealised bilaterian body plan.¹⁷ With an elongated body and a direction of movement the animal has head and tail ends. Sense organs and mouth form the basis of the head. Opposed circular and longitudinal muscles enable peristaltic motion.



Figure 19.14: *Xenoturbella japonica*, a xenacoelomorph member (xenoturbellids).¹⁸

encounters stimuli, such as food, favouring cephalisation, the development of a head with sense organs and a mouth. Many bilaterians have a combination of circular muscles that constrict the body, making it longer, and an opposing set of longitudinal muscles, that shorten the body; these enable soft-bodied animals with a hydrostatic skeleton to move by peristalsis. They also have a gut that extends through the basically cylindrical body from mouth to anus. Many bilaterian phyla have primary larvae which swim with cilia and have an apical organ containing sensory cells. However, there are exceptions to each of these characteristics; for example, adult echinoderms are radially symmetric (unlike their larvae), while some parasitic worms have extremely simplified body structures.

Genetic studies have considerably changed zoologists' understanding of the relationships within the Bilateria. Most appear to belong to two major lineages, the protostomes and the deuterostomes that together form the Nephrozoa. Their sister clade are the Xenacoelomorpha, the basalmost bilaterian phylum of small and very simple animals. All xenacoelomorphs lack a typical stomatogastric system, i.e., they do not have a true gut and lack an excretory system, circulatory and respiratory system. The nervous system is a simple nerve net without a brain.

Table 19.4: Comparison of major characteristics of porifera, cnidaria, ctenophora and bilateria.

	Sponges	Cnidarians	Ctenophores	Bilateria
Cnidocytes	No	Yes	No	No
Colloblasts	No	No	Yes	No
Digestive and circulatory organs	No	No	No	Yes
Number of main cell layers	Two, with jelly-like layer between them	Three	Two or Three	Three
Cells in each layer bound together	cell-adhesion molecules, but no basement membranes except Homoscleromorpha.	inter-cell connections; basement membranes	inter-cell connections; basement membranes	inter-cell connections; basement membranes
Sensory organs	No	Yes	Yes	Yes
Number of cells in middle "jelly" layer	Many	Few	Few	(Not applicable)
Cells in outer layers	Yes	No	No	(Not applicable)
can move inwards and change functions				
Nervous system	No	Yes, simple	Yes, simple	Simple to complex
Muscles	None	Mostly epitheliomuscular	Mostly myoepithelial	Mostly myocytes

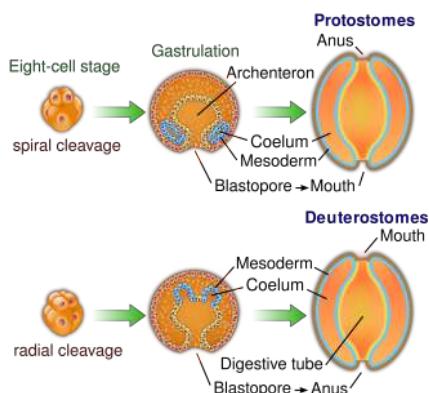


Figure 19.15: The bilaterian gut develops in two ways. In many protostomes, the blastopore develops into the mouth, while in deuterostomes it becomes the anus.¹⁹

19.2.1 Protostomes and deuterostomes

Protostomes and deuterostomes differ in several ways. Early in development, deuterostome embryos undergo radial cleavage during cell division, while many protostomes (the Spiralia) undergo spiral cleavage. Animals from both groups possess a complete digestive tract, but in protostomes the first opening of the embryonic gut develops into the mouth, and the anus forms secondarily. In deuterostomes, the anus forms first while the mouth develops secondarily. Most protostomes have schizocoelous development, where cells simply fill in the interior of the gastrula to form the mesoderm. In deuterostomes, the mesoderm forms by enterocoelic pouching, through invagination of the endoderm.

The main deuterostome phyla are the Echinodermata and the Chordata. Echinoderms are exclusively marine and include starfish, sea urchins, and sea cucumbers. The chordates are dominated by the vertebrates (animals with backbones), which consist of fishes, amphibians, reptiles, birds, and mammals. The deuterostomes also include the Hemichordata (acorn worms).

19.2.2 Echinoderms

Echinoderm²⁰ is the common name given to any member of the phylum Echinodermata (from Ancient Greek, ἔχινος, echinos – “hedgehog” and δέρμα, derma – “skin”) of marine animals. The adults are recognizable by their (usually five-point) radial symmetry, and include starfish, sea urchins, sand dollars, and sea cucumbers, as well as the sea lilies or “stone lilies”. Echinoderms are found at every ocean depth, from the intertidal zone to the abyssal zone. The phylum contains about 7000 living species, making it the second-largest grouping of deuterostomes (a superphylum), after



Figure 19.16: A sea cucumber from Malaysia.²¹



Figure 19.17: Starfish exhibit a wide range of colours.²²

the chordates (which include the vertebrates, such as birds, fishes, mammals, and reptiles). Echinoderms are also the largest phylum that has no freshwater or terrestrial (land-based) representatives.

Aside from the hard-to-classify Arkarua (a Precambrian animal with echinoderm-like pentamerous radial symmetry), the first definitive members of the phylum appeared near the start of the Cambrian. One group of Cambrian echinoderms, the cinctans (Homalozoa), which are close to the base of the echinoderm origin, have been found to possess external gills used for filter feeding, similar to those possessed by chordates and hemichordates.

The echinoderms are important both ecologically and geologically. Ecologically, there are few other groupings so abundant in the biotic desert of the deep sea, as well as shallower oceans. Most echinoderms are able to reproduce asexually and regenerate tissue, organs, and limbs; in some cases, they can undergo complete regeneration from a single limb. Geologically, the value of echinoderms is in their ossified skeletons, which are major contributors to many limestone formations, and can provide valuable clues as to the geological environment. They were the most used species in regenerative research in the 19th and 20th centuries. Further, some scientists hold that the radiation of echinoderms was responsible for the Mesozoic Marine Revolution.

²⁰<https://en.wikipedia.org/wiki/Echinoderm>



Figure 19.18: *Strongylocentrotus purpuratus*, a well-armoured sea urchin.²³

Along with the chordates and hemichordates, echinoderms are deuterostomes, one of the two major divisions of the bilaterians, the other being the protostomes. During the early development of the embryo, in deuterostomes, the blastopore (the first opening to form) becomes the anus whereas in the protostomes, it becomes the mouth. In deuterostomes, the mouth develops at a later stage, at the opposite end of the blastopore from the blastopore, and a gut forms connecting the two. The larvae of echinoderms have bilateral symmetry but this is lost during metamorphosis when their bodies are reorganised and develop the characteristic radial symmetry of the echinoderm, typically pentamerism. The characteristics of adult echinoderms are the possession of a water vascular system with external tube feet and a calcareous endoskeleton consisting of ossicles connected by a mesh of collagen fibres. A 2014 analysis of 219 genes from all classes of echinoderms gives the following phylogenetic tree.

There are a total of about 7,000 extant species of echinoderm as well as about 13,000 extinct species. They are found in habitats ranging from shallow intertidal areas to abyssal depths. Two main subdivisions are traditionally recognised: the more familiar motile Euteutherozoa, which encompasses the Asteroidea (starfish, 1,745 recent species), Ophiozoidea (brittle stars, 2,300 species), Echinozoa (sea urchins and sand dollars, 900 species) and Holothurozoa (sea cucumbers, 1,430 species); and the Pelmatozoa, some of which are sessile while others move around. These consist of the Crinozoa (feather stars and sea lilies, 580 species) and the extinct blastoids and Paracrinoids. A fifth class of Euteutherozoa consisting of just three species, the Concentricyclozoa (sea daisies), were recently merged into the Asterozoa.

Echinoderms evolved from animals with bilateral symmetry. Although adult echinoderms possess pentaradial, or five-sided, symmetry, echinoderm larvae are ciliated,

free-swimming organisms that organize in bilateral symmetry which makes them look like embryonic chordates. Later, the left side of the body grows at the expense of the right side, which is eventually absorbed. The left side then grows in a pentaradially symmetric fashion, in which the body is arranged in five parts around a central axis. Within the Asterozoa, there can be a few exceptions from the rule. The starfish genus *Leptasterias* normally have six arms, although five-armed individuals can occur. Also the Brisidae have six armed species. Amongst the brittle stars, six-armed species such as *Ophiothela danae*, *Ophiactis savignyi*, and *Ophionotus hexactis* exists, and *Ophiancantha vivipara* often has more than six.

Echinoderms exhibit secondary radial symmetry in portions of their body at some stage of life. This, however, is an adaptation to their sessile existence. They developed from other members of the Bilateria and exhibit bilateral symmetry in their larval stage. Many crinoids and some seastars exhibit symmetry in multiples of the basic five, with starfish such as *Labidiaster annulatus* known to possess up to fifty arms, and the sea-lily *Comaster schlegelii* having two hundred.

Echinoderms have a mesodermal skeleton composed of calcareous plates or ossicles. Each one of these, even the articulating spine of a sea urchin, is composed mineralogically of a crystal of calcite. If solid, these would form a heavy skeleton, so they have a sponge-like porous structure known as stereom. Ossicles may be fused together, as in the test of sea urchins, or may articulate with each other as in the arms of sea stars, brittle stars and crinoids. The ossicles may be flat plates or bear external projections in the form of spines, granules or warts and they are supported by a tough epidermis (skin). Skeletal elements are also deployed in some specialized ways, such as the "Aristotle's lantern" mouthparts of sea urchins used for grinding, the supportive stalks of crinoids and the structural "lime ring" of sea cucumbers.

Despite the robustness of the individual skeletal modules complete skeletons of starfish, brittle stars and crinoids are rare in the fossil record. This is because they quickly disarticulate (disconnect from each other) once the encompassing skin rots away, and in the absence of tissue there is nothing to hold the plates together. The modular construction is a result of the growth system employed by echinoderms, which adds new segments at the centre of the radial limbs, pushing the existing plates outwards and lengthening the arms. Sea urchins on the other hand are often well preserved in chalk beds or limestone. During fossilization, the cavities in the stereom are filled in with calcite that is in crystalline continuity with the surrounding material. On fracturing such rock, distinctive cleavage patterns can be seen and sometimes even the intricate

internal and external structure of the test.

The epidermis consists of cells responsible for the support and maintenance of the skeleton, as well as pigment cells, mechanoreceptor cells (which detect motion on the animal's surface), and sometimes gland cells which secrete sticky fluids or even toxins. The varied and often vivid colours of echinoderms are produced by the action of skin pigment cells. These are produced by a variable combination of coloured pigments, such as the dark melanin, red carotinoids, and carotene proteins, which can be blue, green, or violet. These may be light-sensitive, and as a result many echinoderms change appearance completely as night falls. The reaction can happen quickly – the sea urchin *Centrostephanus longispinus* changes from jet black to grey-brown in just fifty minutes when exposed to light.

One characteristic of most echinoderms is a special kind of tissue known as "catch connective tissue". This collagenous material can change its mechanical properties in a few seconds or minutes through nervous control rather than by muscular means. This tissue enables a starfish to change from moving flexibly around the seabed to becoming rigid while prying open a bivalve mollusc or preventing itself from being extracted from a crevice. Similarly, sea urchins can lock their normally mobile spines rigidly as a defensive mechanism when attacked.

Echinoderms possess a unique water vascular system. This is a network of fluid-filled canals derived from the coelom (body cavity) that function in gas exchange, feeding, sensory reception and locomotion. This system varies between different classes of echinoderm but typically opens to the exterior through a sieve-like madreporite on the aboral (upper) surface of the animal. The madreporite is linked to a slender duct, the stone canal, which extends to a ring canal that encircles the mouth or oesophagus. From this, radial canals extend along the arms of asteroids and adjoin the test in the ambulacral areas of echinoids. Short lateral canals branch off the radial canals, each one ending in an ampulla. Part of the ampulla can protrude through a pore (or a pair of pores in sea urchins) to the exterior and is known as a podium or tube feet. The water vascular system assists with the distribution of nutrients throughout the animal's body and is most obviously expressed in the tube feet which can be extended or contracted by the redistribution of fluid between the foot and the internal sac.

The organization of the system is somewhat different in ophiuroids where the madreporite may be on the oral surface and the podia lack suckers. In holothuroids, the podia may be reduced or absent and the madreporite opens into the body cavity so that the circulating liquid is coelomic fluid rather than sea water. The arrangements

in crinoids is similar to asteroids but the tube feet lack suckers and are used to pass food particles captured by the arms towards the central mouth. In the asteroids, the same wafting motion is employed to move the animal across the ground. Sea urchins use their feet to prevent the larvae of encrusting organisms from settling on their surfaces; potential settlers are moved to the urchin's mouth and eaten. Some burrowing sea stars extend their elongated dorsal tube feet to the surface of the sand or mud above and use them to absorb oxygen from the water column.

Echinoderms possess a simple digestive system which varies according to the animal's diet. Starfish are mostly carnivorous and have a mouth, oesophagus, two-part stomach, intestine and rectum, with the anus located in the centre of the aboral body surface. With a few exceptions, the members of the order Paxillosida do not possess an anus. In many species of starfish, the large cardiac stomach can be everted and digest food outside the body. In other species, whole food items such as molluscs may be ingested. Brittle stars have a blind gut with no intestine or anus. They have varying diets and expel food waste through their mouth. Sea urchins are herbivores and use their specialised mouthparts to graze, tear and chew algae and sometimes other animal or vegetable material. They have an oesophagus, a large stomach and a rectum with the anus at the apex of the test. Sea cucumbers are mostly detritivores, sorting through the sediment with their buccal tentacles which are modified tube feet. Sand and mud accompanies their food through their simple gut which has a long coiled intestine and a capacious cloaca. Crinoids are passive suspension feeders, catching plankton with their outstretched arms. Boluses of mucus-trapped food are passed to the mouth which is linked to the anus by a loop consisting of a short oesophagus and longer intestine.

The coelomic cavities of echinoderms are complex. Aside from the water vascular system, echinoderms have a haemal coelom (or haemal system, the "haemal" being a misnomer), a perivisceral coelom, a gonadal coelom and often also a perihæmal coelom (or perihæmal system). During development, echinoderm coelom is divided in metacoel, mesocoel and protoocoel (also called somatocoel, hydrocoel and axocoel, respectively). The water vascular system, haemal system and perihæmal system form the tubular coelomic system. Echinoderms are an exception having both a coelomic circulatory system (i.e., the water vascular system) and a haemal circulatory system (i.e., the haemal and perihæmal systems).

Haemal and perihæmal systems are derived from the coelom and form an open and reduced circulatory system. This usually consists of a central ring and five radial ves-

sels. There is no true heart and the blood often lacks any respiratory pigment. Gaseous exchange occurs via dermal branchiae or papulae in starfish, genital bursae in brittle stars, peristomial gills in sea urchins and cloacal trees in sea cucumbers. Exchange of gases also takes place through the tube feet. Echinoderms lack specialized excretory (waste disposal) organs and so nitrogenous waste, chiefly in the form of ammonia, diffuses out through the respiratory surfaces.

The coelomic fluid contains the coelomocytes, or immune cells. There are several types of immune cells, which vary among classes and species. All classes possess a type of phagocytic amebocyte, which engulf invading particles and infected cells, aggregate or clot, and may be involved in cytotoxicity. These cells are usually larger and granular, and are suggested to be a main line of defense against potential pathogens. Depending on the class, echinoderms may have spherule cells (for cytotoxicity, inflammation, and anti-bacterial activity), vibratile cells (for coelomic fluid movement and clotting), and crystal cells (potential osmoregulatory cells in sea cucumbers). The coelomocytes also secrete Anti-Microbial Peptides (AMPs) against bacteria, and have a set of lectins and complement proteins as part of an innate immune system that is still being characterized.

Echinoderms have a simple radial nervous system that consists of a modified nerve net consisting of interconnecting neurons with no central brain, although some do possess ganglia. Nerves radiate from central rings around the mouth into each arm or along the body wall; the branches of these nerves coordinate the movements of the organism and the synchronisation of the tube feet. Starfish have sensory cells in the epithelium and have simple eyespots and touch-sensitive tentacle-like tube feet at the tips of their arms. Sea urchins have no particular sense organs but do have statocysts that assist in gravitational orientation, and they have sensory cells in their epidermis, particularly in the tube feet, spines and pedicellariae. Brittle stars, crinoids and sea cucumbers in general do not have sensory organs but some burrowing sea cucumbers of the order Apodida have a single statocyst adjoining each radial nerve and some have an eyespot at the base of each tentacle.

The gonads occupy much of the body cavities of sea urchins and sea cucumbers, while the less voluminous crinoids, brittle stars and starfish have two gonads in each arm. While the ancestral condition is considered to be the possession of one genital aperture, many organisms have multiple gonopores through which eggs or sperm may be released.

Many echinoderms have remarkable powers of regeneration. Many species routinely autotomize and regenerate

arms and viscera. Sea cucumbers often discharge parts of their internal organs if they perceive themselves to be threatened. The discharged organs and tissues are regenerated over the course of several months. Sea urchins are constantly replacing spines lost through damage. Sea stars and sea lilies readily lose and regenerate their arms. In most cases, a single severed arm cannot grow into a new starfish in the absence of at least part of the disc. However, in a few species a single arm can survive and develop into a complete individual and in some species, the arms are intentionally detached for the purpose of asexual reproduction. During periods when they have lost their digestive tracts, sea cucumbers live off stored nutrients and absorb dissolved organic matter directly from the water.

Echinoderms become sexually mature after approximately two to three years, depending on the species and the environmental conditions. They are nearly all gonochoric, though a few species are hermaphroditic. The eggs and sperm cells are typically released into open water, where fertilization takes place. The release of sperm and eggs is synchronised in some species, usually with regard to the lunar cycle. In other species, individuals may aggregate during the reproductive season, thereby increasing the likelihood of successful fertilisation. Internal fertilisation has currently been observed in three species of sea star, three brittle stars and a deep water sea cucumber. Even at abyssal depths, where no light penetrates, synchronisation of reproductive activity in echinoderms is surprisingly frequent.

Some echinoderms brood their eggs. This is especially common in cold water species where planktonic larvae might not be able to find sufficient food. These retained eggs are usually few in number and are supplied with large yolks to nourish the developing embryos. In starfish, the female may carry the eggs in special pouches, under her arms, under her arched body or even in her cardiac stomach. Many brittle stars are hermaphrodites. Egg brooding is quite common and usually takes place in special chambers on their oral surfaces, but sometimes the ovary or coelom is used. In these starfish and brittle stars, direct development without passing through a bilateral larval stage usually takes place. A few sea urchins and one species of sand dollar carry their eggs in cavities, or near their anus, holding them in place with their spines. Some sea cucumbers use their buccal tentacles to transfer their eggs to their underside or back where they are retained. In a very small number of species, the eggs are retained in the coelom where they develop viviparously, later emerging through ruptures in the body wall. In some species of crinoid, the embryos develop in special breeding bags, where the eggs are held until sperm released by a male happens to find them.

The development of an echinoderm begins with a bilaterally symmetrical embryo, with a coeloblastula developing first. Gastrulation marks the opening of the “second mouth” that places echinoderms within the deuterostomes, and the mesoderm, which will host the skeleton, migrates inwards. The secondary body cavity, the coelom, forms by the partitioning of three body cavities. The larvae are mostly planktonic but in some species the eggs are retained inside the female and in some, the larvae are also brooded by the female.

The larvae of echinoderms pass through a number of stages and these have specific names derived from the taxonomic names of the adults or from their appearance. For example, a sea urchin has an ‘echinopluteus’ larva while a brittle star has an ‘ophiopluteus’ larva. A starfish has a ‘bipinnaria’ larva but this later develops into a multi-armed ‘brachioria’ larva. A sea cucumber larva is an ‘auricularia’ while a crinoid one is a ‘vitellaria’. All these larvae are bilaterally symmetrical and have bands of cilia with which they swim and some, usually known as ‘pluteus’ larvae, have arms. When fully developed they settle on the seabed to undergo metamorphosis and the larval arms and gut degenerate. The left hand side of the larva develops into the oral surface of the juvenile while the right side becomes the aboral surface. At this stage the bilateral symmetry is lost and radial symmetry develops.

The planktotrophic larva is considered to be the ancestral larval type for echinoderms but after 500 million years of larval evolution, about 68% of species whose development is known have a lecithotrophic larval type. The provision of a yolk-sac means that smaller numbers of eggs are produced, the larvae have a shorter development period, smaller dispersal potential but a greater chance of survival. There seems to be an evolutionary trend towards a “lower-risk-lower-gain” strategy of direct development.

Echinoderms are globally distributed in almost all depths, latitudes and environments in the ocean. They reach highest diversity in reef environments but are also widespread on shallow shores, around the poles – refugia where crinoids are at their most abundant – and throughout the deep ocean, where bottom-dwelling and burrowing sea cucumbers are common – sometimes accounting for up to 90% of organisms. While almost all echinoderms are benthic – that is, they live on the sea floor – some sea-lilies can swim at great velocity for brief periods of time, and a few deep-sea sea cucumbers are fully floating. Some crinoids are pseudo-planktonic, attaching themselves to floating logs and debris, although this behaviour was exercised most extensively in the Paleozoic, before competition from such organisms as barnacles restricted the extent of the behaviour.

The larvae of echinoderms, especially starfish and sea



Figure 19.19: Acorn worm, a hemicordate.²⁵

urchins, are pelagic, and with the aid of ocean currents can be transported for great distances, reinforcing the global distribution of the phylum.

19.2.3 Hemichordata

Hemichordata²⁴ is a phylum of marine deuterostome animals, generally considered the sister group of the echinoderms. They appear in the Lower or Middle Cambrian and include two main classes: Enteropneusta (acorn worms), and Pterobranchia. A third class, Placopeltida, is known only from the larva of a single species, *Placopeltis pelagica*. The extinct class Graptolithina is closely related to the pterobranchs.

Acorn worms are solitary worm-shaped organisms. They generally live in burrows (the earliest secreted tubes) and are deposit feeders, but some species are pharyngeal filter feeders, while the family Torquatoidea are free living detritivores. Many are well known for their production and accumulation of various halogenated phenols and pyrroles. Pterobranchs are filter-feeders, mostly colonial, living in a collagenous tubular structure called a coenecium.

²⁴<https://en.wikipedia.org/wiki/Hemichordata>



Figure 19.20: A Lancelet (*Branchiostoma lanceolatum*).³³



Figure 19.21: Gold-mouth sea squirt (*Polycarpa aurata*), a tunicate.³⁴

19.3 Chordata

A chordate²⁶ is an animal of the phylum Chordata. During some period of their life cycle, chordates possess a notochord, a dorsal nerve cord, pharyngeal slits, and a post-anal tail: these four anatomical features define this phylum. Chordates are also bilaterally symmetric, and have a coelom, metamerism, and circulatory system.

The Chordata and Ambulacraria²⁷ together form the superphylum Deuterostomia. Chordates are divided into three subphyla: Vertebrata²⁸ (fish, amphibians, reptiles, birds, and mammals); Tunicata²⁹ or Urochordata (sea squirts, salps); and Cephalochordata³⁰ (which includes lancelets). There are also extinct taxa such as the Vetulicolia. Hemichordata³¹ (which includes the acorn worms) has been presented as a fourth chordate subphylum, but now is treated as a separate phylum: hemichordates and Echinodermata form the Ambulacraria, the sister phylum of the Chordates. Of the more than 65,000 living species of chordates, about half are bony fish that are members of the superclass Pisces, class Osteichthyes³².

Chordate fossils have been found from as early as the Cambrian explosion, 541 million years ago. Cladistically (phylogenetically), vertebrates – chordates with the notochord replaced by a vertebral column during development – are considered to be a subgroup of the clade Craniata, which consists of chordates with a skull. The Craniata and Tunicata compose the clade Olfactores.

Chordates form a phylum of animals that are defined by having at some stage in their lives all of the following anatomical features:

- A notochord, a fairly stiff rod of cartilage that ex-

tends along the inside of the body. Among the vertebrate sub-group of chordates the notochord develops into the spine, and in wholly aquatic species this helps the animal to swim by flexing its tail.

- A dorsal neural tube. In fish and other vertebrates, this develops into the spinal cord, the main communications trunk of the nervous system.
- Pharyngeal slits. The pharynx is the part of the throat immediately behind the mouth. In fish, the slits are modified to form gills, but in some other chordates they are part of a filter-feeding system that extracts particles of food from the water in which the animals live.
- Post-anal tail. A muscular tail that extends backwards behind the anus.
- An endostyle. This is a groove in the ventral wall of the pharynx. In filter-feeding species it produces mucus to gather food particles, which helps in transporting food to the esophagus. It also stores iodine, and may be a precursor of the vertebrate thyroid gland.

There are soft constraints that separate chordates from certain other biological lineages, but are not part of the formal definition:

- All chordates are deuterostomes. This means that, during the embryo development stage, the anus forms before the mouth.
- All chordates are based on a bilateral body plan.
- All chordates are coelomates, and have a fluid-filled body cavity called a coelom with a complete lining called peritoneum derived from mesoderm (see Brusca and Brusca).

Cephalochordates, one of the three subdivisions of chordates, are small, “vaguely fish-shaped” animals that lack brains, clearly defined heads and specialized sense organs. These burrowing filter-feeders compose the earliest-branching chordate sub-phylum.

²⁶<https://en.wikipedia.org/wiki/Chordate>

²⁷<https://en.wikipedia.org/wiki/Ambulacraria>

²⁸<https://en.wikipedia.org/wiki/Vertebrate>

²⁹<https://en.wikipedia.org/wiki/Tunicate>

³⁰<https://en.wikipedia.org/wiki/Cephalochordate>

³¹<https://en.wikipedia.org/wiki/Hemichordate>

³²<https://en.wikipedia.org/wiki/Osteichthyes>

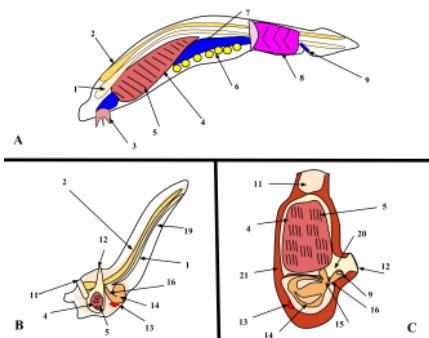


Figure 19.22: A. Lancelet, B. Larval tunicate, C. Adult tunicate³⁵ 1. Notochord, 2. Nerve chord, 3. Buccal cirri, 4. Pharynx, 5. Gill slit, 6. Gonad, 7. Gut, 8. V-shaped muscles, 9. Anus, 10. Inhalant siphon, 11. Exhalant siphon, 12. Heart, 13. Stomach, 14. Esophagus, 15. Intestines, 16. Tail, 17. Atrium, 18. Tunic

Most tunicates appear as adults in two major forms, known as “sea squirts” and salps, both of which are soft-bodied filter-feeders that lack the standard features of chordates. Sea squirts are sessile and consist mainly of water pumps and filter-feeding apparatus; salps float in mid-water, feeding on plankton, and have a two-generation cycle in which one generation is solitary and the next forms chain-like colonies. However, all tunicate larvae have the standard chordate features, including long, tadpole-like tails; they also have rudimentary brains, light sensors and tilt sensors. The third main group of tunicates, Appendicularia (also known as Larvacea), retain tadpole-like shapes and active swimming all their lives, and were for a long time regarded as larvae of sea squirts or salps. The etymology of the term Urochordata (Balfour 1881) is from the ancient Greek οὐρά (oura, “tail”) + Latin *chorda* (“cord”), because the notochord is only found in the tail. The term Tunicata (Lamarck 1816) is recognised as having precedence and is now more commonly used.

Craniates all have distinct skulls. They include the hagfish, which have no vertebrae. Michael J. Benton commented that “craniates are characterized by their heads, just as chordates, or possibly all deuterostomes, are by their tails”.

Most craniates are vertebrates, in which the notochord is replaced by the vertebral column. These consist of a series of bony or cartilaginous cylindrical vertebrae, generally with neural arches that protect the spinal cord, and with projections that link the vertebrae. However hagfish have incomplete braincases and no vertebrae, and are therefore not regarded as vertebrates, but as members of the craniates, the group from which vertebrates are thought to have evolved. However the cladistic exclusion of hagfish from the vertebrates is controversial, as they may be degenerate



Figure 19.23: Actinopterygii – Rayed-Finned Fish⁴²

vertebrates who have lost their vertebral columns.

The position of lampreys is ambiguous. They have complete braincases and rudimentary vertebrae, and therefore may be regarded as vertebrates and true fish. However, molecular phylogenetics, which uses biochemical features to classify organisms, has produced both results that group them with vertebrates and others that group them with hagfish. If lampreys are more closely related to the hagfish than the other vertebrates, this would suggest that they form a clade, which has been named the Cyclostomata.

19.3.1 Fish

Fish³⁶ are gill-bearing aquatic craniate animals that lack limbs with digits. They form a sister group to the tunicates, together forming the olfactores. Included in this definition are the living jawless (Agnatha³⁷), and cartilaginous (Chondrichthyes³⁸) and bony (Osteichthyes³⁹) fish as well as various extinct related groups. The group Osteichthyes is divided into the ray-finned fish (Actinopterygii⁴⁰) and lobe-finned fish (Sarcopterygii⁴¹).

The earliest organisms that can be classified as fish were soft-bodied chordates that first appeared during the Cambrian period. Although they lacked a true spine, they possessed notochords which allowed them to be more agile than their invertebrate counterparts. Fish would continue to evolve through the Paleozoic era, diversifying into

³⁶<https://en.wikipedia.org/wiki/Fish>

³⁷<https://en.wikipedia.org/wiki/Agnatha>

³⁸<https://en.wikipedia.org/wiki/Chondrichthyes>

³⁹<https://en.wikipedia.org/wiki/Osteichthyes>

⁴⁰<https://en.wikipedia.org/wiki/Actinopterygii>

⁴¹<https://en.wikipedia.org/wiki/Sarcopterygii>

a wide variety of forms. Many fish of the Paleozoic developed external armor that protected them from predators. The first fish with jaws appeared in the Silurian period, after which many (such as sharks) became formidable marine predators rather than just the prey of arthropods.

Most fish are ectothermic ("cold-blooded"), allowing their body temperatures to vary as ambient temperatures change, though some of the large active swimmers like white shark and tuna can hold a higher core temperature.

Fish can communicate in their underwater environments through the use of acoustic communication. Acoustic communication in fish involves the transmission of acoustic signals from one individual of a species to another. The production of sounds as a means of communication among fish is most often used in the context of feeding, aggression or courtship behaviour. The sounds emitted by fish can vary depending on the species and stimulus involved. They can produce either stridulatory sounds by moving components of the skeletal system, or can produce non-stridulatory sounds by manipulating specialized organs such as the swimbladder.

Fish are abundant in most bodies of water. They can be found in nearly all aquatic environments, from high mountain streams (e.g., char and gudgeon) to the abyssal and even hadal depths of the deepest oceans (e.g., cusk-eels and snailfish), although no species has yet been documented in the deepest 25% of the ocean. With 34,300 described species, fish exhibit greater species diversity than any other group of vertebrates.

Fish are an important resource for humans worldwide, especially as food. Commercial and subsistence fishers hunt fish in wild fisheries (see fishing) or farm them in ponds or in cages in the ocean (see aquaculture). They are also caught by recreational fishers, kept as pets, raised by fishkeepers, and exhibited in public aquaria. Fish have had a role in culture through the ages, serving as deities, religious symbols, and as the subjects of art, books and movies.

Tetrapods emerged within lobe-finned fishes, so cladistically they are fish as well. However, traditionally fish are rendered paraphyletic by excluding the tetrapods (i.e., the amphibians, reptiles, birds and mammals which all descended from within the same ancestry). Because in this manner the term "fish" is defined negatively as a paraphyletic group, it is not considered a formal taxonomic grouping in systematic biology, unless it is used in the cladistic sense, including tetrapods. The traditional term *pisces* (also *ichthyes*) is considered a typological, but not a phylogenetic classification.



Figure 19.24: Example of Osteichthyes: Queensland lungfish and West Indian Ocean coelacanth (two Sarcopterygii), Iridescent shark and American black sturgeon (two Actinopterygii).⁴³



Figure 19.25: Diversity of amphibians.⁴⁵ Clockwise from top right: Seymouria, Mexican burrowing caecilian, eastern newt and leaf green tree frog.

19.3.2 Amphibians

Amphibians⁴⁴ are ectothermic, tetrapod vertebrates of the class Amphibia. All living amphibians belong to the group Lissamphibia. They inhabit a wide variety of habitats, with most species living within terrestrial, fossorial, arboreal or freshwater aquatic ecosystems. Thus amphibians typically start out as larvae living in water, but some species have developed behavioural adaptations to bypass this.

Amphibians are ectothermic, tetrapod vertebrates of the class Amphibia. All living amphibians belong to the group Lissamphibia. They inhabit a wide variety of habitats, with most species living within terrestrial, fossorial, arboreal or freshwater aquatic ecosystems. Thus amphibians typically start out as larvae living in water, but some species have developed behavioural adaptations to bypass this.

19.3.3 Reptiles

Reptiles⁴⁶ are tetrapod animals in the class Reptilia, comprising today's turtles, crocodilians, snakes, amphisbaenians, lizards, tuatara, and their extinct relatives. The study of these traditional reptile orders, historically combined with that of modern amphibians, is called herpetology.

⁴⁴<https://en.wikipedia.org/wiki/Amphibian>

⁴⁶<https://en.wikipedia.org/wiki/Reptile>



Figure 19.26: Diversity of reptiles.⁴⁷ Clockwise from above left: Green sea turtle (*Chelonia mydas*), Tuatara (*Sphenodon punctatus*), Nile crocodile (*Crocodylus niloticus*), and Sinai agama (*Pseudotrapelus sinaitus*)

Because some reptiles are more closely related to birds than they are to other reptiles (e.g., crocodiles are more closely related to birds than they are to lizards), the traditional groups of “reptiles” listed above do not together constitute a monophyletic grouping or clade (consisting of all descendants of a common ancestor). For this reason, many modern scientists prefer to consider the birds part of Reptilia as well, thereby making Reptilia a monophyletic class, including all living diapsids. The term reptiles is sometimes used as shorthand for ‘non-avian Reptilia’.

The earliest known proto-reptiles originated around 312 million years ago during the Carboniferous period, having evolved from advanced reptiliomorph tetrapods that became increasingly adapted to life on dry land. Some early examples include the lizard-like *Hylonomus* and *Casineria*. In addition to the living reptiles, there are many diverse groups that are now extinct, in some cases due to mass extinction events. In particular, the Cretaceous–Paleogene extinction event wiped out the pterosaurs, plesiosaurs, ornithischians, and sauropods, alongside many species of theropods, crocodyliforms, and squamates (e.g., mosasaurs).

Modern non-avian reptiles inhabit all the continents except Antarctica, although some birds are found on the periphery of Antarctica. Several living subgroups are recognized: Testudines (turtles and tortoises), 350 species; Rhynchocephalia (tuatara from New Zealand), 1 species; Squamata (lizards, snakes, and worm lizards), over 10,200 species; and Crocodilia (crocodiles, gharials, caimans, and alligators), 24 species.

Reptiles are tetrapod vertebrates, creatures that either have four limbs or, like snakes, are descended from four-limbed ancestors. Unlike amphibians, reptiles do not have an aquatic larval stage. Most reptiles are oviparous, although several species of squamates are viviparous,

as were some extinct aquatic clades—the fetus develops within the mother, contained in a placenta rather than an eggshell. As amniotes, reptile eggs are surrounded by membranes for protection and transport, which adapt them to reproduction on dry land. Many of the viviparous species feed their fetuses through various forms of placenta analogous to those of mammals, with some providing initial care for their hatchlings. Extant reptiles range in size from a tiny gecko, *Sphaerodactylus ariasae*, which can grow up to 17 mm (0.7 in) to the saltwater crocodile, *Crocodylus porosus*, which can reach 6 m (19.7 ft) in length and weigh over 1,000 kg (2,200 lb).

19.3.4 Birds

Birds⁴⁸ are a group of warm-blooded vertebrates constituting the class Aves, characterized by feathers, toothless beaked jaws, the laying of hard-shelled eggs, a high metabolic rate, a four-chambered heart, and a strong yet lightweight skeleton. Birds live worldwide and range in size from the 5 cm (2 in) bee hummingbird to the 2.75 m (9 ft) ostrich. There are about ten thousand living species, more than half of which are passerine, or “perching” birds. Birds have wings whose development varies according to species; the only known groups without wings are the extinct moa and elephant birds. Wings, which evolved from forelimbs, gave birds the ability to fly, although further evolution has led to the loss of flight in some birds, including ratites, penguins, and diverse endemic island species. The digestive and respiratory systems of birds are also uniquely adapted for flight. Some bird species of aquatic environments, particularly seabirds and some waterbirds, have further evolved for swimming.

Birds are a group of feathered theropod dinosaurs, and constitute the only living dinosaurs. Likewise, birds are considered reptiles in the modern cladistic sense of the term, and their closest living relatives are the crocodilians. Birds are descendants of the primitive avialans (whose members include *Archaeopteryx*) which first appeared about 160 million years ago (mya) in China. According to DNA evidence, modern birds (Neornithes) evolved in the Middle to Late Cretaceous, and diversified dramatically around the time of the Cretaceous–Paleogene extinction event 66 mya, which killed off the pterosaurs and all non-avian dinosaurs.

Many social species pass on knowledge across generations, which is considered a form of culture. Birds are social, communicating with visual signals, calls, and songs, and participating in such behaviours as cooperative breeding and hunting, flocking, and mobbing of predators. The vast majority of bird species are socially (but not necessarily sexually) monogamous, usually for one breeding season

⁴⁸<https://en.wikipedia.org/wiki/Bird>



Figure 19.27: The diversity of birds.⁴⁹ This image shows 17 biological orders of birds (from top, left to right): Musophagiformes, Pelecaniformes, Phaethontiformes, Accipitriformes, Gruiformes, Galliformes, Columbiformes, Apodiformes, Charadriiformes, Casuariiformes, Psittaciformes, Phoenicopteriformes, Sphenisciformes, Pelecaniformes, Suliformes, Passeriformes, Strigiformes, Piciformes.

at a time, sometimes for years, but rarely for life. Other species have breeding systems that are polygynous (one male with many females) or, rarely, polyandrous (one female with many males). Birds produce offspring by laying eggs which are fertilised through sexual reproduction. They are usually laid in a nest and incubated by the parents. Most birds have an extended period of parental care after hatching.

Many species of birds are economically important as food for human consumption and raw material in manufacturing, with domesticated and undomesticated birds being important sources of eggs, meat, and feathers. Songbirds, parrots, and other species are popular as pets. Guano (bird excrement) is harvested for use as a fertiliser. Birds figure throughout human culture. About 120 to 130 species have become extinct due to human activity since the 17th century, and hundreds more before then. Human activity threatens about 1,200 bird species with extinction, though efforts are underway to protect them. Recreational bird-watching is an important part of the ecotourism industry.

19.3.5 Mammals

Mammals⁵⁰ (from Latin *mamma* “breast”) are vertebrate animals constituting the class *Mammalia*, and characterized by the presence of mammary glands which in females produce milk for feeding (nursing) their young, a neocortex (a region of the brain), fur or hair, and three middle ear bones. These characteristics distinguish them from reptiles and birds, from which they diverged in the late Carboniferous, approximately 300 million years ago. Around 6,400 extant species of mammals have been described. The largest orders are the rodents, bats and Eulipotyphla (hedgehogs, moles, shrews, and others). The next three are the Primates (apes including humans, monkeys, and others), the Cetartiodactyla (cetaceans and even-toed ungulates), and the Carnivora (cats, dogs, seals, and others).

In terms of cladistics, which reflects evolutionary history, mammals are the only living members of the *Synapsida*; this clade, together with *Sauropsida* (reptiles and birds), constitutes the larger *Amniota* clade. The early synapsid mammalian ancestors were sphenacodont pelycosaurs, a group that included the non-mammalian *Dimetrodon*. At the end of the Carboniferous period around 300 million years ago, this group diverged from the sauropsid line that led to today’s reptiles and birds. The line following the stem group *Sphenacodontia* split into several diverse groups of non-mammalian synapsids—sometimes incorrectly referred to as mammal-like reptiles—before giving rise to *Therapsida* in the Early Permian period. The modern mammalian orders arose in the Paleogene and Neogene periods of the Cenozoic

⁵⁰<https://en.wikipedia.org/wiki/Mammal>



Figure 19.28: Diversity of mammals.⁵¹

era, after the extinction of non-avian dinosaurs, and have been the dominant terrestrial animal group from 66 million years ago to the present.

The basic body type is quadruped, and most mammals use their four extremities for terrestrial locomotion; but in some, the extremities are adapted for life at sea, in the air, in trees, underground, or on two legs. Mammals range in size from the 30–40 mm (1.2–1.6 in) bumblebee bat to the 30 m (98 ft) blue whale—possibly the largest animal to have ever lived. Maximum lifespan varies from two years for the shrew to 211 years for the bowhead whale. All modern mammals give birth to live young, except the five species of monotremes, which are egg-laying mammals. The most species-rich group of mammals, the cohort called placentals, have a placenta, which enables the feeding of the fetus during gestation.

Most mammals are intelligent, with some possessing large brains, self-awareness, and tool use. Mammals can communicate and vocalize in several ways, including the production of ultrasound, scent-marking, alarm signals, singing, and echolocation. Mammals can organize themselves into fission-fusion societies, harems, and hierarchies—but can also be solitary and territorial. Most mammals are polygynous, but some can be monogamous or polyandrous.

Domestication of many types of mammals by humans played a major role in the Neolithic revolution, and resulted in farming replacing hunting and gathering as the primary source of food for humans. This led to a major restructuring of human societies from nomadic to sedentary, with

more co-operation among larger and larger groups, and ultimately the development of the first civilizations. Domesticated mammals provided, and continue to provide, power for transport and agriculture, as well as food (meat and dairy products), fur, and leather. Mammals are also hunted and raced for sport, and are used as model organisms in science. Mammals have been depicted in art since Palaeolithic times, and appear in literature, film, mythology, and religion. Decline in numbers and extinction of many mammals is primarily driven by human poaching and habitat destruction, primarily deforestation.

19.3.6 Ecdysozoa

The Ecdysozoa⁵² are protostomes, named after their shared trait of ecdysis, growth by moultling. They include the largest animal phylum, the Arthropoda, which contains insects, spiders, crabs, and their kin. All of these have a body divided into repeating segments, typically with paired appendages. Two smaller phyla, the Onychophora and Tardigrada, are close relatives of the arthropods and share these traits. The ecdysozoans also include the Nematoda or roundworms, perhaps the second largest animal phylum. Roundworms are typically microscopic, and occur in nearly every environment where there is water; some are important parasites. Smaller phyla related to them are the Nematomorpha or horsehair worms, and the Kinorhyncha, Priapulida, and Loricifera. These groups have a reduced coelom, called a pseudocoelom.

19.3.7 Arthropods

An arthropod⁵³ (from Greek ἄρθρον arthon, “joint” and πούς pouς, “foot” (gen. ποδός)) is an invertebrate animal having an exoskeleton, a segmented body, and paired jointed appendages. Arthropods form the phylum Euarthropoda, which includes insects, arachnids, myriapods, and crustaceans. The term Arthropoda as originally proposed refers to a proposed grouping of Euarthropods and the phylum Onychophora.

Arthropods are characterized by their jointed limbs and cuticle made of chitin, often mineralised with calcium carbonate. The arthropod body plan consists of segments, each with a pair of appendages. The rigid cuticle inhibits growth, so arthropods replace it periodically by moultling. Arthropods are bilaterally symmetrical and their body possesses an external skeleton. Some species have wings.

Their versatility has enabled arthropods to become the most species-rich members of all ecological guilds in most environments. They have over a million described species, making up more than 80 percent of all described living

⁵²

⁵³ <https://en.wikipedia.org/wiki/Arthropod>



Figure 19.29: Diversity of arthropods.⁵⁴ From left to right and from top to bottom: Kolihapeltis, Styelonurus, scorpion, crab, centipede, butterfly.

animal species, some of which, unlike most other animals, are very successful in dry environments. Arthropods range in size from the microscopic crustacean *Stygotantulus* up to the Japanese spider crab.

An arthropod's primary internal cavity is a haemocoel, which accommodates its internal organs, and through which its haemolymph – analogue of blood – circulates; it has an open circulatory system. Like their exteriors, the internal organs of arthropods are generally built of repeated segments. Their nervous system is "ladder-like", with paired ventral nerve cords running through all segments and forming paired ganglia in each segment. Their heads are formed by fusion of varying numbers of segments, and their brains are formed by fusion of the ganglia of these segments and encircle the esophagus. The respiratory and excretory systems of arthropods vary, depending as much on their environment as on the subphylum to which they belong.

Their vision relies on various combinations of compound eyes and pigment-pit ocelli: in most species the ocelli can only detect the direction from which light is coming, and the compound eyes are the main source of information, but the main eyes of spiders are ocelli that can form images and, in a few cases, can swivel to track prey. Arthropods also have a wide range of chemical and mechanical sensors, mostly based on modifications of the many bristles known as setae that project through their

cuticles.

Arthropods' methods of reproduction and development are diverse; all terrestrial species use internal fertilization, but this is often by indirect transfer of the sperm via an appendage or the ground, rather than by direct injection. Aquatic species use either internal or external fertilization. Almost all arthropods lay eggs, but scorpions give birth to live young after the eggs have hatched inside the mother. Arthropod hatchlings vary from miniature adults to grubs and caterpillars that lack jointed limbs and eventually undergo a total metamorphosis to produce the adult form. The level of maternal care for hatchlings varies from nonexistent to the prolonged care provided by scorpions.

The evolutionary ancestry of arthropods dates back to the Cambrian period. The group is generally regarded as monophyletic, and many analyses support the placement of arthropods with cycloneuralians (or their constituent clades) in a superphylum Ecdysozoa. Overall, however, the basal relationships of animals are not yet well resolved. Likewise, the relationships between various arthropod groups are still actively debated.

Arthropods belong to phylum Euarthropoda. The phylum is sometimes called Arthropoda, but strictly this term denotes a (putative – see Tactopoda) clade that also encompasses the phylum Onychophora.

Euarthropoda is typically subdivided into five subphyla, of which one is extinct:

- Trilobites are a group of formerly numerous marine animals that disappeared in the Permian-Triassic extinction event, though they were in decline prior to this killing blow, having been reduced to one order in the Late Devonian extinction.
- Chelicерates include horseshoe crabs, spiders, mites, scorpions and related organisms. They are characterised by the presence of chelicerae, appendages just above / in front of the mouth. Chelicerae appear in scorpions and horseshoe crabs as tiny claws that they use in feeding, but those of spiders have developed as fangs that inject venom.
- Myriapods comprise millipedes, centipedes, and their relatives and have many body segments, each segment bearing one or two pairs of legs (or in a few cases being legless). They are sometimes grouped with the hexapods.
- Crustaceans are primarily aquatic (a notable exception being woodlice) and are characterised by having biramous appendages. They include lobsters, crabs, barnacles, crayfish, shrimp and many others.
- Hexapods comprise insects and three small orders of insect-like animals with six thoracic legs. They



Figure 19.30: A Maine Lobster served in Boston.⁵⁵

are sometimes grouped with the myriapods, in a group called Uniramia, though genetic evidence tends to support a closer relationship between hexapods and crustaceans.

Aside from these major groups, there are also a number of fossil forms, mostly from the Early Cambrian, which are difficult to place, either from lack of obvious affinity to any of the main groups or from clear affinity to several of them. Marrella was the first one to be recognized as significantly different from the well-known groups.

The phylogeny of the major extant arthropod groups has been an area of considerable interest and dispute. Recent studies strongly suggest that Crustacea, as traditionally defined, is paraphyletic, with Hexapoda having evolved from within it, so that Crustacea and Hexapoda form a clade, Pancrustacea. The position of Myriapoda, Chelicerata and Pancrustacea remains unclear as of April 2012. In some studies, Myriapoda is grouped with Chelicerata (forming Myriochelata); in other studies, Myriapoda is grouped with Pancrustacea (forming Mandibulata), or Myriapoda may be sister to Chelicerata plus Pancrustacea.

Arthropods contribute to the human food supply both directly as food, and more importantly indirectly as pollinators of crops.

Some species are known to spread severe disease to humans, livestock, and crops.

19.3.8 Nematodes

The nematodes⁵⁶ or roundworms constitute the phylum Nematoda (also called Nemathelminthes), with plant-parasitic nematodes being known as eelworms. The word nematode comes from the Modern Latin compound of



Figure 19.31: *Caenorhabditis elegans*, a free-living transparent nematode about 1 mm in length that lives in temperate soil environments.⁵⁷ In 1963, Sydney Brenner proposed research into *C. elegans*, primarily in the area of neuronal development. In 1974, he began research into the molecular and developmental biology of *C. elegans*, which has since been extensively used as a model organism. It was the first multicellular organism to have its whole genome sequenced, and as of 2019, is the only organism to have its connectome (neuronal “wiring diagram”) completed.

nemat- “thread” (from Greek *nema*, genitive *nematos* “thread,” from stem of *nein* “to spin”; see needle) + -odes “like, of the nature of” (see -oid). They are a diverse animal phylum inhabiting a broad range of environments. Taxonomically, they are classified along with insects and other moulting animals in the clade Ecdysozoa, and unlike flatworms, have tubular digestive systems with openings at both ends. Like tardigrades they have a reduced number of Hox genes, but as their sister phylum Nematomorpha has kept the ancestral protostome Hox genotype, it shows that the reduction has occurred within the nematode phylum.

Nematode species can be difficult to distinguish from one another. Consequently, estimates of the number of nematode species described to date vary by author and may change rapidly over time. A 2013 survey of animal biodiversity published in the mega journal Zootaxa puts this figure at over 25,000. Estimates of the total number of extant species are subject to even greater variation. A widely referenced article published in 1993 estimated there may be over 1 million species of nematode, a claim which has since been repeated in numerous publications. Many other publications have since vigorously refuted this claim on the grounds that it is unsupported by fact. More recent, fact-based estimates have placed the true figure closer to 40,000 species worldwide.

Nematodes have successfully adapted to nearly every ecosystem: from marine (salt) to fresh water, soils, from the polar regions to the tropics, as well as the highest to the lowest of elevations (including mountains). They are ubiquitous in freshwater, marine, and terrestrial environments, where they often outnumber other animals in both individual and species counts, and are found in locations as diverse as mountains, deserts, and oceanic trenches. They are found in every part of the earth's lithosphere, even at

⁵⁶<https://en.wikipedia.org/wiki/Nematode>



Figure 19.32: A Rotifera (wheel animal).⁶²

great depths, 0.9–3.6 km (3,000–12,000 ft) below the surface of the Earth in gold mines in South Africa. They represent 90% of all animals on the ocean floor. In total, 4.4×10^{20} nematodes inhabit the Earth's topsoil, or approximately 60 billion for each human, with the highest densities observed in tundra and boreal forests. Their numerical dominance, often exceeding a million individuals per square meter and accounting for about 80% of all individual animals on earth, their diversity of lifecycles, and their presence at various trophic levels point to an important role in many ecosystems. They have been shown to play crucial roles in polar ecosystems. The roughly 2,271 genera are placed in 256 families. The many parasitic forms include pathogens in most plants and animals. A third of the genera occur as parasites of vertebrates; about 35 nematode species occur in humans.

19.3.9 Spiralia

The Spiralia⁵⁸ are a large group of protostomes that develop by spiral cleavage in the early embryo. The Spiralia's phylogeny has been disputed, but it contains a large clade, the superphylum Lophotrochozoa⁵⁹, and smaller groups of phyla such as the Rousphozoa which includes the gastrotrichs and the flatworms. All of these are grouped as the Platytrichozoa, which has a sister group, the Gnathifera⁶⁰ (from the Greek gnáthos, "jaw", and the Latin -fera, "bearing"), which includes the rotifers.

19.3.10 Rotifera

The rotifers⁶¹ (from Latin rota "wheel" and -fer "bearing"), commonly called wheel animals or wheel animalcules, make up a phylum (Rotifera) of microscopic and near-microscopic pseudocoelomate animals.

They were first described by Rev. John Harris in 1696, and other forms were described by Antonie van Leeuwenhoek in 1703. Most rotifers are around 0.1–0.5 mm long

(although their size can range from 50 µm to over 2 mm), and are common in freshwater environments throughout the world with a few saltwater species.

Some rotifers are free swimming and truly planktonic, others move by inchworming along a substrate, and some are sessile, living inside tubes or gelatinous holdfasts that are attached to a substrate. About 25 species are colonial (e.g., *Sinantherina semibullata*), either sessile or planktonic. Rotifers are an important part of the freshwater zooplankton, being a major foodsource and with many species also contributing to the decomposition of soil organic matter. Most species of the rotifers are cosmopolitan, but there are also some endemic species, like *Cephalodella vitata* to Lake Baikal. Recent barcoding evidence, however, suggests that some 'cosmopolitan' species, such as *Brachionus plicatilis*, *B. calyciflorus*, *Lecane bulla*, among others, are actually species complexes.

Rotifers have bilateral symmetry and a variety of different shapes. The body of a rotifer is divided into a head, trunk, and foot, and is typically somewhat cylindrical. There is a well-developed cuticle, which may be thick and rigid, giving the animal a box-like shape, or flexible, giving the animal a worm-like shape; such rotifers are respectively called loricate and illoricate. Rigid cuticles are often composed of multiple plates, and may bear spines, ridges, or other ornamentation. Their cuticle is nonchitinous and is formed from sclerotized proteins.

The most distinctive feature of rotifers is the presence of a ciliated structure, called the corona, on the head. In the more primitive species, this forms a simple ring of cilia around the mouth from which an additional band of cilia stretches over the back of the head. In the great majority of rotifers, however, this has evolved into a more complex structure.

Modifications to the basic plan of the corona include alteration of the cilia into bristles or large tufts, and either expansion or loss of the ciliated band around the head. In genera such as *Collotheca*, the corona is modified to form a funnel surrounding the mouth. In many species, such as those in the genus *Testudinella*, the cilia around the mouth have disappeared, leaving just two small circular bands on the head. In the bdelloids, this plan is further modified, with the upper band splitting into two rotating wheels, raised up on a pedestal projecting from the upper surface of the head.

The trunk forms the major part of the body, and encloses most of the internal organs. The foot projects from the rear of the trunk, and is usually much narrower, giving the appearance of a tail. The cuticle over the foot often forms rings, making it appear segmented, although the internal structure is uniform. Many rotifers can retract

⁵⁸<https://en.wikipedia.org/wiki/Spiralia>

⁵⁹<https://en.wikipedia.org/wiki/Lophotrochozoa>

⁶⁰[https://en.wikipedia.org/wiki/Gnathifera_\(clade\)](https://en.wikipedia.org/wiki/Gnathifera_(clade))

⁶¹<https://en.wikipedia.org/wiki/Rotifer>

the foot partially or wholly into the trunk. The foot ends in from one to four toes, which, in sessile and crawling species, contain adhesive glands to attach the animal to the substratum. In many free-swimming species, the foot as a whole is reduced in size, and may even be absent.

The coronal cilia create a current that sweeps food into the mouth. The mouth opens into a characteristic chewing pharynx (called the mastax), sometimes via a ciliated tube, and sometimes directly. The pharynx has a powerful muscular wall and contains tiny, calcified, jaw-like structures called trophi, which are the only fossilizable parts of a rotifer. The shape of the trophi varies between different species, depending partly on the nature of their diet. In suspension feeders, the trophi are covered in grinding ridges, while in more actively carnivorous species, they may be shaped like forceps to help bite into prey. In some ectoparasitic rotifers, the mastax is adapted to grip onto the host, although, in others, the foot performs this function instead.

Behind the mastax lies an oesophagus, which opens into a stomach where most of the digestion and absorption occurs. The stomach opens into a short intestine that terminates in a cloaca on the posterior dorsal surface of the animal. Up to seven salivary glands are present in some species, emptying to the mouth in front of the oesophagus, while the stomach is associated with two gastric glands that produce digestive enzymes.

A pair of protonephridia open into a bladder that drains into the cloaca. These organs expel water from the body, helping to maintain osmotic balance.

Rotifers have a small brain, located just above the mastax, from which a number of nerves extend throughout the body. The number of nerves varies among species, although the nervous system usually has a simple layout. Close to the brain lies a retrocerebral organ, consisting of two glands either side of a medial sac. The sac drains into a duct that divides into two before opening through pores on the uppermost part of the head. The function of the retrocerebral organ is unclear.

The nervous system comprises about 25% of the roughly 1,000 cells in a rotifer.

Rotifers typically possess one or two pairs of short antennae and up to five eyes. The eyes are simple in structure, sometimes with just a single photoreceptor cell. In addition, the bristles of the corona are sensitive to touch, and there are also a pair of tiny sensory pits lined by cilia in the head region.

The coronal cilia pull the animal, when unattached, through the water.

Like many other microscopic animals, adult rotifers fre-

quently exhibit eutely—they have a fixed number of cells within a species, usually on the order of 1,000.

Bdelloid rotifer genomes contain two or more divergent copies of each gene, suggesting a long-term asexual evolutionary history. For example, four copies of *hsp82* are found. Each is different and found on a different chromosome excluding the possibility of homozygous sexual reproduction.

Rotifers eat particulate organic detritus, dead bacteria, algae, and protozoans. They eat particles up to 10 micrometres in size. Like crustaceans, rotifers contribute to nutrient recycling. For this reason, they are used in fish tanks to help clean the water, to prevent clouds of waste matter. Rotifers affect the species composition of algae in ecosystems through their choice in grazing. Rotifers may be in competition with cladocera and copepods for planktonic food sources.

Rotifers are dioecious and reproduce sexually or parthenogenetically. They are sexually dimorphic, with the females always being larger than the males. In some species, this is relatively mild, but in others the female may be up to ten times the size of the male. In parthenogenetic species, males may be present only at certain times of the year, or absent altogether.

The female reproductive system consists of one or two ovaries, each with a vitellarium gland that supplies the eggs with yolk. Together, each ovary and vitellarium form a single syncytial structure in the anterior part of the animal, opening through an oviduct into the cloaca.

Males do not usually have a functional digestive system, and are therefore short-lived, often being sexually fertile at birth. They have a single testicle and sperm duct, associated with a pair of glandular structures referred to as prostates (unrelated to the vertebrate prostate). The sperm duct opens into a gonopore at the posterior end of the animal, which is usually modified to form a penis. The gonopore is homologous to the cloaca of females, but in most species has no connection to the vestigial digestive system, which lacks an anus.

The phylum Rotifera encloses three classes that reproduce by three different mechanisms: Seisonidea only reproduce sexually; Bdelloidea reproduce exclusively by asexual parthenogenesis; Monogononta reproduce alternating these two mechanisms ("cyclical parthenogenesis" or "heterogony"). Parthenogenesis (amictic phase) dominates the monogonont life cycle, promoting fast population growth and colonization. In this phase males are absent and amictic females produce diploid eggs by mitosis which develop parthenogenetically into females that are clones of their mothers. Some amictic females can generate mictic females that will produce haploid



Figure 19.33: The turbellarian *Pseudoceros dimidiatus*.⁶⁴

eggs by meiosis. Mixis (meiosis) is induced by different types of stimulus depending on species. Haploid eggs develop into haploid dwarf males if they are not fertilized and into diploid “resting eggs” (or “diapausing eggs”) if they are fertilized by males.

Fertilization is internal. The male either inserts his penis into the female’s cloaca or uses it to penetrate her skin, injecting the sperm into the body cavity. The egg secretes a shell, and is attached either to the substratum, nearby plants, or the female’s own body. A few species, such as members of the Rotaria, are ovoviparous, retaining the eggs inside their body until they hatch.

Most species hatch as miniature versions of the adult. Sessile species, however, are born as free-swimming larvae, which closely resemble the adults of related free-swimming species. Females grow rapidly, reaching their adult size within a few days, while males typically do not grow in size at all.

The life span of monogonont females varies from two days to about three weeks.

19.3.11 Flatworms

The flatworms⁶³, flat worms, Platyhelminthes, Plathelminthes, or platyhelminths (from the Greek πλατύ, *platy*, meaning “flat” and ἔλμιντ, *helmint*-, meaning “worm”) are a phylum of relatively simple bilaterian, unsegmented, soft-bodied invertebrates. Unlike other bilaterians, they are acelomates (having no body cavity), and have no specialized circulatory and respiratory organs, which restricts them to having flattened shapes that allow oxygen and nutrients to pass through their bodies by diffusion. The digestive cavity has only one opening for both ingestion (intake of nutrients) and egestion (removal of undigested wastes); as a result, the food cannot be processed continuously.

In traditional medicinal texts, Platyhelminthes are divided into Turbellaria, which are mostly non-parasitic

animals such as planarians, and three entirely parasitic groups: Cestoda, Trematoda and Monogenea; however, since the turbellarians have since been proven not to be monophyletic, this classification is now deprecated. Free-living flatworms are mostly predators, and live in water or in shaded, humid terrestrial environments, such as leaf litter. Cestodes (tapeworms) and trematodes (flukes) have complex life-cycles, with mature stages that live as parasites in the digestive systems of fish or land vertebrates, and intermediate stages that infest secondary hosts. The eggs of trematodes are excreted from their main hosts, whereas adult cestodes generate vast numbers of hermaphroditic, segment-like proglottids that detach when mature, are excreted, and then release eggs. Unlike the other parasitic groups, the monogeneans are external parasites infesting aquatic animals, and their larvae metamorphose into the adult form after attaching to a suitable host.

Platyhelminthes are bilaterally symmetrical animals: their left and right sides are mirror images of each other; this also implies they have distinct top and bottom surfaces and distinct head and tail ends. Like other bilaterians, they have three main cell layers (endoderm, mesoderm, and ectoderm), while the radially symmetrical cnidarians and ctenophores (comb jellies) have only two cell layers. Beyond that, they are “defined more by what they do not have than by any particular series of specializations.” Unlike other bilaterians, Platyhelminthes have no internal body cavity, so are described as acelomates. They also lack specialized circulatory and respiratory organs, both of these facts are defining features when classifying a flatworm’s anatomy. Their bodies are soft and unsegmented.

Because they do not have internal body cavities, Platyhelminthes were regarded as a primitive stage in the evolution of bilaterians (animals with bilateral symmetry and hence with distinct front and rear ends). However, analyses since the mid-1980s have separated out one subgroup, the Acoelomorpha, as basal bilaterians – closer to the original bilaterians than to any other modern groups. The remaining Platyhelminthes form a monophyletic group, one that contains all and only descendants of a common ancestor that is itself a member of the group. The redefined Platyhelminthes is part of the Lophotrochozoa, one of the three main groups of more complex bilaterians. These analyses had concluded the redefined Platyhelminthes, excluding Acoelomorpha, consists of two monophyletic subgroups, Catenulida and Rhabditophora, with Cestoda, Trematoda and Monogenea forming a monophyletic subgroup within one branch of the Rhabditophora. Hence, the traditional platyhelminth subgroup “Turbellaria” is now regarded as paraphyletic, since it excludes the wholly parasitic groups, although these are descended from one group of “turbel-

⁶³<https://en.wikipedia.org/wiki/Flatworm>

larians".

The lack of circulatory and respiratory organs limits platyhelminths to sizes and shapes that enable oxygen to reach and carbon dioxide to leave all parts of their bodies by simple diffusion. Hence, many are microscopic and the large species have flat ribbon-like or leaf-like shapes. The guts of large species have many branches, allowing nutrients to diffuse to all parts of the body. Respiration through the whole surface of the body makes them vulnerable to fluid loss, and restricts them to environments where dehydration is unlikely: sea and freshwater, moist terrestrial environments such as leaf litter or between grains of soil, and as parasites within other animals.

The space between the skin and gut is filled with mesenchyme, also known as parenchyma, a connective tissue made of cells and reinforced by collagen fibers that act as a type of skeleton, providing attachment points for muscles. The mesenchyme contains all the internal organs and allows the passage of oxygen, nutrients and waste products. It consists of two main types of cell: fixed cells, some of which have fluid-filled vacuoles; and stem cells, which can transform into any other type of cell, and are used in regenerating tissues after injury or asexual reproduction.

Most platyhelminths have no anus and regurgitate undigested material through the mouth. However, some long species have an anus and some with complex, branched guts have more than one anus, since excretion only through the mouth would be difficult for them. The gut is lined with a single layer of endodermal cells that absorb and digest food. Some species break up and soften food first by secreting enzymes in the gut or pharynx (throat).

All animals need to keep the concentration of dissolved substances in their body fluids at a fairly constant level. Internal parasites and free-living marine animals live in environments with high concentrations of dissolved material, and generally let their tissues have the same level of concentration as the environment, while freshwater animals need to prevent their body fluids from becoming too dilute. Despite this difference in environments, most platyhelminths use the same system to control the concentration of their body fluids. Flame cells, so called because the beating of their flagella looks like a flickering candle flame, extract from the mesenchyme water that contains wastes and some reusable material, and drive it into networks of tube cells which are lined with flagella and microvilli. The tube cells' flagella drive the water towards exits called nephridiopores, while their microvilli reabsorb reusable materials and as much water as is needed to keep the body fluids at the right concentration. These combinations of flame cells and tube cells are called protonephridia.

In all platyhelminths, the nervous system is concentrated at the head end. This is least marked in the acoels, which have nerve nets rather like those of cnidarians and ctenophores, but densest around the head. Other platyhelminths have rings of ganglia in the head and main nerve trunks running along their bodies.

19.3.12 Trematoda

These parasites' name refers to the cavities in their holdfasts (Greek τρῆμα, hole), which resemble suckers and anchor them within their hosts. The skin of all species is a syncitium, which is a layer of cells that shares a single external membrane. Trematodes are divided into two groups, Digenea and Aspidogastrea (also known as Aspidothorea)

19.3.13 Digenea

These are often called flukes, as most have flat rhomboid shapes like that of a flounder (Old English flóc). There are about 11,000 species, more than all other platyhelminthes combined, and second only to roundworms among parasites on metazoans. Adults usually have two holdfasts: a ring around the mouth and a larger sucker midway along what would be the underside in a free-living flatworm. Although the name "Digeneans" means "two generations", most have very complex life cycles with up to seven stages, depending on what combinations of environments the early stages encounter - the most important factor being whether the eggs are deposited on land or in water. The intermediate stages transfer the parasites from one host to another. The definitive host in which adults develop is a land vertebrate; the earliest host of juvenile stages is usually a snail that may live on land or in water, whilst in many cases, a fish or arthropod is the second host. For example, (Figure ??) shows the life cycle of the fluke *Schistosoma mansoni*, which causes the devastating tropical disease bilharzia. Infected individuals release *Schistosoma* eggs into water via their fecal material or urine. After larvae hatch from these eggs, the larvae infect a very specific type of freshwater snail. For example, in *S. mansoni* it is snail of the genus *Biomphalaria*. The *Schistosoma* larvae undergo the next phase of their lifecycles in these snails, spending their time reproducing and developing. Once this step has been completed, the parasite leaves the snail and enters the water. The parasite can live in the water for only 48 hours without a mammalian host. Humans encounter larvae of the *Schistosoma* parasite when they enter contaminated water while bathing, playing, swimming, washing, fishing, or walking through the water. Once a host has been found, the worm enters its blood vessels. For several weeks, the worm remains in the vessels, continuing its development into its adult phase. When maturity is

reached, mating occurs and eggs are produced. Eggs enter the bladder/intestine and are excreted through urine and feces and the process repeats. If the eggs do not get excreted, they can become entrained in the body tissues and cause a variety of problems such as immune reactions and organ damage.

Adults range between 0.2 mm (0.0079 in) and 6 mm (0.24 in) in length. Individual adult digenarians are of a single sex, and in some species slender females live in enclosed grooves that run along the bodies of the males, partially emerging to lay eggs. In all species the adults have complex reproductive systems, capable of producing between 10,000 and 100,000 times as many eggs as a free-living flatworm. In addition, the intermediate stages that live in snails reproduce asexually.

Adults of different species infest different parts of the definitive host – for example the intestine, lungs, large blood vessels, and liver. The adults use a relatively large, muscular pharynx to ingest cells, cell fragments, mucus, body fluids or blood. In both the adult and snail-inhabiting stages, the external syncytium absorbs dissolved nutrients from the host. Adult digenarians can live without oxygen for long periods.

19.3.14 Cestoda

These are often called tapeworms⁶⁶ because of their flat, slender but very long bodies – the name “cestode” is derived from the Latin word *cestus*, which means “tape”. The adults of all 3,400 cestode species are internal parasites. Cestodes have no mouths or guts, and the syncytial skin absorbs nutrients – mainly carbohydrates and amino acids – from the host, and also disguises it chemically to avoid attacks by the host’s immune system. Shortage of carbohydrates in the host’s diet stunts the growth of parasites and may even kill them. Their metabolisms generally use simple but inefficient chemical processes, compensating for this inefficiency by consuming large amounts of food relative to their physical size.

In the majority of species, known as eucestodes (“true tapeworms”), the neck produces a chain of segments called proglottids via a process known as strobilation. As a result, the most mature proglottids are furthest from the scolex. Adults of *Taenia saginata*, which infests humans, can form proglottid chains over 20 metres (66 ft) long, although 4 metres (13 ft) is more typical. Each proglottid has both male and female reproductive organs. If the host’s gut contains two or more adults of the same cestode species they generally fertilize each other, however, proglottids of the same worm can fertilize each other and even themselves.

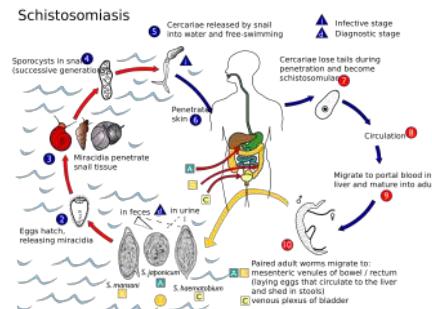


Figure 19.34: Schistosoma life cycle.⁶⁵ Schistosoma eggs are eliminated with feces or urine, depending on species (1). Under appropriate conditions the eggs hatch and release miracidia (2), which swim and penetrate specific snail intermediate hosts (3). The stages in the snail include two generations of sporocysts (4) and the production of cercariae (5). Upon release from the snail, the infective cercariae swim, penetrate the skin of the human host (6), and shed their forked tails, becoming schistosomulae (7). The schistosomulae migrate via venous circulation to lungs, then to the heart, and then develop in the liver, exiting the liver via the portal vein system when mature, (8)(9). Male and female adult worms copulate and reside in the mesenteric venules, the location of which varies by species (with some exceptions) (10). For instance, *S. japonicum* is more frequently found in the superior mesenteric veins draining the small intestine (A), and *S. mansoni* occurs more often in the inferior mesenteric veins draining the large intestine (B). However, both species can occupy either location and are capable of moving between sites. *S. intercalatum* and *S. guineensis* also inhabit the inferior mesenteric plexus but lower in the bowel than *S. mansoni*. *S. haematobium* most often inhabits in the vesicular and pelvic venous plexus of the bladder (C), but it can also be found in the rectal venules. The females (size ranges from 7–28 mm, depending on species) deposit eggs in the small venules of the portal and perivesical systems. The eggs are moved progressively toward the lumen of the intestine (*S. mansoni*, *S. japonicum*, *S. mekongi*, *S. intercalatum/guineensis*) and of the bladder and ureters (*S. haematobium*), and are eliminated with feces or urine, respectively (1).

⁶⁶<https://en.wikipedia.org/wiki/Cestoda>

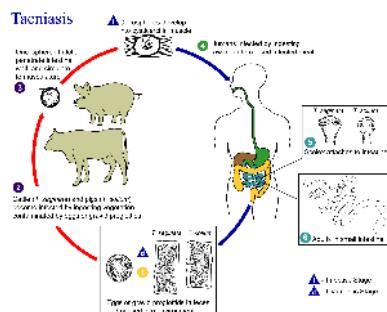


Figure 19.35: Life cycle of the cestode *Taenia*.⁶⁷ *Taeniasis* is the infection of humans with the adult tapeworm of *Taenia saginata* or *Taenia solium*. Humans are the only definitive hosts for *T. saginata* and *T. solium*. Eggs or gravid proglottids are passed with feces (1); the eggs can survive for days to months in the environment. Cattle (*T. saginata*) and pigs (*T. solium*) become infected by ingesting vegetation contaminated with eggs or gravid proglottids (2). In the animal's intestine, the oncospheres hatch (3), invade the intestinal wall, and migrate to the striated muscles, where they develop into cysticerci. A cysticercus can survive for several years in the animal. Humans become infected by ingesting raw or undercooked infected meat (4). In the human intestine, the cysticercus develops over 2 months into an adult tapeworm, which can survive for years. The adult tapeworms attach to the small intestine by their scolex (5) and reside in the small intestine (6). Length of adult worms is usually 5 m or less for *T. saginata* (however it may reach up to 25 m) and 2 to 7 m for *T. solium*. The adults produce proglottids which mature, become gravid, detach from the tapeworm, and migrate to the anus or are passed in the stool (approximately 6 per day). *T. saginata* adults usually have 1,000 to 2,000 proglottids, while *T. solium* adults have an average of 1,000 proglottids. The eggs contained in the gravid proglottids are released after the proglottids are passed with the feces. *T. saginata* may produce up to 100,000 and *T. solium* may produce 50,000 eggs per proglottid respectively.

When the eggs are fully developed, the proglottids separate and are excreted by the host.

19.3.15 Lophotrochozoa

The clade Lophotrochozoa⁶⁸ is named after the two distinct characteristics of its members; the feeding structure lophophore, which is a ciliated crown of tentacles surrounding a mouth, and the developmental stage trophophore larvae. This clade includes the molluscs, annelids, brachiopods, nemerteans (ribbon worms or proboscis worms), bryozoa and entoprocts. The molluscs, the second-largest animal phylum by number of described species, includes snails, clams, and squids, while the annelids are the segmented worms, such as earthworms, lugworms, and leeches. These two groups have long been considered close relatives because they share trophophore larvae.

19.3.16 Brachiopods

Brachiopods⁶⁹ (from the Ancient Greek words brachion ("arm") and podos ("foot")) are a group of lophotrochozoan animals that have hard "valves" (shells) on the upper and lower surfaces, unlike the left and right arrangement in bivalve molluscs. They are often known as "lamp shells", since the curved shells of the class Terebratulida resemble pottery oil-lamps. Brachiopod valves are hinged at the rear end, while the front can be opened for feeding or closed for protection. Two major groups are recognized, articulate and inarticulate. The word "articulate" is used to describe the tooth-and-groove features of the valve-hinge which is present in the articulate group, and absent from the inarticulate group. This is the leading diagnostic feature (fossilizable), by which the two main groups can be readily distinguished. Articulate brachiopods have toothed hinges and simple opening and closing muscles, while inarticulate brachiopods have untoothed hinges and a more complex system of muscles used to keep the two valves aligned. In a typical brachiopod a stalk-like pedicle projects from an opening in one of the valves near the hinges, known as the pedicle valve, keeping the animal anchored to the seabed but clear of silt that would obstruct the opening.

19.3.17 Bryozoa

Bryozoa⁷¹ (also known as the Polyzoa, Ectoprocta or commonly as moss animals) are a phylum of aquatic invertebrate animals. Typically about 0.5 millimetres (1/64 inch) long, they are filter feeders that sieve food particles out of

⁶⁸<https://en.wikipedia.org/wiki/Lophotrochozoa>

⁶⁹<https://en.wikipedia.org/wiki/Brachiopod>

⁷¹<https://en.wikipedia.org/wiki/Bryozoa>



Figure 19.36: *Lingula anatina*, a brachipod, from Stradbroke Island, Australia.⁷⁰

the water using a retractable lophophore, a “crown” of tentacles lined with cilia. Most marine species live in tropical waters, but a few occur in oceanic trenches, and others are found in polar waters. One class lives only in a variety of freshwater environments, and a few members of a mostly marine class prefer brackish water. 5869 living species are known. One genus is solitary and the rest are colonial.



Figure 19.37: An entoproct: *Barentsia discreta*.⁷³

19.3.18 Nemertea

Nemertea⁷² is a phylum of invertebrate animals also known as ribbon worms or proboscis worms. Many have patterns of yellow, orange, red and green coloration. The foregut, stomach and intestine run a little below the midline of the body, the anus is at the tip of the tail, and the mouth is under the front. A little above the gut is the rhynchocoel, a cavity which mostly runs above the midline and ends a little short of the rear of the body. All species have a proboscis which lies in the rhynchocoel when inactive but evicts (turns inside-out) to emerge just above the mouth and capture the animal’s prey with venom. A highly extensible muscle in the back of the rhynchocoel pulls the proboscis in when an attack ends. A few species with stubby bodies filter feed and have suckers at the front and back ends, with which they attach to a host.

Entoprocta, whose name means “anus inside”, or Kamptozoa, is a phylum of mostly sessile aquatic animals, ranging from 0.1 to 7 millimetres (0.004 to 0.3 in) long. Mature individuals are goblet-shaped, on relatively long stalks. They have a “crown” of solid tentacles whose cilia generate water currents that draw food particles towards the mouth, and both the mouth and anus lie inside the “crown”. The superficially similar Bryozoa (Ectoprocta) have the anus outside a “crown” of hollow tentacles. Most families of entoprocts are colonial, and all but 2 of the 150 species are marine. A few solitary species can move slowly.



Figure 19.38: *Cornu aspersum* (formerly *Helix aspersa*) – a common land snail, a gastropod.⁷⁵

19.3.19 Molluscs

Mollusca⁷⁴ is the second-largest phylum of invertebrate animals after the Arthropoda. The members are known as molluscs or mollusks. Around 85,000 extant species of molluscs are recognized. The number of fossil species is estimated between 60,000 and 100,000 additional species. The proportion of undescribed species is very high. Many taxa remain poorly studied.

Molluscs are the largest marine phylum, comprising about 23% of all the named marine organisms. Numerous molluscs also live in freshwater and terrestrial habitats. They are highly diverse, not just in size and anatomical structure, but also in behaviour and habitat. The phylum is typically divided into 8 or 9 taxonomic classes, of which

⁷²<https://en.wikipedia.org/wiki/Nemertea>

⁷⁴<https://en.wikipedia.org/wiki/Mollusca>



Figure 19.39: The common cuttlefish *Sepia officinalis*, a cephalopod.⁷⁶

two are entirely extinct. Cephalopod molluscs, such as squid, cuttlefish, and octopuses, are among the most neurologically advanced of all invertebrates—and either the giant squid or the colossal squid is the largest known invertebrate species. The gastropods (snails and slugs) are by far the most numerous molluscs and account for 80% of the total classified species.

The three most universal features defining modern molluscs are a mantle with a significant cavity used for breathing and excretion, the presence of a radula (except for bivalves), and the structure of the nervous system. Other than these common elements, molluscs express great morphological diversity, so many textbooks base their descriptions on a “hypothetical ancestral mollusc” (see image below). This has a single, “limpet-like” shell on top, which is made of proteins and chitin reinforced with calcium carbonate, and is secreted by a mantle covering the whole upper surface. The underside of the animal consists of a single muscular “foot”. Although molluscs are coelomates, the coelom tends to be small. The main body cavity is a hemocoel through which blood circulates; as such, their circulatory systems are mainly open. The “generalized” mollusc’s feeding system consists of a rasping “tongue”, the radula, and a complex digestive system in which exuded mucus and microscopic, muscle-powered “hairs” called cilia play various important roles. The generalized mollusc has two paired nerve cords, or three in bivalves. The brain, in species that have one, encircles the esophagus. Most molluscs have eyes, and all have sensors to detect chemicals, vibrations, and touch. The simplest type of molluscan reproductive system relies on external fertilization, but more complex variations occur. Nearly all produce eggs, from which may emerge trochophore larvae, more complex veliger larvae, or miniature adults. The coelomic cavity is reduced. They have an open circulatory system and kidney-like organs for excretion.

Good evidence exists for the appearance of gastropods, cephalopods, and bivalves in the Cambrian period, 541–



Figure 19.40: *Argopecten irradians*, the Atlantic bay scallop, a bivalve.⁷⁷

485.4 million years ago. However, the evolutionary history both of molluscs’ emergence from the ancestral Lophotrochozoa and of their diversification into the well-known living and fossil forms are still subjects of vigorous debate among scientists.

Fossilized ammonite displayed at the National Museum of the Philippines Molluscs have been and still are an important food source for anatomically modern humans. A risk of food poisoning exists from toxins that can accumulate in certain molluscs under specific conditions, however, and because of this, many countries have regulations to reduce this risk. Molluscs have, for centuries, also been the source of important luxury goods, notably pearls, mother of pearl, Tyrian purple dye, and sea silk. Their shells have also been used as money in some preindustrial societies.

Mollusc species can also represent hazards or pests for human activities. The bite of the blue-ringed octopus is often fatal, and that of Octopus apollyon causes inflammation that can last over a month. Stings from a few species of large tropical cone shells can also kill, but their sophisticated, though easily produced, venoms have become important tools in neurological research. Schistosomiasis (also known as bilharzia, bilharziosis, or snail fever) is transmitted to humans by water snail hosts, and affects about 200 million people. Snails and slugs can also be serious agricultural pests, and accidental or deliberate introduction of some snail species into new environments has seriously damaged some ecosystems.

Table 19.5: The commonly recognized classes of living molluscs.

Class	Major organisms	Described living species	Distribution
Gastropoda (p300)	all snails and slugs including abalone, limpets, conch, nudibranchs, sea hares, sea butterflies	70,000	marine, freshwater, land
Bivalvia (p367)	clams, oysters, scallops, geoducks, mussels, rudists†	20,000	marine, freshwater
Polyplacophora (pp292–298)	chitons	1,000	rocky tidal zone and seabed
Cephalopoda (p343)	squid, octopuses, cuttlefish, nautiluses, Spirula, belemnites†, ammonites†	900	marine
Scaphopoda (pp403–407)	tusk shells	500	marine 6–7,000 metres (20–22,966 ft)
Aplacophora (pp291–292)	worm-like molluscs	320	seabed 200–3,000 metres (660–9,840 ft)
Monoplacophora (pp298–300)	ancient lineage of molluscs with cap-like shells	31	seabed 1,800–7,000 metres (5,900–23,000 ft); one species 200 metres (660 ft)



Figure 19.41: The earthworm *Lumbricus terrestris* is a representative of the annelids.⁷⁹



Figure 19.42: The leech *Hirudo medicinalis* sucking blood.⁸⁰ Although blood-letting is used less frequently by doctors, some leech species are regarded as endangered species because they have been over-harvested for this purpose in the last few centuries.

19.3.20 Annelids

The annelids⁷⁸ (Annelida, from Latin *anellus*, “little ring”[a]), also known as the ringed worms or segmented worms, are a large phylum, with over 22,000 extant species including ragworms, earthworms, and leeches. The species exist in and have adapted to various ecologies – some in marine environments as distinct as tidal zones and hydrothermal vents, others in fresh water, and yet others in moist terrestrial environments.

The Annelids are bilaterally symmetrical, triploblastic, coelomate, invertebrate organisms. They also have parapodia for locomotion. Most textbooks still use the traditional division into polychaetes (almost all marine), oligochaetes (which include earthworms) and leech-like species. Cladistic research since 1997 has radically changed this scheme, viewing leeches as a sub-group of oligochaetes and oligochaetes as a sub-group of polychaetes. In addition, the Pogonophora, Echiura and Sipuncula, previously regarded as separate phyla, are now regarded as sub-groups of polychaetes. Annelids are considered members of the Lophotrochozoa, a “super-phylum” of protostomes that also includes molluscs, brachiopods, and nemerteans.

The basic annelid form consists of multiple segments. Each segment has the same sets of organs and, in most polychaetes, has a pair of parapodia that many species use for locomotion. Septa separate the segments of many species, but are poorly defined or absent in others, and Echiura and Sipuncula show no obvious signs of segmentation. In species with well-developed septa, the blood circulates entirely within blood vessels, and the vessels in segments near the front ends of these species are often built up with muscles that act as hearts. The



Figure 19.43: *Nereis virens*, a polychaete annelid.⁸¹ These sandworms eat seaweed and microorganisms and can be longer than four feet.

septa of such species also enable them to change the shapes of individual segments, which facilitates movement by peristalsis (“ripples” that pass along the body) or by undulations that improve the effectiveness of the parapodia. In species with incomplete septa or none, the blood circulates through the main body cavity without any kind of pump, and there is a wide range of locomotory techniques – some burrowing species turn their pharynges inside out to drag themselves through the sediment.

Earthworms are oligochaetes that support terrestrial food chains both as prey and in some regions are important in aeration and enriching of soil. Oligochaetes have few setae (chaetae) or “bristles” on their outer body surfaces, and lack parapodia, unlike polychaeta. The body segments of polychaetes or bristle worms have a pair of fleshy protrusions called parapodia that bear many bristles, called chaetae, which are made of chitin. More than 10,000 species are described in this class. The burrowing of marine polychaetes, which may constitute up to a third of all species in near-shore environments, encourages the development of ecosystems by enabling water and oxygen to penetrate the sea floor.

⁷⁸<https://en.wikipedia.org/wiki/Annelid>

Since annelids are soft-bodied, their fossils are rare – mostly jaws and the mineralized tubes that some of the species secreted. Although some late Ediacaran fossils may represent annelids, the oldest known fossil that is identified with confidence comes from about 518 million years ago in the early Cambrian period. Fossils of most modern mobile polychaete groups appeared by the end of the Carboniferous, about 299 million years ago. Palaeontologists disagree about whether some body fossils from the mid Ordovician, about 472 to 461 million years ago, are the remains of oligochaetes, and the earliest indisputable fossils of the group appear in the Tertiary period, which began 66 million years ago.

19.4 Tissues, Organs And Organ Systems

In biology, tissue is a cellular organizational level between cells and a complete organ. A tissue is an ensemble of similar cells and their extracellular matrix from the same origin that together carry out a specific function. Organs are then formed by the functional grouping together of multiple tissues. The English word “tissue” derives from the French word “tissu”, meaning that something that is “woven”, from the verb *tisser*, “to weave”.

An organ is a group of tissues with similar functions. Plant life and animal life rely on many organs that coexist in organ systems.

The study of human and animal tissues is known as histology or, in connection with disease, as histopathology. For plants, the discipline is called plant anatomy. The classical tools for studying tissues are the paraffin block in which tissue is embedded and then sectioned, the histological stain, and the optical microscope. Developments in electron microscopy, immunofluorescence, and the use of frozen tissue-sections have enhanced the detail that can be observed in tissues. With these tools, the classical appearances of tissues can be examined in health and disease, enabling considerable refinement of medical diagnosis and prognosis.

Animal tissues are grouped into four basic types: connective, muscle, nervous, and epithelial. Collections of tissues joined in units to serve a common function compose organs. While all animals can generally be considered to contain the four tissue types, the manifestation of these tissues can differ depending on the type of organism. For example, the origin of the cells comprising a particular tissue type may differ developmentally for different classifications of animals.

The epithelium in all animals is derived from the ectoderm and endoderm, with a small contribution from the

mesoderm, forming the endothelium, a specialized type of epithelium that composes the vasculature. By contrast, a true epithelial tissue is present only in a single layer of cells held together via occluding junctions called tight junctions, to create a selectively permeable barrier. This tissue covers all organismal surfaces that come in contact with the external environment such as the skin, the airways, and the digestive tract. It serves functions of protection, secretion, and absorption, and is separated from other tissues below by a basal lamina.

19.4.1 Epithelial tissue

The epithelial tissues are formed by cells that cover the organ surfaces, such as the surface of skin, the airways, surfaces of soft organs, the reproductive tract, and the inner lining of the digestive tract. The cells comprising an epithelial layer are linked via semi-permeable, tight junctions; hence, this tissue provides a barrier between the external environment and the organ it covers. In addition to this protective function, epithelial tissue may also be specialized to function in secretion, excretion and absorption. Epithelial tissue helps to protect organs from microorganisms, injury, and fluid loss.

Functions of epithelial tissue:

- The principle function of epithelial tissues are covering and lining of free surface
- The cells of the body's surface form the outer layer of skin.
- Inside the body, epithelial cells form the lining of the mouth and alimentary canal and protect these organs.
- Epithelial tissues help in absorption of water and nutrients.
- Epithelial tissues help in the elimination of waste.
- Epithelial tissues secrete enzymes and/or hormones in the form of glands.
- Some epithelial tissue perform secretory functions. They secrete a variety of substances including sweat, saliva, mucus, enzymes.

There are many kinds of epithelium, and nomenclature is somewhat variable. Most classification schemes combine a description of the cell-shape in the upper layer of the epithelium with a word denoting the number of layers: either simple (one layer of cells) or stratified (multiple layers of cells). However, other cellular features such as cilia may also be described in the classification system. Some common kinds of epithelium are listed below:

- Simple squamous epithelium
- Stratified squamous epithelium
- Simple cuboidal epithelium
- Transitional epithelium

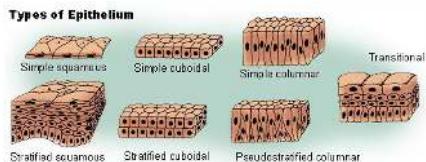


Figure 19.44: Types of epithelia⁸²

- Pseudostratified columnar epithelium (also known as ciliated columnar epithelium)
- Columnar epithelium
- Glandular epithelium

19.4.2 Connective tissue

Connective tissues are fibrous tissues made up of cells separated by non-living material, which is called an extracellular matrix. This matrix can be liquid or rigid. For example, blood contains plasma as its matrix and bone's matrix is rigid. Connective tissue gives shape to organs and holds them in place. Blood, bone, tendon, ligament, adipose, and areolar tissues are examples of connective tissues. One method of classifying connective tissues is to divide them into three types: fibrous connective tissue, skeletal connective tissue, and fluid connective tissue.

19.4.3 Muscular tissue

Muscle cells form the active contractile tissue of the body known as muscle tissue or muscular tissue. Muscle tissue functions to produce force and cause motion, either locomotion or movement within internal organs. Muscle tissue is separated into three distinct categories: visceral or smooth muscle, found in the inner linings of organs; skeletal muscle, typically attached to bones, which generate gross movement; and cardiac muscle, found in the heart, where it contracts to pump blood throughout an organism.

19.4.4 Nervous tissue

Cells comprising the central nervous system and peripheral nervous system are classified as nervous (or neural) tissue. In the central nervous system, neural tissues form the brain and spinal cord. In the peripheral nervous system, neural tissues form the cranial nerves and spinal nerves, inclusive of the motor neurons.

A given organ's tissues can be broadly categorized as parenchyma, the tissue peculiar to (or at least archetypal of) the organ and that does the organ's specialized job, and stroma, the tissues with supportive, structural, connective, or ancillary functions. For example, in a gland, the tissue that makes the hormones is the parenchyma, whereas the

stroma includes the nerves that innervate the parenchyma, the blood vessels that oxygenate and nourish it and carry away its metabolic wastes, and the connective tissues that provide a suitable place for it to be situated and anchored. The main tissues that make up an organ tend to have common embryologic origins, such as arising from the same germ layer. Functionally related organs often cooperate to form whole organ systems. Organs exist in most multicellular organisms. In single-celled organisms such as bacteria, the functional analogue of an organ is known as an organelle. In plants, there are three main organs. A hollow organ is an internal organ that forms a hollow tube, or pouch such as the stomach, intestine, or bladder.

In the study of anatomy, the term viscera refers to an internal organ. Viscera is the plural form.

The number of organs in any organism depends on which precise definition of the term one uses. By one widely used definition, 79 organs have been identified in the human body.

Two or more organs working together in the execution of a specific body function form an organ system, also called a biological system or body system. The functions of organ systems often share significant overlap. For instance, the nervous and endocrine system both operate via a shared organ, the hypothalamus. For this reason, the two systems are combined and studied as the neuroendocrine system. The same is true for the musculoskeletal system because of the relationship between the muscular and skeletal systems.

The organ level of organisation in animals can be first detected in flatworms and the more derived phyla. The less-advanced taxa (like Placozoa, Sponges and Radiata) do not show consolidation of their tissues into organs.

More complex animals are composed of different organs, which have been evolving over time. For example, the liver evolved in the stem vertebrates more than 500 million years ago, while the gut and brain are even more ancient, arising in the ancestor of vertebrates, insects, and worms more than 600 million years ago.

Given the ancient origin of most vertebrate organs, researchers have looked for model systems, where organs have evolved more recently, and ideally have evolved multiple times independently. An outstanding model for this kind of research is the placenta, which has evolved more than 100 times independently in vertebrates, has evolved relatively recently in some lineages, and exists in intermediate forms in extant taxa. Studies on the evolution of the placenta have identified a variety of genetic and physiological processes that contribute to the origin and evolution of organs, these include the re-purposing of existing animal tissues, the acquisition of new functional properties

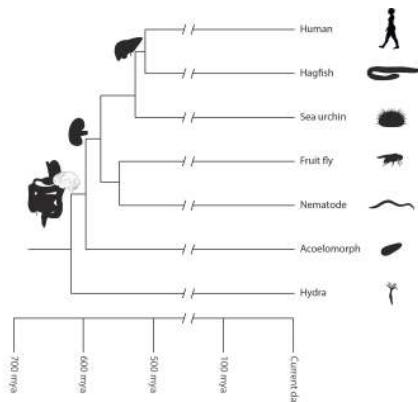


Figure 19.45: Relationship of major animal lineages with indication of how long ago these animals shared a common ancestor. On the left, important organs are shown, which allows us to determine how long ago these may have evolved.⁸³

by these tissues, and novel interactions of distinct tissue types.

Bilateral animals have a variety of organ systems:

- Cardiovascular system: pumping and channeling blood to and from the body and lungs with heart, blood and blood vessels.
- Digestive system: digestion and processing food with salivary glands, esophagus, stomach, liver, gallbladder, pancreas, intestines, colon, rectum and anus.
- Endocrine system: communication within the body using hormones made by endocrine glands such as the hypothalamus, pituitary gland, pineal body or pineal gland, thyroid, parathyroids and adrenals, i.e., adrenal glands.
- Excretory system: kidneys, ureters, bladder and urethra involved in fluid balance, electrolyte balance and excretion of urine.
- Lymphatic system: structures involved in the transfer of lymph between tissues and the blood stream, the lymph and the nodes and vessels that transport it including the Immune system: defending against disease-causing agents with leukocytes, tonsils, adenoids, thymus and spleen.
- Integumentary system: skin, hair and nails of mammals. Also scales of fish, reptiles, and birds, and feathers of birds.
- Muscular system: movement with muscles.
- Nervous system: collecting, transferring and processing information with brain, spinal cord and nerves.
- Reproductive system: the sex organs, such as ovaries, fallopian tubes, uterus, vulva, vagina,

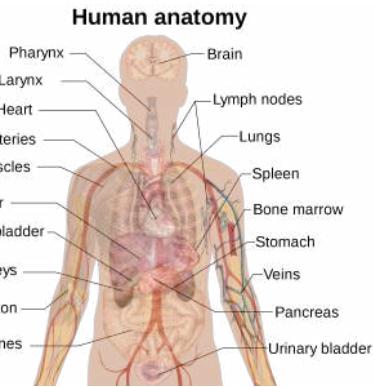


Figure 19.46: Selection of internal organs in human anatomy.⁸⁴

testes, vas deferens, seminal vesicles, prostate and penis.

- Respiratory system: the organs used for breathing, the pharynx, larynx, trachea, bronchi, lungs and diaphragm.
- Skeletal system: structural support and protection with bones, cartilage, ligaments and tendons.

19.4.5 The Integumentary System

The integumentary system⁸⁵ comprises the skin and its appendages acting to protect the body from various kinds of damage, such as loss of water or damages from outside. The integumentary system includes hair, scales, feathers, hooves, and nails. It has a variety of additional functions; it may serve to waterproof, and protect the deeper tissues, excrete wastes, and regulate body temperature, and is the attachment site for sensory receptors to detect pain, sensation, pressure, and temperature. In most land vertebrates with significant exposure to sunlight, the integumentary system also provides for vitamin D synthesis.

The skin is the largest organ of the body. In humans, it accounts for about 12 to 15 percent of total body weight and covers 1.5–2m² of surface area.

The human skin (integument) is composed of at least two major layers of tissue: the epidermis and dermis. (The hypodermis or subcutaneous layer is not part of the skin[citation needed].) The epidermis is the outermost layer, providing the initial barrier to the external environment. It is separated from the dermis by the basement membrane. The epidermis contains melanocytes and gives color to the skin. The deepest layer of epidermis also contains nerve endings. Beneath this, the dermis comprises two sections, the papillary and reticular layers, and contains connective tissues, vessels, glands, follicles,

⁸⁵ https://en.wikipedia.org/wiki/Integumentary_system

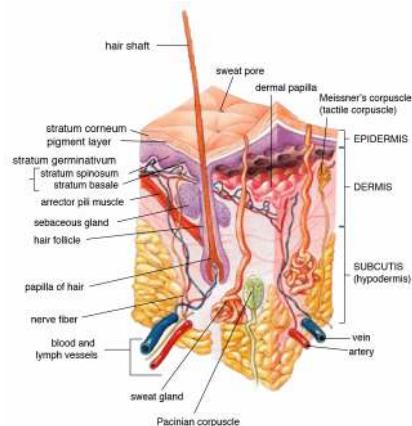


Figure 19.47: Anatomy of human skin.⁸⁶

hair roots, sensory nerve endings, and muscular tissue. The deepest layer, the hypodermis, is primarily made up of adipose tissue. Substantial collagen bundles anchor the dermis to the hypodermis in a way that permits most areas of the skin to move freely over the deeper tissue layers

The epidermis is the top layer of skin made up of epithelial cells. It contains blood vessels. Its main functions are protection, absorption of nutrients, and homeostasis. In structure, it consists of a keratinized stratified squamous epithelium; four types of cells: keratinocytes, melanocytes, Merkel cells, and Langerhans cells. The major cell of the epidermis is the keratinocyte, which produces keratin, a fibrous protein that aids in skin protection. An overwhelming amount of keratin can cause disease by giving rise to eruptions from the skin that will protrude outwards and lead to infection.[citation needed] Keratin is also a waterproofing protein. Millions of dead keratinocytes rub off daily. The majority of the skin on the body is keratinized. The only skin on the body that is non-keratinized is the lining of mucous membranes, such as the inside of the mouth. Non-keratinized cells allow water to "stay" atop the structure.

The protein keratin stiffens epidermal tissue to form fingernails. Nails grow from a thin area called the nail matrix at an average of 1 mm per week. The lunula is the crescent-shape area at the base of the nail, lighter in color as it mixes with matrix cells. Also, the stratum corneum is the top part of the epidermis.

The dermis is the middle layer of skin, composed of dense irregular connective tissue and areolar connective tissue such as a collagen with elastin arranged in a diffusely bundled and woven pattern. The dermis has two layers. One is the papillary layer which is the superficial layer and consists of the areolar connective tissue. The

other is the reticular layer which is the deep layer of the dermis and consists of the dense irregular connective tissue. These layers serve to give elasticity to the integument, allowing stretching and conferring flexibility, while also resisting distortions, wrinkling, and sagging. The dermal layer provides a site for the endings of blood vessels and nerves. Many chromatophores are also stored in this layer, as are the bases of integumental structures such as hair, feathers, and glands.

The hypodermis, otherwise known as the subcutaneous layer, is a layer beneath the skin. It invaginates into the dermis and is attached to the latter, immediately above it, by collagen and elastin fibers. It is essentially composed of a type of cell known as adipocytes specialized in accumulating and storing fats. These cells are grouped together in lobules separated by connective tissue.

The hypodermis acts as an energy reserve. The fats contained in the adipocytes can be put back into circulation, via the venous route, during intense effort or when there is a lack of energy-providing substances, and are then transformed into energy. The hypodermis participates, passively at least, in thermoregulation since fat is a heat insulator.

The integumentary system has multiple roles in maintaining the body's equilibrium. All body systems work in an interconnected manner to maintain the internal conditions essential to the function of the body. The skin has an important job of protecting the body and acts as the body's first line of defense against infection, temperature change, and other challenges to homeostasis. Functions include:

- Protect the body's internal living tissues and organs
- Protect against invasion by infectious organisms
- Protect the body from dehydration
- Protect the body against abrupt changes in temperature, maintain homeostasis
- Help excrete waste materials through perspiration
- Act as a receptor for touch, pressure, pain, heat, and cold (see Somatosensory system)
- Protect the body against sunburns by secreting melanin
- Generate vitamin D through exposure to ultraviolet light
- Store water, fat, glucose, vitamin D
- Maintenance of the body form
- Formation of new cells from stratum germinativum to repair minor injuries
- Protect from UV rays.
- Regulates body temperature

It distinguishes, separates, and protects the organism from its surroundings. Small-bodied invertebrates of



Figure 19.48: All extant species of hominid:⁸⁸ [Top) Human, Middle) Common chimpanzee, bonobo, western gorilla & eastern gorilla, Bottom) Bornean, Sumatran & Tapanuli orangutan]

aquatic or continually moist habitats respire using the outer layer (integument). This gas exchange system, where gases simply diffuse into and out of the interstitial fluid, is called integumentary exchange.

19.5 Human evolution

Human evolution⁸⁷ is the evolutionary process that led to the emergence of anatomically modern humans, beginning with the evolutionary history of primates—in particular genus *Homo*—and leading to the emergence of *Homo sapiens* as a distinct species of the hominid family, which includes the great apes. This process involved the gradual development of traits such as human bipedalism and language, as well as interbreeding with other hominins, which indicate that human evolution was not linear but a web.

The study of human evolution involves several scientific disciplines, including physical anthropology, primatol-

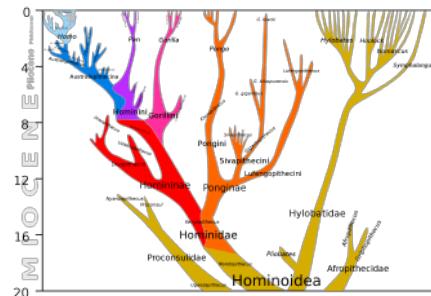


Figure 19.49: Model of the phylogeny of Hominidae, with adjacent branches of Hominoidea, over the past 20 million years.⁸⁹

ogy, archaeology, paleontology, neurobiology, ethology, linguistics, evolutionary psychology, embryology and genetics. Genetic studies show that primates diverged from other mammals about 85 million years ago, in the Late Cretaceous period, and the earliest fossils appear in the Paleocene, around 55 million years ago.

Within the superfamily Hominoidea, the family Hominidae⁸⁹ diverged from the family Hylobatidae some 15–20 million years ago; subfamily Homininae (African apes) diverged from Ponginae (orangutans) about 14 million years ago; the tribe Hominini (including humans, *Australopithecus*, and chimpanzees) parted from the tribe Gorillini (gorillas) between 8–9 million years ago; and, in turn, the subtribes Hominina (humans and extinct biped ancestors) and Panina (chimpanzees) separated 4–7 million years ago.

Human evolution from its first separation from the last common ancestor of humans and chimpanzees is characterized by a number of morphological, developmental, physiological, and behavioral changes. The most significant of these adaptations are bipedalism, increased brain size, lengthened ontogeny (gestation and infancy), and decreased sexual dimorphism. The relationship between these changes is the subject of ongoing debate. Other significant morphological changes included the evolution of a power and precision grip, a change first occurring in *H. erectus*.

Bipedalism is the basic adaptation of the hominid and is considered the main cause behind a suite of skeletal changes shared by all bipedal hominids. The earliest hominin, of presumably primitive bipedalism, is considered to be either *Sahelanthropus* or *Orrorin*, both of which arose some 6 to 7 million years ago. The non-bipedal knuckle-walkers, the gorillas and chimpanzees, diverged from the hominin line over a period covering the same time, so either *Sahelanthropus* or *Orrorin* may be our last shared an-

⁸⁷ https://en.wikipedia.org/wiki/Human_evolution

⁸⁹ <https://en.wikipedia.org/wiki/Hominidae>

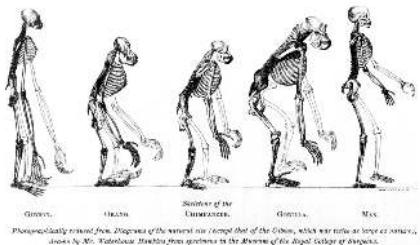


Figure 19.50: From Thomas Huxley's 1863 Evidence as to Man's Place in Nature, the compared skeletons of apes to humans.⁹¹

cestor. *Ardipithecus*, a full biped, arose approximately 5.6 million years ago.

The early bipeds eventually evolved into the australopithecines and still later into the genus *Homo*. There are several theories of the adaptation value of bipedalism. It is possible that bipedalism was favored because it freed the hands for reaching and carrying food, saved energy during locomotion, enabled long-distance running and hunting, provided an enhanced field of vision, and helped avoid hyperthermia by reducing the surface area exposed to direct sun; features all advantageous for thriving in the new savanna and woodland environment created as a result of the East African Rift Valley uplift versus the previous closed forest habitat. A 2007 study provides support for the hypothesis that walking on two legs, or bipedalism, evolved because it used less energy than quadrupedal knuckle-walking. However, recent studies suggest that bipedality without the ability to use fire would not have allowed global dispersal. This change in gait saw a lengthening of the legs proportionately when compared to the length of the arms, which were shortened through the removal of the need for brachiation. Another change is the shape of the big toe. Recent studies suggest that australopithecines still lived part of the time in trees as a result of maintaining a grasping big toe. This was progressively lost in *habilines*.

Anatomically, the evolution of bipedalism has been accompanied by a large number of skeletal changes, not just to the legs and pelvis, but also to the vertebral column, feet and ankles, and skull. The femur evolved into a slightly more angular position to move the center of gravity toward the geometric center of the body. The knee and ankle joints became increasingly robust to better support increased weight. To support the increased weight on each vertebra in the upright position, the human vertebral column became S-shaped and the lumbar vertebrae became shorter and wider. In the feet the big toe moved into alignment with the other toes to help in forward locomotion. The arms and forearms shortened relative to the legs mak-

ing it easier to run. The foramen magnum migrated under the skull and more anterior.

The most significant changes occurred in the pelvic region, where the long downward facing iliac blade was shortened and widened as a requirement for keeping the center of gravity stable while walking; bipedal hominids have a shorter but broader, bowl-like pelvis due to this. A drawback is that the birth canal of bipedal apes is smaller than in knuckle-walking apes, though there has been a widening of it in comparison to that of australopithecine and modern humans, permitting the passage of newborns due to the increase in cranial size but this is limited to the upper portion, since further increase can hinder normal bipedal movement.

The shortening of the pelvis and smaller birth canal evolved as a requirement for bipedalism and had significant effects on the process of human birth which is much more difficult in modern humans than in other primates. During human birth, because of the variation in size of the pelvic region, the fetal head must be in a transverse position (compared to the mother) during entry into the birth canal and rotate about 90 degrees upon exit. The smaller birth canal became a limiting factor to brain size increases in early humans and prompted a shorter gestation period leading to the relative immaturity of human offspring, who are unable to walk much before 12 months and have greater neoteny, compared to other primates, who are mobile at a much earlier age. The increased brain growth after birth and the increased dependency of children on mothers had a major effect upon the female reproductive cycle, and the more frequent appearance of allo-parenting in humans when compared with other hominids. Delayed human sexual maturity also led to the evolution of menopause with one explanation providing that elderly women could better pass on their genes by taking care of their daughter's offspring, as compared to having more children of their own.

The human species eventually developed a much larger brain than that of other primates—typically 1,330 cm³ (81 cu in) in modern humans, nearly three times the size of a chimpanzee or gorilla brain. After a period of stasis with *Australopithecus anamensis* and *Ardipithecus*, species which had smaller brains as a result of their bipedal locomotion, the pattern of encephalization started with *Homo habilis*, whose 600 cm³ (37 cu in) brain was slightly larger than that of chimpanzees. This evolution continued in *Homo erectus* with 800–1,100 cm³ (49–67 cu in), and reached a maximum in Neanderthals with 1,200–1,900 cm³ (73–116 cu in), larger even than modern *Homo sapiens*. This brain increase manifested during postnatal brain growth, far exceeding that of other apes (heterochrony). It also allowed for extended periods

of social learning and language acquisition in juvenile humans, beginning as much as 2 million years ago.

Furthermore, the changes in the structure of human brains may be even more significant than the increase in size.

The temporal lobes, which contain centers for language processing, have increased disproportionately, as has the prefrontal cortex, which has been related to complex decision-making and moderating social behavior. Encephalization has been tied to increased meat and starches in the diet, and the development of cooking, and it has been proposed that intelligence increased as a response to an increased necessity for solving social problems as human society became more complex. Changes in skull morphology, such as smaller mandibles and mandible muscle attachments, allowed more room for the brain to grow.

The increase in volume of the neocortex also included a rapid increase in size of the cerebellum. Its function has traditionally been associated with balance and fine motor control, but more recently with speech and cognition. The great apes, including hominids, had a more pronounced cerebellum relative to the neocortex than other primates. It has been suggested that because of its function of sensory-motor control and learning complex muscular actions, the cerebellum may have underpinned human technological adaptations, including the preconditions of speech.

The immediate survival advantage of encephalization is difficult to discern, as the major brain changes from *Homo erectus* to *Homo heidelbergensis* were not accompanied by major changes in technology. It has been suggested that the changes were mainly social and behavioural, including increased empathic abilities, increases in size of social groups, and increased behavioural plasticity. Encephalization may be due to a dependency on calorie-dense, difficult-to-acquire food.

The reduced degree of sexual dimorphism in humans is visible primarily in the reduction of the male canine tooth relative to other ape species (except gibbons) and reduced brow ridges and general robustness of males. Another important physiological change related to sexuality in humans was the evolution of hidden estrus. Humans are the only hominoids in which the female is fertile year round and in which no special signals of fertility are produced by the body (such as genital swelling or overt changes in proceptivity during estrus).

Nonetheless, humans retain a degree of sexual dimorphism in the distribution of body hair and subcutaneous fat, and in the overall size, males being around 15% larger than females. These changes taken together have been interpreted as a result of an increased emphasis on pair

bonding as a possible solution to the requirement for increased parental investment due to the prolonged infancy of offspring.

The ulnar opposition—the contact between the thumb and the tip of the little finger of the same hand—is unique to the genus *Homo*, including Neanderthals, the Sima de los Huesos hominins and anatomically modern humans. In other primates, the thumb is short and unable to touch the little finger. The ulnar opposition facilitates the precision grip and power grip of the human hand, underlying all the skilled manipulations.

A number of other changes have also characterized the evolution of humans, among them an increased importance on vision rather than smell; a longer juvenile developmental period and higher infant dependency; a smaller gut; faster basal metabolism; loss of body hair; evolution of sweat glands; a change in the shape of the dental arcade from being u-shaped to being parabolic; development of a chin (found in *Homo sapiens* alone); development of styloid processes; and the development of a descended larynx.

The word *homo*, the name of the biological genus to which humans belong, is Latin for "human". It was chosen originally by Carl Linnaeus in his classification system. The word "human" is from the Latin *humanus*, the adjectival form of *homo*. The Latin "*homo*" derives from the Indo-European root *dhghem, or "earth". Linnaeus and other scientists of his time also considered the great apes to be the closest relatives of humans based on morphological and anatomical similarities.

The possibility of linking humans with earlier apes by descent became clear only after 1859 with the publication of Charles Darwin's *On the Origin of Species*, in which he argued for the idea of the evolution of new species from earlier ones. Darwin's book did not address the question of human evolution, saying only that "Light will be thrown on the origin of man and his history."

The first debates about the nature of human evolution arose between Thomas Henry Huxley and Richard Owen. Huxley argued for human evolution from apes by illustrating many of the similarities and differences between humans and apes, and did so particularly in his 1863 book *Evidence as to Man's Place in Nature*. However, many of Darwin's early supporters (such as Alfred Russel Wallace and Charles Lyell) did not initially agree that the origin of the mental capacities and the moral sensibilities of humans could be explained by natural selection, though this later changed. Darwin applied the theory of evolution and sexual selection to humans when he published *The Descent of Man* in 1871.

A major problem in the 19th century was the lack of fossil intermediaries. Neanderthal remains were discov-

ered in a limestone quarry in 1856, three years before the publication of *On the Origin of Species*, and Neanderthal fossils had been discovered in Gibraltar even earlier, but it was originally claimed that these were human remains of a creature suffering some kind of illness. Despite the 1891 discovery by Eugène Dubois of what is now called *Homo erectus* at Trinil, Java, it was only in the 1920s when such fossils were discovered in Africa, that intermediate species began to accumulate. In 1925, Raymond Dart described *Australopithecus africanus*. The type specimen was the Taung Child, an australopithecine infant which was discovered in a cave. The child's remains were a remarkably well-preserved tiny skull and an endocast of the brain.

Although the brain was small (410 cm^3), its shape was rounded, unlike that of chimpanzees and gorillas, and more like a modern human brain. Also, the specimen showed short canine teeth, and the position of the foramen magnum (the hole in the skull where the spine enters) was evidence of bipedal locomotion. All of these traits convinced Dart that the Taung Child was a bipedal human ancestor, a transitional form between apes and humans.

During the 1960s and 1970s, hundreds of fossils were found in East Africa in the regions of the Olduvai Gorge and Lake Turkana. These searches were carried out by the Leakey family, with Louis Leakey and his wife Mary Leakey, and later their son Richard and daughter-in-law Meave, fossil hunters and paleoanthropologists. From the fossil beds of Olduvai and Lake Turkana they amassed specimens of the early hominins: the australopithecines and *Homo* species, and even *Homo erectus*.

These finds cemented Africa as the cradle of humankind. In the late 1970s and the 1980s, Ethiopia emerged as the new hot spot of paleoanthropology after "Lucy", the most complete fossil member of the species *Australopithecus afarensis*, was found in 1974 by Donald Johanson near Hadar in the desertic Afar Triangle region of northern Ethiopia. Although the specimen had a small brain, the pelvis and leg bones were almost identical in function to those of modern humans, showing with certainty that these hominins had walked erect. Lucy was classified as a new species, *Australopithecus afarensis*, which is thought to be more closely related to the genus *Homo* as a direct ancestor, or as a close relative of an unknown ancestor, than any other known hominid or hominin from this early time range; see terms "hominid" and "hominin". (The specimen was nicknamed "Lucy" after the Beatles' song "Lucy in the Sky with Diamonds", which was played loudly and repeatedly in the camp during the excavations.) The Afar Triangle area would later yield discovery of many more hominin fossils, particularly those uncovered or described by teams headed by Tim D.

White in the 1990s, including *Ardipithecus ramidus* and *Ardipithecus kadabba*.

In 2013, fossil skeletons of *Homo naledi*, an extinct species of hominin assigned (provisionally) to the genus *Homo*, were found in the Rising Star Cave system, a site in South Africa's Cradle of Humankind region in Gauteng province near Johannesburg. As of September 2015, fossils of at least fifteen individuals, amounting to 1,550 specimens, have been excavated from the cave. The species is characterized by a body mass and stature similar to small-bodied human populations, a smaller endocranial volume similar to *Australopithecus*, and a cranial morphology (skull shape) similar to early *Homo* species. The skeletal anatomy combines primitive features known from australopithecines with features known from early hominins. The individuals show signs of having been deliberately disposed of within the cave near the time of death. The fossils were dated close to 250,000 years ago, and thus are not a direct ancestor but a contemporary with the first appearance of larger-brained anatomically modern humans.

The genetic revolution in studies of human evolution started when Vincent Sarich and Allan Wilson measured the strength of immunological cross-reactions of blood serum albumin between pairs of creatures, including humans and African apes (chimpanzees and gorillas). The strength of the reaction could be expressed numerically as an immunological distance, which was in turn proportional to the number of amino acid differences between homologous proteins in different species. By constructing a calibration curve of the ID of species' pairs with known divergence times in the fossil record, the data could be used as a molecular clock to estimate the times of divergence of pairs with poorer or unknown fossil records.

In their seminal 1967 paper in *Science*, Sarich and Wilson estimated the divergence time of humans and apes as four to five million years ago, at a time when standard interpretations of the fossil record gave this divergence as at least 10 to as much as 30 million years. Subsequent fossil discoveries, notably "Lucy", and reinterpretation of older fossil materials, notably *Ramapithecus*, showed the younger estimates to be correct and validated the albumin method.

Progress in DNA sequencing, specifically mitochondrial DNA (mtDNA) and then Y-chromosome DNA (Y-DNA) advanced the understanding of human origins. Application of the molecular clock principle revolutionized the study of molecular evolution.

On the basis of a separation from the orangutan between 10 and 20 million years ago, earlier studies of the molecular clock suggested that there were about 76 mu-

tations per generation that were not inherited by human children from their parents; this evidence supported the divergence time between hominins and chimpanzees noted above. However, a 2012 study in Iceland of 78 children and their parents suggests a mutation rate of only 36 mutations per generation; this datum extends the separation between humans and chimpanzees to an earlier period greater than 7 million years ago (Ma). Additional research with 226 offspring of wild chimpanzee populations in eight locations suggests that chimpanzees reproduce at age 26.5 years on average; which suggests the human divergence from chimpanzees occurred between 7 and 13 million years ago. And these data suggest that *Ardipithecus* (4.5 Ma), *Orrorin* (6 Ma) and *Sahelanthropus* (7 Ma) all may be on the hominid lineage, and even that the separation may have occurred outside the East African Rift region.

Furthermore, analysis of the two species' genes in 2006 provides evidence that after human ancestors had started to diverge from chimpanzees, interspecies mating between "proto-human" and "proto-chimpanzees" nonetheless occurred regularly enough to change certain genes in the new gene pool:

A new comparison of the human and chimpanzee genomes suggests that after the two lineages separated, they may have begun interbreeding... A principal finding is that the X chromosomes of humans and chimpanzees appear to have diverged about 1.2 million years more recently than the other chromosomes.

The research suggests:

There were in fact two splits between the human and chimpanzee lineages, with the first being followed by interbreeding between the two populations and then a second split. The suggestion of a hybridization has startled paleoanthropologists, who nonetheless are treating the new genetic data seriously.

In the 1990s, several teams of paleoanthropologists were working throughout Africa looking for evidence of the earliest divergence of the hominin lineage from the great apes. In 1994, Meave Leakey discovered *Australopithecus anamensis*. The find was overshadowed by Tim D. White's 1995 discovery of *Ardipithecus ramidus*, which pushed back the fossil record to 4.2 million years ago.

In 2000, Martin Pickford and Brigitte Senut discovered, in the Tugen Hills of Kenya, a 6-million-year-old bipedal hominin which they named *Orrorin tugenensis*. And in 2001, a team led by Michel Brunet discovered the skull of *Sahelanthropus tchadensis* which was dated as 7.2 million years ago, and which Brunet argued was a bipedal, and therefore a hominid—that is, a hominin (cf Hominidae; terms "hominids" and hominins).

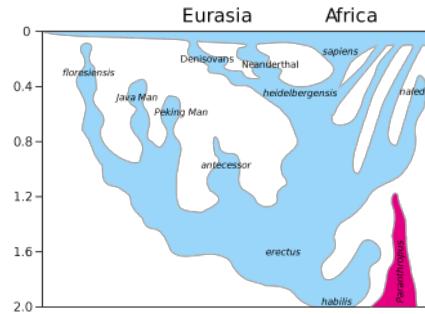


Figure 19.51: A model of the evolution of the genus *Homo*⁹² over the last 2 million years (vertical axis). The rapid "Out of Africa" expansion of *H. sapiens* is indicated at the top of the diagram, with admixture indicated with Neanderthals, Denisovans, and unspecified archaic African hominins. Late survival of robust australopithecines (*Paranthropus*) alongside *Homo* until 1.2 Mya is indicated in purple.

19.5.1 Human Dispersal

Anthropologists in the 1980s were divided regarding some details of reproductive barriers and migratory dispersals of the genus *Homo*. Subsequently, genetics has been used to investigate and resolve these issues. According to the Sahara pump theory evidence suggests that the genus *Homo* have migrated out of Africa at least three and possibly four times (e.g. *Homo erectus*, *Homo heidelbergensis* and two or three times for *Homo sapiens*). Recent evidence suggests these dispersals are closely related to fluctuating periods of climate change.

Recent evidence suggests that humans may have left Africa half a million years earlier than previously thought. A joint Franco-Indian team has found human artifacts in the Siwalk Hills north of New Delhi dating back at least 2.6 million years. This is earlier than the previous earliest finding of genus *Homo* at Dmanisi, in Georgia, dating to 1.85 million years. Although controversial, tools found at a Chinese cave strengthen the case that humans used tools as far back as 2.48 million years ago. This suggests that the Asian "Chopper" tool tradition, found in Java and northern China may have left Africa before the appearance of the Acheulian hand axe.

Up until the genetic evidence became available, there were two dominant models for the dispersal of modern humans. The multiregional hypothesis proposed that the genus *Homo* contained only a single interconnected population as it does today (not separate species), and that its evolution took place worldwide continuously over the last couple of million years. This model was proposed in 1988 by Milford H. Wolpoff. In contrast, the "out of Africa" model proposed that modern *H. sapiens* speciated in Africa

recently (that is, approximately 200,000 years ago) and the subsequent migration through Eurasia resulted in the nearly complete replacement of other *Homo* species. This model has been developed by Chris B. Stringer and Peter Andrews.

Sequencing mtDNA and Y-DNA sampled from a wide range of indigenous populations revealed ancestral information relating to both male and female genetic heritage, and strengthened the “out of Africa” theory and weakened the views of multiregional evolutionism. Aligned in genetic tree differences were interpreted as supportive of a recent single origin. Analyses have shown a greater diversity of DNA patterns throughout Africa, consistent with the idea that Africa is the ancestral home of mitochondrial Eve and Y-chromosomal Adam, and that modern human dispersal out of Africa has only occurred over the last 55,000 years.

“Out of Africa” has thus gained much support from research using female mitochondrial DNA and the male Y chromosome. After analysing genealogy trees constructed using 133 types of mtDNA, researchers concluded that all were descended from a female African progenitor, dubbed Mitochondrial Eve. “Out of Africa” is also supported by the fact that mitochondrial genetic diversity is highest among African populations.

A broad study of African genetic diversity, headed by Sarah Tishkoff, found the San people had the greatest genetic diversity among the 113 distinct populations sampled, making them one of 14 “ancestral population clusters”. The research also located a possible origin of modern human migration in southwestern Africa, near the coastal border of Namibia and Angola. The fossil evidence was insufficient for archaeologist Richard Leakey to resolve the debate about exactly where in Africa modern humans first appeared. Studies of haplogroups in Y-chromosomal DNA and mitochondrial DNA have largely supported a recent African origin. All the evidence from autosomal DNA also predominantly supports a Recent African origin. However, evidence for archaic admixture in modern humans, both in Africa and later, throughout Eurasia has recently been suggested by a number of studies.

Recent sequencing of Neanderthal and Denisovan genomes shows that some admixture with these populations has occurred. All modern human groups outside Africa have 1–4% or (according to more recent research) about 1.5–2.6% Neanderthal alleles in their genome, and some Melanesians have an additional 4–6% of Denisovan alleles. These new results do not contradict the “out of Africa” model, except in its strictest interpretation, although they make the situation more complex. After recovery from a genetic bottleneck that some researchers speculate might be linked to the Toba supervolcano



Figure 19.52: Dermoplastics reconstruction of a Neanderthal⁹³

catastrophe, a fairly small group left Africa and interbred with Neanderthals, probably in the Middle East, on the Eurasian steppe or even in North Africa before their departure. Their still predominantly African descendants spread to populate the world. A fraction in turn interbred with Denisovans, probably in southeastern Asia, before populating Melanesia. HLA haplotypes of Neanderthal and Denisova origin have been identified in modern Eurasian and Oceanian populations. The Denisovan EPAS1 gene has also been found in Tibetan populations. Studies of the human genome using machine learning have identified additional genetic contributions in Eurasians from an “unknown” ancestral population potentially related to the Neanderthal-Denisovan lineage.

There are still differing theories on whether there was a single exodus from Africa or several. A multiple dispersal model involves the Southern Dispersal theory, which has gained support in recent years from genetic, linguistic and archaeological evidence. In this theory, there was a coastal dispersal of modern humans from the Horn of Africa crossing the Bab el Mandib to Yemen at a lower sea level around 70,000 years ago. This group helped to populate Southeast Asia and Oceania, explaining the discovery of early human sites in these areas much earlier than those in the Levant. This group seems to have been dependent upon marine resources for their survival.

Recent genetic evidence suggests that all modern non-African populations, including those of Eurasia and Oceania, are descended from a single wave that left Africa between 65,000 and 50,000 years ago.

19.5.2 Evidence of Human Evolution

The evidence on which scientific accounts of human evolution are based comes from many fields of natural science. The main source of knowledge about the evolutionary process has traditionally been the fossil record, but since the development of genetics beginning in the 1970s, DNA analysis has come to occupy a place of comparable importance. The studies of ontogeny, phylogeny and especially evolutionary developmental biology of both vertebrates and invertebrates offer considerable insight into the evolution of all life, including how humans evolved. The specific study of the origin and life of humans is anthropology, particularly paleoanthropology which focuses on the study of human prehistory.

The closest living relatives of humans are bonobos and chimpanzees (both genus *Pan*) and gorillas (genus *Gorilla*). With the sequencing of both the human and chimpanzee genome, as of 2012 estimates of the similarity between their DNA sequences range between 95% and 99%. By using the technique called the molecular clock which estimates the time required for the number of divergent mutations to accumulate between two lineages, the approximate date for the split between lineages can be calculated.

The gibbons (family *Hylobatidae*) and then the orangutans (genus *Pongo*) were the first groups to split from the line leading to the hominins, including humans—followed by gorillas (genus *Gorilla*), and, ultimately, by the chimpanzees (genus *Pan*). The splitting date between hominin and chimpanzee lineages is placed by some between 4 to 8 million years ago, that is, during the Late Miocene. Speciation, however, appears to have been unusually drawn out. Initial divergence occurred sometime between 7 to 13 million years ago, but ongoing hybridization blurred the separation and delayed complete separation during several millions of years. Patterson (2006) dated the final divergence at 5 to 6 million years ago.

Genetic evidence has also been employed to resolve the question of whether there was any gene flow between early modern humans and Neanderthals, and to enhance our understanding of the early human migration patterns and splitting dates. By comparing the parts of the genome that are not under natural selection and which therefore accumulate mutations at a fairly steady rate, it is possible to reconstruct a genetic tree incorporating the entire human species since the last shared ancestor.

Each time a certain mutation (single-nucleotide polymorphism) appears in an individual and is passed on to his or her descendants, a haplogroup is formed including all of the descendants of the individual who will also carry that mutation. By comparing mitochondrial DNA which is inher-

ited only from the mother, geneticists have concluded that the last female common ancestor whose genetic marker is found in all modern humans, the so-called mitochondrial Eve, must have lived around 200,000 years ago.

Human evolutionary genetics studies how one human genome differs from the other, the evolutionary past that gave rise to it, and its current effects. Differences between genomes have anthropological, medical and forensic implications and applications. Genetic data can provide important insight into human evolution.

There is little fossil evidence for the divergence of the gorilla, chimpanzee and hominin lineages. The earliest fossils that have been proposed as members of the hominin lineage are *Sahelanthropus tchadensis* dating from 7 million years ago, *Orrorin tugenensis* dating from 5.7 million years ago, and *Ardipithecus kadabba* dating to 5.6 million years ago. Each of these have been argued to be a bipedal ancestor of later hominins but, in each case, the claims have been contested. It is also possible that one or more of these species are ancestors of another branch of African apes, or that they represent a shared ancestor between hominins and other apes.

The question then of the relationship between these early fossil species and the hominin lineage is still to be resolved. From these early species, the australopithecines arose around 4 million years ago and diverged into robust (also called *Paranthropus*) and gracile branches, one of which (possibly *A. garhi*) probably went on to become ancestors of the genus *Homo*. The australopithecine species that is best represented in the fossil record is *Australopithecus afarensis* with more than 100 fossil individuals represented, found from Northern Ethiopia (such as the famous "Lucy"), to Kenya, and South Africa. Fossils of robust australopithecines such as *Au. robustus* (or alternatively *Paranthropus robustus*) and *Au./P. boisei* are particularly abundant in South Africa at sites such as Kromdraai and Swartkrans, and around Lake Turkana in Kenya.

The earliest member of the genus *Homo* is *Homo habilis* which evolved around 2.8 million years ago. *Homo habilis* is the first species for which we have positive evidence of the use of stone tools. They developed the Oldowan lithic technology, named after the Olduvai Gorge in which the first specimens were found. Some scientists consider *Homo rudolfensis*, a larger bodied group of fossils with similar morphology to the original *H. habilis* fossils, to be a separate species, while others consider them to be part of *H. habilis*—simply representing intraspecies variation, or perhaps even sexual dimorphism. The brains of these early hominins were about the same size as that of a chimpanzee, and their main adaptation was bipedalism as an adaptation to terrestrial living.



Figure 19.53: Replica of fossil skull of *Homo ergaster* (African *Homo erectus*). Fossil number Khm-Heu 3733 discovered in 1975 in Kenya.⁹⁴

During the next million years, a process of encephalization began and, by the arrival (about 1.9 million years ago) of *Homo erectus* in the fossil record, cranial capacity had doubled. *Homo erectus* were the first of the hominins to emigrate from Africa, and, from 1.8 to 1.3 million years ago, this species spread through Africa, Asia, and Europe. One population of *H. erectus*, also sometimes classified as a separate species *Homo ergaster*, remained in Africa and evolved into *Homo sapiens*. It is believed that these species, *H. erectus* and *H. ergaster*, were the first to use fire and complex tools.

The earliest transitional fossils between *H. ergaster/erectus* and archaic *H. sapiens* are from Africa, such as *Homo rhodesiensis*. These descendants of African *H. erectus* spread through Eurasia from ca. 500,000 years ago, evolving into *H. antecessor*, *H. heidelbergensis* and *H. neanderthalensis*. The earliest fossils of anatomically modern humans are from the Middle Paleolithic, about 300–200,000 years ago such as the Herto and Omo remains of Ethiopia, Jebel Irhoud remains of Morocco, and Florisbad remains of South Africa; later fossils from Es Skhul cave in Israel and Southern Europe begin around 90,000 years ago (0.09 million years ago).

As modern humans spread out from Africa, they encountered other hominins such as *Homo neanderthalensis* and the Denisovans, who may have evolved from populations of *Homo erectus* that had left Africa around 2 million years ago. The nature of interaction between early humans and these sister species has been a long-standing source of controversy, the question being whether humans replaced these earlier species or whether they were in fact

similar enough to interbreed, in which case these earlier populations may have contributed genetic material to modern humans.

This migration out of Africa is estimated to have begun about 70–50,000 years BP and modern humans subsequently spread globally, replacing earlier hominins either through competition or hybridization. They inhabited Eurasia and Oceania by 40,000 years BP, and the Americas by at least 14,500 years BP.

The hypothesis of interbreeding, also known as hybridization, admixture or hybrid-origin theory, has been discussed ever since the discovery of Neanderthal remains in the 19th century. The linear view of human evolution began to be abandoned in the 1970s as different species of humans were discovered that made the linear concept increasingly unlikely. In the 21st century with the advent of molecular biology techniques and computerization, whole-genome sequencing of Neanderthal and human genome were performed, confirming recent admixture between different human species. In 2010, evidence based on molecular biology was published, revealing unambiguous examples of interbreeding between archaic and modern humans during the Middle Paleolithic and early Upper Paleolithic. It has been demonstrated that interbreeding happened in several independent events that included Neanderthals and Denisovans, as well as several unidentified hominins. Today, approximately 2% of DNA from all non-African populations (including Europeans, Asians, and Oceanians) is Neanderthal, with traces of Denisovan heritage. Also, 4–6% of modern Melanesian genetics are Denisovan. Comparisons of the human genome to the genomes of Neandertals, Denisovans and apes can help identify features that set modern humans apart from other hominin species. In a 2016 comparative genomics study, a Harvard Medical School/UCLA research team made a world map on the distribution and made some predictions about where Denisovan and Neanderthal genes may be impacting modern human biology.

For example, comparative studies in the mid-2010s found several traits related to neurological, immunological, developmental, and metabolic phenotypes, that were developed by archaic humans to European and Asian environments and inherited to modern humans through admixture with local hominins.

Although the narratives of human evolution are often contentious, several discoveries since 2010 show that human evolution should not be seen as a simple linear or branched progression, but a mix of related species. In fact, genomic research has shown that hybridization between substantially diverged lineages is the rule, not the exception, in human evolution. Furthermore, it is argued that hybridization was an essential creative force in the emer-

gence of modern humans.

19.5.3 Early Evolution of Primates

The evolutionary history of the primates can be traced back 65 million years. One of the oldest known primate-like mammal species, the *Plesiadapis*, came from North America; another, *Archicebus*, came from China. Other similar basal primates were widespread in Eurasia and Africa during the tropical conditions of the Paleocene and Eocene.

David R. Begun concluded that early primates flourished in Eurasia and that a lineage leading to the African apes and humans, including to *Dryopithecus*, migrated south from Europe or Western Asia into Africa. The surviving tropical population of primates—which is seen most completely in the Upper Eocene and lowermost Oligocene fossil beds of the Faiyum depression southwest of Cairo—gave rise to all extant primate species, including the lemurs of Madagascar, lorises of Southeast Asia, galagos or “bush babies” of Africa, and to the anthropoids, which are the Platyrrhines or New World monkeys, the Catarrhines or Old World monkeys, and the great apes, including humans and other hominids.

The earliest known catarrhine is *Kamoyapithecus* from uppermost Oligocene at Eragaleit in the northern Great Rift Valley in Kenya, dated to 24 million years ago. Its ancestry is thought to be species related to *Aegyptopithecus*, *Propliopithecus*, and *Parapithecus* from the Faiyum, at around 35 million years ago. In 2010, *Saadanius* was described as a close relative of the last common ancestor of the crown catarrhines, and tentatively dated to 29–28 million years ago, helping to fill an 11-million-year gap in the fossil record.

In the Early Miocene, about 22 million years ago, the many kinds of arboreally adapted primitive catarrhines from East Africa suggest a long history of prior diversification. Fossils at 20 million years ago include fragments attributed to *Victoriapithecus*, the earliest Old World monkey. Among the genera thought to be in the ape lineage leading up to 13 million years ago are *Proconsul*, *Rangwapithecus*, *Dendropithecus*, *Limnopithecus*, *Nacholapithecus*, *Equatorius*, *Nyanzapithecus*, *Afropithecus*, *Heliopithecus*, and *Kenyapithecus*, all from East Africa.

The presence of other generalized non-cercopithecids of Middle Miocene from sites far distant—*Otavipithecus* from cave deposits in Namibia, and *Pierolapithecus* and *Dryopithecus* from France, Spain and Austria—is evidence of a wide diversity of forms across Africa and the Mediterranean basin during the relatively warm and equable climatic regimes of the Early and Middle Miocene. The youngest of the Miocene hominoids, *Oreopithecus*, is from coal beds in Italy that have been dated to 9 million

years ago.

Molecular evidence indicates that the lineage of gibbons (family Hylobatidae) diverged from the line of great apes some 18–12 million years ago, and that of orangutans (subfamily Ponginae) diverged from the other great apes at about 12 million years; there are no fossils that clearly document the ancestry of gibbons, which may have originated in a so-far-unknown Southeast Asian hominoid population, but fossil proto-orangutans may be represented by *Sivapithecus* from India and *Grighopithecus* from Turkey, dated to around 10 million years ago.

19.5.4 Divergence of The Human Clade From Other Great Apes

Species close to the last common ancestor of gorillas, chimpanzees and humans may be represented by *Nakalipithecus* fossils found in Kenya and *Ouranopithecus* found in Greece. Molecular evidence suggests that between 8 and 4 million years ago, first the gorillas, and then the chimpanzees (genus *Pan*) split off from the line leading to the humans. Human DNA is approximately 98.4% identical to that of chimpanzees when comparing single nucleotide polymorphisms (see human evolutionary genetics). The fossil record, however, of gorillas and chimpanzees is limited; both poor preservation — rain forest soils tend to be acidic and dissolve bone — and sampling bias probably contribute to this problem.

Other hominins probably adapted to the drier environments outside the equatorial belt; and there they encountered antelope, hyenas, dogs, pigs, elephants, horses, and others. The equatorial belt contracted after about 8 million years ago, and there is very little fossil evidence for the split—thought to have occurred around that time—of the hominin lineage from the lineages of gorillas and chimpanzees. The earliest fossils argued by some to belong to the human lineage are *Sahelanthropus tchadensis* (7 Ma) and *Orrorin tugenensis* (6 Ma), followed by *Ardipithecus* (5.5–4.4 Ma), with species *Ar. kadabba* and *Ar. ramidus*.

It has been argued in a study of the life history of *Ar. ramidus* that the species provides evidence for a suite of anatomical and behavioral adaptations in very early hominins unlike any species of extant great ape. This study demonstrated affinities between the skull morphology of *Ar. ramidus* and that of infant and juvenile chimpanzees, suggesting the species evolved a juvenilised or paedomorphic craniofacial morphology via heterochronic dissociation of growth trajectories. It was also argued that the species provides support for the notion that very early hominins, akin to bonobos (*Pan paniscus*) the less aggressive species of the genus *Pan*, may have evolved via the process of self-domestication. Consequently, arguing against

the so-called “chimpanzee referential model” the authors suggest it is no longer tenable to use chimpanzee (*Pan troglodytes*) social and mating behaviors in models of early hominin social evolution. When commenting on the absence of aggressive canine morphology in *Ar. ramidus* and the implications this has for the evolution of hominin social psychology, they wrote:

Of course *Ar. ramidus* differs significantly from bonobos, bonobos having retained a functional canine honing complex. However, the fact that *Ar. ramidus* shares with bonobos reduced sexual dimorphism, and a more paedomorphic form relative to chimpanzees, suggests that the developmental and social adaptations evident in bonobos may be of assistance in future reconstructions of early hominin social and sexual psychology. In fact the trend towards increased maternal care, female mate selection and self-domestication may have been stronger and more refined in *Ar. ramidus* than what we see in bonobos.:128

The authors argue that many of the basic human adaptations evolved in the ancient forest and woodland ecosystems of late Miocene and early Pliocene Africa. Consequently, they argue that humans may not represent evolution from a chimpanzee-like ancestor as has traditionally been supposed. This suggests many modern human adaptations represent phylogenetically deep traits and that the behavior and morphology of chimpanzees may have evolved subsequent to the split with the common ancestor they share with humans.

19.5.5 Genus Australopithecus

The genus *Australopithecus* evolved in eastern Africa around 4 million years ago before spreading throughout the continent and eventually becoming extinct 2 million years ago. During this time period various forms of australopiths existed, including *Australopithecus anamensis*, *Au. afarensis*, *Au. sediba*, and *Au. africanus*. There is still some debate among academics whether certain African hominid species of this time, such as *Au. robustus* and *Au. boisei*, constitute members of the same genus; if so, they would be considered to be *Au. robustus* australopiths whilst the others would be considered *Au. gracilis* australopiths. However, if these species do indeed constitute their own genus, then they may be given their own name, *Paranthropus*.

- *Australopithecus* (4–1.8 Ma), with species *Au. anamensis*, *Au. afarensis*, *Au. africanus*, *Au. bahrelghazali*, *Au. garhi*, and *Au. sediba*;
- *Kenyanthropus* (3–2.7 Ma), with species *K. platyops*;
- *Paranthropus* (3–1.2 Ma), with species *P. aethiopicus*, *P. boisei*, and *P. robustus*

A new proposed species *Australopithecus deyiremeda*

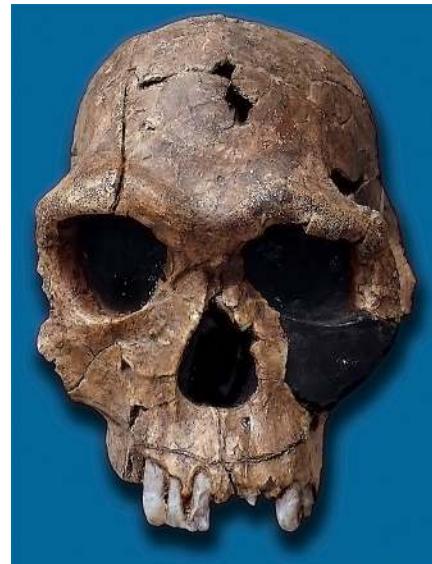


Figure 19.54: Replica of fossil skull of *Homo habilis*. Fossil number KNM ER 1813, found at Koobi Fora, Kenya.⁹⁵

is claimed to have been discovered living at the same time period of *Au. afarensis*. There is debate if *Au. deyiremeda* is a new species or is *Au. afarensis*. *Australopithecus prometheus*, otherwise known as Little Foot has recently been dated at 3.67 million years old through a new dating technique, making the genus *Australopithecus* as old as *afarensis*. Given the opposable big toe found on Little Foot, it seems that he was a good climber, and it is thought given the night predators of the region, he probably, like gorillas and chimpanzees, built a nesting platform at night, in the trees.

19.5.6 Evolution of Genus *Homo*

The earliest documented representative of the genus *Homo* is *Homo habilis*, which evolved around 2.8 million years ago, and is arguably the earliest species for which there is positive evidence of the use of stone tools. The brains of these early hominins were about the same size as that of a chimpanzee, although it has been suggested that this was the time in which the human SRGAP2 gene doubled, producing a more rapid wiring of the frontal cortex. During the next million years a process of rapid encephalization occurred, and with the arrival of *Homo erectus* and *Homo ergaster* in the fossil record, cranial capacity had doubled to 850 cm³. (Such an increase in human brain size is equivalent to each generation having 125,000 more neurons than their parents.) It is believed that *Homo erectus* and *Homo ergaster* were the first to use fire and complex tools, and were the first of the hominin line to leave Africa, spreading throughout Africa, Asia, and Europe between 1.3 to 1.8 million years ago.

According to the recent African origin of modern humans theory, modern humans evolved in Africa possibly from *Homo heidelbergensis*, *Homo rhodesiensis* or *Homo antecessor* and migrated out of the continent some 50,000 to 100,000 years ago, gradually replacing local populations of *Homo erectus*, Denisova hominins, *Homo floresiensis*, *Homo luzonensis* and *Homo neanderthalensis*. Archaic *Homo sapiens*, the forerunner of anatomically modern humans, evolved in the Middle Paleolithic between 400,000 and 250,000 years ago. Recent DNA evidence suggests that several haplotypes of Neanderthal origin are present among all non-African populations, and Neanderthals and other hominins, such as Denisovans, may have contributed up to 6% of their genome to present-day humans, suggestive of a limited interbreeding between these species. The transition to behavioral modernity with the development of symbolic culture, language, and specialized lithic technology happened around 50,000 years ago, according to some anthropologists, although others point to evidence that suggests that a gradual change in behavior took place over a longer time span.

Homo sapiens is the only extant species of its genus, *Homo*. While some (extinct) *Homo* species might have been ancestors of *Homo sapiens*, many, perhaps most, were likely "cousins", having speciated away from the ancestral hominin line. There is yet no consensus as to which of these groups should be considered a separate species and which should be a subspecies; this may be due to the dearth of fossils or to the slight differences used to classify species in the genus *Homo*. The Sahara pump theory (describing an occasionally passable "wet" Sahara desert) provides one possible explanation of the early variation in the genus *Homo*.

Based on archaeological and paleontological evidence, it has been possible to infer, to some extent, the ancient dietary practices of various *Homo* species and to study the role of diet in physical and behavioral evolution within *Homo*.

Some anthropologists and archaeologists subscribe to the Toba catastrophe theory, which posits that the supereruption of Lake Toba on Sumatran island in Indonesia some 70,000 years ago caused global consequences, killing the majority of humans and creating a population bottleneck that affected the genetic inheritance of all humans today. The genetic and archaeological evidence for this remains in question however.

19.5.7 *H. habilis* and *H. gautengensis*

Homo habilis lived from about 2.8 to 1.4 Ma. The species evolved in South and East Africa in the Late Pliocene or Early Pleistocene, 2.5–2 Ma, when it diverged from the australopithecines. *Homo habilis* had smaller molars and larger

brains than the australopithecines, and made tools from stone and perhaps animal bones. One of the first known hominins was nicknamed 'handy man' by discoverer Louis Leakey due to its association with stone tools. Some scientists have proposed moving this species out of *Homo* and into *Australopithecus* due to the morphology of its skeleton being more adapted to living on trees rather than to moving on two legs like *Homo sapiens*.

In May 2010, a new species, *Homo gautengensis*, was discovered in South Africa.

19.5.8 *H. rudolfensis* and *H. georgicus*

These are proposed species names for fossils from about 1.9–1.6 Ma, whose relation to *Homo habilis* is not yet clear.

- *Homo rudolfensis* refers to a single, incomplete skull from Kenya. Scientists have suggested that this was another *Homo habilis*, but this has not been confirmed.
- *Homo georgicus*, from Georgia, may be an intermediate form between *Homo habilis* and *Homo erectus*, or a subspecies of *Homo erectus*.

19.5.9 *H. ergaster* and *H. erectus*

The first fossils of *Homo erectus* were discovered by Dutch physician Eugene Dubois in 1891 on the Indonesian island of Java. He originally named the material *Anthropopithecus erectus* (1892–1893, considered at this point as a chimpanzee-like fossil primate) and *Pithecanthropus erectus* (1893–1894, changing his mind as of based on its morphology, which he considered to be intermediate between that of humans and apes). Years later, in the 20th century, the German physician and paleoanthropologist Franz Weidenreich (1873–1948) compared in detail the characters of Dubois' Java Man, then named *Pithecanthropus erectus*, with the characters of the Peking Man, then named *Sinanthropus pekinensis*. Weidenreich concluded in 1940 that because of their anatomical similarity with modern humans it was necessary to gather all these specimens of Java and China in a single species of the genus *Homo*, the species *Homo erectus*. *Homo erectus* lived from about 1.8 Ma to about 70,000 years ago — which would indicate that they were probably wiped out by the Toba catastrophe; however, nearby *Homo floresiensis* survived it. The early phase of *Homo erectus*, from 1.8 to 1.25 Ma, is considered by some to be a separate species, *Homo ergaster*, or as *Homo erectus ergaster*, a subspecies of *Homo erectus*.

In Africa in the Early Pleistocene, 1.5–1 Ma, some populations of *Homo habilis* are thought to have evolved larger brains and to have made more elaborate stone tools; these differences and others are sufficient for anthropologists to classify them as a new species, *Homo*

erectus—in Africa. The evolution of locking knees and the movement of the foramen magnum are thought to be likely drivers of the larger population changes. This species also may have used fire to cook meat. Richard Wrangham suggests that the fact that *Homo* seems to have been ground dwelling, with reduced intestinal length, smaller dentition, “and swelled our brains to their current, horrendously fuel-inefficient size”, suggest that control of fire and releasing increased nutritional value through cooking was the key adaptation that separated *Homo* from tree-sleeping Australopithecines.

A famous example of *Homo erectus* is Peking Man; others were found in Asia (notably in Indonesia), Africa, and Europe. Many paleoanthropologists now use the term *Homo ergaster* for the non-Asian forms of this group, and reserve *Homo erectus* only for those fossils that are found in Asia and meet certain skeletal and dental requirements which differ slightly from *H. ergaster*.

19.5.10 *H. cepranensis* and *H. antecessor*

These are proposed as species that may be intermediate between *H. erectus* and *H. heidelbergensis*.

H. antecessor is known from fossils from Spain and England that are dated 1.2 Ma–500 ka. *H. cepranensis* refers to a single skull cap from Italy, estimated to be about 800,000 years old.

19.5.11 *H. heidelbergensis*

H. heidelbergensis (“Heidelberg Man”) lived from about 800,000 to about 300,000 years ago. Also proposed as *Homo sapiens heidelbergensis* or *Homo sapiens paleohungaricus*.

19.5.12 *H. rhodesiensis*, and the Gawis cranium

H. rhodesiensis, estimated to be 300,000–125,000 years old. Most current researchers place Rhodesian Man within the group of *Homo heidelbergensis*, though other designations such as archaic *Homo sapiens* and *Homo sapiens rhodesiensis* have been proposed. In February 2006 a fossil, the Gawis cranium, was found which might possibly be a species intermediate between *H. erectus* and *H. sapiens* or one of many evolutionary dead ends. The skull from Gawis, Ethiopia, is believed to be 500,000–250,000 years old. Only summary details are known, and the finders have not yet released a peer-reviewed study. Gawis man’s facial features suggest its being either an intermediate species or an example of a “Bodo man” female.

19.5.13 Neanderthal and Denisovan

Homo neanderthalensis, alternatively designated as *Homo sapiens neanderthalensis*, lived in Europe and Asia from 400,000 to about 28,000 years ago. There are a number of clear anatomical differences between anatomically modern humans (AMH) and Neanderthal populations. Many of these relate to the superior adaptation to cold environments possessed by the Neanderthal populations. Their surface to volume ratio is an extreme version of that found amongst Inuit populations, indicating that they were less inclined to lose body heat than were AMH. From brain Endocasts, Neanderthals also had significantly larger brains. This would seem to indicate that the intellectual superiority of AMH populations may be questionable. More recent research by Eiluned Pearce, Chris Stringer, R.I.M. Dunbar, however, have shown important differences in brain architecture. For example, in both the orbital chamber size and in the size of the occipital lobe, the larger size suggests that the Neanderthal had a better visual acuity than modern humans. This would give a superior vision in the inferior light conditions found in Glacial Europe. It also seems that the higher body mass of Neanderthals had a correspondingly larger brain mass required for body care and control.

The Neanderthal populations seem to have been physically superior to AMH populations. These differences may have been sufficient to give Neanderthal populations an environmental superiority to AMH populations from 75,000 to 45,000 years BP. With these differences, Neanderthal brains show a smaller area was available for social functioning. Plotting group size possible from endocranial volume, suggests that AMH populations (minus occipital lobe size), had a Dunbars number of 144 possible relationships. Neanderthal populations seem to have been limited to about 120 individuals. This would show up in a larger number of possible mates for AMH humans, with increased risks of inbreeding amongst Neanderthal populations. It also suggests that humans had larger trade catchment areas than Neanderthals (confirmed in the distribution of stone tools). With larger populations, social and technological innovations were easier to fix in human populations, which may have all contributed to the fact that modern *Homo sapiens* replaced the Neanderthal populations by 28,000 BP.

Earlier evidence from sequencing mitochondrial DNA suggested that no significant gene flow occurred between *H. neanderthalensis* and *H. sapiens*, and that the two were separate species that shared a common ancestor about 660,000 years ago. However, a sequencing of the Neanderthal genome in 2010 indicated that Neanderthals did indeed interbreed with anatomically modern humans circa 45,000 to 80,000 years ago (at the approximate time that modern humans migrated out from Africa, but

before they dispersed into Europe, Asia and elsewhere). The genetic sequencing of a 40,000 year old human skeleton from Romania showed that 11% of its genome was Neanderthal, and it was estimated that the individual had a Neanderthal ancestor 4–6 generations previously, in addition to a contribution from earlier interbreeding in the Middle East. Though this interbred Romanian population seems not to have been ancestral to modern humans, the finding indicates that interbreeding happened repeatedly.

All modern non-African humans have about 1% to 4% or, according to more recent data, about 1.5% to 2.6% of their DNA derived from Neanderthal DNA, and this finding is consistent with recent studies indicating that the divergence of some human alleles dates to one Ma, although the interpretation of these studies has been questioned. Neanderthals and *Homo sapiens* could have co-existed in Europe for as long as 10,000 years, during which human populations exploded vastly outnumbering Neanderthals, possibly outcompeting them by sheer numerical strength.

In 2008, archaeologists working at the site of Denisova Cave in the Altai Mountains of Siberia uncovered a small bone fragment from the fifth finger of a juvenile member of Denisovans. Artifacts, including a bracelet, excavated in the cave at the same level were carbon dated to around 40,000 BP. As DNA had survived in the fossil fragment due to the cool climate of the Denisova Cave, both mtDNA and nuclear DNA were sequenced.

Alleles thought to have originated in Neanderthals and Denisovans have been identified at several genetic loci in the genomes of modern humans outside of Africa. HLA haplotypes from Denisovans and Neanderthal represent more than half the HLA alleles of modern Eurasians, indicating strong positive selection for these introgressed alleles. Corinne Simonet at Vanderbilt University, in Nashville and her team have found from medical records of 28,000 people of European descent that the presence of Neanderthal DNA segments may be associated with a likelihood to suffer depression more frequently.

The flow of genes from Neanderthal populations to modern humans was not all one way. Sergi Castellano of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, has in 2016 reported that while Denisovan and Neanderthal genomes are more related to each other than they are to us, Siberian Neanderthal genomes show similarity to the modern human gene pool, more so than to European Neanderthal populations. The evidence suggests that the Neanderthal populations interbred with modern humans possibly 100,000 years ago, probably somewhere in the Near East.

Studies of a Neanderthal child at Gibraltar show from brain development and teeth eruption that Neanderthal

children may have matured more rapidly than is the case for *Homo sapiens*.

19.5.14 *H. floresiensis*

H. floresiensis, which lived from approximately 190,000 to 50,000 years before present (BP), has been nicknamed the hobbit for its small size, possibly a result of insular dwarfism. *H. floresiensis* is intriguing both for its size and its age, being an example of a recent species of the genus *Homo* that exhibits derived traits not shared with modern humans. In other words, *H. floresiensis* shares a common ancestor with modern humans, but split from the modern human lineage and followed a distinct evolutionary path. The main find was a skeleton believed to be a woman of about 30 years of age. Found in 2003, it has been dated to approximately 18,000 years old. The living woman was estimated to be one meter in height, with a brain volume of just 380 cm³ (considered small for a chimpanzee and less than a third of the *H. sapiens* average of 1400 cm³).

However, there is an ongoing debate over whether *H. floresiensis* is indeed a separate species. Some scientists hold that *H. floresiensis* was a modern *H. sapiens* with pathological dwarfism. This hypothesis is supported in part, because some modern humans who live on Flores, the Indonesian island where the skeleton was found, are pygmies. This, coupled with pathological dwarfism, could have resulted in a significantly diminutive human. The other major attack on *H. floresiensis* as a separate species is that it was found with tools only associated with *H. sapiens*.

The hypothesis of pathological dwarfism, however, fails to explain additional anatomical features that are unlike those of modern humans (diseased or not) but much like those of ancient members of our genus. Aside from cranial features, these features include the form of bones in the wrist, forearm, shoulder, knees, and feet. Additionally, this hypothesis fails to explain the find of multiple examples of individuals with these same characteristics, indicating they were common to a large population, and not limited to one individual.

19.5.15 *H. luzonensis*

A small number of specimens from the island of Luzon, dated 50,000 to 67,000 years ago, have recently been assigned by their discoverers, based on dental characteristics, to a novel human species, *H. luzonensis*.

19.5.16 *H. sapiens*

H. sapiens (the adjective *sapiens* is Latin for “wise” or “intelligent”) emerged in Africa around 300,000 years ago, likely

derived from *Homo heidelbergensis* or a related lineage. In September 2019, scientists reported the computerized determination, based on 260 CT scans, of a virtual skull shape of the last common human ancestor to modern humans/*H. sapiens*, representative of the earliest modern humans, and suggested that modern humans arose between 260,000 and 350,000 years ago through a merging of populations in East and South Africa.

Between 400,000 years ago and the second interglacial period in the Middle Pleistocene, around 250,000 years ago, the trend in intra-cranial volume expansion and the elaboration of stone tool technologies developed, providing evidence for a transition from *H. erectus* to *H. sapiens*. The direct evidence suggests there was a migration of *H. erectus* out of Africa, then a further speciation of *H. sapiens* from *H. erectus* in Africa. A subsequent migration (both within and out of Africa) eventually replaced the earlier dispersed *H. erectus*. This migration and origin theory is usually referred to as the “recent single-origin hypothesis” or “out of Africa” theory. *H. sapiens* interbred with archaic humans both in Africa and in Eurasia, in Eurasia notably with Neanderthals and Denisovans.

The Toba catastrophe theory, which postulates a population bottleneck for *H. sapiens* about 70,000 years ago, was controversial from its first proposal in the 1990s and by the 2010s had very little support. Distinctive human genetic variability has arisen as the result of the founder effect, by archaic admixture and by recent evolutionary pressures.

The use of tools has been interpreted as a sign of intelligence, and it has been theorized that tool use may have stimulated certain aspects of human evolution, especially the continued expansion of the human brain. Paleontology has yet to explain the expansion of this organ over millions of years despite being extremely demanding in terms of energy consumption. The brain of a modern human consumes about 13 watts (260 kilocalories per day), a fifth of the body's resting power consumption. Increased tool use would allow hunting for energy-rich meat products, and would enable processing more energy-rich plant products. Researchers have suggested that early hominins were thus under evolutionary pressure to increase their capacity to create and use tools.

Precisely when early humans started to use tools is difficult to determine, because the more primitive these tools are (for example, sharp-edged stones) the more difficult it is to decide whether they are natural objects or human artifacts. There is some evidence that the australopithecines (4 Ma) may have used broken bones as tools, but this is debated.

Many species make and use tools, but it is the human



Figure 19.55: Oldowan-tradition stone chopper.⁹⁶

genus that dominates the areas of making and using more complex tools. The oldest known tools are flakes from West Turkana, Kenya, which date to 3.3 million years ago. The next oldest stone tools are from Gona, Ethiopia, and are considered the beginning of the Oldowan technology. These tools date to about 2.6 million years ago. A *Homo* fossil was found near some Oldowan tools, and its age was noted at 2.3 million years old, suggesting that maybe the *Homo* species did indeed create and use these tools. It is a possibility but does not yet represent solid evidence. The third metacarpal styloid process enables the hand bone to lock into the wrist bones, allowing for greater amounts of pressure to be applied to the wrist and hand from a grasping thumb and fingers. It allows humans the dexterity and strength to make and use complex tools. This unique anatomical feature separates humans from apes and other nonhuman primates, and is not seen in human fossils older than 1.8 million years.

Bernard Wood noted that *Paranthropus* co-existed with the early *Homo* species in the area of the “Oldowan Industrial Complex” over roughly the same span of time. Although there is no direct evidence which identifies *Paranthropus* as the tool makers, their anatomy lends to indirect evidence of their capabilities in this area. Most paleoanthropologists agree that the early *Homo* species were indeed responsible for most of the Oldowan tools found. They argue that when most of the Oldowan tools were found in association with human fossils, *Homo* was always present, but *Paranthropus* was not.

In 1994, Randall Susman used the anatomy of opposable thumbs as the basis for his argument that both the *Homo* and *Paranthropus* species were toolmakers. He compared bones and muscles of human and chimpanzee thumbs, finding that humans have 3 muscles which are lacking in chimpanzees. Humans also have thicker metacarpals with broader heads, allowing more precise

grasping than the chimpanzee hand can perform. Susman posited that modern anatomy of the human opposable thumb is an evolutionary response to the requirements associated with making and handling tools and that both species were indeed toolmakers.

Stone tools are first attested around 2.6 million years ago, when hominins in Eastern Africa used so-called core tools, choppers made out of round cores that had been split by simple strikes. This marks the beginning of the Paleolithic, or Old Stone Age; its end is taken to be the end of the last Ice Age, around 10,000 years ago. The Paleolithic is subdivided into the Lower Paleolithic (Early Stone Age), ending around 350,000–300,000 years ago, the Middle Paleolithic (Middle Stone Age), until 50,000–30,000 years ago, and the Upper Paleolithic, (Late Stone Age), 50,000–10,000 years ago.

Archaeologists working in the Great Rift Valley in Kenya have discovered the oldest known stone tools in the world. Dated to around 3.3 million years ago, the implements are some 700,000 years older than stone tools from Ethiopia that previously held this distinction.

The period from 700,000–300,000 years ago is also known as the Acheulean, when *H. ergaster* (or *erectus*) made large stone hand axes out of flint and quartzite, at first quite rough (Early Acheulian), later “retouched” by additional, more-subtle strikes at the sides of the flakes. After 350,000 BP the more refined so-called Levallois technique was developed, a series of consecutive strikes, by which scrapers, slicers (“racloirs”), needles, and flattened needles were made. Finally, after about 50,000 BP, ever more refined and specialized flint tools were made by the Neanderthals and the immigrant Cro-Magnons (knives, blades, skimmers). Bone tools were also made by *H. sapiens* in Africa by 90–70,000 years ago and are also known from early *H. sapiens* sites in Eurasia by about 50,000 years ago.

Until about 50,000–40,000 years ago, the use of stone tools seems to have progressed stepwise. Each phase (*H. habilis*, *H. ergaster*, *H. neanderthalensis*) started at a higher level than the previous one, but after each phase started, further development was slow. Currently paleoanthropologists are debating whether these *Homo* species possessed some or many of the cultural and behavioral traits associated with modern humans such as language, complex symbolic thinking, technological creativity etc. It seems that they were culturally conservative maintaining simple technologies and foraging patterns over very long periods.

Around 50,000 BP, modern human culture started to evolve more rapidly. The transition to behavioral modernity has been characterized by some as a “Great Leap For-



Figure 19.56: Aurignacian Culture bone tools (needle, points and tools for punching holes), Hayonim Cave, 30000 BP (Before Present).⁹⁷ HaYonim Cave (Hebrew: מערת הינום, Me'arat HaYonim, lit. Cave of the Pigeons) is a cave located in a limestone bluff about 250 meters above modern sea level, in the Upper Galilee, Israel.

ward”, or as the “Upper Palaeolithic Revolution”, due to the sudden appearance of distinctive signs of modern behavior and big game hunting in the archaeological record. Evidence of behavioral modernity significantly earlier also exists from Africa, with older evidence of abstract imagery, widened subsistence strategies, more sophisticated tools and weapons, and other “modern” behaviors, and many scholars have recently argued that the transition to modernity occurred sooner than previously believed. Some other scholars consider the transition to have been more gradual, noting that some features had already appeared among archaic African *Homo sapiens* since 300–200,000 years ago. Recent evidence suggests that the Australian Aboriginal population separated from the African population 75,000 years ago, and that they made a sea journey of up to 160 km 60,000 years ago, which may diminish the evidence of the Upper Paleolithic Revolution.

Modern humans started burying their dead, using animal hides to make clothing, hunting with more sophisticated techniques (such as using trapping pits or driving animals off cliffs), and engaging in cave painting. As human culture advanced, different populations of humans introduced novelty to existing technologies: artifacts such as fish hooks, buttons, and bone needles show signs of variation among different populations of humans, something that had not been seen in human cultures prior to 50,000 BP. Typically, *H. neanderthalensis* populations do not vary in their technologies, although the Chatelperronian assemblages have been found to be Neanderthal innovations produced as a result of exposure to the *Homo sapiens* Aurignacian technologies.

Among concrete examples of modern human behavior, anthropologists include specialization of tools, use of jewellery and images (such as cave drawings), organization of living space, rituals (for example, burials with grave gifts), specialized hunting techniques, exploration of less hospitable geographical areas, and barter trade networks. Debate continues as to whether a "revolution" led to modern humans ("the big bang of human consciousness"), or whether the evolution was more "gradual".

Chapter 20

Circulatory Systems

The circulatory system, also called the cardiovascular system or the vascular system, is an organ system that permits blood to circulate and transport nutrients (such as amino acids and electrolytes), oxygen, carbon dioxide, hormones, and blood cells to and from the cells in the body to provide nourishment and help in fighting diseases, stabilize temperature and pH, and maintain homeostasis. The cardiovascular (from Latin words meaning "heart" and "vessel") system comprises the blood, heart, and blood vessels.

While humans, as well as other vertebrates, have a closed cardiovascular system (meaning that the blood never leaves the network of arteries, veins and capillaries), some invertebrate groups have an open cardiovascular system. The more primitive, diploblastic animal phyla lack circulatory systems.

20.1 Closed Circulatory System

The circulatory systems of all vertebrates, as well as of annelids (for example, earthworms) and cephalopods (squids, octopuses and relatives) are closed, just as in humans. Still, the systems of fish, amphibians, reptiles, and birds show various stages of the evolution of the circulatory system.

Fish have the simplest circulatory system. It has only one circuit, with the blood being pumped through the capillaries of the gills and on to the capillaries of the body tissues. This is known as single cycle circulation. The heart of fish is, therefore, only a single pump consisting of one atrium to receive blood and one ventricle to pump it, in contrast to three chambers (two atria, one ventricle) of amphibian and most reptile hearts and four chambers (two atria, two ventricles) of mammal and bird hearts. However, the fish heart has entry and exit compartments that may be called chambers, so it is also sometimes described as three-chambered or four-chambered, depending on what is counted as a chamber. The atrium and ventricle are sometimes considered "true chambers", while the others are considered "accessory chambers".

The four compartments are arranged sequentially:

- Sinus venosus, a thin-walled sac or reservoir with some cardiac muscle that collects deoxygenated blood through the incoming hepatic and cardinal veins.
- Atrium, a thicker-walled, muscular chamber that sends blood to the ventricle.
- Ventricle, a thick-walled, muscular chamber that pumps the blood to the fourth part, the outflow tract. The shape of the ventricle varies considerably, usually tubular in fish with elongated bodies, pyramidal with a triangular base in others, or sometimes sac-like in some marine fish.
- The outflow tract (OFT) to the ventral aorta, consisting of the tubular conus arteriosus, bulbus arteriosus, or both. The conus arteriosus, typically found in more primitive species of fish, contracts to assist blood flow to the aorta, while the bulbus arteriosus does not.

Ostial valves, consisting of flap-like connective tissues, prevent blood from flowing backward through the compartments. The ostial valve between the sinus venosus and atrium is called the sino-atrial valve, which closes during ventricular contraction. Between the atrium and ventricle is an ostial valve called the atrio-ventricular valve, and between the bulbus arteriosus and ventricle is an ostial valve called the bulbo-ventricular valve. The conus arteriosus has a variable number of semilunar valves.

The ventral aorta delivers blood to the gills where it is oxygenated and flows, through the dorsal aorta, into the rest of the body. (In tetrapods, the ventral aorta has divided in two; one half forms the ascending aorta, while the other forms the pulmonary artery).

In amphibians and most reptiles, a double circulatory system is used, but the heart is not always completely separated into two pumps. Amphibians have a three-chambered heart.

In reptiles, the ventricular septum of the heart is incomplete and the pulmonary artery is equipped with a sphincter muscle. This allows a second possible route of blood flow.

Instead of blood flowing through the pulmonary artery to the lungs, the sphincter may be contracted to divert this blood flow through the incomplete ventricular septum into the left ventricle and out through the aorta. This means the blood flows from the capillaries to the heart and back to the capillaries instead of to the lungs. This process is useful to ectothermic (cold-blooded) animals in the regulation of their body temperature.

Birds, mammals, and crocodilians show complete separation of the heart into two pumps, for a total of four heart chambers; it is thought that the four-chambered heart of birds and crocodilians evolved independently from that of mammals.

20.2 Open Circulatory Systems

In arthropods, the open circulatory system is a system in which a fluid in a cavity called the hemocoel bathes the organs directly with oxygen and nutrients and there is no distinction between blood and interstitial fluid; this combined fluid is called hemolymph or haemolymph. Muscular movements by the animal during locomotion can facilitate hemolymph movement, but diverting flow from one area to another is limited. When the heart relaxes, blood is drawn back toward the heart through open-ended pores (ostia).

Hemolymph fills all of the interior hemocoel of the body and surrounds all cells. Hemolymph is composed of water, inorganic salts (mostly sodium, chloride, potassium, magnesium, and calcium), and organic compounds (mostly carbohydrates, proteins, and lipids). The primary oxygen transporter molecule is hemocyanin.

There are free-floating cells, the hemocytes, within the hemolymph. They play a role in the arthropod immune system.

20.3 No Circulatory System

Circulatory systems are absent in some animals, including flatworms. Their body cavity has no lining or enclosed fluid. Instead a muscular pharynx leads to an extensively branched digestive system that facilitates direct diffusion of nutrients to all cells. The flatworm's dorso-ventrally flattened body shape also restricts the distance of any cell from the digestive system or the exterior of the organism. Oxygen can diffuse from the surrounding water into the cells, and carbon dioxide can diffuse out. Consequently, every cell is able to obtain nutrients, water and oxygen without the need of a transport system.

Some animals, such as jellyfish, have more extensive branching from their gastrovascular cavity (which functions as both a place of digestion and a form of

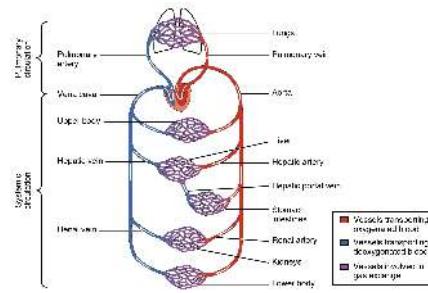


Figure 20.1: The systemic circulation and capillary networks shown and also as separate from the pulmonary circulation.¹

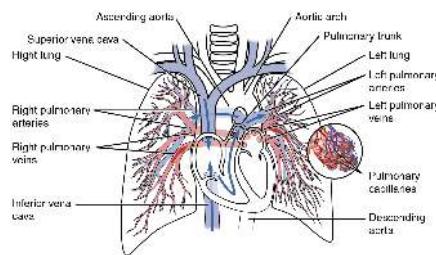


Figure 20.2: The pulmonary circulation as it passes from the heart. Showing both the pulmonary and bronchial arteries.²

circulation), this branching allows for bodily fluids to reach the outer layers, since the digestion begins in the inner layers.

20.4 The Human Cardiovascular System

The essential components of the human cardiovascular system are the heart, blood and blood vessels. Blood is a fluid consisting of plasma, red blood cells, white blood cells, and platelets that is circulated by the heart through the vertebrate vascular system, carrying oxygen and nutrients to and waste materials away from all body tissues. The circulatory system of the blood is seen as having two components, a systemic circulation and a pulmonary circulation.

The pulmonary circulation forms a "loop" through the lungs where blood is oxygenated; and the systemic circulation forms a "loop" through the rest of the body to provide oxygenated blood. The systemic circulation can also be seen to function in two parts – a macrocirculation and a microcirculation. An average adult contains five to six quarts (roughly 4.7 to 5.7 liters) of blood, accounting for approximately 7% of their total body weight. Blood consists of plasma, red blood cells, white blood cells, and platelets.

Also, the digestive system works with the circulatory system to provide the nutrients the system needs to keep the heart pumping.

The cardiovascular system of humans is closed, meaning that the blood never leaves the network of blood vessels. In contrast, oxygen and nutrients diffuse across the blood vessel layers and enter interstitial fluid, which carries oxygen and nutrients to the target cells, and carbon dioxide and wastes in the opposite direction.

20.5 The Heart

In humans, the heart is approximately the size of a closed fist and is located between the lungs, in the middle compartment of the chest.

In humans, other mammals, and birds, the heart is divided into four chambers: upper left and right atria and lower left and right ventricles. Commonly the right atrium and ventricle are referred together as the right heart and their left counterparts as the left heart. Fish, in contrast, have two chambers, an atrium and a ventricle, while reptiles have three chambers. In a healthy heart blood flows one way through the heart due to heart valves, which prevent backflow. The heart is enclosed in a protective sac, the pericardium, which also contains a small amount of fluid. The wall of the heart is made up of three layers: epicardium, myocardium, and endocardium.

The heart pumps blood with a rhythm determined by a group of pacemaking cells in the sinoatrial node. These generate a current that causes contraction of the heart, traveling through the atrioventricular node and along the conduction system of the heart. The heart receives blood low in oxygen from the systemic circulation, which enters the right atrium from the superior and inferior venae cavae and passes to the right ventricle. From here it is pumped into the pulmonary circulation, through the lungs where it receives oxygen and gives off carbon dioxide. Oxygenated blood then returns to the left atrium, passes through the left ventricle and is pumped out through the aorta to the systemic circulation—where the oxygen is used and metabolized to carbon dioxide. The heart beats at a resting rate close to 72 beats per minute. Exercise temporarily increases the rate, but lowers resting heart rate in the long term, and is good for heart health.

Cardiovascular diseases (CVD) are the most common cause of death globally as of 2008, accounting for 30% of deaths. Of these more than three quarters are a result of coronary artery disease and stroke. Risk factors include: smoking, being overweight, little exercise, high cholesterol, high blood pressure, and poorly controlled diabetes, among others. Cardiovascular diseases frequently

do not have symptoms or may cause chest pain or shortness of breath. Diagnosis of heart disease is often done by the taking of a medical history, listening to the heart-sounds with a stethoscope, ECG, and ultrasound. Specialists who focus on diseases of the heart are called cardiologists, although many specialties of medicine may be involved in treatment.

The human heart is situated in the middle mediastinum, at the level of thoracic vertebrae T5–T8. A double-membraned sac called the pericardium surrounds the heart and attaches to the mediastinum. The back surface of the heart lies near the vertebral column, and the front surface sits behind the sternum and rib cartilages. The upper part of the heart is the attachment point for several large blood vessels—the venae cavae, aorta and pulmonary trunk. The upper part of the heart is located at the level of the third costal cartilage. The lower tip of the heart, the apex, lies to the left of the sternum (8 to 9 cm from the midsternal line) between the junction of the fourth and fifth ribs near their articulation with the costal cartilages.

The largest part of the heart is usually slightly offset to the left side of the chest (though occasionally it may be offset to the right) and is felt to be on the left because the left heart is stronger and larger, since it pumps to all body parts. Because the heart is between the lungs, the left lung is smaller than the right lung and has a cardiac notch in its border to accommodate the heart. The heart is cone-shaped, with its base positioned upwards and tapering down to the apex. An adult heart has a mass of 250–350 grams (9–12 oz). The heart is often described as the size of a fist: 12 cm (5 in) in length, 8 cm (3.5 in) wide, and 6 cm (2.5 in) in thickness, although this description is disputed, as the heart is likely to be slightly larger. Well-trained athletes can have much larger hearts due to the effects of exercise on the heart muscle, similar to the response of skeletal muscle.

20.5.1 Structure of The Heart

The heart has four chambers, two upper atria, the receiving chambers, and two lower ventricles, the discharging chambers. The atria open into the ventricles via the atrioventricular valves, present in the atrioventricular septum. This distinction is visible also on the surface of the heart as the coronary sulcus. There is an ear-shaped structure in the upper right atrium called the right atrial appendage, or auricle, and another in the upper left atrium, the left atrial appendage. The right atrium and the right ventricle together are sometimes referred to as the right heart. Similarly, the left atrium and the left ventricle together are sometimes referred to as the left heart. The ventricles are separated from each other by the interventricular septum,

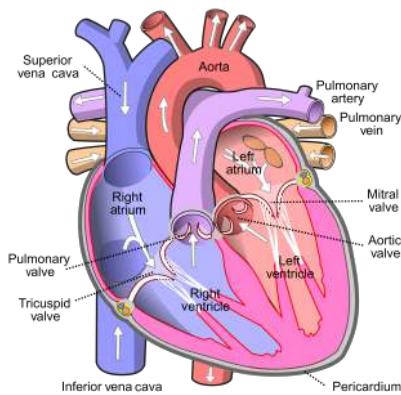


Figure 20.3: Diagram of the human heart³ 1. Superior vena cava 2. Right atrium 3. Right ventricle 4. Mitral valve 5. Aortic valve 6. Left ventricle 7. Left atrium 8. Pulmonary vein 9. Inferior vena cava 10. Aorta 11. Pulmonary artery 12. Tricuspid valve 13. Pericardium

visible on the surface of the heart as the anterior longitudinal sulcus and the posterior interventricular sulcus.

The cardiac skeleton is made of dense connective tissue and this gives structure to the heart. It forms the atrioventricular septum which separates the atria from the ventricles, and the fibrous rings which serve as bases for the four heart valves. The cardiac skeleton also provides an important boundary in the heart's electrical conduction system since collagen cannot conduct electricity. The interatrial septum separates the atria and the interventricular septum separates the ventricles. The interventricular septum is much thicker than the interatrial septum, since the ventricles need to generate greater pressure when they contract.

The heart has four valves, which separate its chambers. One valve lies between each atrium and ventricle, and one valve rests at the exit of each ventricle.

The valves between the atria and ventricles are called the atrioventricular valves. Between the right atrium and the right ventricle is the tricuspid valve. The tricuspid valve has three cusps, which connect to chordae tendinae and three papillary muscles named the anterior, posterior, and septal muscles, after their relative positions. The mitral valve lies between the left atrium and left ventricle. It is also known as the bicuspid valve due to its having two cusps, an anterior and a posterior cusp. These cusps are also attached via chordae tendinae to two papillary muscles projecting from the ventricular wall.

The papillary muscles extend from the walls of the heart to valves by cartilaginous connections called chordae tendinae. These muscles prevent the valves from falling

too far back when they close. During the relaxation phase of the cardiac cycle, the papillary muscles are also relaxed and the tension on the chordae tendinae is slight. As the heart chambers contract, so do the papillary muscles. This creates tension on the chordae tendinae, helping to hold the cusps of the atrioventricular valves in place and preventing them from being blown back into the atria. [g]

Two additional semilunar valves sit at the exit of each of the ventricles. The pulmonary valve is located at the base of the pulmonary artery. This has three cusps which are not attached to any papillary muscles. When the ventricle relaxes blood flows back into the ventricle from the artery and this flow of blood fills the pocket-like valve, pressing against the cusps which close to seal the valve. The semilunar aortic valve is at the base of the aorta and also is not attached to papillary muscles. This too has three cusps which close with the pressure of the blood flowing back from the aorta.

The right heart consists of two chambers, the right atrium and the right ventricle, separated by a valve, the tricuspid valve.

The right atrium receives blood almost continuously from the body's two major veins, the superior and inferior venae cavae. A small amount of blood from the coronary circulation also drains into the right atrium via the coronary sinus, which is immediately above and to the middle of the opening of the inferior vena cava. In the wall of the right atrium is an oval-shaped depression known as the fossa ovalis, which is a remnant of an opening in the fetal heart known as the foramen ovale. Most of the internal surface of the right atrium is smooth, the depression of the fossa ovalis is medial, and the anterior surface has prominent ridges of pectinate muscles, which are also present in the right atrial appendage.

The right atrium is connected to the right ventricle by the tricuspid valve. The walls of the right ventricle are lined with trabeculae carneae, ridges of cardiac muscle covered by endocardium. In addition to these muscular ridges, a band of cardiac muscle, also covered by endocardium, known as the moderator band reinforces the thin walls of the right ventricle and plays a crucial role in cardiac conduction. It arises from the lower part of the interventricular septum and crosses the interior space of the right ventricle to connect with the inferior papillary muscle. The right ventricle tapers into the pulmonary trunk, into which it ejects blood when contracting. The pulmonary trunk branches into the left and right pulmonary arteries that carry the blood to each lung. The pulmonary valve lies between the right heart and the pulmonary trunk.

The left heart has two chambers: the left atrium and the left ventricle, separated by the mitral valve.

The left atrium receives oxygenated blood back from the lungs via one of the four pulmonary veins. The left atrium has an outpouching called the left atrial appendage. Like the right atrium, the left atrium is lined by pectinate muscles. The left atrium is connected to the left ventricle by the mitral valve.

The left ventricle is much thicker as compared with the right, due to the greater force needed to pump blood to the entire body. Like the right ventricle, the left also has trabeculae carneae, but there is no moderator band. The left ventricle pumps blood to the body through the aortic valve and into the aorta. Two small openings above the aortic valve carry blood to the heart itself, the left main coronary artery and the right coronary artery.

The heart wall is made up of three layers: the inner endocardium, middle myocardium and outer epicardium. These are surrounded by a double-membraned sac called the pericardium.

The innermost layer of the heart is called the endocardium. It is made up of a lining of simple squamous epithelium and covers heart chambers and valves. It is continuous with the endothelium of the veins and arteries of the heart, and is joined to the myocardium with a thin layer of connective tissue. The endocardium, by secreting endothelins, may also play a role in regulating the contraction of the myocardium.

The middle layer of the heart wall is the myocardium, which is the cardiac muscle—a layer of involuntary striated muscle tissue surrounded by a framework of collagen. The cardiac muscle pattern is elegant and complex, as the muscle cells swirl and spiral around the chambers of the heart, with the outer muscles forming a figure 8 pattern around the atria and around the bases of the great vessels and the inner muscles, forming a figure 8 around the two ventricles and proceeding toward the apex. This complex swirling pattern allows the heart to pump blood more effectively.

There are two types of cells in cardiac muscle: muscle cells which have the ability to contract easily, and pacemaker cells of the conducting system. The muscle cells make up the bulk (99%) of cells in the atria and ventricles. These contractile cells are connected by intercalated discs which allow a rapid response to impulses of action potential from the pacemaker cells. The intercalated discs allow the cells to act as a syncytium and enable the contractions that pump blood through the heart and into the major arteries. The pacemaker cells make up 1% of cells and form the conduction system of the heart. They are generally much smaller than the contractile cells and have few myofibrils which gives them limited contractility. Their function is similar in many respects to neurons. Cardiac muscle tissue has autorhythmicity, the unique ability to initiate a

cardiac action potential at a fixed rate—spreading the impulse rapidly from cell to cell to trigger the contraction of the entire heart.

20.5.2 The Pericardium

The pericardium is the sack that surrounds the heart. The tough outer surface of the pericardium is called the fibrous membrane. This is lined by a double inner membrane called the serous membrane that produces pericardial fluid to lubricate the surface of the heart. The part of the serous membrane attached to the fibrous membrane is called the parietal pericardium, while the part of the serous membrane attached to the heart is known as the visceral pericardium. The pericardium is present in order to lubricate its movement against other structures within the chest, to keep the heart's position stabilised within the chest, and to protect the heart from infection.

Heart tissue, like all cells in the body, needs to be supplied with oxygen, nutrients and a way of removing metabolic wastes. This is achieved by the coronary circulation, which includes arteries, veins, and lymphatic vessels. Blood flow through the coronary vessels occurs in peaks and troughs relating to the heart muscle's relaxation or contraction.

Heart tissue receives blood from two arteries which arise just above the aortic valve. These are the left main coronary artery and the right coronary artery. The left main coronary artery splits shortly after leaving the aorta into two vessels, the left anterior descending and the left circumflex artery. The left anterior descending artery supplies heart tissue and the front, outer side, and the septum of the left ventricle. It does this by branching into smaller arteries—diagonal and septal branches. The left circumflex supplies the back and underneath of the left ventricle. The right coronary artery supplies the right atrium, right ventricle, and lower posterior sections of the left ventricle. The right coronary artery also supplies blood to the atrioventricular node (in about 90% of people) and the sinoatrial node (in about 60% of people). The right coronary artery runs in a groove at the back of the heart and the left anterior descending artery runs in a groove at the front. There is significant variation between people in the anatomy of the arteries that supply the heart. The arteries divide at their further reaches into smaller branches that join together at the edges of each arterial distribution.

The coronary sinus is a large vein that drains into the right atrium, and receives most of the venous drainage of the heart. It receives blood from the great cardiac vein (receiving the left atrium and both ventricles), the posterior cardiac vein (draining the back of the left ventricle), the middle cardiac vein (draining the bottom of the left and right ventricles), and small cardiac veins. The anterior car-

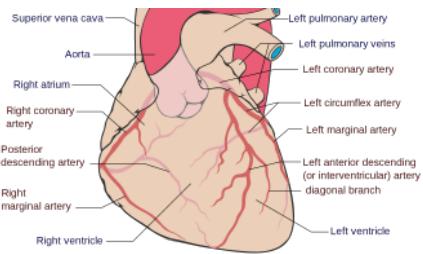


Figure 20.4: Arterial supply to the heart (red), with other areas labelled (blue).⁴

diac veins drain the front of the right ventricle and drain directly into the right atrium.

The heart receives nerve signals from the vagus nerve and from nerves arising from the sympathetic trunk. These nerves act to influence, but not control, the heart rate. Sympathetic nerves also influence the force of heart contraction. Signals that travel along these nerves arise from two paired cardiovascular centres in the medulla oblongata. The vagus nerve of the parasympathetic nervous system acts to decrease the heart rate, and nerves from the sympathetic trunk act to increase the heart rate. These nerves form a network of nerves that lies over the heart called the cardiac plexus.

The vagus nerve is a long, wandering nerve that emerges from the brainstem and provides parasympathetic stimulation to a large number of organs in the thorax and abdomen, including the heart. The nerves from the sympathetic trunk emerge through the T1-T4 thoracic ganglia and travel to both the sinoatrial and atrioventricular nodes, as well as to the atria and ventricles. The ventricles are more richly innervated by sympathetic fibers than parasympathetic fibers. Sympathetic stimulation causes the release of the neurotransmitter norepinephrine (also known as noradrenaline) at the neuromuscular junction of the cardiac nerves. This shortens the repolarization period, thus speeding the rate of depolarization and contraction, which results in an increased heart rate. It opens chemical or ligand-gated sodium and calcium ion channels, allowing an influx of positively charged ions. Norepinephrine binds to the beta-1 receptor.

20.6 The Blood Flow

The heart functions as a pump in the circulatory system to provide a continuous flow of blood throughout the body. This circulation consists of the systemic circulation to and from the body and the pulmonary circulation to and from the lungs. Blood in the pulmonary circulation exchanges carbon dioxide for oxygen in the lungs through the process of respiration. The systemic circulation then trans-

ports oxygen to the body and returns carbon dioxide and relatively deoxygenated blood to the heart for transfer to the lungs.

The right heart collects deoxygenated blood from two large veins, the superior and inferior venae cavae. Blood collects in the right and left atrium continuously. The superior vena cava drains blood from above the diaphragm and empties into the upper back part of the right atrium. The inferior vena cava drains the blood from below the diaphragm and empties into the back part of the atrium below the opening for the superior vena cava. Immediately above and to the middle of the opening of the inferior vena cava is the opening of the thin-walled coronary sinus. Additionally, the coronary sinus returns deoxygenated blood from the myocardium to the right atrium. The blood collects in the right atrium. When the right atrium contracts, the blood is pumped through the tricuspid valve into the right ventricle. As the right ventricle contracts, the tricuspid valve closes and the blood is pumped into the pulmonary trunk through the pulmonary valve. The pulmonary trunk divides into pulmonary arteries and progressively smaller arteries throughout the lungs, until it reaches capillaries. As these pass by alveoli carbon dioxide is exchanged for oxygen. This happens through the passive process of diffusion.

In the left heart, oxygenated blood is returned to the left atrium via the pulmonary veins. It is then pumped into the left ventricle through the mitral valve and into the aorta through the aortic valve for systemic circulation. The aorta is a large artery that branches into many smaller arteries, arterioles, and ultimately capillaries. In the capillaries, oxygen and nutrients from blood are supplied to body cells for metabolism, and exchanged for carbon dioxide and waste products. Capillary blood, now deoxygenated, travels into venules and veins that ultimately collect in the superior and inferior vena cavae, and into the right heart.

20.6.1 The Cardiac Cycle

The cardiac cycle refers to the sequence of events in which the heart contracts and relaxes with every heartbeat. The period of time during which the ventricles contract, forcing blood out into the aorta and main pulmonary artery, is known as systole, while the period during which the ventricles relax and refill with blood is known as diastole. The atria and ventricles work in concert, so in systole when the ventricles are contracting, the atria are relaxed and collecting blood. When the ventricles are relaxed in diastole, the atria contract to pump blood to the ventricles. This coordination ensures blood is pumped efficiently to the body.

At the beginning of the cardiac cycle, the ventricles are relaxing. As they do so, they are filled by blood passing through the open mitral and tricuspid valves. After the ven-

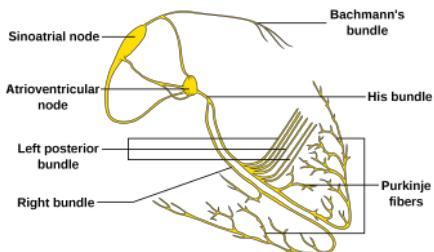


Figure 20.5: Conduction system of the heart.⁵

tricles have completed most of their filling, the atria contract, forcing further blood into the ventricles and priming the pump. Next, the ventricles start to contract. As the pressure rises within the cavities of the ventricles, the mitral and tricuspid valves are forced shut. As the pressure within the ventricles rises further, exceeding the pressure within the aorta and pulmonary arteries, the aortic and pulmonary valves open. Blood is ejected from the heart, causing the pressure within the ventricles to fall. Simultaneously, the atria refill as blood flows into the right atrium through the superior and inferior vena cavae, and into the left atrium through the pulmonary veins. Finally, when the pressure within the ventricles falls below the pressure within the aorta and pulmonary arteries, the aortic and pulmonary valves close. The ventricles start to relax, the mitral and tricuspid valves open, and the cycle begins again.

20.6.2 Control of The Heart Rate

The normal resting heart rate is called the sinus rhythm, created and sustained by the sinoatrial node, a group of pacemaking cells found in the wall of the right atrium. Cells in the sinoatrial node do this by creating an action potential. The cardiac action potential is created by the movement of specific electrolytes into and out of the pacemaker cells. The action potential then spreads to nearby cells.

When the sinoatrial cells are resting, they have a negative charge on their membranes. However a rapid influx of sodium ions causes the membrane's charge to become positive. This is called depolarisation and occurs spontaneously. Once the cell has a sufficiently high charge, the sodium channels close and calcium ions then begin to enter the cell, shortly after which potassium begins to leave it. All the ions travel through ion channels in the membrane of the sinoatrial cells. The potassium and calcium start to move out of and into the cell only once it has a sufficiently high charge, and so are called voltage-gated. Shortly after this, the calcium channels close and potassium channels open, allowing potassium to leave the cell. This causes the cell to have a negative resting charge and is called repolarization. When the membrane potential reaches ap-

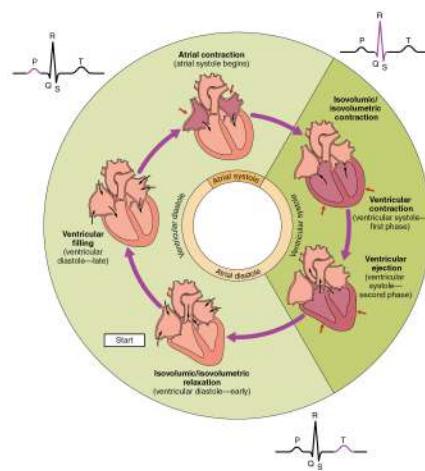


Figure 20.6: The cardiac cycle as correlated to the EKG⁶

proximately -60 mV, the potassium channels close and the process may begin again.

The ions move from areas where they are concentrated to where they are not. For this reason sodium moves into the cell from outside, and potassium moves from within the cell to outside the cell. Calcium also plays a critical role. Their influx through slow channels means that the sinoatrial cells have a prolonged "plateau" phase when they have a positive charge. A part of this is called the absolute refractory period. Calcium ions also combine with the regulatory protein troponin C in the troponin complex to enable contraction of the cardiac muscle, and separate from the protein to allow relaxation.

The adult resting heart rate ranges from 60 to 100 bpm. The resting heart rate of a newborn can be 129 beats per minute (bpm) and this gradually decreases until maturity. An athlete's heart rate can be lower than 60 bpm. During exercise the rate can be 150 bpm with maximum rates reaching from 200 to 220 bpm.

20.6.3 The Electrocardiogram (EKG)

Using surface electrodes on the body, it is possible to record the electrical activity of the heart. This tracing of the electrical signal is the electrocardiogram (ECG) or (EKG). An ECG is a bedside test and involves the placement of ten leads on the body. This produces a "12 lead" ECG (three extra leads are calculated mathematically, and one lead is a ground).

There are five prominent features on the ECG: the P wave (atrial depolarisation), the QRS complex (ventricular depolarisation) and the T wave (ventricular repolarisation). As the heart cells contract, they create a current that travels through the heart. A downward deflection on the ECG

implies cells are becoming more positive in charge ("depolarising") in the direction of that lead, whereas an upward inflection implies cells are becoming more negative ("repolarising") in the direction of the lead. This depends on the position of the lead, so if a wave of depolarising moved from left to right, a lead on the left would show a negative deflection, and a lead on the right would show a positive deflection. The ECG is a useful tool in detecting rhythm disturbances and in detecting insufficient blood supply to the heart. Sometimes abnormalities are suspected, but not immediately visible on the ECG. Testing when exercising can be used to provoke an abnormality, or an ECG can be worn for a longer period such as a 24-hour Holter monitor if a suspected rhythm abnormality is not present at the time of assessment.

20.7 Arteries

Oxygenated blood enters the systemic circulation when leaving the left ventricle, through the aortic semilunar valve. The first part of the systemic circulation is the aorta, a massive and thick-walled artery. The aorta arches and gives branches supplying the upper part of the body after passing through the aortic opening of the diaphragm at the level of thoracic ten vertebra, it enters the abdomen. Later it descends down and supplies branches to abdomen, pelvis, perineum and the lower limbs. The walls of aorta are elastic. This elasticity helps to maintain the blood pressure throughout the body. When the aorta receives almost five litres of blood from the heart, it recoils and is responsible for pulsating blood pressure. Moreover, as aorta branches into smaller arteries, their elasticity goes on decreasing and their compliance goes on increasing.

20.8 Capillaries

Arteries branch into small passages called arterioles and then into the capillaries. The capillaries merge to bring blood into the venous system.

20.9 Veins

Capillaries merge into venules, which merge into veins. The venous system feeds into the two major veins: the superior vena cava - which mainly drains tissues above the heart - and the inferior vena cava - which mainly drains tissues below the heart. These two large veins empty into the right atrium of the heart.

20.9.1 Portal veins

The general rule is that arteries from the heart branch out into capillaries, which collect into veins leading back to

the heart. Portal veins are a slight exception to this. In humans the only significant example is the hepatic portal vein which combines from capillaries around the gastrointestinal tract where the blood absorbs the various products of digestion; rather than leading directly back to the heart, the hepatic portal vein branches into a second capillary system in the liver.

20.10 The Systemic Circulation

Systemic circulation is the portion of the cardiovascular system which transports oxygenated blood away from the heart through the aorta from the left ventricle where the blood has been previously deposited from pulmonary circulation, to the rest of the body, and returns oxygen-depleted blood back to the heart.

The brain has a dual blood supply that comes from arteries at its front and back. These are called the "anterior" and "posterior" circulation respectively. The anterior circulation arises from the internal carotid arteries and supplies the front of the brain. The posterior circulation arises from the vertebral arteries, and supplies the back of the brain and brainstem. The circulation from the front and the back join together (anastomose) at the Circle of Willis.

The renal circulation receives around 20% of the cardiac output. It branches from the abdominal aorta and returns blood to the ascending vena cava. It is the blood supply to the kidneys, and contains many specialized blood vessels.

20.11 The Pulmonary Circulation

The circulatory system of the lungs is the portion of the cardiovascular system in which oxygen-depleted blood is pumped away from the heart, via the pulmonary artery, to the lungs and returned, oxygenated, to the heart via the pulmonary vein.

Oxygen deprived blood from the superior and inferior vena cava enters the right atrium of the heart and flows through the tricuspid valve (right atrioventricular valve) into the right ventricle, from which it is then pumped through the pulmonary semilunar valve into the pulmonary artery to the lungs. Gas exchange occurs in the lungs, whereby CO₂ is released from the blood, and oxygen is absorbed. The pulmonary vein returns the now oxygen-rich blood to the left atrium.

A separate system known as the bronchial circulation supplies blood to the tissue of the larger airways of the lung.

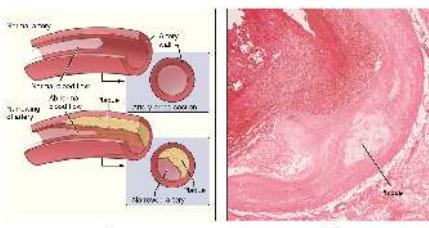


Figure 20.7: Atherosclerosis is a condition affecting the circulatory system. If the coronary arteries are affected, angina pectoris may result or at worse a heart attack.⁷

20.12 Cardiovascular Disease

Diseases affecting the cardiovascular system are called cardiovascular disease.

Many of these diseases are called “lifestyle diseases” because they develop over time and are related to a person’s exercise habits, diet, whether they smoke, and other lifestyle choices a person makes. Atherosclerosis is the precursor to many of these diseases. It is where small atheromatous plaques build up in the walls of medium and large arteries. This may eventually grow or rupture to occlude the arteries. It is also a risk factor for acute coronary syndromes, which are diseases that are characterised by a sudden deficit of oxygenated blood to the heart tissue. Atherosclerosis is also associated with problems such as aneurysm formation or splitting (“dissection”) of arteries.

Another major cardiovascular disease involves the creation of a clot, called a “thrombus”. These can originate in veins or arteries. Deep venous thrombosis, which mostly occurs in the legs, is one cause of clots in the veins of the legs, particularly when a person has been stationary for a long time. These clots may embolise, meaning travel to another location in the body. The results of this may include pulmonary embolus, transient ischaemic attacks, or stroke.

Cardiovascular diseases may also be congenital in nature, such as heart defects or persistent fetal circulation, where the circulatory changes that are supposed to happen after birth do not. Not all congenital changes to the circulatory system are associated with diseases, a large number are anatomical variations.

20.13 History

The earliest known writings on the circulatory system are found in the Ebers Papyrus (16th century BCE), an ancient Egyptian medical papyrus containing over 700 prescriptions and remedies, both physical and spiritual. In the papyrus, it acknowledges the connection of the heart to the

arteries. The Egyptians thought air came in through the mouth and into the lungs and heart. From the heart, the air travelled to every member through the arteries. Although this concept of the circulatory system is only partially correct, it represents one of the earliest accounts of scientific thought.

In the 6th century BCE, the knowledge of circulation of vital fluids through the body was known to the Ayurvedic physician Sushruta in ancient India. He also seems to have possessed knowledge of the arteries, described as ‘channels’ by Dwivedi & Dwivedi (2007). The valves of the heart were discovered by a physician of the Hippocratean school around the 4th century BCE. However their function was not properly understood then. Because blood pools in the veins after death, arteries look empty. Ancient anatomists assumed they were filled with air and that they were for transport of air.

The Greek physician, Herophilus, distinguished veins from arteries but thought that the pulse was a property of arteries themselves. Greek anatomist Erasistratus observed that arteries that were cut during life bleed. He ascribed the fact to the phenomenon that air escaping from an artery is replaced with blood that entered by very small vessels between veins and arteries. Thus he apparently postulated capillaries but with reversed flow of blood.

In 2nd century AD Rome, the Greek physician Galen knew that blood vessels carried blood and identified venous (dark red) and arterial (brighter and thinner) blood, each with distinct and separate functions. Growth and energy were derived from venous blood created in the liver from chyle, while arterial blood gave vitality by containing pneuma (air) and originated in the heart. Blood flowed from both creating organs to all parts of the body where it was consumed and there was no return of blood to the heart or liver. The heart did not pump blood around, the heart’s motion sucked blood in during diastole and the blood moved by the pulsation of the arteries themselves.

Galen believed that the arterial blood was created by venous blood passing from the left ventricle to the right by passing through ‘pores’ in the interventricular septum, air passed from the lungs via the pulmonary artery to the left side of the heart. As the arterial blood was created ‘sooty’ vapors were created and passed to the lungs also via the pulmonary artery to be exhaled.

In 1025, The Canon of Medicine by the Persian physician, Avicenna, “erroneously accepted the Greek notion regarding the existence of a hole in the ventricular septum by which the blood traveled between the ventricles.” Despite this, Avicenna “correctly wrote on the cardiac cycles and valvular function”, and “had a vision of blood circulation” in his Treatise on Pulse. While also refining Galen’s

erroneous theory of the pulse, Avicenna provided the first correct explanation of pulsation: "Every beat of the pulse comprises two movements and two pauses. Thus, expansion : pause : contraction : pause. [...] The pulse is a movement in the heart and arteries ... which takes the form of alternate expansion and contraction."

In 1242, the Arabian physician, Ibn al-Nafis, became the first person to accurately describe the process of pulmonary circulation, for which he is sometimes considered the father of circulatory physiology.[failed verification] Ibn al-Nafis stated in his Commentary on Anatomy in Avicenna's Canon:

"...the blood from the right chamber of the heart must arrive at the left chamber but there is no direct pathway between them. The thick septum of the heart is not perforated and does not have visible pores as some people thought or invisible pores as Galen thought. The blood from the right chamber must flow through the vena arteriosa (pulmonary artery) to the lungs, spread through its substances, be mingled there with air, pass through the arteria venosa (pulmonary vein) to reach the left chamber of the heart and there form the vital spirit..."

In addition, Ibn al-Nafis had an insight into what would become a larger theory of the capillary circulation. He stated that "there must be small communications or pores (manafidh in Arabic) between the pulmonary artery and vein," a prediction that preceded the discovery of the capillary system by more than 400 years. Ibn al-Nafis' theory, however, was confined to blood transit in the lungs and did not extend to the entire body.

Michael Servetus was the first European to describe the function of pulmonary circulation, although his achievement was not widely recognized at the time, for a few reasons. He firstly described it in the "Manuscript of Paris" (near 1546), but this work was never published. And later he published this description, but in a theological treatise, *Christianismi Restitutio*, not in a book on medicine. Only three copies of the book survived but these remained hidden for decades, the rest were burned shortly after its publication in 1553 because of persecution of Servetus by religious authorities.

Better known discovery of pulmonary circulation was by Vesalius's successor at Padua, Realdo Colombo, in 1559.

Finally, the English physician William Harvey, a pupil of Hieronymus Fabricius (who had earlier described the valves of the veins without recognizing their function), performed a sequence of experiments and published his *Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus* in 1628, which "demonstrated that there had to be a direct connection between the venous and arterial systems throughout the body, and not just the lungs. Most

importantly, he argued that the beat of the heart produced a continuous circulation of blood through minute connections at the extremities of the body. This is a conceptual leap that was quite different from Ibn al-Nafis' refinement of the anatomy and bloodflow in the heart and lungs." This work, with its essentially correct exposition, slowly convinced the medical world. However, Harvey was not able to identify the capillary system connecting arteries and veins; these were later discovered by Marcello Malpighi in 1661.

In 1956, André Frédéric Cournand, Werner Forssmann and Dickinson W. Richards were awarded the Nobel Prize in Medicine "for their discoveries concerning heart catheterization and pathological changes in the circulatory system." In his Nobel lecture, Forssmann credits Harvey as birthing cardiology with the publication of his book in 1628.

In the 1970s, Diana McSherry developed computer-based systems to create images of the circulatory system and heart without the need for surgery.

Chapter 21

The Lymphatic And Immune Systems

The lymphatic system consists of a network of lymphatic vessels and lymph capillaries, lymph nodes and organs, and lymphatic tissues and circulating lymph. Lymph is essentially recycled excess blood plasma after it has been filtered from the interstitial fluid (between cells) and returned to the lymphatic system. One of its major functions is to carry the lymph, draining and returning interstitial fluid back towards the heart for return to the cardiovascular system, by emptying into the lymphatic ducts. Its other main function is in the adaptive immune system. The lymphatic system is an open system providing an accessory route for excess interstitial fluid to be returned to the blood.

The immune system is a host defense system comprising many biological structures and processes within an organism that protects against disease. To function properly, an immune system must detect a wide variety of agents, known as pathogens, from viruses to parasitic worms, and distinguish them from the organism's own healthy tissue. In many species, there are two major subsystems of the immune system: the innate immune system and the adaptive immune system. Both subsystems use humoral immunity and cell-mediated immunity to perform their functions. In humans, the blood-brain barrier, blood-cerebrospinal fluid barrier, and similar fluid-brain barriers separate the peripheral immune system from the neuroimmune system, which protects the brain.

21.1 The Lymphatic System

The lymphatic system, or lymphoid system, is an organ system in vertebrates that is part of the circulatory system and the immune system. It is made up of a large network of lymphatic vessels, lymphatic or lymphoid organs, and lymphoid tissues. The vessels carry a clear fluid called lymph (the Latin word *lympha* refers to the deity of fresh water, "Lympha") towards the heart.

The adjective used for the lymph-transporting system

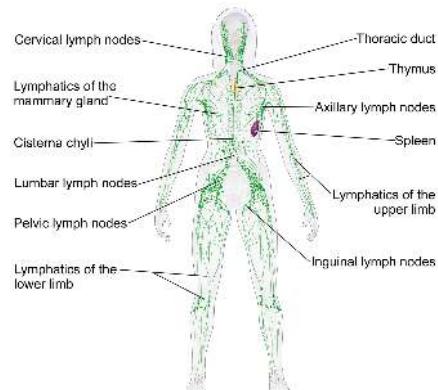


Figure 21.1: The human lymphatic system.¹

is lymphatic. The adjective used for the tissues where lymphocytes are formed is lymphoid. Lymphatic comes from the Latin word *lymphaticus*, meaning "connected to water."

Unlike the cardiovascular system, the lymphatic system is not a closed system. The human circulatory system processes an average of 20 litres of blood per day through capillary filtration, which removes plasma from the blood. Roughly 17 litres of the filtered plasma is reabsorbed directly into the blood vessels, while the remaining three litres remain in the interstitial fluid. One of the main functions of the lymphatic system is to provide an accessory return route to the blood for the surplus three litres.

The other main function is that of immune defense. Lymph is very similar to blood plasma, in that it contains waste products and cellular debris, together with bacteria and proteins. The cells of the lymph are mostly lymphocytes. Associated lymphoid organs are composed of lymphoid tissue, and are the sites either of lymphocyte production or of lymphocyte activation. These include the lymph nodes (where the highest lymphocyte concentration is found), the spleen, the thymus, and the tonsils. Lymphocytes are initially generated in the bone marrow. The lymphoid organs also contain other types of cells such as stromal cells for support. Lymphoid tissue is also associ-

ated with mucosas such as mucosa-associated lymphoid tissue (MALT).

Fluid from circulating blood leaks into the tissues of the body by capillary action, carrying nutrients to the cells. The fluid bathes the tissues as interstitial fluid, collecting waste products, bacteria, and damaged cells, and then drains as lymph into the lymphatic capillaries and lymphatic vessels. These vessels carry the lymph throughout the body, passing through numerous lymph nodes which filter out unwanted materials such as bacteria and damaged cells. Lymph then passes into much larger lymph vessels known as lymph ducts. The right lymphatic duct drains the right side of the region and the much larger left lymphatic duct, known as the thoracic duct, drains the left side of the body. The ducts empty into the subclavian veins to return to the blood circulation. Lymph is moved through the system by muscle contractions. In some vertebrates, a lymph heart is present that pumps the lymph to the veins.

The lymphatic system was first described in the 17th century independently by Olaus Rudbeck and Thomas Bartholin.

The lymphatic system consists of a conducting network of lymphatic vessels, lymphoid organs, lymphoid tissues, and the circulating lymph.

21.1.1 Primary lymphoid organs

The primary (or central) lymphoid organs generate lymphocytes from immature progenitor cells. The thymus and the bone marrow constitute the primary lymphoid organs involved in the production and early clonal selection of lymphocyte tissues.

21.1.2 The Bone Marrow

Bone marrow is responsible for both the creation of T cells and the production and maturation of B cells, which are important cell types of the immune system. From the bone marrow, B cells immediately join the circulatory system and travel to secondary lymphoid organs in search of pathogens. T cells, on the other hand, travel from the bone marrow to the thymus, where they develop further and mature. Mature T cells then join B cells in search of pathogens. The other 95% of T cells begin a process of apoptosis, a form of programmed cell death.

21.1.3 The Thymus

The thymus increases in size from birth in response to postnatal antigen stimulation. It is most active during the neonatal and pre-adolescent periods. At puberty, by the early teens, the thymus begins to atrophy and regress, with

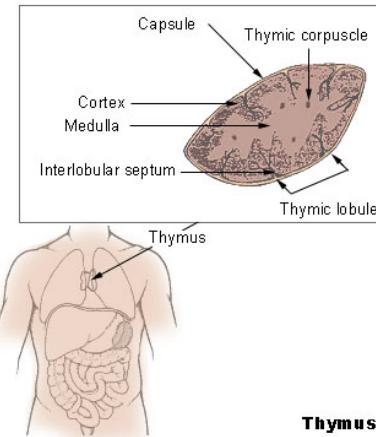


Figure 21.2: Location and microscopic anatomy of the human thymus²

adipose tissue mostly replacing the thymic stroma. However, residual T lymphopoiesis continues throughout adult life. The loss or lack of the thymus results in severe immunodeficiency and subsequent high susceptibility to infection. In most species, the thymus consists of lobules divided by septa which are made up of epithelium; it is therefore often considered an epithelial organ. T cells mature from thymocytes, proliferate, and undergo a selection process in the thymic cortex before entering the medulla to interact with epithelial cells.

The thymus provides an inductive environment for the development of T cells from hematopoietic progenitor cells. In addition, thymic stromal cells allow for the selection of a functional and self-tolerant T cell repertoire. Therefore, one of the most important roles of the thymus is the induction of central tolerance.

21.1.4 The Secondary Lymphoid Organs

The secondary (or peripheral) lymphoid organs (SLO), which include lymph nodes and the spleen, maintain mature naive lymphocytes and initiate an adaptive immune response. The peripheral lymphoid organs are the sites of lymphocyte activation by antigens. Activation leads to clonal expansion and affinity maturation. Mature lymphocytes recirculate between the blood and the peripheral lymphoid organs until they encounter their specific antigen.

21.1.5 The Spleen

The main functions of the spleen are:

- to produce immune cells to fight antigens

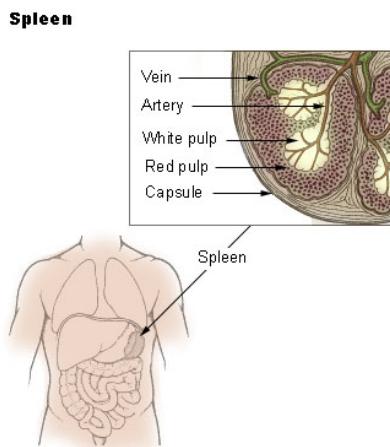


Figure 21.3: Location and microscopic anatomy of the human spleen.³

- to remove particulate matter and aged blood cells, mainly red blood cells
- to produce blood cells during fetal life.

The spleen synthesizes antibodies in its white pulp and removes antibody-coated bacteria and antibody-coated blood cells by way of blood and lymph node circulation. A study published in 2009 using mice found that the spleen contains, in its reserve, half of the body's monocytes within the red pulp. These monocytes, upon moving to injured tissue (such as the heart), turn into dendritic cells and macrophages while promoting tissue healing. The spleen is a center of activity of the mononuclear phagocyte system and can be considered analogous to a large lymph node, as its absence causes a predisposition to certain infections.

Like the thymus, the spleen has only efferent lymphatic vessels. Both the short gastric arteries and the splenic artery supply it with blood. The germinal centers are supplied by arterioles called penicillary radicles.

Until the fifth month of prenatal development, the spleen creates red blood cells; after birth, the bone marrow is solely responsible for hematopoiesis. As a major lymphoid organ and a central player in the reticuloendothelial system, the spleen retains the ability to produce lymphocytes. The spleen stores red blood cells and lymphocytes. It can store enough blood cells to help in an emergency. Up to 25% of lymphocytes can be stored at any one time.

21.1.6 The Lymph Nodes

A lymph node is an organized collection of lymphoid tissue, through which the lymph passes on its way back to

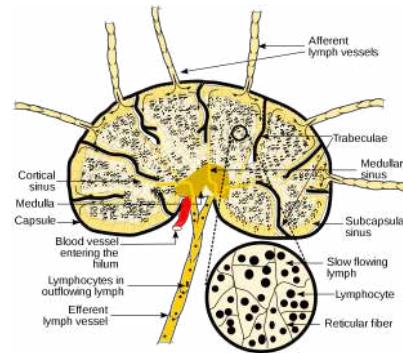


Figure 21.4: Schematic diagram of a lymph node showing flow of lymph through lymph sinuses⁴

the blood. Lymph nodes are located at intervals along the lymphatic system. Several afferent lymph vessels bring in lymph, which percolates through the substance of the lymph node, and is then drained out by an efferent lymph vessel. Of the nearly 800 lymph nodes in the human body, about 300 are located in the head and neck. Many are grouped in clusters in different regions, as in the under-arm and abdominal areas. Lymph node clusters are commonly found at the proximal ends of limbs (groin, armpits) and in the neck, where lymph is collected from regions of the body likely to sustain pathogen contamination from injuries. Lymph nodes are particularly numerous in the mediastinum in the chest, neck, pelvis, axilla, inguinal region, and in association with the blood vessels of the intestines.

The substance of a lymph node consists of lymphoid follicles in an outer portion called the cortex. The inner portion of the node is called the medulla, which is surrounded by the cortex on all sides except for a portion known as the hilum. The hilum presents as a depression on the surface of the lymph node, causing the otherwise spherical lymph node to be bean-shaped or ovoid. The efferent lymph vessel directly emerges from the lymph node at the hilum. The arteries and veins supplying the lymph node with blood enter and exit through the hilum. The region of the lymph node called the paracortex immediately surrounds the medulla. Unlike the cortex, which has mostly immature T cells, or thymocytes, the paracortex has a mixture of immature and mature T cells. Lymphocytes enter the lymph nodes through specialised high endothelial venules found in the paracortex.

A lymph follicle is a dense collection of lymphocytes, the number, size, and configuration of which change in accordance with the functional state of the lymph node. For example, the follicles expand significantly when encountering a foreign antigen. The selection of B cells, or B lymphocytes, occurs in the germinal centre of the lymph nodes.

Secondary lymphoid tissue provides the environment for the foreign or altered native molecules (antigens) to interact with the lymphocytes. It is exemplified by the lymph nodes, and the lymphoid follicles in tonsils, Peyer's patches, spleen, adenoids, skin, etc. that are associated with the mucosa-associated lymphoid tissue (MALT).

In the gastrointestinal wall, the appendix has mucosa resembling that of the colon, but here it is heavily infiltrated with lymphocytes.

Lymphoid tissue associated with the lymphatic system is concerned with immune functions in defending the body against infections and the spread of tumours. It consists of connective tissue formed of reticular fibers, with various types of leukocytes (white blood cells), mostly lymphocytes enmeshed in it, through which the lymph passes. Regions of the lymphoid tissue that are densely packed with lymphocytes are known as lymphoid follicles. Lymphoid tissue can either be structurally well organized as lymph nodes or may consist of loosely organized lymphoid follicles known as the mucosa-associated lymphoid tissue (MALT).

21.1.7 The Lymphatic Vessels

The lymphatic vessels, also called lymph vessels, are thin-walled vessels that conduct lymph between different parts of the body. They include the tubular vessels of the lymph capillaries, and the larger collecting vessels—the right lymphatic duct and the thoracic duct (the left lymphatic duct). The lymph capillaries are mainly responsible for the absorption of interstitial fluid from the tissues, while lymph vessels propel the absorbed fluid forward into the larger collecting ducts, where it ultimately returns to the bloodstream via one of the subclavian veins.

The tissues of the lymphatic system are responsible for maintaining the balance of the body fluids. Its network of capillaries and collecting lymphatic vessels work to efficiently drain and transport extravasated fluid, along with proteins and antigens, back to the circulatory system. Numerous intraluminal valves in the vessels ensure a unidirectional flow of lymph without reflux. Two valve systems, a primary and a secondary valve system, are used to achieve this unidirectional flow. The capillaries are blind-ended, and the valves at the ends of capillaries use specialised junctions together with anchoring filaments to allow a unidirectional flow to the primary vessels. The collecting lymphatics, however, act to propel the lymph by the combined actions of the intraluminal valves and lymphatic muscle cells.

The lymphatic system has multiple interrelated functions:

- It is responsible for the removal of interstitial fluid from tissues

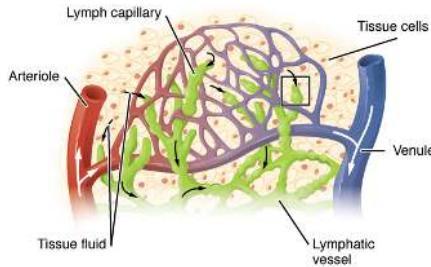


Figure 21.5: Lymph capillaries in the tissue spaces.⁵

- It absorbs and transports fatty acids and fats as chyle from the digestive system
- It transports white blood cells to and from the lymph nodes into the bones
- The lymph transports antigen-presenting cells, such as dendritic cells, to the lymph nodes where an immune response is stimulated.

Nutrients in food are absorbed via intestinal villi (greatly enlarged in the picture) to blood and lymph. Long-chain fatty acids (and other lipids with similar fat solubility like some medicines) are absorbed to the lymph and move in it enveloped inside chylomicrons. They move via the thoracic duct of the lymphatic system and finally enter the blood via the left subclavian vein, thus bypassing the liver's first-pass metabolism completely. Lymph vessels called lacteals are at the beginning of the gastrointestinal tract, predominantly in the small intestine. While most other nutrients absorbed by the small intestine are passed on to the portal venous system to drain via the portal vein into the liver for processing, fats (lipids) are passed on to the lymphatic system to be transported to the blood circulation via the thoracic duct. (There are exceptions, for example medium-chain triglycerides are fatty acid esters of glycerol that passively diffuse from the GI tract to the portal system.) The enriched lymph originating in the lymphatics of the small intestine is called chyle. The nutrients that are released into the circulatory system are processed by the liver, having passed through the systemic circulation.

The lymphatic system plays a major role in the body's immune system, as the primary site for cells relating to adaptive immune system including T-cells and B-cells. Cells in the lymphatic system react to antigens presented or found by the cells directly or by other dendritic cells. When an antigen is recognized, an immunological cascade begins involving the activation and recruitment of more and more cells, the production of antibodies and cytokines and the recruitment of other immunological cells such as macrophages.

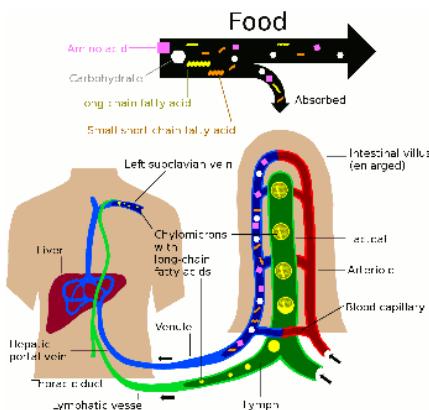


Figure 21.6: Nutrients in food are absorbed via intestinal vili (greatly enlarged in the picture) to blood and lymph. Long-chain fatty acids (and other lipids with similar fat solubility like some medicines) are absorbed to the lymph and move in it enveloped inside chylomicrons. They move via the thoracic duct of the lymphatic system and finally enter the blood via the left subclavian vein, thus bypassing the liver's first-pass metabolism completely.⁶

21.2 The Immune System

Pathogens can rapidly evolve and adapt, and thereby avoid detection and neutralization by the immune system; however, multiple defense mechanisms have also evolved to recognize and neutralize pathogens. Even simple unicellular organisms such as bacteria possess a rudimentary immune system in the form of enzymes that protect against bacteriophage infections. Other basic immune mechanisms evolved in ancient eukaryotes and remain in their modern descendants, such as plants and invertebrates. These mechanisms include phagocytosis, antimicrobial peptides called defensins, and the complement system. Jawed vertebrates, including humans, have even more sophisticated defense mechanisms, including the ability to adapt over time to recognize specific pathogens more efficiently. Adaptive (or acquired) immunity creates immunological memory after an initial response to a specific pathogen, leading to an enhanced response to subsequent encounters with that same pathogen. This process of acquired immunity is the basis of vaccination.

Disorders of the immune system can result in autoimmune diseases, inflammatory diseases and cancer. Immunodeficiency occurs when the immune system is less active than normal, resulting in recurring and life-threatening infections. In humans, immunodeficiency can either be the result of a genetic disease such as severe combined immunodeficiency, acquired conditions such as HIV/AIDS, or the use of immunosuppressive

medication. In contrast, autoimmunity results from a hyperactive immune system attacking normal tissues as if they were foreign organisms. Common autoimmune diseases include Hashimoto's thyroiditis, rheumatoid arthritis, diabetes mellitus type 1, and systemic lupus erythematosus. Immunology covers the study of all aspects of the immune system.

The immune system protects its host from infection with layered defenses of increasing specificity. In simple terms, physical barriers prevent pathogens such as bacteria and viruses from entering the organism. If a pathogen breaches these barriers, the innate immune system provides an immediate, but non-specific response. Innate immune systems are found in all plants and animals. If pathogens successfully evade the innate response, vertebrates possess a second layer of protection, the adaptive immune system, which is activated by the innate response. Here, the immune system adapts its response during an infection to improve its recognition of the pathogen. This improved response is then retained after the pathogen has been eliminated, in the form of an immunological memory, and allows the adaptive immune system to mount faster and stronger attacks each time this pathogen is encountered.

Table 21.1: Comparison of major characteristics of the innate and adaptive immune systems.

Innate immune system	Adaptive immune system
Response is non-specific	Pathogen and antigen specific response
Composed of leukocytes	Composed of antigens, B cells, T cells
Exposure leads to immediate maximal response	Lag time between exposure and maximal response
Cellular & humoral components	Cellular & humoral components
No immunological memory	Exposure leads to immunological memory
Found in nearly all forms of life	Found only in jawed vertebrates

Both innate and adaptive immunity depend on the ability of the immune system to distinguish between self and non-self molecules. In immunology, self molecules are those components of an organism's body that can be distinguished from foreign substances by the immune system. Conversely, non-self molecules are those recognized as foreign molecules. One class of non-self molecules are called antigens (short for antibody generators) and are defined as substances that bind to specific immune receptors

and elicit an immune response.

Newborn infants have no prior exposure to microbes and are particularly vulnerable to infection. Several layers of passive protection are provided by the mother. During pregnancy, a particular type of antibody, called IgG, is transported from mother to baby directly through the placenta, so human babies have high levels of antibodies even at birth, with the same range of antigen specificities as their mother. Breast milk or colostrum also contains antibodies that are transferred to the gut of the infant and protect against bacterial infections until the newborn can synthesize its own antibodies. This is passive immunity because the fetus does not actually make any memory cells or antibodies—it only borrows them. This passive immunity is usually short-term, lasting from a few days up to several months. In medicine, protective passive immunity can also be transferred artificially from one individual to another via antibody-rich serum.

21.3 The Innate Immune System

Microorganisms or toxins that successfully enter an organism encounter the cells and mechanisms of the innate immune system. The innate response is usually triggered when microbes are identified by pattern recognition receptors, which recognize components that are conserved among broad groups of microorganisms, or when damaged, injured or stressed cells send out alarm signals, many of which (but not all) are recognized by the same receptors as those that recognize pathogens. Innate immune defenses are non-specific, meaning these systems respond to pathogens in a generic way. This system does not confer long-lasting immunity against a pathogen. The innate immune system is the dominant system of host defense in most organisms.

Cells in the innate immune system use pattern recognition receptors (PRRs) to recognize molecular structures that are produced by microbial pathogens. PRRs are germline-encoded host sensors, which detect molecules typical for the pathogens. They are proteins expressed, mainly, by cells of the innate immune system, such as dendritic cells, macrophages, monocytes, neutrophils and epithelial cells, to identify two classes of molecules: pathogen-associated molecular patterns (PAMPs), which are associated with microbial pathogens, and damage-associated molecular patterns (DAMPs), which are associated with components of host's cells that are released during cell damage or death.

Recognition of extracellular or endosomal pathogen-associated molecular patterns (PAMPs) is mediated by transmembrane proteins known as toll-like receptors (TLRs). TLRs share a typical structural motif, the Leucine

rich repeats (LRR), which give them their specific appearance and are also responsible for TLR functionality. Toll-like receptors were first discovered in *Drosophila* and trigger the synthesis and secretion of cytokines and activation of other host defense programs that are necessary for both innate or adaptive immune responses. To date, ten functional members of the TLR family have been described in humans.

Cells in the innate immune system have pattern recognition receptors that detect infection or cell damage in the cytosol. Three major classes of these cytosolic receptors are NOD-like receptors, RIG (retinoic acid-inducible gene)-like receptors, and cytosolic DNA sensors.

21.3.1 Surface Barriers

Several barriers protect organisms from infection, including mechanical, chemical, and biological barriers. The waxy cuticle of most leaves, the exoskeleton of insects, the shells and membranes of externally deposited eggs, and skin are examples of mechanical barriers that are the first line of defense against infection. However, as organisms cannot be completely sealed from their environments, other systems act to protect body openings such as the lungs, intestines, and the genitourinary tract. In the lungs, coughing and sneezing mechanically eject pathogens and other irritants from the respiratory tract. The flushing action of tears and urine also mechanically expels pathogens, while mucus secreted by the respiratory and gastrointestinal tract serves to trap and entangle microorganisms.

Chemical barriers also protect against infection. The skin and respiratory tract secrete antimicrobial peptides such as the β -defensins. Enzymes such as lysozyme and phospholipase A2 in saliva, tears, and breast milk are also antibacterials. Vaginal secretions serve as a chemical barrier following menarche, when they become slightly acidic, while semen contains defensins and zinc to kill pathogens. In the stomach, gastric acid serves as a powerful chemical defense against ingested pathogens.

Within the genitourinary and gastrointestinal tracts, commensal flora serve as biological barriers by competing with pathogenic bacteria for food and space and, in some cases, by changing the conditions in their environment, such as pH or available iron. As a result of the symbiotic relationship between commensals and the immune system, the probability that pathogens will reach sufficient numbers to cause illness is reduced. However, since most antibiotics non-specifically target bacteria and do not affect fungi, oral antibiotics can lead to an “overgrowth” of fungi and cause conditions such as a vaginal candidiasis (a yeast infection). There is good evidence that re-introduction of probiotic flora, such

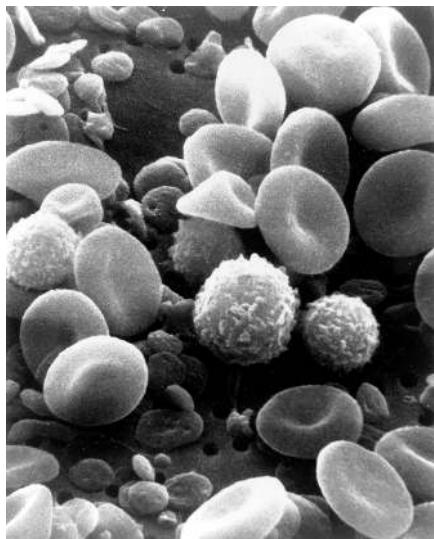


Figure 21.7: A scanning electron microscope image of normal circulating human blood.⁷ One can see red blood cells, several knobby white blood cells including lymphocytes, a monocyte, a neutrophil, and many small disc-shape platelets.

as pure cultures of the lactobacilli normally found in unpasteurized yogurt, helps restore a healthy balance of microbial populations in intestinal infections in children and encouraging preliminary data in studies on bacterial gastroenteritis, inflammatory bowel diseases, urinary tract infection and post-surgical infections.

21.4 Cellular Components of The Innate Immune System

All white blood cells (WBCs) are known as leukocytes. Most leukocytes differ from other cells of the body in that they are not tightly associated with a particular organ or tissue; thus, their function is similar to that of independent, single-cell organisms. Most leukocytes are able to move freely and interact with and capture cellular debris, foreign particles, and invading microorganisms (although macrophages, mast cells, and dendritic cells are less mobile). Unlike many other cells in the body, most innate immune leukocytes cannot divide or reproduce on their own, but are the products of multipotent hematopoietic stem cells present in the bone marrow.

The innate leukocytes include: natural killer cells, mast cells, eosinophils, basophils; and the phagocytic cells include macrophages, neutrophils, and dendritic cells, and function within the immune system by identifying and eliminating pathogens that might cause infection.

Phagocytosis is an important feature of cellular innate

immunity performed by cells called phagocytes that engulf, or eat, pathogens or particles. Phagocytes generally patrol the body searching for pathogens, but can be called to specific locations by cytokines. Once a pathogen has been engulfed by a phagocyte, it becomes trapped in an intracellular vesicle called a phagosome, which subsequently fuses with another vesicle called a lysosome to form a phagolysosome. The pathogen is killed by the activity of digestive enzymes or following a respiratory burst that releases free radicals into the phagolysosome. Phagocytosis evolved as a means of acquiring nutrients, but this role was extended in phagocytes to include engulfment of pathogens as a defense mechanism. Phagocytosis probably represents the oldest form of host defense, as phagocytes have been identified in both vertebrate and invertebrate animals.

21.4.1 Mast cells

Mast cells are a type of innate immune cell that reside in connective tissue and in the mucous membranes. They are intimately associated with wound healing and defense against pathogens, but are also often associated with allergy and anaphylaxis (serious allergic reactions that can cause death). When activated, mast cells rapidly release characteristic granules, rich in histamine and heparin, along with various hormonal mediators and chemokines, or chemotactic cytokines into the environment. Histamine dilates blood vessels, causing the characteristic signs of inflammation, and recruits neutrophils and macrophages.

21.4.2 Phagocytes

Phagocytes are cells that protect the body by ingesting harmful foreign particles, bacteria, and dead or dying cells. Their name comes from the Greek *phagein*, “to eat” or “devour”, and “-cyte”, the suffix in biology denoting “cell”, from the Greek *kutos*, “hollow vessel”. They are essential for fighting infections and for subsequent immunity. Phagocytes are important throughout the animal kingdom and are highly developed within vertebrates. One litre of human blood contains about six billion phagocytes. They were discovered in 1882 by Ilya Ilyich Mechnikov while he was studying starfish larvae. Mechnikov was awarded the 1908 Nobel Prize in Physiology or Medicine for his discovery. Phagocytes occur in many species; some amoebae behave like macrophage phagocytes, which suggests that phagocytes appeared early in the evolution of life.

Phagocytes of humans and other animals are called “professional” or “non-professional” depending on how effective they are at phagocytosis. The professional phagocytes include many types of white blood cells (such as neutrophils, monocytes, macrophages, mast cells, and dendritic cells). The main difference between professional and non-professional phagocytes is that the professional

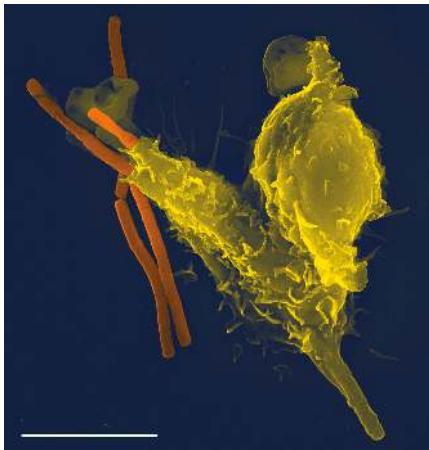


Figure 21.8: (ref.)

phagocytes have molecules called receptors on their surfaces that can detect harmful objects, such as bacteria, that are not normally found in the body. Phagocytes are crucial in fighting infections, as well as in maintaining healthy tissues by removing dead and dying cells that have reached the end of their lifespan.

During an infection, chemical signals attract phagocytes to places where the pathogen has invaded the body. These chemicals may come from bacteria or from other phagocytes already present. The phagocytes move by a method called chemotaxis. When phagocytes come into contact with bacteria, the receptors on the phagocyte's surface will bind to them. This binding will lead to the engulfing of the bacteria by the phagocyte. Some phagocytes kill the ingested pathogen with oxidants and nitric oxide. After phagocytosis, macrophages and dendritic cells can also participate in antigen presentation, a process in which a phagocyte moves parts of the ingested material back to its surface. This material is then displayed to other cells of the immune system. Some phagocytes then travel to the body's lymph nodes and display the material to white blood cells called lymphocytes. This process is important in building immunity, and many pathogens have evolved methods to evade attacks by phagocytes.

21.4.3 Macrophages

Macrophages, from the Greek, meaning “large eaters,” are large phagocytic leukocytes, which are able to move outside of the vascular system by migrating through the walls of capillary vessels and entering the areas between cells in pursuit of invading pathogens. In tissues, organ-specific macrophages are differentiated from phagocytic cells present in the blood called monocytes. Macrophages are the most efficient phagocytes and can phagocytose substantial numbers of bacteria or other cells or microbes.

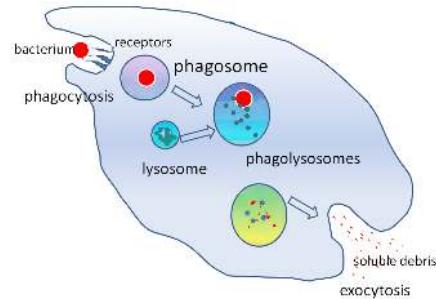


Figure 21.9: Simplified diagram of the phagocytosis and destruction of a bacterial cell.⁸

The binding of bacterial molecules to receptors on the surface of a macrophage triggers it to engulf and destroy the bacteria through the generation of a “respiratory burst”, causing the release of reactive oxygen species. Pathogens also stimulate the macrophage to produce chemokines, which summon other cells to the site of infection.

21.4.4 Neutrophils

Neutrophils, along with two other cell types (eosinophils and basophils; see below), are known as granulocytes due to the presence of granules in their cytoplasm, or as polymorphonuclear cells (PMNs) due to their distinctive lobed nuclei. Neutrophil granules contain a variety of toxic substances that kill or inhibit growth of bacteria and fungi. Similar to macrophages, neutrophils attack pathogens by activating a respiratory burst. The main products of the neutrophil respiratory burst are strong oxidizing agents including hydrogen peroxide, free oxygen radicals and hypochlorite. Neutrophils are the most abundant type of phagocyte, normally representing 50–60% of the total circulating leukocytes, and are usually the first cells to arrive at the site of an infection. The bone marrow of a normal healthy adult produces more than 100 billion neutrophils per day, and more than 10 times that many per day during acute inflammation.

21.4.5 Dendritic Cells

Dendritic cells (DC) are phagocytes in tissues that are in contact with the external environment; therefore, they are located mainly in the skin, nose, lungs, stomach, and intestines. They are named for their resemblance to neuronal dendrites, as both have many spine-like projections, but dendritic cells are in no way connected to the nervous system. Dendritic cells serve as a link between the bodily tissues and the innate and adaptive immune systems, as they present antigens to T cells, one of the key cell types of the adaptive immune system.

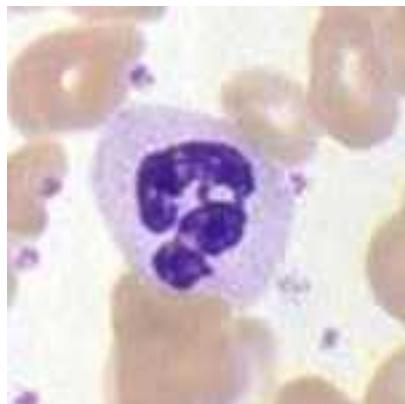


Figure 21.10: A neutrophil.⁹

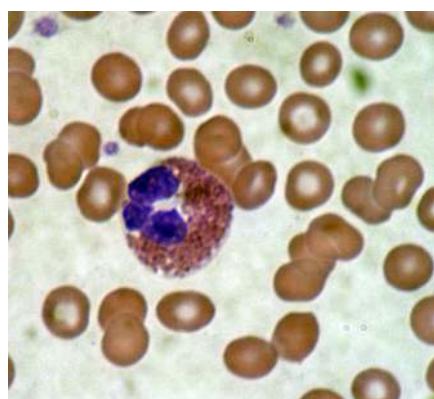


Figure 21.11: An eosinophil.¹⁰ Red blood cells surround the eosinophil, two platelets at the top left corner.

21.4.6 Basophils And Eosinophils

Basophils and eosinophils are cells related to the neutrophil. When activated by a pathogen encounter, histamine-releasing basophils are important in the defense against parasites and play a role in allergic reactions, such as asthma. Upon activation, eosinophils secrete a range of highly toxic proteins and free radicals that are highly effective in killing parasites, but may also damage tissue during an allergic reaction. Activation and release of toxins by eosinophils are, therefore, tightly regulated to prevent any inappropriate tissue destruction.

21.4.7 Innate Lymphoid Cells

Innate lymphoid cells (ILCs) are a group of innate immune cells that are derived from common lymphoid progenitor (CLP) and belong to the lymphoid lineage. These cells are defined by absence of antigen specific B or T cell receptor because of the lack of recombination activating gene (RAG). ILCs do not express myeloid or dendritic cell markers.

Natural killer cells, one of member ILCs, are lympho-

cytes and a component of the innate immune system which does not directly attack invading microbes. Rather, NK cells destroy compromised host cells, such as tumor cells or virus-infected cells, recognizing such cells by a condition known as “missing self.” This term describes cells with low levels of a cell-surface marker called MHC I (major histocompatibility complex)—a situation that can arise in viral infections of host cells. They were named “natural killer” because of the initial notion that they do not require activation in order to kill cells that are “missing self.” For many years it was unclear how NK cells recognize tumor cells and infected cells. It is now known that the MHC makeup on the surface of those cells is altered and the NK cells become activated through recognition of “missing self”. Normal body cells are not recognized and attacked by NK cells because they express intact self MHC antigens. Those MHC antigens are recognized by killer cell immunoglobulin receptors (KIR) which essentially put the brakes on NK cells.

21.5 Inflammation

Inflammation is one of the first responses of the immune system to infection or irritation. Inflammation is stimulated by chemical factors released by injured cells and serves to establish a physical barrier against the spread of infection, and to promote healing of any damaged tissue following the clearance of pathogens.

The process of acute inflammation is initiated by cells already present in all tissues, mainly resident macrophages, dendritic cells, histiocytes, Kupffer cells, and mast cells. These cells present receptors contained on the surface or within the cell, named pattern recognition receptors (PRRs), which recognize molecules that are broadly shared by pathogens but distinguishable from host molecules, collectively referred to as pathogen-associated molecular patterns (PAMPs). At the onset of an infection, burn, or other injuries, these cells undergo activation (one of their PRRs recognizes a PAMP) and release inflammatory mediators responsible for the clinical signs of inflammation.

Chemical factors produced during inflammation (histamine, bradykinin, serotonin, leukotrienes, and prostaglandins) sensitize pain receptors, cause local vasodilation of the blood vessels, and attract phagocytes, especially neutrophils. Neutrophils then trigger other parts of the immune system by releasing factors that summon additional leukocytes and lymphocytes. Cytokines produced by macrophages and other cells of the innate immune system mediate the inflammatory response. These cytokines include TNF, HMGB1, and IL-1.

The inflammatory response is characterized by the following symptoms:

- redness of the skin, due to locally increased blood circulation;
- heat, either increased local temperature, such as a warm feeling around a localized infection, or a systemic fever;
- swelling of affected tissues, such as the upper throat during the common cold or joints affected by rheumatoid arthritis;
- increased production of mucus, which can cause symptoms like a runny nose or a productive cough;
- pain, either local pain, such as painful joints or a sore throat, or affecting the whole body, such as body aches; and
- possible dysfunction of the organs or tissues involved.

21.6 The Complement System

The complement system is a biochemical cascade that attacks the surfaces of foreign cells. It contains over 20 different proteins and is named for its ability to “complement” the killing of pathogens by antibodies. Complement is the major humoral component of the innate immune response. Many species have complement systems, including non-mammals like plants, fish, and some invertebrates.

In humans, this response is activated by complement binding to antibodies that have attached to these microbes or the binding of complement proteins to carbohydrates on the surfaces of microbes. This recognition signal triggers a rapid killing response. The speed of the response is a result of signal amplification that occurs after sequential proteolytic activation of complement molecules, which are also proteases. After complement proteins initially bind to the microbe, they activate their protease activity, which in turn activates other complement proteases, and so on. This produces a catalytic cascade that amplifies the initial signal by controlled positive feedback. The cascade results in the production of peptides that attract immune cells, increase vascular permeability, and opsonize (coat) the surface of a pathogen, marking it for destruction. This deposition of complement can also kill cells directly by disrupting their plasma membrane.

Bacteria (and perhaps other prokaryotic organisms), utilize a unique defense mechanism, called the restriction modification system to protect themselves from pathogens, such as bacteriophages. In this system, bacteria produce enzymes, called restriction endonucleases, that attack and destroy specific regions of the viral DNA of invading bacteriophages. Methylation of the host's own DNA marks it as “self” and prevents it from being attacked by endonucleases. Restriction endonucleases and the restriction modification system exist exclusively

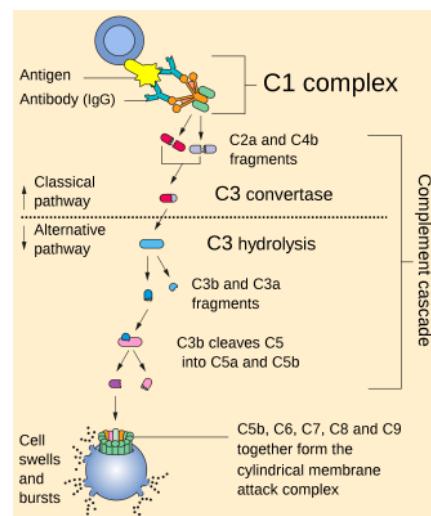


Figure 21.12: The complement system¹¹ is made up of about 25 proteins that work together to “complement” the action of antibodies in destroying bacteria. Complement proteins circulate in the blood in an inactive form. When the first protein in the complement series is activated—typically by antibody that has locked onto an antigen—it sets in motion a domino effect. Each component takes its turn in a precise chain of steps known as the complement cascade. The end product is a cylinder inserted into—and puncturing a hole in—the cell’s wall. With fluids and molecules flowing in and out, the cell swells and bursts.

in prokaryotes.

Invertebrates do not possess lymphocytes or an antibody-based humoral immune system, and it is likely that a multicomponent, adaptive immune system arose with the first vertebrates. Nevertheless, invertebrates possess mechanisms that appear to be precursors of these aspects of vertebrate immunity. Pattern recognition receptors are proteins used by nearly all organisms to identify molecules associated with microbial pathogens. Toll-like receptors are a major class of pattern recognition receptor, that exists in all coelomates (animals with a body-cavity), including humans. The complement system, as discussed above, is a biochemical cascade of the immune system that helps clear pathogens from an organism, and exists in most forms of life. Some invertebrates, including various insects, crabs, and worms utilize a modified form of the complement response known as the prophenoloxidase (proPO) system.

Antimicrobial peptides are an evolutionarily conserved component of the innate immune response found among all classes of life and represent the main form of invertebrate systemic immunity. Several species of insect produce antimicrobial peptides known as defensins and cecropins.

In invertebrates, pattern recognition proteins (PRPs) trigger proteolytic cascades that degrade proteins and control many of the mechanisms of the innate immune system of invertebrates—including hemolymph coagulation and melanization. Proteolytic cascades are important components of the invertebrate immune system because they are turned on more rapidly than other innate immune reactions because they do not rely on gene changes. Proteolytic cascades have been found to function the same in both vertebrate and invertebrates, even though different proteins are used throughout the cascades.

In the hemolymph, which makes up the fluid in the circulatory system of arthropods, a gel-like fluid surrounds pathogen invaders, similar to the way blood does in other animals. There are various different proteins and mechanisms that are involved in invertebrate clotting. In crustaceans, transglutaminase from blood cells and mobile plasma proteins make up the clotting system, where the transglutaminase polymerizes 210 kDa sub-units of a plasma-clotting protein. On the other hand, in the horseshoe crab species clotting system, components of proteolytic cascades are stored as inactive forms in granules of hemocytes, which are released when foreign molecules, like lipopolysaccharides enter.

Members of every class of pathogen that infect humans also infect plants. Although the exact pathogenic species vary with the infected species, bacteria, fungi, viruses, nematodes, and insects can all cause plant disease. As with

animals, plants attacked by insects or other pathogens use a set of complex metabolic responses which lead to the formation of defensive chemical compounds that fight infection or make the plant less attractive to insects and other herbivores. (see: plant defense against herbivory).

Like invertebrates, plants neither generate antibody or T-cell responses nor possess mobile cells that detect and attack pathogens. In addition, in case of infection, parts of some plants are treated as disposable and replaceable, in ways that very few animals are able to do. Walling off or discarding a part of a plant helps stop spread of an infection.

Most plant immune responses involve systemic chemical signals sent throughout a plant. Plants use pattern-recognition receptors to recognize conserved microbial signatures. This recognition triggers an immune response. The first plant receptors of conserved microbial signatures were identified in rice (XA21, 1995) and in Arabidopsis (FLS2, 2000). Plants also carry immune receptors that recognize highly variable pathogen effectors. These include the NBS-LRR class of proteins. When a part of a plant becomes infected with a microbial or viral pathogen, in case of an incompatible interaction triggered by specific elicitors, the plant produces a localized hypersensitive response (HR), in which cells at the site of infection undergo rapid programmed cell death to prevent the spread of the disease to other parts of the plant. HR has some similarities to animal pyroptosis, such as a requirement of caspase-1-like proteolytic activity of VPE γ , a cysteine protease that regulates cell disassembly during cell death.

“Resistance” (R) proteins, encoded by R genes, are widely present in plants and detect pathogens. These proteins contain domains similar to the NOD Like Receptors and Toll-like receptors utilized in animal innate immunity. Systemic acquired resistance (SAR) is a type of defensive response that renders the entire plant resistant to a broad spectrum of infectious agents. SAR involves the production of chemical messengers, such as salicylic acid or jasmonic acid. Some of these travel through the plant and signal other cells to produce defensive compounds to protect uninfected parts, e.g., leaves. Salicylic acid itself, although indispensable for expression of SAR, is not the translocated signal responsible for the systemic response. Recent evidence indicates a role for jasmonates in transmission of the signal to distal portions of the plant. RNA silencing mechanisms are also important in the plant systemic response, as they can block virus replication. The jasmonic acid response, is stimulated in leaves damaged by insects, and involves the production of methyl jasmonate.

21.7 The Adaptive Immune System

The adaptive immune system, also referred as the acquired immune system, is a subsystem of the immune system that is composed of specialized, systemic cells and processes that eliminates pathogens by preventing their growth. The acquired immune system is one of the two main immunity strategies found in vertebrates (the other being the innate immune system). The adaptive immune system evolved in early vertebrates.

Acquired immunity creates immunological memory after an initial response to a specific pathogen, and leads to an enhanced response to subsequent encounters with that pathogen. This process of acquired immunity is the basis of vaccination. Like the innate system, the acquired system includes both humoral immunity components and cell-mediated immunity components.

Unlike the innate immune system, the acquired immune system is highly specific to a particular pathogen. Acquired immunity can also provide long-lasting protection; for example, someone who recovers from measles is now protected against measles for their lifetime. In other cases it does not provide lifetime protection; for example, chickenpox. The acquired system response destroys invading pathogens and any toxic molecules they produce. Sometimes the acquired system is unable to distinguish harmful from harmless foreign molecules; the effects of this may be hayfever, asthma or any other allergy.

Antigens are any substances that elicit the acquired immune response (whether adaptive or maladaptive to the organism).

The cells that carry out the acquired immune response are white blood cells known as lymphocytes. Two main activities—antibody responses and cell mediated immune response—are also carried out by two different lymphocytes (B cells and T cells). In antibody responses, B cells are activated to secrete antibodies, which are proteins also known as immunoglobulins. Antibodies travel through the bloodstream and bind to the foreign antigen causing it to inactivate, which does not allow the antigen to bind to the host.

In acquired immunity, pathogen-specific receptors are “acquired” during the lifetime of the organism (whereas in innate immunity pathogen-specific receptors are already encoded in the germline). The acquired response is called “adaptive” because it prepares the body’s immune system for future challenges (though it can actually also be maladaptive when it results in autoimmunity).

The system is highly adaptable because of somatic hypermutation (a process of accelerated somatic mutations), and V(D)J recombination (an irreversible genetic

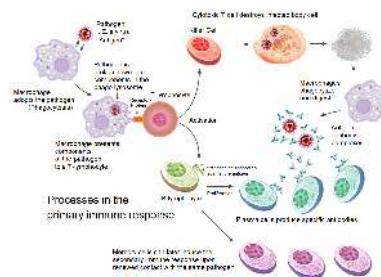


Figure 21.13: Overview of the processes involved in the primary immune response¹²

recombination of antigen receptor gene segments). This mechanism allows a small number of genes to generate a vast number of different antigen receptors, which are then uniquely expressed on each individual lymphocyte. Since the gene rearrangement leads to an irreversible change in the DNA of each cell, all progeny (offspring) of that cell inherit genes that encode the same receptor specificity, including the memory B cells and memory T cells that are the keys to long-lived specific immunity.

Acquired immunity is triggered in vertebrates when a pathogen evades the innate immune system and (1) generates a threshold level of antigen and (2) generates “stranger” or “danger” signals activating dendritic cells.

The major functions of the acquired immune system include:

- Recognition of specific “non-self” antigens in the presence of “self”, during the process of antigen presentation.
- Generation of responses that are tailored to maximally eliminate specific pathogens or pathogen-infected cells.
- Development of immunological memory, in which pathogens are “remembered” through memory B cells and memory T cells.

In humans, it takes 4–7 days for the adaptive immune system to mount a significant response.

21.7.1 The Lymphocytes

The cells of the acquired immune system are T and B lymphocytes; lymphocytes are a subset of leukocyte. B cells and T cells are the major types of lymphocytes. The human body has about 2 trillion lymphocytes, constituting 20–40% of white blood cells (WBCs); their total mass is about the same as the brain or liver. The peripheral blood contains 2% of circulating lymphocytes; the rest move within the tissues and lymphatic system.

B cells and T cells are derived from the same multi-potent hematopoietic stem cells, and are morphologically indistinguishable from one another until after they are activated. B cells play a large role in the humoral immune response, whereas T cells are intimately involved in cell-mediated immune responses. In all vertebrates except Ag-natha, B cells and T cells are produced by stem cells in the bone marrow.

T progenitors migrate from the bone marrow to the thymus where they are called thymocytes and where they develop into T cells. In humans, approximately 1–2% of the lymphocyte pool recirculates each hour to optimize the opportunities for antigen-specific lymphocytes to find their specific antigen within the secondary lymphoid tissues. In an adult animal, the peripheral lymphoid organs contain a mixture of B and T cells in at least three stages of differentiation:

- naive B and naive T cells (cells that have not matured), left the bone marrow or thymus, have entered the lymphatic system, but have yet to encounter their cognate antigen,
- effector cells that have been activated by their cognate antigen, and are actively involved in eliminating a pathogen.
- memory cells – the survivors of past infections.

21.7.2 Antigen Presentation to T Lymphocytes

Acquired immunity relies on the capacity of immune cells to distinguish between the body's own cells and unwanted invaders. The host's cells express "self" antigens. These antigens are different from those on the surface of bacteria or on the surface of virus-infected host cells ("non-self" or "foreign" antigens). The acquired immune response is triggered by recognizing foreign antigen in the cellular context of an activated dendritic cell.

With the exception of non-nucleated cells (including erythrocytes), all cells are capable of presenting antigen through the function of major histocompatibility complex (MHC) molecules. Some cells are specially equipped to present antigen, and to prime naive T cells. Dendritic cells, B-cells, and macrophages are equipped with special "co-stimulatory" ligands recognized by co-stimulatory receptors on T cells, and are termed professional antigen-presenting cells (APCs).

Several T cells subgroups can be activated by professional APCs, and each type of T cell is specially equipped to deal with each unique toxin or microbial pathogen. The type of T cell activated, and the type of response generated, depends, in part, on the context in which the APC first encountered the antigen.

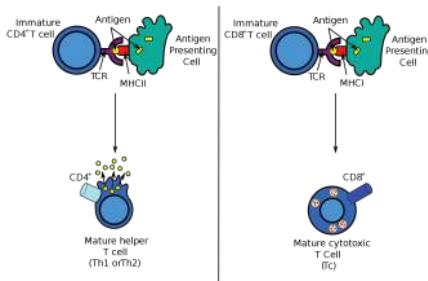


Figure 21.14: Antigen presentation stimulates T cells to become either "cytotoxic" CD8+ cells or "helper" CD4+ cells.¹³

Dendritic cells engulf exogenous pathogens, such as bacteria, parasites or toxins in the tissues and then migrate, via chemotactic signals, to the T cell-enriched lymph nodes. During migration, dendritic cells undergo a process of maturation in which they lose most of their ability to engulf other pathogens, and develop an ability to communicate with T-cells. The dendritic cell uses enzymes to chop the pathogen into smaller pieces, called antigens. In the lymph node, the dendritic cell displays these non-self antigens on its surface by coupling them to a receptor called the major histocompatibility complex, or MHC (also known in humans as human leukocyte antigen (HLA)). This MHC: antigen complex is recognized by T-cells passing through the lymph node. Exogenous antigens are usually displayed on MHC class II molecules, which activate CD4+T helper cells.

Endogenous antigens are produced by intracellular bacteria and viruses replicating within a host cell. The host cell uses enzymes to digest virally associated proteins, and displays these pieces on its surface to T-cells by coupling them to MHC. Endogenous antigens are typically displayed on MHC class I molecules, and activate CD8+ cytotoxic T-cells. With the exception of non-nucleated cells (including erythrocytes), MHC class I is expressed by all host cells.

21.7.3 Cytotoxic T lymphocytes

Cytotoxic T cells (also known as TC, killer T cell, or cytotoxic T-lymphocyte (CTL)) are a sub-group of T cells that induce the death of cells that are infected with viruses (and other pathogens), or are otherwise damaged or dysfunctional.

Naive cytotoxic T cells are activated when their T-cell receptor (TCR) strongly interacts with a peptide-bound MHC class I molecule. This affinity depends on the type and orientation of the antigen/MHC complex, and is what keeps the CTL and infected cell bound together. Once

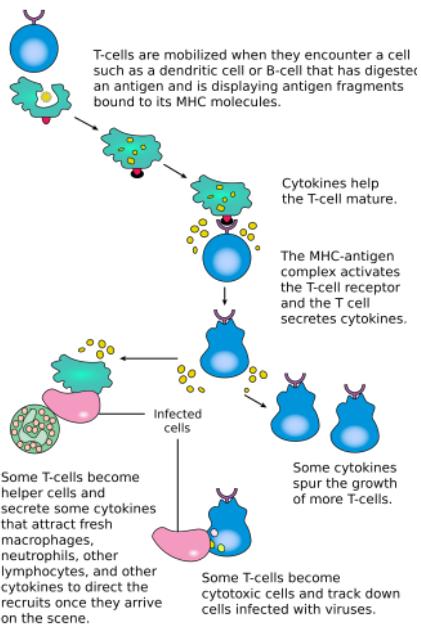


Figure 21.15: The T lymphocyte activation pathway.¹⁴ The T lymphocyte activation pathway is triggered when a T cell encounters its cognate antigen, coupled to a MHC molecule, on the surface of an infected cell or a phagocyte. T cells contribute to immune defenses in two major ways: some direct and regulate immune responses; others directly attack infected or cancerous cells.

activated, the CTL undergoes a process called clonal selection, in which it gains functions and divides rapidly to produce an army of “armed” effector cells. Activated CTL then travels throughout the body searching for cells that bear that unique MHC Class I + peptide.[citation needed]

When exposed to these infected or dysfunctional somatic cells, effector CTL release perforin and granzysin: cytotoxins that form pores in the target cell's plasma membrane, allowing ions and water to flow into the infected cell, and causing it to burst or lyse. CTL release granzyme, a serine protease encapsulated in a granule that enters cells via pores to induce apoptosis (cell death). To limit extensive tissue damage during an infection, CTL activation is tightly controlled and in general requires a very strong MHC/antigen activation signal, or additional activation signals provided by “helper” T-cells (see below).

On resolution of the infection, most effector cells die and phagocytes clear them away—but a few of these cells remain as memory cells. On a later encounter with the same antigen, these memory cells quickly differentiate into effector cells, dramatically shortening the time required to mount an effective response.[citation needed]

CD4+ lymphocytes, also called “helper” T cells, are immune response mediators, and play an important role in establishing and maximizing the capabilities of the acquired immune response. These cells have no cytotoxic or phagocytic activity; and cannot kill infected cells or clear pathogens, but, in essence “manage” the immune response, by directing other cells to perform these tasks.

21.7.4 T Helper Cells

Helper T cells express T cell receptors (TCR) that recognize antigen bound to Class II MHC molecules. The activation of a naive helper T-cell causes it to release cytokines, which influences the activity of many cell types, including the APC (Antigen-Presenting Cell) that activated it. Helper T-cells require a much milder activation stimulus than cytotoxic T cells. Helper T cells can provide extra signals that “help” activate cytotoxic cells.

Classically, two types of effector CD4+ T helper cell responses can be induced by a professional APC, designated Th1 and Th2, each designed to eliminate different types of pathogens. The factors that dictate whether an infection triggers a Th1 or Th2 type response are not fully understood, but the response generated does play an important role in the clearance of different pathogens.

The Th1 response is characterized by the production of Interferon-gamma, which activates the bactericidal activities of macrophages, and induces B cells to make opsonizing (marking for phagocytosis) and complement-fixing antibodies, and leads to cell-mediated immunity. In gen-

eral, Th1 responses are more effective against intracellular pathogens (viruses and bacteria that are inside host cells).

The Th2 response is characterized by the release of Interleukin 5, which induces eosinophils in the clearance of parasites. Th2 also produce Interleukin 4, which facilitates B cell isotype switching. In general, Th2 responses are more effective against extracellular bacteria, parasites including helminths and toxins. Like cytotoxic T cells, most of the CD4+ helper cells die on resolution of infection, with a few remaining as CD4+ memory cells.

Increasingly, there is strong evidence from mouse and human-based scientific studies of a broader diversity in CD4+ effector T helper cell subsets. Regulatory T (Treg) cells, have been identified as important negative regulators of adaptive immunity as they limit and suppresses the immune system to control aberrant immune responses to self-antigens; an important mechanism in controlling the development of autoimmune diseases. Follicular helper T (Tfh) cells are another distinct population of effector CD4+ T cells that develop from naive T cells post-antigen activation. Tfh cells are specialized in helping B cell humoral immunity as they are uniquely capable of migrating to follicular B cells in secondary lymphoid organs and provide them positive paracrine signals to enable the generation and recall production of high-quality affinity-matured antibodies. Similar to Tregs, Tfh cells also play a role in immunological tolerance as an abnormal expansion of Tfh cell numbers can lead to unrestricted autoreactive antibody production causing severe systemic autoimmune disorders.

The relevance of CD4+ T helper cells is highlighted during an HIV infection. HIV is able to subvert the immune system by specifically attacking the CD4+ T cells, precisely the cells that could drive the clearance of the virus, but also the cells that drive immunity against all other pathogens encountered during an organism's lifetime.

21.7.5 Gamma delta T cells

Gamma delta T cells ($\gamma\delta$ T cells) possess an alternative T cell receptor (TCR) as opposed to CD4+ and CD8+ $\alpha\beta$ T cells and share characteristics of helper T cells, cytotoxic T cells and natural killer cells. Like other 'unconventional' T cell subsets bearing invariant TCRs, such as CD1d-restricted natural killer T cells, $\gamma\delta$ T cells exhibit characteristics that place them at the border between innate and acquired immunity. On one hand, $\gamma\delta$ T cells may be considered a component of adaptive immunity in that they rearrange TCR genes via V(D)J recombination, which also produces junctional diversity, and develop a memory phenotype. On the other hand, however, the various subsets may also be considered part of the innate immune system where a restricted TCR or NK receptors may be used as a pattern recognition receptor. For example, according to

this paradigm, large numbers of V γ 9/V δ 2 T cells respond within hours to common molecules produced by microbes, and highly restricted intraepithelial V δ 1 T cells respond to stressed epithelial cells.

21.7.6 B Lymphocytes And Antibody Production

The B lymphocyte activation pathway. B cells function to protect the host by producing antibodies that identify and neutralize foreign objects like bacteria and viruses. B Cells are the major cells involved in the creation of antibodies that circulate in blood plasma and lymph, known as humoral immunity. Antibodies (also known as immunoglobulin, Ig), are large Y-shaped proteins used by the immune system to identify and neutralize foreign objects. In mammals, there are five types of antibody: IgA, IgD, IgE, IgG, and IgM, differing in biological properties; each has evolved to handle different kinds of antigens. Upon activation, B cells produce antibodies, each of which recognize a unique antigen, and neutralizing specific pathogens.

Antigen and antibody binding would cause five different protective mechanisms:

- Agglutination: Reduces number of infectious units to be dealt with
- Activation of complement: Cause inflammation and cell lysis
- Opsonization: Coating antigen with antibody enhances phagocytosis
- Antibody-dependent cell-mediated cytotoxicity: Antibodies attached to target cell cause destruction by macrophages, eosinophils, and NK cells
- Neutralization: Blocks adhesion of bacteria and viruses to mucosa

Like the T cell, B cells express a unique B cell receptor (BCR), in this case, a membrane-bound antibody molecule. All the BCR of any one clone of B cells recognizes and binds to only one particular antigen. A critical difference between B cells and T cells is how each cell "sees" an antigen. T cells recognize their cognate antigen in a processed form – as a peptide in the context of an MHC molecule, whereas B cells recognize antigens in their native form. Once a B cell encounters its cognate (or specific) antigen (and receives additional signals from a helper T cell (predominately Th2 type)), it further differentiates into an effector cell, known as a plasma cell.

Plasma cells are short-lived cells (2–3 days) that secrete antibodies. These antibodies bind to antigens, making them easier targets for phagocytes, and trigger the complement cascade. About 10% of plasma cells survive to become long-lived antigen-specific memory B cells. Already

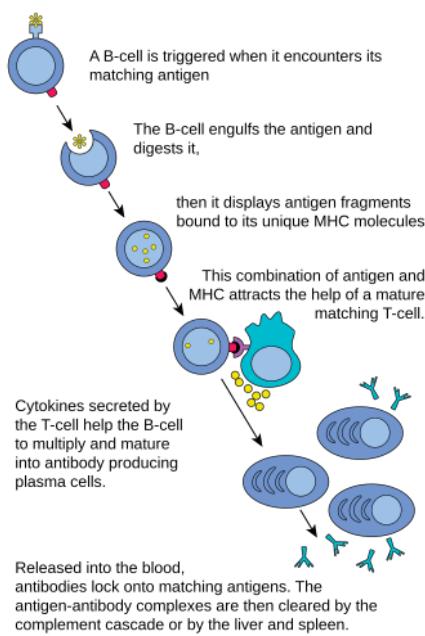


Figure 21.16: The B lymphocyte activation pathway.¹⁵ The B lymphocyte activation pathway. B cells function to protect the host by producing antibodies that identify and neutralize foreign objects like bacteria and viruses.

primed to produce specific antibodies, these cells can be called upon to respond quickly if the same pathogen re-infects the host, while the host experiences few, if any, symptoms.

21.8 The humoral Immune Response

A B cell identifies pathogens when antibodies on its surface bind to a specific foreign antigen. This antigen/antibody complex is taken up by the B cell and processed by proteolysis into peptides. The B cell then displays these antigenic peptides on its surface MHC class II molecules. This combination of MHC and antigen attracts a matching helper T cell, which releases lymphokines and activates the B cell. As the activated B cell then begins to divide, its offspring (plasma cells) secrete millions of copies of the antibody that recognizes this antigen. These antibodies circulate in blood plasma and lymph, bind to pathogens expressing the antigen and mark them for destruction by complement activation or for uptake and destruction by phagocytes. Antibodies can also neutralize challenges directly, by binding to bacterial toxins or by interfering with the receptors that viruses and bacteria use to infect cells.

21.8.1 Antibodies

An antibody (Ab), also known as an immunoglobulin (Ig), is a large, Y-shaped protein produced mainly by plasma cells that is used by the immune system to neutralize pathogens such as pathogenic bacteria and viruses. The antibody recognizes a unique molecule of the pathogen, called an antigen, via the fragment antigen-binding (Fab) variable region. Each tip of the "Y" of an antibody contains a paratope (analogous to a lock) that is specific for one particular epitope (analogous to a key) on an antigen, allowing these two structures to bind together with precision. Using this binding mechanism, an antibody can tag a microbe or an infected cell for attack by other parts of the immune system, or can neutralize its target directly (for example, by inhibiting a part of a microbe that is essential for its invasion and survival). Depending on the antigen, the binding may impede the biological process causing the disease or may activate macrophages to destroy the foreign substance. The ability of an antibody to communicate with the other components of the immune system is mediated via its Fc region (located at the base of the "Y"), which contains a conserved glycosylation site involved in these interactions. The production of antibodies is the main function of the humoral immune system.

Antibodies are secreted by B cells of the adaptive immune system, mostly by differentiated B cells called plasma cells. Antibodies can occur in two physical forms, a soluble form that is secreted from the cell to be free in the blood plasma, and a membrane-bound form that is attached to the surface of a B cell and is referred to as the B-cell receptor (BCR). The BCR is found only on the surface of B cells and facilitates the activation of these cells and their subsequent differentiation into either antibody factories called plasma cells or memory B cells that will survive in the body and remember that same antigen so the B cells can respond faster upon future exposure. In most cases, interaction of the B cell with a T helper cell is necessary to produce full activation of the B cell and, therefore, antibody generation following antigen binding. Soluble antibodies are released into the blood and tissue fluids, as well as many secretions to continue to survey for invading microorganisms.

Antibodies are glycoproteins belonging to the immunoglobulin superfamily. They constitute most of the gamma globulin fraction of the blood proteins. They are typically made of basic structural units—each with two large heavy chains and two small light chains. There are several different types of antibody heavy chains that define the five different types of crystallisable fragments (Fc) that may be attached to the antigen-binding fragments (Fab). The five different types of Fc regions allow antibodies to be grouped into five isotypes. Each Fc region of a particular antibody isotype is able to bind to its specific

Fc Receptor (FcR), except for IgD, which is essentially the BCR, thus allowing the antigen–antibody complex to mediate different roles depending on which FcR it binds. The ability of an antibody to bind to its corresponding FcR is further modulated by the structure of the glycan(s) present at conserved sites within its Fc region. The ability of antibodies to bind to FcRs helps to direct the appropriate immune response for each different type of foreign object they encounter. For example, IgE is responsible for an allergic response consisting of mast cell degranulation and histamine release. IgE's Fab paratope binds to allergic antigen, for example house dust mite particles, while its Fc region binds to Fc receptor ϵ . The allergen–IgE–Fc ϵ R interaction mediates allergic signal transduction to induce conditions such as asthma.

Though the general structure of all antibodies is very similar, a small region at the tip of the protein is extremely variable, allowing millions of antibodies with slightly different tip structures, or antigen–binding sites, to exist. This region is known as the hypervariable region. Each of these variants can bind to a different antigen. This enormous diversity of antibody paratopes on the antigen-binding fragments allows the immune system to recognize an equally wide variety of antigens. The large and diverse population of antibody paratope is generated by random recombination events of a set of gene segments that encode different antigen–binding sites (or paratopes), followed by random mutations in this area of the antibody gene, which create further diversity. This recombinational process that produces clonal antibody paratope diversity is called V(D) or VJ recombination. The antibody paratope is polygenic, made up of three genes, V, D, and J. Each paratope locus is also polymorphic, such that during antibody production, one allele of V, one of D, and one of J is chosen. These gene segments are then joined together using random genetic recombination to produce the paratope. The regions where the genes are randomly recombined together is the hypervariable region used to recognise different antigens on a clonal basis.

Antibody genes also re-organize in a process called class switching that changes the one type of heavy chain Fc fragment to another, creating a different isotype of the antibody that retains the antigen-specific variable region. This allows a single antibody to be used by different types of Fc receptors, expressed on different parts of the immune system.

Antibodies can come in different varieties known as isotypes or classes. In placental mammals there are five antibody isotypes known as IgA, IgD, IgE, IgG, and IgM. They are each named with an “Ig” prefix that stands for immunoglobulin (a name sometimes used interchangeably with antibody) and differ in their biological properties,

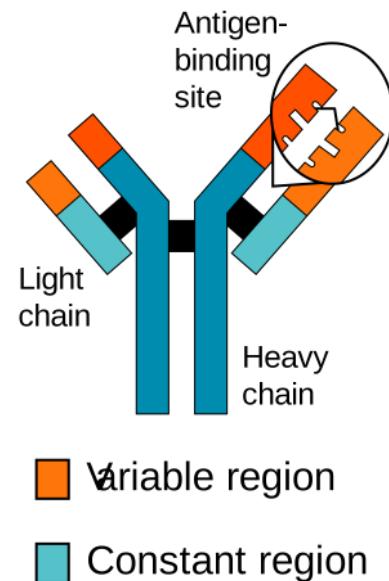


Figure 21.17: An antibody is made up of two heavy chains and two light chains. The unique variable region allows an antibody to recognize its matching antigen.¹⁶

functional locations and ability to deal with different antigens, as depicted in the table. The different suffixes of the antibody isotypes denote the different types of heavy chains the antibody contains, with each heavy chain class named alphabetically: α (alpha), γ (gamma), δ (delta), ϵ (epsilon), and μ (mu). This gives rise to IgA, IgG, IgD, IgE, and IgM, respectively.

Antibodies are heavy (~150 kDa) globular plasma proteins. The size of an antibody molecule is about 10 nm. They have sugar chains (glycans) added to conserved amino acid residues. In other words, antibodies are glycoproteins. The attached glycans are critically important to the structure and function of the antibody. Among other things the expressed glycans can modulate an antibody's affinity for its corresponding FcR(s).

The basic functional unit of each antibody is an immunoglobulin (Ig) monomer (containing only one Ig unit); secreted antibodies can also be dimeric with two Ig units as with IgA, tetrameric with four Ig units like teleost fish IgM, or pentameric with five Ig units, like mammalian IgM.

The variable parts of an antibody are its V regions, and the constant part is its C region.

21.8.2 The Immunological Memory

When B cells and T cells are activated and begin to replicate, some of their offspring become long-lived memory

cells. Throughout the lifetime of an animal, these memory cells remember each specific pathogen encountered and can mount a strong response if the pathogen is detected again. This is “adaptive” because it occurs during the lifetime of an individual as an adaptation to infection with that pathogen and prepares the immune system for future challenges. Immunological memory can be in the form of either passive short-term memory or active long-term memory.

It is likely that a multicomponent, adaptive immune system arose with the first vertebrates, as invertebrates do not generate lymphocytes or an antibody-based humoral response. Many species, however, utilize mechanisms that appear to be precursors of these aspects of vertebrate immunity. Immune systems appear even in the structurally most simple forms of life, with bacteria using a unique defense mechanism, called the restriction modification system to protect themselves from viral pathogens, called bacteriophages. Prokaryotes also possess acquired immunity, through a system that uses CRISPR sequences to retain fragments of the genomes of phage that they have come into contact with in the past, which allows them to block virus replication through a form of RNA interference. Prokaryotes also possess other defense mechanisms. Offensive elements of the immune systems are also present in unicellular eukaryotes, but studies of their roles in defense are few.

Pattern recognition receptors are proteins used by nearly all organisms to identify molecules associated with pathogens. Antimicrobial peptides called defensins are an evolutionarily conserved component of the innate immune response found in all animals and plants, and represent the main form of invertebrate systemic immunity. The complement system and phagocytic cells are also used by most forms of invertebrate life. Ribonucleases and the RNA interference pathway are conserved across all eukaryotes, and are thought to play a role in the immune response to viruses.

Unlike animals, plants lack phagocytic cells, but many plant immune responses involve systemic chemical signals that are sent through a plant. Individual plant cells respond to molecules associated with pathogens known as Pathogen-associated molecular patterns or PAMPs. When a part of a plant becomes infected, the plant produces a localized hypersensitive response, whereby cells at the site of infection undergo rapid apoptosis to prevent the spread of the disease to other parts of the plant. Systemic acquired resistance (SAR) is a type of defensive response used by plants that renders the entire plant resistant to a particular infectious agent. RNA silencing mechanisms are particularly important in this systemic response as they can block virus replication.

Evolution of the adaptive immune system occurred

in an ancestor of the jawed vertebrates. Many of the classical molecules of the adaptive immune system (e.g., immunoglobulins and T-cell receptors) exist only in jawed vertebrates. However, a distinct lymphocyte-derived molecule has been discovered in primitive jawless vertebrates, such as the lamprey and hagfish. These animals possess a large array of molecules called Variable lymphocyte receptors (VLRs) that, like the antigen receptors of jawed vertebrates, are produced from only a small number (one or two) of genes. These molecules are believed to bind pathogenic antigens in a similar way to antibodies, and with the same degree of specificity.

Chapter 22

Respiratory Systems

The respiratory system (also respiratory apparatus, ventilatory system) is a biological system consisting of specific organs and structures used for gas exchange in animals and plants. The anatomy and physiology that make this happen varies greatly, depending on the size of the organism, the environment in which it lives and its evolutionary history. In land animals the respiratory surface is internalized as linings of the lungs. Gas exchange in the lungs occurs in millions of small air sacs called alveoli in mammals and reptiles, but atria in birds. These microscopic air sacs have a very rich blood supply, thus bringing the air into close contact with the blood. These air sacs communicate with the external environment via a system of airways, or hollow tubes, of which the largest is the trachea, which branches in the middle of the chest into the two main bronchi. These enter the lungs where they branch into progressively narrower secondary and tertiary bronchi that branch into numerous smaller tubes, the bronchioles. In birds the bronchioles are termed parabronchi. It is the bronchioles, or parabronchi that generally open into the microscopic alveoli in mammals and atria in birds. Air has to be pumped from the environment into the alveoli or atria by the process of breathing which involves the muscles of respiration.

In most fish, and a number of other aquatic animals (both vertebrates and invertebrates) the respiratory system consists of gills, which are either partially or completely external organs, bathed in the watery environment. This water flows over the gills by a variety of active or passive means. Gas exchange takes place in the gills which consist of thin or very flat filaments and lamellae which expose a very large surface area of highly vascularized tissue to the water.

Other animals, such as insects, have respiratory systems with very simple anatomical features, and in amphibians even the skin plays a vital role in gas exchange. Plants also have respiratory systems but the directionality of gas exchange can be opposite to that in animals. The respiratory system in plants includes anatomical features such as

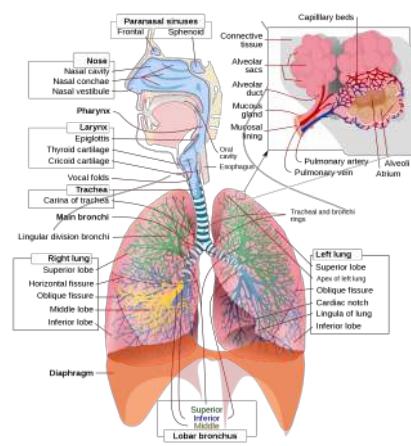


Figure 22.1: The respiratory system consists of the airways, the lungs, and the respiratory muscles that mediate the movement of air into and out of the body.¹

stomata, that are found in various parts of the plant.

22.1 The Mammalian Respiratory System

In humans and other mammals, the anatomy of a typical respiratory system is the respiratory tract. The tract is divided into an upper and a lower respiratory tract. The upper tract includes the nose, nasal cavities, sinuses, pharynx and the part of the larynx above the vocal folds. The lower tract (Fig. 2.) includes the lower part of the larynx, the trachea, bronchi, bronchioles and the alveoli.

The branching airways of the lower tract are often described as the respiratory tree or tracheobronchial tree. The intervals between successive branch points along the various branches of "tree" are often referred to as branching "generations", of which there are, in the adult human about 23. The earlier generations (approximately generations 0–16), consisting of the trachea and the bronchi, as well as the larger bronchioles which simply act

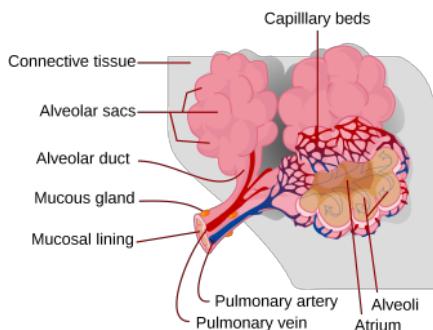


Figure 22.2: Alveoli and their capillary networks.²

as air conduits, bringing air to the respiratory bronchioles, alveolar ducts and alveoli (approximately generations 17–23), where gas exchange takes place. Bronchioles are defined as the small airways lacking any cartilagenous support.

The first bronchi to branch from the trachea are the right and left main bronchi. Second only in diameter to the trachea (1.8 cm), these bronchi (1 –1.4 cm in diameter) enter the lungs at each hilum, where they branch into narrower secondary bronchi known as lobar bronchi, and these branch into narrower tertiary bronchi known as segmental bronchi. Further divisions of the segmental bronchi (1 to 6 mm in diameter) are known as 4th order, 5th order, and 6th order segmental bronchi, or grouped together as subsegmental bronchi.

The alveoli are the dead end terminals of the “tree”, meaning that any air that enters them has to exit via the same route. A system such as this creates dead space, a volume of air (about 150 ml in the adult human) that fills the airways after exhalation and is breathed back into the alveoli before environmental air reaches them. At the end of inhalation the airways are filled with environmental air, which is exhaled without coming in contact with the gas exchanger.

The lungs expand and contract during the breathing cycle, drawing air in and out of the lungs. The volume of air moved in or out of the lungs under normal resting circumstances (the resting tidal volume of about 500 ml), and volumes moved during maximally forced inhalation and maximally forced exhalation are measured in humans by spirometry.

Not all the air in the lungs can be expelled during maximally forced exhalation. This is the residual volume of about 1.0–1.5 liters which cannot be measured by spirometry. Volumes that include the residual volume (i.e. functional residual capacity of about 2.5–3.0 liters, and total lung capacity of about 6 liters) can therefore also not be measured by spirometry. Their measurement requires spe-

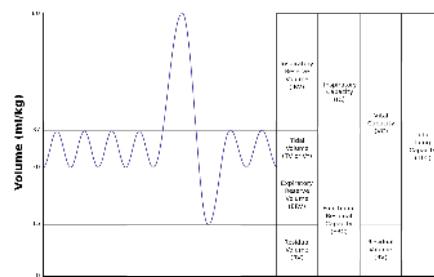


Figure 22.3: Output of a ‘spirometer’. Upward movement of the graph, read from the left, indicates the intake of air; downward movements represent exhalation.³

cial techniques.

The rates at which air is breathed in or out, either through the mouth or nose, or into or out of the alveoli are tabulated below, together with how they are calculated. The number of breath cycles per minute is known as the respiratory rate.

In mammals, inhalation at rest is primarily due to the contraction of the diaphragm. This is an upwardly domed sheet of muscle that separates the thoracic cavity from the abdominal cavity. When it contracts the sheet flattens, (i.e. moves downwards as shown in Fig. 7) increasing the volume of the thoracic cavity. The contracting diaphragm pushes the abdominal organs downwards. But because the pelvic floor prevents the lowermost abdominal organs moving in that direction, the pliable abdominal contents cause the belly to bulge outwards to the front and sides, because the relaxed abdominal muscles do not resist this movement (Fig. 7). This entirely passive bulging (and shrinking during exhalation) of the abdomen during normal breathing is sometimes referred to as “abdominal breathing”, although it is, in fact, “diaphragmatic breathing”, which is not visible on the outside of the body. Mammals only use their abdominal muscles only during forceful exhalation (see Fig. 8, and discussion below). Never during any form of inhalation.

The enlargement of the thoracic cavity’s vertical dimension by the contraction of the diaphragm, and its two horizontal dimensions by the lifting of the front and sides of the ribs, causes the intrathoracic pressure to fall. The lungs’ interiors are open to the outside air, and being elastic, therefore expand to fill the increased space. The inflow of air into the lungs occurs via the respiratory airways. In health, these airways begin with the nose. It is possible to begin with the mouth, which is the backup breathing system. However, chronic mouth breathing leads to, or is a sign of, illness. They end in the microscopic dead-end sacs called alveoli) are always open, though the diameters of the various sections can be changed by the sympathetic and parasympathetic nervous systems. The alveolar air pres-

sure is therefore always close to atmospheric air pressure (about 100 kPa at sea level) at rest, with the pressure gradients that cause air to move in and out of the lungs during breathing rarely exceeding 2–3 kPa.

During exhalation the diaphragm and intercostal muscles relax. This returns the chest and abdomen to a position determined by their anatomical elasticity. This is the “resting mid-position” of the thorax and abdomen when the lungs contain their functional residual capacity of air (the light blue area in the right hand illustration of Fig. 7), which in the adult human has a volume of about 2.5–3.0 liters. Resting exhalation lasts about twice as long as inhalation because the diaphragm relaxes passively more gently than it contracts actively during inhalation.

The volume of air that moves in or out (at the nose or mouth) during a single breathing cycle is called the tidal volume. In a resting adult human it is about 500 ml per breath. At the end of exhalation the airways contain about 150 ml of alveolar air which is the first air that is breathed back into the alveoli during inhalation. This volume air that is breathed out of the alveoli and back in again is known as dead space ventilation, which has the consequence that of the 500 ml breathed into the alveoli with each breath only 350 ml ($500\text{ ml} - 150\text{ ml} = 350\text{ ml}$) is fresh warm and moistened air. Since this 350 ml of fresh air is thoroughly mixed and diluted by the air that remains in the alveoli after normal exhalation (i.e. the functional residual capacity of about 2.5–3.0 liters), it is clear that the composition of the alveolar air changes very little during the breathing cycle (see Fig. 9). The oxygen tension (or partial pressure) remains close to 13–14 kPa (about 100 mm Hg), and that of carbon dioxide very close to 5.3 kPa (or 40 mm Hg). This contrasts with composition of the dry outside air at sea level, where the partial pressure of oxygen is 21 kPa (or 160 mm Hg) and that of carbon dioxide 0.04 kPa (or 0.3 mmHg).

During heavy breathing (hyperpnea), as, for instance, during exercise, inhalation is brought about by a more powerful and greater excursion of the contracting diaphragm than at rest. In addition the “accessory muscles of inhalation” exaggerate the actions of the intercostal muscles (Fig. 8). These accessory muscles of inhalation are muscles that extend from the cervical vertebrae and base of the skull to the upper ribs and sternum, sometimes through an intermediary attachment to the clavicles. When they contract the rib cage’s internal volume is increased to a far greater extent than can be achieved by contraction of the intercostal muscles alone. Seen from outside the body the lifting of the clavicles during strenuous or labored inhalation is sometimes called clavicular breathing, seen especially during asthma attacks and in people with chronic obstructive pulmonary disease.

During heavy breathing, exhalation is caused by relaxation of all the muscles of inhalation. But now, the abdominal muscles, instead of remaining relaxed (as they do at rest), contract forcibly pulling the lower edges of the rib cage downwards (front and sides). This not only drastically decreases the size of the rib cage, but also pushes the abdominal organs upwards against the diaphragm which consequently bulges deeply into the thorax. The end-exhalatory lung volume is now well below the resting mid-position and contains far less air than the resting “functional residual capacity”. However, in a normal mammal, the lungs cannot be emptied completely. In an adult human there is always still at least 1 liter of residual air left in the lungs after maximum exhalation.

The automatic rhythmical breathing in and out, can be interrupted by coughing, sneezing (forms of very forceful exhalation), by the expression of a wide range of emotions (laughing, sighing, crying out in pain, exasperated intakes of breath) and by such voluntary acts as speech, singing, whistling and the playing of wind instruments. All of these actions rely on the muscles described above, and their effects on the movement of air in and out of the lungs.

Although not a form of breathing, the Valsalva maneuver involves the respiratory muscles. It is, in fact, a very forceful exhalatory effort against a tightly closed glottis, so that no air can escape from the lungs. Instead abdominal contents are evacuated in the opposite direction, through orifices in the pelvic floor. The abdominal muscles contract very powerfully, causing the pressure inside the abdomen and thorax to rise to extremely high levels. The Valsalva maneuver can be carried out voluntarily, but is more generally a reflex elicited when attempting to empty the abdomen during, for instance, difficult defecation, or during childbirth. Breathing ceases during this maneuver.

22.2 Gas Exchange

The primary purpose of the respiratory system is the equilibration of the partial pressures of the respiratory gases in the alveolar air with those in the pulmonary capillary blood. This process occurs by simple diffusion, across a very thin membrane (known as the blood-air barrier), which forms the walls of the pulmonary alveoli (Figure 22.4). It consisting of the alveolar epithelial cells, their basement membranes and the endothelial cells of the alveolar capillaries (Figure 22.4). This blood gas barrier is extremely thin (in humans, on average, 2.2 μm thick). It is folded into about 300 million small air sacs called alveoli (each between 75 and 300 μm in diameter) branching off from the respiratory bronchioles in the lungs, thus providing an extremely large surface area (approximately 145 m²) for gas exchange to occur.

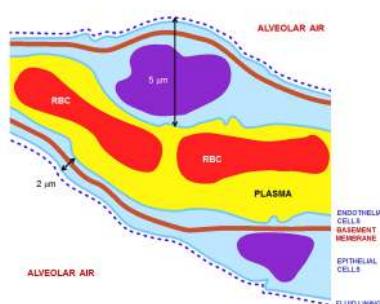


Figure 22.4: (ref:alve)

(ref:alve) A histological cross-section through an alveolar wall showing the layers through which the gases have to move between the blood plasma and the alveolar air. The dark blue objects are the nuclei of the capillary endothelial and alveolar type I epithelial cells (or type 1 pneumocytes). The two red objects labeled "RBC" are red blood cells in the pulmonary capillary blood.⁴

The air contained within the alveoli has a semi-permanent volume of about 2.5–3.0 liters which completely surrounds the alveolar capillary blood. This ensures that equilibration of the partial pressures of the gases in the two compartments is very efficient and occurs very quickly. The blood leaving the alveolar capillaries and is eventually distributed throughout the body therefore has a partial pressure of oxygen of 13–14 kPa (100 mmHg), and a partial pressure of carbon dioxide of 5.3 kPa (40 mmHg) (i.e. the same as the oxygen and carbon dioxide gas tensions as in the alveoli). As mentioned in the section above, the corresponding partial pressures of oxygen and carbon dioxide in the ambient (dry) air at sea level are 21 kPa (160 mmHg) and 0.04 kPa (0.3 mmHg) respectively.

This marked difference between the composition of the alveolar air and that of the ambient air can be maintained because the functional residual capacity is contained in dead-end sacs connected to the outside air by fairly narrow and relatively long tubes (the airways: nose, pharynx, larynx, trachea, bronchi and their branches down to the bronchioles), through which the air has to be breathed both in and out (i.e. there is no unidirectional through-flow as there is in the bird lung). This typical mammalian anatomy combined with the fact that the lungs are not emptied and re-inflated with each breath (leaving a substantial volume of air, of about 2.5–3.0 liters, in the alveoli after exhalation), ensures that the composition of the alveolar air is only minimally disturbed when the 350 ml of fresh air



Figure 22.5: A highly diagrammatic illustration of the process of gas exchange in the mammalian lungs, emphasizing the differences between the gas compositions of the ambient air, the alveolar air (light blue) with which the pulmonary capillary blood equilibrates, and the blood gas tensions in the pulmonary arterial (blue blood entering the lung on the left) and venous blood (red blood leaving the lung on the right). All the gas tensions are in kPa. To convert to mm Hg, multiply by 7.5.⁵

is mixed into it with each inhalation. Thus the animal is provided with a very special "portable atmosphere", whose composition differs significantly from the present-day ambient air. It is this portable atmosphere (the functional residual capacity) to which the blood and therefore the body tissues are exposed – not to the outside air.

The resulting arterial partial pressures of oxygen and carbon dioxide are homeostatically controlled. A rise in the arterial partial pressure of CO₂) and, to a lesser extent, a fall in the arterial partial pressure of O₂), will reflexly cause deeper and faster breathing till the blood gas tensions in the lungs, and therefore the arterial blood, return to normal. The converse happens when the carbon dioxide tension falls, or, again to a lesser extent, the oxygen tension rises: the rate and depth of breathing are reduced till blood gas normality is restored.

Since the blood arriving in the alveolar capillaries has a partial pressure of O₂) of, on average, 6 kPa (45 mmHg), while the pressure in the alveolar air is 13–14 kPa (100 mmHg), there will be a net diffusion of oxygen into the capillary blood, changing the composition of the 3 liters of alveolar air slightly. Similarly, since the blood arriving in the alveolar capillaries has a partial pressure of CO₂) of also about 6 kPa (45 mmHg), whereas that of the alveolar air is 5.3 kPa (40 mmHg), there is a net movement of carbon dioxide out of the capillaries into the alveoli. The changes brought about by these net flows of individual gases into and out of the alveolar air necessitate the replacement of about 15% of the alveolar air with ambient air every 5 seconds or so. This is very tightly controlled by the monitoring of the arterial blood gases (which accurately reflect composition of the alveolar air) by the aortic and carotid bodies, as well as by the blood gas and pH sensor on the anterior surface of the medulla oblongata in the brain. There are also oxygen and carbon dioxide sensors in the lungs, but they primarily determine the diameters of the bronchioles and pulmonary capillaries, and are therefore responsible for directing the flow of air and blood to different parts of the lungs.

⁴https://upload.wikimedia.org/wikipedia/commons/c/c8/Alveolar_wall.jpg

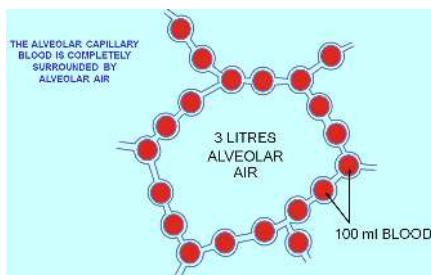


Figure 22.6: A diagrammatic histological cross-section through a portion of lung tissue showing a normally inflated alveolus (at the end of a normal exhalation), and its walls containing the pulmonary capillaries (shown in cross-section). This illustrates how the pulmonary capillary blood is completely surrounded by alveolar air. In a normal human lung all the alveoli together contain about 3 liters of alveolar air. All the pulmonary capillaries contain about 100 ml blood.⁶

It is only as a result of accurately maintaining the composition of the 3 liters of alveolar air that with each breath some carbon dioxide is discharged into the atmosphere and some oxygen is taken up from the outside air. If more carbon dioxide than usual has been lost by a short period of hyperventilation, respiration will be slowed down or halted until the alveolar partial pressure of carbon dioxide has returned to 5.3 kPa (40 mmHg). It is therefore strictly speaking untrue that the primary function of the respiratory system is to rid the body of carbon dioxide “waste”. The carbon dioxide that is breathed out with each breath could probably be more correctly be seen as a byproduct of the body’s extracellular fluid carbon dioxide and pH homeostats

If these homeostats are compromised, then a respiratory acidosis, or a respiratory alkalosis will occur. In the long run these can be compensated by renal adjustments to the H⁺ and HCO₃⁻ concentrations in the plasma; but since this takes time, the hyperventilation syndrome can, for instance, occur when agitation or anxiety cause a person to breathe fast and deeply thus causing a distressing respiratory alkalosis through the blowing off of too much CO₂ from the blood into the outside air.

Oxygen has a very low solubility in water, and is therefore carried in the blood loosely combined with hemoglobin. The oxygen is held on the hemoglobin by four ferrous iron-containing heme groups per hemoglobin molecule. When all the heme groups carry one O₂ molecule each the blood is said to be “saturated” with oxygen, and no further increase in the partial pressure of oxygen will meaningfully increase the oxygen concentration of the blood. Most of the carbon dioxide in the blood is carried as bicarbonate ions (HCO₃⁻) in the plasma.

However the conversion of dissolved CO₂ into HCO₃⁻ (through the addition of water) is too slow for the rate at which the blood circulates through the tissues on the one hand, and through alveolar capillaries on the other. The reaction is therefore catalyzed by carbonic anhydrase, an enzyme inside the red blood cells. The reaction can go in both directions depending on the prevailing partial pressure of CO₂). A small amount of carbon dioxide is carried on the protein portion of the hemoglobin molecules as carbamino groups. The total concentration of carbon dioxide (in the form of bicarbonate ions, dissolved CO₂, and carbamino groups) in arterial blood (i.e. after it has equilibrated with the alveolar air) is about 26 mM (or 58 ml/100 ml), compared to the concentration of oxygen in saturated arterial blood of about 9 mM (or 20 ml/100 ml blood).

22.3 Control of Ventilation

Ventilation of the lungs in mammals occurs via the respiratory centers in the medulla oblongata and the pons of the brainstem. These areas form a series of neural pathways which receive information about the partial pressures of oxygen and carbon dioxide in the arterial blood. This information determines the average rate of ventilation of the alveoli of the lungs, to keep these pressures constant. The respiratory center does so via motor nerves which activate the diaphragm and other muscles of respiration.

The breathing rate increases when the partial pressure of carbon dioxide in the blood increases. This is detected by central blood gas chemoreceptors on the anterior surface of the medulla oblongata. The aortic and carotid bodies, are the peripheral blood gas chemoreceptors which are particularly sensitive to the arterial partial pressure of O₂ (though they also respond, but less strongly, to the partial pressure of CO₂). At sea level, under normal circumstances, the breathing rate and depth, is determined primarily by the arterial partial pressure of carbon dioxide rather than by the arterial partial pressure of oxygen, which is allowed to vary within a fairly wide range before the respiratory centers in the medulla oblongata and pons respond to it to change the rate and depth of breathing.

Exercise increases the breathing rate due to the extra carbon dioxide produced by the enhanced metabolism of the exercising muscles. In addition passive movements of the limbs also reflexively produce an increase in the breathing rate.

Information received from stretch receptors in the lungs limits tidal volume (the depth of inhalation and exhalation).

The respiratory system of birds differs significantly

from that found in mammals. Firstly, they have rigid lungs which do not expand and contract during the breathing cycle. Instead an extensive system of air sacs (Fig. 15) distributed throughout their bodies act as the bellows drawing environmental air into the sacs, and expelling the spent air after it has passed through the lungs. Birds also do not have diaphragms or pleural cavities.

Bird lungs are smaller than those in mammals of comparable size, but the air sacs account for 15% of the total body volume, compared to the 7% devoted to the alveoli which act as the bellows in mammals.

Inhalation and exhalation are brought about by alternately increasing and decreasing the volume of the entire thoraco-abdominal cavity (or coelom) using both their abdominal and costal muscles. During inhalation the muscles attached to the vertebral ribs (Fig. 17) contract angling them forwards and outwards. This pushes the sternal ribs, to which they are attached at almost right angles, downwards and forwards, taking the sternum (with its prominent keel) in the same direction (Fig. 17). This increases both the vertical and transverse diameters of thoracic portion of the trunk. The forward and downward movement of, particularly, the posterior end of the sternum pulls the abdominal wall downwards, increasing the volume of that region of the trunk as well. The increase in volume of the entire trunk cavity reduces the air pressure in all the thoraco-abdominal air sacs, causing them to fill with air as described below.

During exhalation the external oblique muscle which is attached to the sternum and vertebral ribs anteriorly, and to the pelvis (pubis and ilium in Fig. 17) posteriorly (forming part of the abdominal wall) reverses the inhalatory movement, while compressing the abdominal contents, thus increasing the pressure in all the air sacs. Air is therefore expelled from the respiratory system in the act of exhalation.

During inhalation air enters the trachea via the nostrils and mouth, and continues to just beyond the syrinx at which point the trachea branches into two primary bronchi, going to the two lungs (Fig. 16). The primary bronchi enter the lungs to become the intrapulmonary bronchi, which give off a set of parallel branches called ventrobronchi and, a little further on, an equivalent set of dorsobronchi (Fig. 16). The ends of the intrapulmonary bronchi discharge air into the posterior air sacs at the caudal end of the bird. Each pair of dorso-ventrobronchi is connected by a large number of parallel microscopic air capillaries (or parabronchi) where gas exchange occurs (Fig. 16). As the bird inhales, tracheal air flows through the intrapulmonary bronchi into the posterior air sacs, as well as into the dorsobronchi, but not into the ventrobronchi (Fig. 18). This is due to the bronchial

architecture which directs the inhaled air away from the openings of the ventrobronchi, into the continuation of the intrapulmonary bronchus towards the dorsobronchi and posterior air sacs. From the dorsobronchi the inhaled air flows through the parabronchi (and therefore the gas exchanger) to the ventrobronchi from where the air can only escape into the expanding anterior air sacs. So, during inhalation, both the posterior and anterior air sacs expand, the posterior air sacs filling with fresh inhaled air, while the anterior air sacs fill with "spent" (oxygen-poor) air that has just passed through the lungs.

During exhalation the pressure in the posterior air sacs (which were filled with fresh air during inhalation) increases due to the contraction of the oblique muscle described above. The aerodynamics of the interconnecting openings from the posterior air sacs to the dorsobronchi and intrapulmonary bronchi ensures that the air leaves these sacs in the direction of the lungs (via the dorsobronchi), rather than returning down the intrapulmonary bronchi (Fig. 18). From the dorsobronchi the fresh air from the posterior air sacs flows through the parabronchi (in the same direction as occurred during inhalation) into ventrobronchi. The air passages connecting the ventrobronchi and anterior air sacs to the intrapulmonary bronchi direct the "spent", oxygen poor air from these two organs to the trachea from where it escapes to the exterior. Oxygenated air therefore flows constantly (during the entire breathing cycle) in a single direction through the parabronchi.

The blood flow through the bird lung is at right angles to the flow of air through the parabronchi, forming a cross-current flow exchange system (Fig. 19). The partial pressure of oxygen in the parabronchi declines along their lengths as O_2 diffuses into the blood. The blood capillaries leaving the exchanger near the entrance of airflow take up more O_2 than do the capillaries leaving near the exit end of the parabronchi. When the contents of all capillaries mix, the final partial pressure of oxygen of the mixed pulmonary venous blood is higher than that of the exhaled air, but is nevertheless less than half that of the inhaled air, thus achieving roughly the same systemic arterial blood partial pressure of oxygen as mammals do with their bellows-type lungs.

The trachea is an area of dead space: the oxygen-poor air it contains at the end of exhalation is the first air to re-enter the posterior air sacs and lungs. In comparison to the mammalian respiratory tract, the dead space volume in a bird is, on average, 4.5 times greater than it is in mammals of the same size. Birds with long necks will inevitably have long tracheae, and must therefore take deeper breaths than mammals do to make allowances for their greater dead space volumes. In some birds (e.g. the

whooper swan, *Cygnus cygnus*, the white spoonbill, *Platalea leucorodia*, the whooping crane, *Grus americana*, and the helmeted curassow, *Pauxi pauxi*) the trachea, which some cranes can be 1.5 m long, is coiled back and forth within the body, drastically increasing the dead space ventilation. The purpose of this extraordinary feature is unknown.

22.4 The Respiratory System of Birds

Due to the high metabolic rate required for flight, birds have a high oxygen demand. Their highly effective respiratory system helps them meet that demand.

Although birds have lungs, theirs are fairly rigid structures that do not expand and contract as they do in mammals, reptiles and many amphibians. Instead, the structures that act as the bellows that ventilate the lungs are the air sacs, which are distributed throughout much of the birds' bodies. The air sacs move air unidirectionally through the parabronchi of the rigid lungs. Although bird lungs are smaller than those of mammals of comparable size, the air sacs account for 15% of the total body volume, whereas in mammals, the alveoli, which act as the bellows, constitute only 7% of the total body volume. The walls of the air sacs do not have a good blood supply and so do not play a direct role in gas exchange.

Birds lack a diaphragm, and therefore use their intercostal and abdominal muscles to expand and contract their entire thoraco-abdominal cavities, thus rhythmically changing the volumes of all their air sacs in unison (illustration on the right). The active phase of respiration in birds is exhalation, requiring contraction of their muscles of respiration. Relaxation of these muscles causes inhalation.

Three distinct sets of organs perform respiration — the anterior air sacs (interclavicular, cervicals, and anterior thoracics), the lungs, and the posterior air sacs (posterior thoracics and abdominals). Typically there are nine air sacs within the system; however, that number can range between seven and twelve, depending on the species of bird. Passerines possess seven air sacs, as the clavicular air sacs may interconnect or be fused with the anterior thoracic sacs.

During inhalation, environmental air initially enters the bird through the nostrils from where it is heated, humidified, and filtered in the nasal passages and upper parts of the trachea. From there, the air enters the lower trachea and continues to just beyond the syrinx, at which point the trachea branches into two primary bronchi, going to the two lungs. The primary bronchi enter the lungs to become the intrapulmonary bronchi, which give

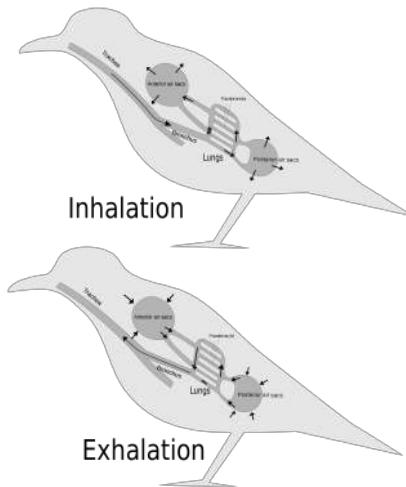


Figure 22.7: The respiratory system of birds.⁷ On inhalation, air travels to air sacs near the back of a bird. The air then passes through the lungs to air sacs near the front of the bird, from where the air is exhaled.

off a set of parallel branches called venterbronchi and, a little further on, an equivalent set of dorsobronchi. The ends of the intrapulmonary bronchi discharge air into the posterior air sacs at the caudal end of the bird. Each pair of dorso-venterbronchi is connected by a large number of parallel microscopic air capillaries (or parabronchi) where gas exchange occurs. As the bird inhales, tracheal air flows through the intrapulmonary bronchi into the posterior air sacs, as well as into the dorsobronchi (but not into the venterbronchi whose openings into the intrapulmonary bronchi were previously believed to be tightly closed during inhalation). However, more recent studies have shown that the aerodynamics of the bronchial architecture directs the inhaled air away from the openings of the venterbronchi, into the continuation of the intrapulmonary bronchus towards the dorsobronchi and posterior air sacs). From the dorsobronchi the air flows through the parabronchi (and therefore the gas exchanger) to the venterbronchi from where the air can only escape into the expanding anterior air sacs. So, during inhalation, both the posterior and anterior air sacs expand, the posterior air sacs filling with fresh inhaled air, while the anterior air sacs fill with "spent" (oxygen-poor) air that has just passed through the lungs.

During exhalation the intrapulmonary bronchi were believed to be tightly constricted between the region where the venterbronchi branch off and the region where the dorsobronchi branch off. But it is now believed that more intricate aerodynamic features have the same effect. The contracting posterior air sacs can therefore only empty into the dorsobronchi. From there the fresh air from the posterior

air sacs flows through the parabronchi (in the same direction as occurred during inhalation) into venterbronchi. The air passages connecting the venterbronchi and anterior air sacs to the intrapulmonary bronchi open up during exhalation, thus allowing oxygen-poor air from these two organs to escape via the trachea to the exterior. Oxygenated air therefore flows constantly (during the entire breathing cycle) in a single direction through the parabronchi.

The cross-current respiratory gas exchanger in the lungs of birds. Air is forced from the air sacs unidirectionally (from right to left in the diagram) through the parabronchi. The pulmonary capillaries surround the parabronchi in the manner shown (blood flowing from below the parabronchus to above it in the diagram). Blood or air with a high oxygen content is shown in red; oxygen-poor air or blood is shown in various shades of purple-blue. The blood flow through the bird lung is at right angles to the flow of air through the parabronchi, forming a cross-current flow exchange system (see illustration on the left). The partial pressure of oxygen in the parabronchi declines along their lengths as O₂ diffuses into the blood. The blood capillaries leaving the exchanger near the entrance of airflow take up more O₂ than do the capillaries leaving near the exit end of the parabronchi. When the contents of all capillaries mix, the final partial pressure of oxygen of the mixed pulmonary venous blood is higher than that of the exhaled air, but is nevertheless less than half that of the inhaled air, thus achieving roughly the same systemic arterial blood partial pressure of oxygen as mammals do with their bellows-type lungs.

The trachea is an area of dead space: the oxygen-poor air it contains at the end of exhalation is the first air to re-enter the posterior air sacs and lungs. In comparison to the mammalian respiratory tract, the dead space volume in a bird is, on average, 4.5 times greater than it is in mammals of the same size. Birds with long necks will inevitably have long tracheae, and must therefore take deeper breaths than mammals do to make allowances for their greater dead space volumes. In some birds (e.g. the whooper swan, *Cygnus cygnus*, the white spoonbill, *Platalea leucorodia*, the whooping crane, *Grus americana*, and the helmeted curassow, *Pauxi pauxi*) the trachea, which some cranes can be 1.5 m long, is coiled back and forth within the body, drastically increasing the dead space ventilation. The purpose of this extraordinary feature is unknown.

Air passes unidirectionally through the lungs during both exhalation and inspiration, causing, except for the oxygen-poor dead space air left in the trachea after exhalation and breathed in at the beginning of inhalation, little to no mixing of new oxygen-rich air with spent oxygen-poor

air (as occurs in mammalian lungs), changing only (from oxygen-rich to oxygen-poor) as it moves (unidirectionally) through the parabronchi.

Avian lungs do not have alveoli as mammalian lungs do. Instead they contain millions of narrow passages known as parabronchi, connecting the dorsobronchi to the venterbronchi at either ends of the lungs. Air flows anteriorly (caudal to cranial) through the parallel parabronchi. These parabronchi have honeycombed walls. The cells of the honeycomb are dead-end air vesicles, called atria, which project radially from the parabronchi. The atria are the site of gas exchange by simple diffusion. The blood flow around the parabronchi (and their atria), forms a cross-current gas exchanger (see diagram on the left).

All species of birds with the exception of the penguin, have a small region of their lungs devoted to "neopulmonic parabronchi". This unorganized network of microscopic tubes branches off from the posterior air sacs, and open haphazardly into both the dorso- and venterbronchi, as well as directly into the intrapulmonary bronchi. Unlike the parabronchi, in which the air moves unidirectionally, the air flow in the neopulmonic parabronchi is bidirectional. The neopulmonic parabronchi never make up more than 25% of the total gas exchange surface of birds.

The syrinx is the sound-producing vocal organ of birds, located at the base of a bird's trachea. As with the mammalian larynx, sound is produced by the vibration of air flowing across the organ. The syrinx enables some species of birds to produce extremely complex vocalizations, even mimicking human speech. In some songbirds, the syrinx can produce more than one sound at a time.

22.5 The Respiratory System of Reptiles

The anatomical structure of the lungs is less complex in reptiles than in mammals, with reptiles lacking the very extensive airway tree structure found in mammalian lungs. Gas exchange in reptiles still occurs in alveoli however. Reptiles do not possess a diaphragm. Thus, breathing occurs via a change in the volume of the body cavity which is controlled by contraction of intercostal muscles in all reptiles except turtles. In turtles, contraction of specific pairs of flank muscles governs inhalation and exhalation.

22.6 The Respiratory System of Amphibians

Both the lungs and the skin serve as respiratory organs in amphibians. The ventilation of the lungs in amphibians



Figure 22.8: The axolotl (*Ambystoma mexicanum*) retains its larval form with gills into adulthood.⁸

relies on positive pressure ventilation. Muscles lower the floor of the oral cavity, enlarging it and drawing in air through the nostrils into the oral cavity. With the nostrils and mouth closed, the floor of the oral cavity is then pushed up, which forces air down the trachea into the lungs. The skin of these animals is highly vascularized and moist, with moisture maintained via secretion of mucus from specialised cells, and is involved in cutaneous respiration. While the lungs are of primary organs for gas exchange between the blood and the environmental air (when out of the water), the skin's unique properties aid rapid gas exchange when amphibians are submerged in oxygen-rich water. Some amphibians have gills, either in the early stages of their development (e.g. tadpoles of frogs), while others retain them into adulthood (e.g. some salamanders).

22.7 The Respiratory System of Fish

Oxygen is poorly soluble in water. Fully aerated fresh water therefore contains only 8–10 ml O₂/liter compared to the O₂ concentration of 210 ml/liter in the air at sea level. Furthermore, the coefficient of diffusion (i.e. the rate at which a substance diffuses from a region of high concentration to one of low concentration, under standard conditions) of the respiratory gases is typically 10,000 faster in air than in water. Thus oxygen, for instance, has a diffusion coefficient of 17.6 mm²/s in air, but only 0.0021 mm²/s in water. The corresponding values for carbon dioxide are 16 mm²/s in air and 0.0016 mm²/s in water. This means that when oxygen is taken up from the water in contact with a gas exchanger, it is replaced considerably more slowly by the oxygen from the oxygen-rich regions small distances away from the exchanger than would have occurred in air. Fish have developed gills deal with these problems. Gills are specialized organs containing filaments, which further divide into lamellae. The lamellae contain a dense thin walled capillary network that exposes a large gas exchange surface area to the very large volumes of water passing over them.

Gills use a countercurrent exchange system that in-

creases the efficiency of oxygen-uptake from the water. Fresh oxygenated water taken in through the mouth is uninterrupted "pumped" through the gills in one direction, while the blood in the lamellae flows in the opposite direction, creating the countercurrent blood and water flow (Fig. 22), on which the fish's survival depends.

Water is drawn in through the mouth by closing the operculum (gill cover), and enlarging the mouth cavity (Fig. 23). Simultaneously the gill chambers enlarge, producing a lower pressure there than in the mouth causing water to flow over the gills. The mouth cavity then contracts inducing the closure of the passive oral valves, thereby preventing the back-flow of water from the mouth (Fig. 23). The water in the mouth is, instead, forced over the gills, while the gill chambers contract emptying the water they contain through the opercular openings (Fig. 23). Back-flow into the gill chamber during the inhalatory phase is prevented by a membrane along the ventroposterior border of the operculum (diagram on the left in Fig. 23). Thus the mouth cavity and gill chambers act alternately as suction pump and pressure pump to maintain a steady flow of water over the gills in one direction. Since the blood in the lamellar capillaries flows in the opposite direction to that of the water, the consequent countercurrent flow of blood and water maintains steep concentration gradients for oxygen and carbon dioxide along the entire length of each capillary (lower diagram in Fig. 22). Oxygen is, therefore, able to continually diffuse down its gradient into the blood, and the carbon dioxide down its gradient into the water. Although countercurrent exchange systems theoretically allow an almost complete transfer of a respiratory gas from one side of the exchanger to the other, in fish less than 80% of the oxygen in the water flowing over the gills is generally transferred to the blood.

In certain active pelagic sharks, water passes through the mouth and over the gills while they are moving, in a process known as "ram ventilation". While at rest, most sharks pump water over their gills, as most bony fish do, to ensure that oxygenated water continues to flow over their gills. But a small number of species have lost the ability to pump water through their gills and must swim without rest. These species are obligate ram ventilators and would presumably asphyxiate if unable to move. Obligate ram ventilation is also true of some pelagic bony fish species.

There are a few fish that can obtain oxygen for brief periods of time from air swallowed from above the surface of the water. Thus Lungfish possess one or two lungs, and the labyrinth fish have developed a special "labyrinth organ", which characterizes this suborder of fish. The labyrinth organ is a much-folded suprabranchial accessory breathing organ. It is formed by a vascularized expansion of the epibranchial bone of the first gill arch, and is used for res-

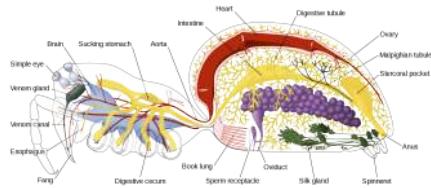


Figure 22.9: The book lungs of a spider.⁹

piration in air.

This organ allows labyrinth fish to take in oxygen directly from the air, instead of taking it from the water in which they reside through use of gills. The labyrinth organ helps the oxygen in the inhaled air to be absorbed into the bloodstream. As a result, labyrinth fish can survive for a short period of time out of water, as they can inhale the air around them, provided they stay moist.

Labyrinth fish are not born with functional labyrinth organs. The development of the organ is gradual and most juvenile labyrinth fish breathe entirely with their gills and develop the labyrinth organs when they grow older.

22.8 The Respiratory System of Arthropods

Some species of crab use a respiratory organ called a branchiostegal lung. Its gill-like structure increases the surface area for gas exchange which is more suited to taking oxygen from the air than from water. Some of the smallest spiders and mites can breathe simply by exchanging gas through the surface of the body. Larger spiders, scorpions and other arthropods use a primitive book lung.

Insects

Most insects breath passively through their spiracles (special openings in the exoskeleton) and the air reaches every part of the body by means of a series of smaller and smaller tubes called 'trachea' when their diameters are relatively large, and 'tracheoles' when their diameters are very small. The tracheoles make contact with individual cells throughout the body. They are partially filled with fluid, which can be withdrawn from the individual tracheoles when the tissues, such as muscles, are active and have a high demand for oxygen, bringing the air closer to the active cells. This is probably brought about by the buildup of lactic acid in the active muscles causing an osmotic gradient, moving the water out of the tracheoles and into the active cells. Diffusion of gases is effective over small distances but not over larger ones, this is one of the reasons insects are all relatively small. Insects which do not have spiracles and trachea, such as some Collembola, breathe

directly through their skins, also by diffusion of gases.

The number of spiracles an insect has is variable between species, however, they always come in pairs, one on each side of the body, and usually one pair per segment. Some of the Diplura have eleven, with four pairs on the thorax, but in most of the ancient forms of insects, such as Dragonflies and Grasshoppers there are two thoracic and eight abdominal spiracles. However, in most of the remaining insects, there are fewer. It is at the level of the tracheoles that oxygen is delivered to the cells for respiration.

Insects were once believed to exchange gases with the environment continuously by the simple diffusion of gases into the tracheal system. More recently, however, large variation in insect ventilatory patterns has been documented and insect respiration appears to be highly variable. Some small insects do not demonstrate continuous respiratory movements and may lack muscular control of the spiracles. Others, however, utilize muscular contraction of the abdomen along with coordinated spiracle contraction and relaxation to generate cyclical gas exchange patterns and to reduce water loss into the atmosphere. The most extreme form of these patterns is termed discontinuous gas exchange cycles.

22.9 The Respiratory System of Molluscs

Molluscs generally possess gills that allow gas exchange between the aqueous environment and their circulatory systems. These animals also possess a heart that pumps blood containing hemocyanin as its oxygen-capturing molecule. Hence, this respiratory system is similar to that of vertebrate fish. The respiratory system of gastropods can include either gills or a lung.

22.10 The Respiratory System of Plants

Plants use carbon dioxide gas in the process of photosynthesis, and exhale oxygen gas as waste. The chemical equation of photosynthesis is 6 CO_2 (carbon dioxide) and $6 \text{ H}_2\text{O}$ (water), which in the presence of sunlight makes $\text{C}_6\text{H}_{12}\text{O}_6$ (glucose) and 6 O_2 (oxygen). Photosynthesis uses electrons on the carbon atoms as the repository for the energy obtained from sunlight. Respiration is the opposite of photosynthesis. It reclaims the energy to power chemical reactions in cells. In so doing the carbon atoms and their electrons are combined with oxygen forming CO_2 which is easily removed from both the cells and

the organism. Plants use both processes, photosynthesis to capture the energy and oxidative metabolism to use it.

Plant respiration is limited by the process of diffusion. Plants take in carbon dioxide through holes, known as stomata, that can open and close on the undersides of their leaves and sometimes other parts of their anatomy. Most plants require some oxygen for catabolic processes (break-down reactions that release energy). But the quantity of O₂) used per hour is small as they are not involved in activities that require high rates of aerobic metabolism. Their requirement for air, however, is very high as they need CO₂) for photosynthesis, which constitutes only 0.04% of the environmental air. Thus, to make 1 g of glucose requires the removal of all the CO₂) from at least 18.7 liters of air at sea level. But inefficiencies in the photosynthetic process cause considerably greater volumes of air to be used.

Chapter 23

Digestive Systems

Digestion is the breakdown of large insoluble food molecules into small water-soluble food molecules so that they can be absorbed into the watery blood plasma. In certain organisms, these smaller substances are absorbed through the small intestine into the blood stream. Digestion is a form of catabolism that is often divided into two processes based on how food is broken down: mechanical and chemical digestion. The term mechanical digestion refers to the physical breakdown of large pieces of food into smaller pieces which can subsequently be accessed by digestive enzymes. In chemical digestion, enzymes break down food into the small molecules the body can use.

Digestive systems take many forms. There is a fundamental distinction between internal and external digestion. External digestion developed earlier in evolutionary history, and most fungi still rely on it. In this process, enzymes are secreted into the environment surrounding the organism, where they break down an organic material, and some of the products diffuse back to the organism. Animals have a tube (gastrointestinal tract) in which internal digestion occurs, which is more efficient because more of the broken down products can be captured, and the internal chemical environment can be more efficiently controlled.

Some organisms, including nearly all spiders, simply secrete biotoxins and digestive chemicals (e.g., enzymes) into the extracellular environment prior to ingestion of the consequent "soup". In others, once potential nutrients or food is inside the organism, digestion can be conducted to a vesicle or a sac-like structure, through a tube, or through several specialized organs aimed at making the absorption of nutrients more efficient.

In the human digestive system, food enters the mouth and mechanical digestion of the food starts by the action of mastication (chewing), a form of mechanical digestion, and the wetting contact of saliva. Saliva, a liquid secreted by the salivary glands, contains salivary amylase, an enzyme which starts the digestion of starch in the food; the saliva also contains mucus, which lubricates the food, and hydro-

gen carbonate, which provides the ideal conditions of pH (alkaline) for amylase to work. After undergoing mastication and starch digestion, the food will be in the form of a small, round slurry mass called a bolus. It will then travel down the esophagus and into the stomach by the action of peristalsis. Gastric juice in the stomach starts protein digestion. Gastric juice mainly contains hydrochloric acid and pepsin. In infants and toddlers gastric juice also contains rennin. As the first two chemicals may damage the stomach wall, mucus is secreted by the stomach, providing a slimy layer that acts as a shield against the damaging effects of the chemicals. At the same time protein digestion is occurring, mechanical mixing occurs by peristalsis, which is waves of muscular contractions that move along the stomach wall. This allows the mass of food to further mix with the digestive enzymes.

After some time (typically 1-2 hours in humans, 4-6 hours in dogs, 3-4 hours in house cats), the resulting thick liquid is called chyme. When the pyloric sphincter valve opens, chyme enters the duodenum where it mixes with digestive enzymes from the pancreas and bile juice from the liver and then passes through the small intestine, in which digestion continues. When the chyme is fully digested, it is absorbed into the blood. 95% of nutrient absorption occurs in the small intestine. Water and minerals are reabsorbed back into the blood in the colon (large intestine) where the pH is slightly acidic about 5.6 ~ 6.9. Some vitamins, such as biotin and vitamin K produced by bacteria in the colon are also absorbed into the blood in the colon. Waste material is eliminated from the rectum during defecation.

In most vertebrates, digestion is a multistage process in the digestive system, starting from ingestion of raw materials, most often other organisms. Ingestion usually involves some type of mechanical and chemical processing. Digestion is separated into four steps:

- Ingestion: placing food into the mouth (entry of food in the digestive system),
- Mechanical and chemical breakdown: mastication and the mixing of the resulting bolus with water,

acids, bile and enzymes in the stomach and intestine to break down complex molecules into simple structures,

- Absorption: of nutrients from the digestive system to the circulatory and lymphatic capillaries through osmosis, active transport, and diffusion, and
- Egestion (Excretion): Removal of undigested materials from the digestive tract through defecation.

Underlying the process is muscle movement throughout the system through swallowing and peristalsis. Each step in digestion requires energy, and thus imposes an “overhead charge” on the energy made available from absorbed substances. Differences in that overhead cost are important influences on lifestyle, behavior, and even physical structures. Examples may be seen in humans, who differ considerably from other hominids (lack of hair, smaller jaws and musculature, different dentition, length of intestines, cooking, etc.).

The major part of digestion takes place in the small intestine. The large intestine primarily serves as a site for fermentation of indigestible matter by gut bacteria and for resorption of water from digests before excretion.

Different phases of digestion take place including: the cephalic phase, gastric phase, and intestinal phase.

The cephalic phase occurs at the sight, thought and smell of food, which stimulate the cerebral cortex. Taste and smell stimuli are sent to the hypothalamus and medulla oblongata. After this it is routed through the vagus nerve and release of acetylcholine. Gastric secretion at this phase rises to 40% of maximum rate. Acidity in the stomach is not buffered by food at this point and thus acts to inhibit parietal (secretes acid) and G cell (secretes gastrin) activity via D cell secretion of somatostatin.

The gastric phase takes 3 to 4 hours. It is stimulated by distension of the stomach, presence of food in stomach and decrease in pH. Distention activates long and myenteric reflexes. This activates the release of acetylcholine, which stimulates the release of more gastric juices. As protein enters the stomach, it binds to hydrogen ions, which raises the pH of the stomach. Inhibition of gastrin and gastric acid secretion is lifted. This triggers G cells to release gastrin, which in turn stimulates parietal cells to secrete gastric acid. Gastric acid is about 0.5% hydrochloric acid (HCl), which lowers the pH to the desired pH of 1–3. Acid release is also triggered by acetylcholine and histamine.

The intestinal phase has two parts, the excitatory and the inhibitory. Partially digested food fills the duodenum. This triggers intestinal gastrin to be released. Enterogastric reflex inhibits vagal nuclei, activating sympathetic fibers causing the pyloric sphincter to tighten to prevent more food from entering, and inhibits local reflexes.

In mammals, preparation for digestion begins with the cephalic phase in which saliva is produced in the mouth and digestive enzymes are produced in the stomach. Mechanical and chemical digestion begin in the mouth where food is chewed, and mixed with saliva to begin enzymatic processing of starches. The stomach continues to break food down mechanically and chemically through churning and mixing with both acids and enzymes. Absorption occurs in the stomach and gastrointestinal tract, and the process finishes with defecation.

Herbivores have evolved cecums (or an abomasum in the case of ruminants). Ruminants have a fore-stomach with four chambers. These are the rumen, reticulum, omasum, and abomasum. In the first two chambers, the rumen and the reticulum, the food is mixed with saliva and separates into layers of solid and liquid material. Solids clump together to form the cud (or bolus). The cud is then regurgitated, chewed slowly to completely mix it with saliva and to break down the particle size.

Fibre, especially cellulose and hemi-cellulose, is primarily broken down into the volatile fatty acids, acetic acid, propionic acid and butyric acid in these chambers (the reticulo-rumen) by microbes: (bacteria, protozoa, and fungi). In the omasum, water and many of the inorganic mineral elements are absorbed into the blood stream.

The abomasum is the fourth and final stomach compartment in ruminants. It is a close equivalent of a monogastric stomach (e.g., those in humans or pigs), and digesta is processed here in much the same way. It serves primarily as a site for acid hydrolysis of microbial and dietary protein, preparing these protein sources for further digestion and absorption in the small intestine. Digesta is finally moved into the small intestine, where the digestion and absorption of nutrients occurs. Microbes produced in the reticulo-rumen are also digested in the small intestine.

23.1 The Human Digestive System

The gastrointestinal tract, (GI tract, GIT, digestive tract, digestion tract, alimentary canal) is the tract from the mouth to the anus which includes all the organs of the digestive system in humans and other animals. Food taken in through the mouth is digested to extract nutrients and absorb energy, and expelled in the remaining waste as feces. The mouth, esophagus, stomach and intestines are all part of the gastrointestinal tract. Gastrointestinal is an adjective meaning of or pertaining to the stomach and intestines. A tract is a collection of related anatomic structures or a series of connected body organs.

All vertebrates and most invertebrates have a digestive tract. The sponges, cnidarians, and ctenophores are the

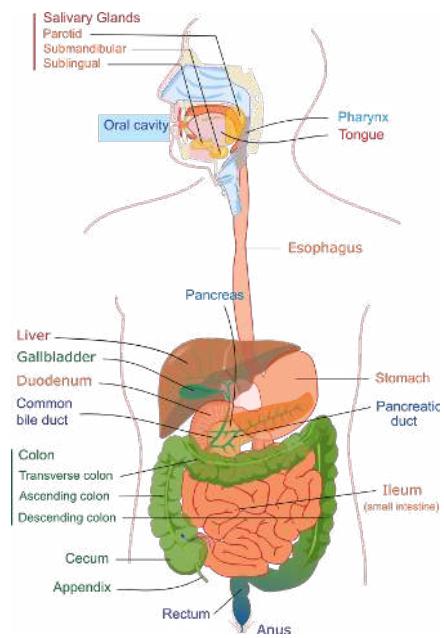


Figure 23.1: Upper and lower human gastrointestinal tract.¹

early invertebrates with an incomplete digestive tract having just one opening instead of two, where food is taken in and waste expelled.

The human gastrointestinal tract is around 9 meters long. Food digestion physiology varies between individuals and upon other factors such as the characteristics of the food and size of the meal, and the process of digestion normally takes between 24 and 72 hours.

Digestion begins in the mouth with the secretion of saliva and its digestive enzymes. Food is formed into a bolus by the mechanical mastication and swallowed into the esophagus from where it enters the stomach through the action of peristalsis. Gastric juice contains hydrochloric acid and pepsin which would damage the walls of the stomach and mucus is secreted for protection. In the stomach further release of enzymes break down the food further and this is combined with the churning action of the stomach. The partially digested food enters the duodenum as a thick semi-liquid chyme. In the small intestine, the larger part of digestion takes place and this is helped by the secretions of bile, pancreatic juice and intestinal juice. The intestinal walls are lined with villi, and their epithelial cells are covered with numerous microvilli to improve the absorption of nutrients by increasing the surface area of the intestine.

In the large intestine the passage of food is slower to enable fermentation by the gut flora to take place. Here water is absorbed and waste material stored as feces to be

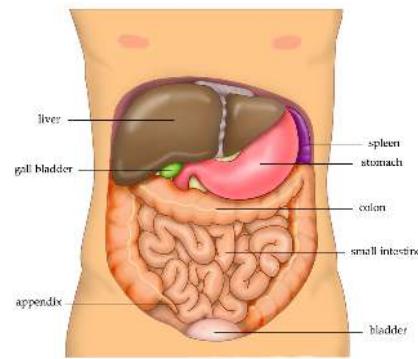


Figure 23.2: Location of the liver, gallbladder, spleen, stomach, colon, appendix, and small intestine in the human abdomen.²

removed by defecation via the anal canal and anus.

The human digestive system consists of the gastrointestinal tract plus the accessory organs of digestion (the tongue, salivary glands, pancreas, liver, and gallbladder). Digestion involves the breakdown of food into smaller and smaller components, until they can be absorbed and assimilated into the body. The process of digestion has three stages. The first stage is the cephalic phase of digestion which begins with gastric secretions in response to the sight and smell of food. This stage includes the mechanical breakdown of food by chewing, and the chemical breakdown by digestive enzymes, that takes place in the mouth.

Saliva contains digestive enzymes called amylase, and lingual lipase, secreted by the salivary glands and serous glands on the tongue. The enzymes start to break down the food in the mouth. Chewing, in which the food is mixed with saliva, begins the mechanical process of digestion. This produces a bolus which can be swallowed down the esophagus to enter the stomach. In the stomach the gastric phase of digestion takes place. The food is further broken down by mixing with gastric acid until it passes into the duodenum, in the third intestinal phase of digestion, where it is mixed with a number of enzymes produced by the pancreas. Digestion is helped by the chewing of food carried out by the muscles of mastication, the tongue, and the teeth, and also by the contractions of peristalsis, and segmentation. Gastric acid, and the production of mucus in the stomach, are essential for the continuation of digestion.

Peristalsis is the rhythmic contraction of muscles that begins in the esophagus and continues along the wall of the stomach and the rest of the gastrointestinal tract. This initially results in the production of chyme which when fully broken down in the small intestine is absorbed as chyle

into the lymphatic system. Most of the digestion of food takes place in the small intestine. Water and some minerals are reabsorbed back into the blood in the colon of the large intestine. The waste products of digestion (feces) are defecated from the anus via the rectum.

There are several organs and other components involved in the digestion of food. The organs known as the accessory digestive organs are the liver, gall bladder and pancreas. Other components include the mouth, salivary glands, tongue, teeth and epiglottis.

The largest structure of the digestive system is the gastrointestinal tract (GI tract). This starts at the mouth and ends at the anus, covering a distance of about nine (9) metres.

The largest part of the GI tract is the colon or large intestine. Water is absorbed here and the remaining waste matter is stored prior to defecation.

Most of the digestion of food takes place in the small intestine which is the longest part of the GI tract.

A major digestive organ is the stomach. Within its mucosa are millions of embedded gastric glands. Their secretions are vital to the functioning of the organ.

There are many specialised cells of the GI tract. These include the various cells of the gastric glands, taste cells, pancreatic duct cells, enterocytes and microfold cells.

Some parts of the digestive system are also part of the excretory system, including the large intestine.

The mouth is the first part of the upper gastrointestinal tract and is equipped with several structures that begin the first processes of digestion. These include salivary glands, teeth and the tongue. The mouth consists of two regions; the vestibule and the oral cavity proper. The vestibule is the area between the teeth, lips and cheeks, and the rest is the oral cavity proper. Most of the oral cavity is lined with oral mucosa, a mucous membrane that produces a lubricating mucus, of which only a small amount is needed. Mucous membranes vary in structure in the different regions of the body but they all produce a lubricating mucus, which is either secreted by surface cells or more usually by underlying glands. The mucous membrane in the mouth continues as the thin mucosa which lines the bases of the teeth. The main component of mucus is a glycoprotein called mucin and the type secreted varies according to the region involved. Mucin is viscous, clear, and clinging. Underlying the mucous membrane in the mouth is a thin layer of smooth muscle tissue and the loose connection to the membrane gives it its great elasticity. It covers the cheeks, inner surfaces of the lips, and floor of the mouth, and the mucin produced is highly protective against tooth decay.

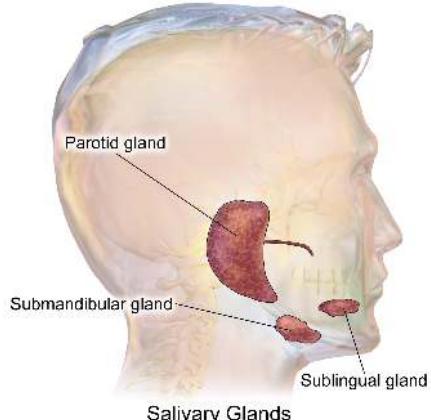


Figure 23.3: The main salivary glands.³

The roof of the mouth is termed the palate and it separates the oral cavity from the nasal cavity. The palate is hard at the front of the mouth since the overlying mucosa is covering a plate of bone; it is softer and more pliable at the back being made of muscle and connective tissue, and it can move to swallow food and liquids. The soft palate ends at the uvula. The surface of the hard palate allows for the pressure needed in eating food, to leave the nasal passage clear. The opening between the lips is termed the oral fissure, and the opening into the throat is called the fauces.

At either side of the soft palate are the palatoglossus muscles which also reach into regions of the tongue. These muscles raise the back of the tongue and also close both sides of the fauces to enable food to be swallowed.:1208 Mucus helps in the mastication of food in its ability to soften and collect the food in the formation of the bolus.

23.1.1 The Salivary Glands

There are three pairs of main salivary glands and between 800 and 1,000 minor salivary glands, all of which mainly serve the digestive process, and also play an important role in the maintenance of dental health and general mouth lubrication, without which speech would be impossible. The main glands are all exocrine glands, secreting via ducts. All of these glands terminate in the mouth. The largest of these are the parotid glands—their secretion is mainly serous. The next pair are underneath the jaw, the submandibular glands, these produce both serous fluid and mucus. The serous fluid is produced by serous glands in these salivary glands which also produce lingual lipase. They produce about 70% of the oral cavity saliva. The third pair are the sublingual glands located underneath the tongue and their secretion is mainly mucous with a small percentage of saliva.

Within the oral mucosa, and also on the tongue, palates, and floor of the mouth, are the minor salivary glands; their secretions are mainly mucous and they are innervated by the facial nerve (CN7). The glands also secrete amylase a first stage in the breakdown of food acting on the carbohydrate in the food to transform the starch content into maltose. There are other serous glands on the surface of the tongue that encircle taste buds on the back part of the tongue and these also produce lingual lipase. Lipase is a digestive enzyme that catalyses the hydrolysis of lipids (fats). These glands are termed Von Ebner's glands which have also been shown to have another function in the secretion of histatins which offer an early defense (outside of the immune system) against microbes in food, when it makes contact with these glands on the tongue tissue. Sensory information can stimulate the secretion of saliva providing the necessary fluid for the tongue to work with and also to ease swallowing of the food.

Saliva moistens and softens food, and along with the chewing action of the teeth, transforms the food into a smooth bolus. The bolus is further helped by the lubrication provided by the saliva in its passage from the mouth into the esophagus. Also of importance is the presence in saliva of the digestive enzymes amylase and lipase. Amylase starts to work on the starch in carbohydrates, breaking it down into the simple sugars of maltose and dextrose that can be further broken down in the small intestine. Saliva in the mouth can account for 30% of this initial starch digestion. Lipase starts to work on breaking down fats. Lipase is further produced in the pancreas where it is released to continue this digestion of fats. The presence of salivary lipase is of prime importance in young babies whose pancreatic lipase has yet to be developed.

As well as its role in supplying digestive enzymes, saliva has a cleansing action for the teeth and mouth. It also has an immunological role in supplying antibodies to the system, such as immunoglobulin A. This is seen to be key in preventing infections of the salivary glands, importantly that of parotitis.

Saliva also contains a glycoprotein called haptocorrin which is a binding protein to vitamin B12. It binds with the vitamin in order to carry it safely through the acidic content of the stomach. When it reaches the duodenum, pancreatic enzymes break down the glycoprotein and free the vitamin which then binds with intrinsic factor.

23.1.2 The Tongue

Food enters the mouth where the first stage in the digestive process takes place, with the action of the tongue and the secretion of saliva. The tongue is a fleshy and muscular sensory organ, and the very first sensory information is re-

ceived via the taste buds in the papillae on its surface. If the taste is agreeable, the tongue will go into action, manipulating the food in the mouth which stimulates the secretion of saliva from the salivary glands. The liquid quality of the saliva will help in the softening of the food and its enzyme content will start to break down the food whilst it is still in the mouth. The first part of the food to be broken down is the starch of carbohydrates (by the enzyme amylase in the saliva).

The tongue is attached to the floor of the mouth by a ligamentous band called the frenum and this gives it great mobility for the manipulation of food (and speech); the range of manipulation is optimally controlled by the action of several muscles and limited in its external range by the stretch of the frenum. The tongue's two sets of muscles, are four intrinsic muscles that originate in the tongue and are involved with its shaping, and four extrinsic muscles originating in bone that are involved with its movement.

Taste is a form of chemoreception that takes place in the specialised taste receptors, contained in structures called taste buds in the mouth. Taste buds are mainly on the upper surface (dorsum) of the tongue. The function of taste perception is vital to help prevent harmful or rotten foods from being consumed. There are also taste buds on the epiglottis and upper part of the esophagus. The taste buds are innervated by a branch of the facial nerve the chorda tympani, and the glossopharyngeal nerve. Taste messages are sent via these cranial nerves to the brain. The brain can distinguish between the chemical qualities of the food. The five basic tastes are referred to as those of saltiness, sourness, bitterness, sweetness, and umami. The detection of saltiness and sourness enables the control of salt and acid balance. The detection of bitterness warns of poisons—many of a plant's defences are of poisonous compounds that are bitter. Sweetness guides to those foods that will supply energy; the initial breakdown of the energy-giving carbohydrates by salivary amylase creates the taste of sweetness since simple sugars are the first result. The taste of umami is thought to signal protein-rich food. Sour tastes are acidic which is often found in bad food. The brain has to decide very quickly whether the food should be eaten or not. It was the findings in 1991, describing the first olfactory receptors that helped to prompt the research into taste. The olfactory receptors are located on cell surfaces in the nose which bind to chemicals enabling the detection of smells. It is assumed that signals from taste receptors work together with those from the nose, to form an idea of complex food flavours.

23.1.3 The Teeth

Teeth are complex structures made of materials specific to them. They are made of a bone-like material called dentin, which is covered by the hardest tissue in the body—enamel. Teeth have different shapes to deal with different aspects of mastication employed in tearing and chewing pieces of food into smaller and smaller pieces. This results in a much larger surface area for the action of digestive enzymes. The teeth are named after their particular roles in the process of mastication—incisors are used for cutting or biting off pieces of food; canines, are used for tearing, premolars and molars are used for chewing and grinding. Mastication of the food with the help of saliva and mucus results in the formation of a soft bolus which can then be swallowed to make its way down the upper gastrointestinal tract to the stomach. The digestive enzymes in saliva also help in keeping the teeth clean by breaking down any lodged food particles.

23.1.4 The Epiglottis

The epiglottis is a flap of elastic cartilage attached to the entrance of the larynx. It is covered with a mucous membrane and there are taste buds on its lingual surface which faces into the mouth. Its laryngeal surface faces into the larynx. The epiglottis functions to guard the entrance of the glottis, the opening between the vocal folds. It is normally pointed upward during breathing with its underside functioning as part of the pharynx, but during swallowing, the epiglottis folds down to a more horizontal position, with its upper side functioning as part of the pharynx. In this manner it prevents food from going into the trachea and instead directs it to the esophagus, which is behind. During swallowing, the backward motion of the tongue forces the epiglottis over the glottis' opening to prevent any food that is being swallowed from entering the larynx which leads to the lungs; the larynx is also pulled upwards to assist this process. Stimulation of the larynx by ingested matter produces a strong cough reflex in order to protect the lungs.

23.1.5 The Pharynx

The pharynx is a part of the conducting zone of the respiratory system and also a part of the digestive system. It is the part of the throat immediately behind the nasal cavity at the back of the mouth and above the esophagus and larynx. The pharynx is made up of three parts. The lower two parts—the oropharynx and the laryngopharynx are involved in the digestive system. The laryngopharynx connects to the esophagus and it serves as a passageway for both air and food. Air enters the larynx anteriorly but anything swallowed has priority and the passage of air is temporarily blocked. The pharynx is innervated by the pharyngeal plexus of the vagus nerve.:1465 Muscles in the pharynx push the food into the esophagus. The pharynx joins the esophagus at the oesophageal inlet which is located behind the cricoid cartilage.

23.1.6 The Esophagus

The esophagus, commonly known as the foodpipe or gullet, consists of a muscular tube through which food passes from the pharynx to the stomach. The esophagus is continuous with the laryngopharynx. It passes through the posterior mediastinum in the thorax and enters the stomach through a hole in the thoracic diaphragm—the esophageal hiatus, at the level of the tenth thoracic vertebra (T10). Its length averages 25 cm, varying with an individual's height. It is divided into cervical, thoracic and abdominal parts. The pharynx joins the esophagus at the esophageal inlet which is behind the cricoid cartilage.

At rest the esophagus is closed at both ends, by the upper and lower esophageal sphincters. The opening of the upper sphincter is triggered by the swallowing reflex so that food is allowed through. The sphincter also serves to prevent back flow from the esophagus into the pharynx. The esophagus has a mucous membrane and the epithelium which has a protective function is continuously replaced due to the volume of food that passes inside the esophagus. During swallowing, food passes from the mouth through the pharynx into the esophagus. The epiglottis folds down to a more horizontal position to direct the food into the esophagus, and away from the trachea.

Once in the esophagus, the bolus travels down to the stomach via rhythmic contraction and relaxation of muscles known as peristalsis. The lower esophageal sphincter is a muscular sphincter surrounding the lower part of the esophagus. The gastroesophageal junction between the esophagus and the stomach is controlled by the lower esophageal sphincter, which remains constricted at all times other than during swallowing and vomiting to prevent the contents of the stomach from entering the esophagus. As the esophagus does not have the same protection from acid as the stomach, any failure of this sphincter can lead to heartburn.

23.1.7 The Diaphragm

The diaphragm is an important part of the body's digestive system. The muscular diaphragm separates the thoracic cavity from the abdominal cavity where most of the digestive organs are located. The suspensory muscle attaches the ascending duodenum to the diaphragm. This muscle is thought to be of help in the digestive system in that its attachment offers a wider angle to the duodenoje-

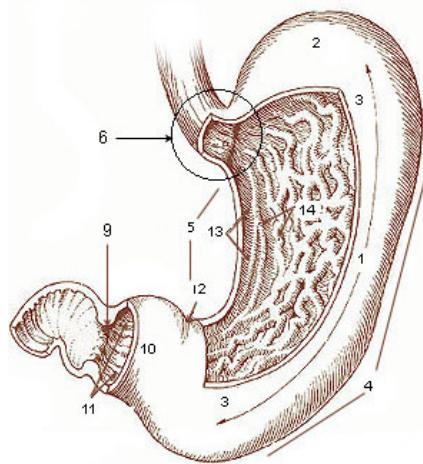


Figure 23.4: The human stomach.⁴ 1. Body of stomach
2. Fundus 3. Anterior wall 4. Greater curvature 5. Lesser curvature 6. Cardia 9. Pyloric sphincter 10. Pyloric antrum 11. Pyloric canal 12. Angular incisure 13. Gastric canal 14. Rugae

junal flexure for the easier passage of digesting material. The diaphragm also attaches to, and anchors the liver at its bare area. The esophagus enters the abdomen through a hole in the diaphragm at the level of T10.

23.1.8 The Stomach

The stomach is a major organ of the gastrointestinal tract and digestive system. It is a consistently J-shaped organ joined to the esophagus at its upper end and to the duodenum at its lower end. Gastric acid (informally gastric juice), produced in the stomach plays a vital role in the digestive process, and mainly contains hydrochloric acid and sodium chloride. A peptide hormone, gastrin, produced by G cells in the gastric glands, stimulates the production of gastric juice which activates the digestive enzymes. Pepsinogen is a precursor enzyme (zymogen) produced by the gastric chief cells, and gastric acid activates this to the enzyme pepsin which begins the digestion of proteins. As these two chemicals would damage the stomach wall, mucus is secreted by innumerable gastric glands in the stomach, to provide a slimy protective layer against the damaging effects of the chemicals on the inner layers of the stomach.

At the same time that protein is being digested, mechanical churning occurs through the action of peristalsis, waves of muscular contractions that move along the stomach wall. This allows the mass of food to further mix with the digestive enzymes. Gastric lipase secreted by the chief cells in the fundic glands in the gastric mucosa of the stomach, is an acidic lipase, in contrast with the alkaline pancreatic lipase. This breaks down fats to some degree though

is not as efficient as the pancreatic lipase.

The pylorus, the lowest section of the stomach which attaches to the duodenum via the pyloric canal, contains countless glands which secrete digestive enzymes including gastrin. After an hour or two, a thick semi-liquid called chyme is produced. When the pyloric sphincter, or valve opens, chyme enters the duodenum where it mixes further with digestive enzymes from the pancreas, and then passes through the small intestine, where digestion continues. When the chyme is fully digested, it is absorbed into the blood. 95% of absorption of nutrients occurs in the small intestine. Water and minerals are reabsorbed back into the blood in the colon of the large intestine, where the environment is slightly acidic. Some vitamins, such as biotin and vitamin K produced by bacteria in the gut flora of the colon are also absorbed.

The parietal cells in the fundus of the stomach, produce a glycoprotein called intrinsic factor which is essential for the absorption of vitamin B12. Vitamin B12 (cobalamin), is carried to, and through the stomach, bound to a glycoprotein secreted by the salivary glands – transcobalamin I also called haptocorrin, which protects the acid-sensitive vitamin from the acidic stomach contents. Once in the more neutral duodenum, pancreatic enzymes break down the protective glycoprotein. The freed vitamin B12 then binds to intrinsic factor which is then absorbed by the enterocytes in the ileum.

The stomach is a distensible organ and can normally expand to hold about one litre of food. This expansion is enabled by a series of gastric folds in the inner walls of the stomach. The stomach of a newborn baby will only be able to expand to retain about 30 ml.

23.1.9 The Spleen

The spleen is the largest lymphoid organ in the body but has other functions. It breaks down both red and white blood cells that are spent. This is why it is sometimes known as the 'graveyard of red blood cells'. A product of this digestion is the pigment bilirubin, which is sent to the liver and secreted in the bile. Another product is iron, which is used in the formation of new blood cells in the bone marrow. Medicine treats the spleen solely as belonging to the lymphatic system, though it is acknowledged that the full range of its important functions is not yet understood.:1751

23.1.10 The Liver, Bile And Gallbladder

The liver is an organ only found in vertebrates. In humans, it is located in the right upper quadrant of the abdomen, below the diaphragm. The liver is the second largest organ (after the skin) and is an accessory digestive gland which

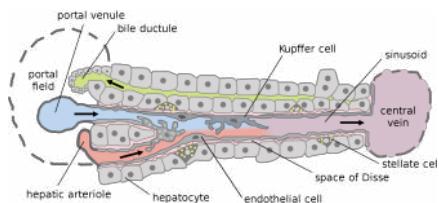


Figure 23.5: Diagram of the microscopic structure of the liver.⁵

plays a role in the body's metabolism. The liver has many functions some of which are important to digestion. The liver can detoxify various metabolites; synthesise proteins and produce biochemicals needed for digestion. It regulates the storage of glycogen which it can form from glucose (glycogenesis). The liver can also synthesise glucose from certain amino acids. Its digestive functions are largely involved with the breaking down of carbohydrates. It also maintains protein metabolism in its synthesis and degradation. In lipid metabolism it synthesises cholesterol. Fats are also produced in the process of lipogenesis. The liver synthesises the bulk of lipoproteins. The liver is located in the upper right quadrant of the abdomen and below the diaphragm to which it is attached at one part, the bare area of the liver. This is to the right of the stomach and it overlies the gall bladder. The liver synthesises bile acids and lecithin to promote the digestion of fat.

The liver is grossly divided into two parts when viewed from above – a right and a left lobe – and four parts when viewed from below (left, right, caudate, and quadrate lobes). Microscopically, each liver lobe is seen to be made up of hepatic lobules. The lobules are roughly hexagonal, and consist of plates of hepatocytes, and sinusoids radiating from a central vein towards an imaginary perimeter of interlobular portal triads. The central vein joins to the hepatic vein to carry blood out from the liver. A distinctive component of a lobule is the portal triad, which can be found running along each of the lobule's corners. The portal triad, consists of the hepatic artery, the portal vein, and the common bile duct. The triad may be seen on a liver ultrasound, as a Mickey Mouse sign with the portal vein as the head, and the hepatic artery, and the common bile duct as the ears.

Bile produced by the liver is made up of water (97%), bile salts, mucus and pigments, 1% fats and inorganic salts. Bilirubin is its major pigment. Bile acts partly as a surfactant which lowers the surface tension between either two liquids or a solid and a liquid and helps to emulsify the fats in the chyme. Food fat is dispersed by the action of bile into smaller units called micelles. The breaking down into micelles creates a much larger surface area for the pancreatic enzyme, lipase to work on. Lipase digests the

triglycerides which are broken down into two fatty acids and a monoglyceride. These are then absorbed by villi on the intestinal wall. If fats are not absorbed in this way in the small intestine problems can arise later in the large intestine which is not equipped to absorb fats. Bile also helps in the absorption of vitamin K from the diet. Bile is collected and delivered through the common hepatic duct. This duct joins with the cystic duct to connect in a common bile duct with the gallbladder. Bile is stored in the gallbladder for release when food is discharged into the duodenum and also after a few hours.

The gallbladder is a hollow part of the biliary tract that sits just beneath the liver, with the gallbladder body resting in a small depression. It is a small organ where the bile produced by the liver is stored, before being released into the small intestine. Bile flows from the liver through the bile ducts and into the gall bladder for storage. The bile is released in response to cholecystokinin (CCK) a peptide hormone released from the duodenum. The production of CCK (by endocrine cells of the duodenum) is stimulated by the presence of fat in the duodenum.

It is divided into three sections, a fundus, body and neck. The neck tapers and connects to the biliary tract via the cystic duct, which then joins the common hepatic duct to form the common bile duct. At this junction is a mucosal fold called Hartmann's pouch, where gallstones commonly get stuck. The muscular layer of the body is of smooth muscle tissue that helps the gallbladder contract, so that it can discharge its bile into the bile duct. The gallbladder needs to store bile in a natural, semi-liquid form at all times. Hydrogen ions secreted from the inner lining of the gallbladder keep the bile acidic enough to prevent hardening. To dilute the bile, water and electrolytes from the digestion system are added. Also, salts attach themselves to cholesterol molecules in the bile to keep them from crystallising. If there is too much cholesterol or bilirubin in the bile, or if the gallbladder doesn't empty properly the systems can fail. This is how gallstones form when a small piece of calcium gets coated with either cholesterol or bilirubin and the bile crystallises and forms a gallstone. The main purpose of the gallbladder is to store and release bile, or gall. Bile is released into the small intestine in order to help in the digestion of fats by breaking down larger molecules into smaller ones. After the fat is absorbed, the bile is also absorbed and transported back to the liver for reuse.

23.1.11 The Pancreas

The pancreas is a major organ functioning as an accessory digestive gland in the digestive system. It is both an endocrine gland and an exocrine gland. The endocrine part secretes insulin when the blood sugar becomes high; in-

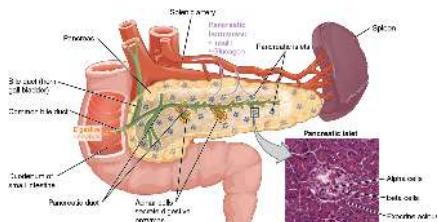


Figure 23.6: The pancreas has a role in digestion, highlighted here. Ducts in the pancreas (green) conduct digestive enzymes into the duodenum. This image also shows a pancreatic islet, part of the endocrine pancreas, which contains cells responsible for secretion of insulin and glucagon.⁶

sulin moves glucose from the blood into the muscles and other tissues for use as energy. The endocrine part releases glucagon when the blood sugar is low; glucagon allows stored sugar to be broken down into glucose by the liver in order to re-balance the sugar levels. The pancreas produces and releases important digestive enzymes in the pancreatic juice that it delivers to the duodenum. The pancreas lies below and at the back of the stomach. It connects to the duodenum via the pancreatic duct which it joins near to the bile duct's connection where both the bile and pancreatic juice can act on the chyme that is released from the stomach into the duodenum. Aqueous pancreatic secretions from pancreatic duct cells contain bicarbonate ions which are alkaline and help with the bile to neutralise the acidic chyme that is churned out by the stomach.

The pancreas is also the main source of enzymes for the digestion of fats and proteins. Some of these are released in response to the production of CKK in the duodenum. (The enzymes that digest polysaccharides, by contrast, are primarily produced by the walls of the intestines.) The cells are filled with secretory granules containing the precursor digestive enzymes. The major proteases, the pancreatic enzymes which work on proteins, are trypsinogen and chymotrypsinogen. Elastase is also produced. Smaller amounts of lipase and amylase are secreted. The pancreas also secretes phospholipase A2, lysophospholipase, and cholesterol esterase. The precursor zymogens, are inactive variants of the enzymes; which avoids the onset of pancreatitis caused by autodegradation. Once released in the intestine, the enzyme enteropeptidase present in the intestinal mucosa activates trypsinogen by cleaving it to form trypsin; further cleavage results in chymotrypsin.

23.1.12 Lower Gastrointestinal Tract

The lower gastrointestinal tract (GI), includes the small intestine and all of the large intestine. The intestine is also

called the bowel or the gut. The lower GI starts at the pyloric sphincter of the stomach and finishes at the anus. The small intestine is subdivided into the duodenum, the jejunum and the ileum. The cecum marks the division between the small and large intestine. The large intestine includes the rectum and anal canal.

23.1.13 The Small Intestine

Partially digested food starts to arrive in the small intestine as semi-liquid chyme, one hour after it is eaten.[citation needed] The stomach is half empty after an average of 1.2 hours. After four or five hours the stomach has emptied.

In the small intestine, the pH becomes crucial; it needs to be finely balanced in order to activate digestive enzymes. The chyme is very acidic, with a low pH, having been released from the stomach and needs to be made much more alkaline. This is achieved in the duodenum by the addition of bile from the gall bladder combined with the bicarbonate secretions from the pancreatic duct and also from secretions of bicarbonate-rich mucus from duodenal glands known as Brunner's glands. The chyme arrives in the intestines having been released from the stomach through the opening of the pyloric sphincter. The resulting alkaline fluid mix neutralises the gastric acid which would damage the lining of the intestine. The mucus component lubricates the walls of the intestine.

When the digested food particles are reduced enough in size and composition, they can be absorbed by the intestinal wall and carried to the bloodstream. The first receptacle for this chyme is the duodenal bulb. From here it passes into the first of the three sections of the small intestine, the duodenum. (The next section is the jejunum and the third is the ileum). The duodenum is the first and shortest section of the small intestine. It is a hollow, jointed C-shaped tube connecting the stomach to the jejunum. It starts at the duodenal bulb and ends at the suspensory muscle of duodenum. The attachment of the suspensory muscle to the diaphragm is thought to help the passage of food by making a wider angle at its attachment.

Most food digestion takes place in the small intestine. Segmentation contractions act to mix and move the chyme more slowly in the small intestine allowing more time for absorption (and these continue in the large intestine). In the duodenum, pancreatic lipase is secreted together with a co-enzyme, colipase to further digest the fat content of the chyme. From this breakdown, smaller particles of emulsified fats called chylomicrons are produced. There are also digestive cells called enterocytes lining the intestines (the majority being in the small intestine). They are unusual cells in that they have villi on their surface which in turn have innumerable microvilli on their surface. All these villi make for a greater surface area, not only for

the absorption of chyme but also for its further digestion by large numbers of digestive enzymes present on the microvilli.

The chylomicrons are small enough to pass through the enterocyte villi and into their lymph capillaries called lacteals. A milky fluid called chyle, consisting mainly of the emulsified fats of the chylomicrons, results from the absorbed mix with the lymph in the lacteals.[clarification needed] Chyle is then transported through the lymphatic system to the rest of the body.

The suspensory muscle marks the end of the duodenum and the division between the upper gastrointestinal tract and the lower GI tract. The digestive tract continues as the jejunum which continues as the ileum. The jejunum, the midsection of the small intestine contains circular folds, flaps of doubled mucosal membrane which partially encircle and sometimes completely encircle the lumen of the intestine. These folds together with villi serve to increase the surface area of the jejunum enabling an increased absorption of digested sugars, amino acids and fatty acids into the bloodstream. The circular folds also slow the passage of food giving more time for nutrients to be absorbed.

The last part of the small intestine is the ileum. This also contains villi and vitamin B12; bile acids and any residue nutrients are absorbed here. When the chyme is exhausted of its nutrients the remaining waste material changes into the semi-solids called feces, which pass to the large intestine, where bacteria in the gut flora further break down residual proteins and starches.

Transit time through the small intestine is an average of 4 hours. Half of the food residues of a meal have emptied from the small intestine by an average of 5.4 hours after ingestion. Emptying of the small intestine is complete after an average of 8.6 hours.

23.1.14 The Cecum

The cecum is a pouch marking the division between the small intestine and the large intestine. It lies below the ileocecal valve in the lower right quadrant of the abdomen. The cecum receives chyme from the last part of the small intestine, the ileum, and connects to the ascending colon of the large intestine. At this junction there is a sphincter or valve, the ileocecal valve which slows the passage of chyme from the ileum, allowing further digestion. It is also the site of the appendix attachment.

23.1.15 The Large Intestine

In the large intestine, the passage of the digesting food in the colon is a lot slower, taking from 30 to 40 hours until it is removed by defecation. The colon mainly serves

as a site for the fermentation of digestible matter by the gut flora. The time taken varies considerably between individuals. The remaining semi-solid waste is termed feces and is removed by the coordinated contractions of the intestinal walls, termed peristalsis, which propels the excreta forward to reach the rectum and exit via defecation from the anus. The wall has an outer layer of longitudinal muscles, the taeniae coli, and an inner layer of circular muscles. The circular muscle keeps the material moving forward and also prevents any back flow of waste. Also of help in the action of peristalsis is the basal electrical rhythm that determines the frequency of contractions. The taeniae coli can be seen and are responsible for the bulges (hastra) present in the colon. Most parts of the GI tract are covered with serous membranes and have a mesentery. Other more muscular parts are lined with adventitia.

23.1.16 Protein Digestion

Protein digestion occurs in the stomach and duodenum in which 3 main enzymes, pepsin secreted by the stomach and trypsin and chymotrypsin secreted by the pancreas, break down food proteins into polypeptides that are then broken down by various exopeptidases and dipeptidases into amino acids. The digestive enzymes however are mostly secreted as their inactive precursors, the zymogens. For example, trypsin is secreted by pancreas in the form of trypsinogen, which is activated in the duodenum by enterokinase to form trypsin. Trypsin then cleaves proteins to smaller polypeptides.

23.1.17 Fat Digestion

Digestion of some fats can begin in the mouth where lingual lipase breaks down some short chain lipids into diglycerides. However fats are mainly digested in the small intestine. The presence of fat in the small intestine produces hormones that stimulate the release of pancreatic lipase from the pancreas and bile from the liver which helps in the emulsification of fats for absorption of fatty acids. Complete digestion of one molecule of fat (a triglyceride) results a mixture of fatty acids, mono- and di-glycerides, as well as some undigested triglycerides, but no free glycerol molecules.

23.1.18 Carbohydrate Digestion

In humans, dietary starches are composed of glucose units arranged in long chains called amylose, a polysaccharide. During digestion, bonds between glucose molecules are broken by salivary and pancreatic amylase, resulting in progressively smaller chains of glucose. This results in simple sugars glucose and maltose (2 glucose molecules) that can be absorbed by the small intestine.

Lactase is an enzyme that breaks down the disaccharide lactose to its component parts, glucose and galactose. Glucose and galactose can be absorbed by the small intestine. Approximately 65 percent of the adult population produce only small amounts of lactase and are unable to eat unfermented milk-based foods. This is commonly known as lactose intolerance. Lactose intolerance varies widely by genetic heritage; more than 90 percent of peoples of east Asian descent are lactose intolerant, in contrast to about 5 percent of people of northern European descent.

Sucrase is an enzyme that breaks down the disaccharide sucrose, commonly known as table sugar, cane sugar, or beet sugar. Sucrose digestion yields the sugars fructose and glucose which are readily absorbed by the small intestine.

23.1.19 DNA And RNA Digestion

DNA and RNA are broken down into mononucleotides by the nucleases deoxyribonuclease and ribonuclease (DNase and RNase) from the pancreas.

23.1.20 Non-destructive digestion

Some nutrients are complex molecules (for example vitamin B12) which would be destroyed if they were broken down into their functional groups. To digest vitamin B12 non-destructively, haptocorrin in saliva strongly binds and protects the B12 molecules from stomach acid as they enter the stomach and are cleaved from their protein complexes.

After the B12-haptocorrin complexes pass from the stomach via the pylorus to the duodenum, pancreatic proteases cleave haptocorrin from the B12 molecules which re-bind to intrinsic factor (IF). These B12-IF complexes travel to the ileum portion of the small intestine where cubilin receptors enable assimilation and circulation of B12-IF complexes in the blood.

23.1.21 The Digestive System of Birds

Many birds possess a muscular pouch along the esophagus called a crop. The crop functions to both soften food and regulate its flow through the system by storing it temporarily. The size and shape of the crop is quite variable among the birds. Members of the family Columbidae, such as pigeons, produce a nutritious crop milk which is fed to their young by regurgitation.

The avian stomach is composed of two organs, the proventriculus and the gizzard that work together during digestion. The proventriculus is a rod shaped tube, which is found between the esophagus and the gizzard, that secretes hydrochloric acid and pepsinogen into the digestive

tract. The acid converts the inactive pepsinogen into the active proteolytic enzyme, pepsin, which breaks down specific peptide bonds found in proteins, to produce a set of peptides, which are amino acid chains that are shorter than the original dietary protein. The gastric juices (hydrochloric acid and pepsinogen) are mixed with the stomach contents through the muscular contractions of the gizzard.

The gizzard is composed of four muscular bands that rotate and crush food by shifting the food from one area to the next within the gizzard. The gizzard of some species of herbivorous birds, like turkey and quails, contains small pieces of grit or stone called gastroliths that are swallowed by the bird to aid in the grinding process, serving the function of teeth. The use of gizzard stones is a similarity found between birds and dinosaurs, which left gastroliths as trace fossils.

The partially digested and pulverized gizzard contents, now called a bolus, are passed into the intestine, where pancreatic and intestinal enzymes complete the digestion of the digestible food. The digestion products are then absorbed through the intestinal mucosa into the blood. The intestine ends via the large intestine in the vent or cloaca which serves as the common exit for renal and intestinal excrements as well as for the laying of eggs. However, unlike mammals, many birds do not excrete the bulky portions (roughage) of their undigested food (e.g. feathers, fur, bone fragments, and seed husks) via the cloaca, but regurgitate them as food pellets.

23.2 The Digestive System of Fish

As with other vertebrates, the intestines of fish consist of two segments, the small intestine and the large intestine. In most higher vertebrates, the small intestine is further divided into the duodenum and other parts. In fish, the divisions of the small intestine are not as clear, and the terms anterior intestine or proximal intestine may be used instead of duodenum. In bony fish, the intestine is relatively short, typically around one and a half times the length of the fish's body. It commonly has a number of pyloric caeca, small pouch-like structures along its length that help to increase the overall surface area of the organ for digesting food. There is no ileocaecal valve in teleosts, with the boundary between the small intestine and the rectum being marked only by the end of the digestive epithelium. There is no small intestine as such in non-teleost fish, such as sharks, sturgeons, and lungfish. Instead, the digestive part of the gut forms a spiral intestine, connecting the stomach to the rectum. In this type of gut, the intestine itself is relatively straight, but has a long fold running along the inner surface in a spiral fashion, sometimes for dozens of turns. This fold creates a valve-like structure that greatly

increases both the surface area and the effective length of the intestine. The lining of the spiral intestine is similar to that of the small intestine in teleosts and non-mammalian tetrapods. In lampreys, the spiral valve is extremely small, possibly because their diet requires little digestion. Hagfish have no spiral valve at all, with digestion occurring for almost the entire length of the intestine, which is not subdivided into different regions.

The pyloric caecum is a pouch, usually peritoneal, at the beginning of the large intestine. It receives faecal material from the ileum, and connects to the ascending colon of the large intestine. It is present in most amniotes, and also in lungfish. Many fish in addition have a number of small outpocketings, also called pyloric caeca, along their intestine; despite the name they are not homologous with the caecum of amniotes. Their purpose is to increase the overall surface area of the digestive epithelium, therefore optimizing the absorption of sugars, amino acids, and dipeptides, among other nutrients.

As with other vertebrates, the relative positions of the esophageal and duodenal openings to the stomach remain relatively constant. As a result, the stomach always curves somewhat to the left before curving back to meet the pyloric sphincter. However, lampreys, hagfishes, chimaeras, lungfishes, and some teleost fish have no stomach at all, with the esophagus opening directly into the intestine. These fish consume diets that either require little storage of food, or no pre-digestion with gastric juices, or both.

23.3 The Digestive System of Annelids

An earthworm's digestive system consists of a mouth, pharynx, esophagus, crop, gizzard, and intestine. The mouth is surrounded by strong lips, which act like a hand to grab pieces of dead grass, leaves, and weeds, with bits of soil to help chew. The lips break the food down into smaller pieces. In the pharynx, the food is lubricated by mucus secretions for easier passage. The esophagus adds calcium carbonate to neutralize the acids formed by food matter decay. Temporary storage occurs in the crop where food and calcium carbonate are mixed. The powerful muscles of the gizzard churn and mix the mass of food and dirt. When the churning is complete, the glands in the walls of the gizzard add enzymes to the thick paste, which helps chemically breakdown the organic matter. By peristalsis, the mixture is sent to the intestine where friendly bacteria continue chemical breakdown. This releases carbohydrates, protein, fat, and various vitamins and minerals for absorption into the body.

Chapter 24

Osmoregulation And Excretory Systems

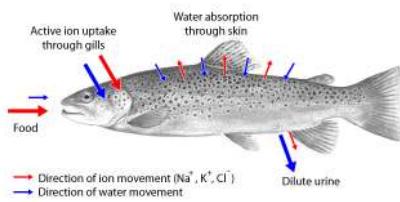


Figure 24.1: Movement of water and ions in freshwater fish.¹

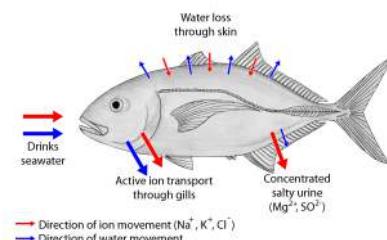


Figure 24.2: Movement of water and ions in saltwater fish.²

Osmoregulation is the active regulation of the osmotic pressure of an organism's body fluids, detected by osmoreceptors, to maintain the homeostasis of the organism's water content; that is, it maintains the fluid balance and the concentration of electrolytes (salts in solution which in this case is represented by body fluid) to keep the body fluids from becoming too diluted or concentrated. Osmotic pressure is a measure of the tendency of water to move into one solution from another by osmosis. The higher the osmotic pressure of a solution, the more water tends to move into it. Pressure must be exerted on the hypertonic side of a selectively permeable membrane to prevent diffusion of water by osmosis from the side containing pure water.

Two major types of osmoregulation are osmoconformers and osmoregulators. Osmoconformers match their body osmolarity to their environment actively or passively. Most marine invertebrates are osmoconformers, although their ionic composition may be different from that of seawater.

Osmoregulators tightly regulate their body osmolarity, maintaining constant internal conditions. They are more common in the animal kingdom. Osmoregulators actively control salt concentrations despite the salt concentrations in the environment. An example is freshwater fish. The gills actively uptake salt from the environment by the use of mitochondria-rich cells. Water will diffuse into the fish, so it excretes a very hypotonic (dilute) urine to expel all the excess water.

A marine fish has an internal osmotic concentration lower than that of the surrounding seawater, so it tends to lose water and gain salt. It actively excretes salt out from the gills. Most fish are stenohaline, which means they are restricted to either salt or fresh water and cannot survive in water with a different salt concentration than they are adapted to.

However, some fish show a tremendous ability to effectively osmoregulate across a broad range of salinities; fish with this ability are known as euryhaline species, e.g., Flounder. Flounder have been observed to inhabit two utterly disparate environments—marine and fresh water—and it is inherent to adapt to both by bringing in behavioral and physiological modifications.

Some marine fish, like sharks, have adopted a different, efficient mechanism to conserve water, i.e., osmoregulation. They retain urea in their blood in relatively higher concentration. Urea damages living tissues so, to cope with this problem, some fish retain trimethylamine oxide. This provides a better solution to urea's toxicity. Sharks, having slightly higher solute concentration (i.e., above 1000 mOsm which is sea solute concentration), do not drink water like fresh water fish. Organisms in aquatic and terrestrial environments must maintain the right concentration of solutes and amount of water in their body fluids; this involves excretion (getting rid of metabolic nitrogen wastes and other substances such as hormones that would be toxic if allowed to accumulate in the blood) through

organs such as the skin and the kidneys. The excretory system is a passive biological system that removes excess, unnecessary materials from the body fluids of an organism, so as to help maintain internal chemical homeostasis and prevent damage to the body. The dual function of excretory systems is the elimination of the waste products of metabolism and to drain the body of used up and broken down components in a liquid and gaseous state. In humans and other amniotes (mammals, birds and reptiles) most of these substances leave the body as urine and to some degree exhalation, mammals also expel them through sweating.

Excretion is a process by which metabolic waste is eliminated from an organism. In vertebrates this is primarily carried out by the lungs, kidneys and skin. Excretion is an essential process in all forms of life. For example, in mammals urine is expelled through the urethra, which is part of the excretory system. In unicellular organisms, waste products are discharged directly through the surface of the cell.

During life activities such as cellular respiration, several chemical reactions take place in the body. These are known as metabolism. These chemical reactions produce waste products such as carbon dioxide, water, salts, urea and uric acid. Accumulation of these wastes beyond a level inside the body is harmful to the body. The excretory organs remove these wastes. This process of removal of metabolic waste from the body is known as excretion.

Green plants produce carbon dioxide and water as respiratory products. In green plants, the carbon dioxide released during respiration gets utilized during photosynthesis. Oxygen is a by product generated during photosynthesis, and exits through stomata, root cell walls, and other routes. Plants can get rid of excess water by transpiration and guttation. It has been shown that the leaf acts as an 'excretophore' and, in addition to being a primary organ of photosynthesis, is also used as a method of excreting toxic wastes via diffusion. Other waste materials that are exuded by some plants — resin, saps, latex, etc. are forced from the interior of the plant by hydrostatic pressures inside the plant and by absorptive forces of plant cells. These latter processes do not need added energy, they act passively. However, during the pre-abscission phase, the metabolic levels of a leaf are high. Plants also excrete some waste substances into the soil around them.

In animals, the main excretory products are carbon dioxide, ammonia (in many aquatic invertebrates), urea (in amphibians and mammals), uric acid (in insects, birds and other reptiles), guanine (in spiders) and creatine. The liver and kidneys clear many substances from the blood (for example, in renal excretion), and the cleared substances are then excreted from the body in the urine and feces.

24.1 Nitrogenous Waste

The nitrogen compounds through which excess nitrogen is eliminated from organisms are called nitrogenous wastes or nitrogen wastes. They are ammonia, urea, uric acid, and creatinine. All of these substances are produced from protein metabolism. In many animals, the urine is the main route of excretion for such wastes; in some, the feces is.

Aquatic animals usually excrete ammonia directly into the external environment, as this compound has high solubility and there is ample water available for dilution. In terrestrial animals ammonia-like compounds are converted into other nitrogenous materials as there is less water in the environment and ammonia itself is toxic.

Birds excrete their nitrogenous wastes as uric acid in the form of a paste. Although this process is metabolically more expensive, it allows more efficient water retention and it can be stored more easily in the egg. Many avian species, especially seabirds, can also excrete salt via specialized nasal salt glands, the saline solution leaving through nostrils in the beak.

In insects, a system involving Malpighian tubules is utilized to excrete metabolic waste. Metabolic waste diffuses or is actively transported into the tubule, which transports the wastes to the intestines. The metabolic waste is then released from the body along with fecal matter.

24.2 Excretory Systems

The excretory system is a passive biological system that removes excess, unnecessary materials from the body fluids of an organism, so as to help maintain internal chemical homeostasis and prevent damage to the body. The dual function of excretory systems is the elimination of the waste products of metabolism and to drain the body of used up and broken down components in a liquid and gaseous state. In humans and other amniotes (mammals, birds and reptiles) most of these substances leave the body as urine and to some degree exhalation, mammals also expel them through sweating.

Only the organs specifically used for the excretion are considered a part of the excretory system. In the narrow sense, the term refers to the urinary system. However, as excretion involves several functions that are only superficially related, it is not usually used in more formal classifications of anatomy or function.

As most healthy functioning organs produce metabolic and other wastes, the entire organism depends on the function of the system. Breaking down of one or more of the systems is a serious health condition, for example kidney failure.

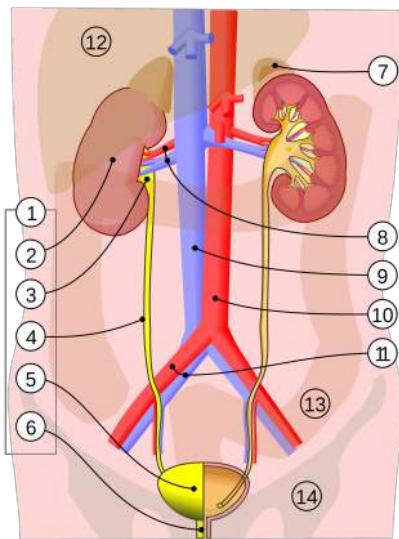


Figure 24.3: The human urinary system.³ 1. Human urinary system: 2. Kidney, 3. Renal pelvis, 4. Ureter, 5. Urinary bladder, 6. Urethra. (Left side with frontal section) 7. Adrenal gland Vessels: 8. Renal artery and vein, 9. Inferior vena cava, 10. Abdominal aorta, 11. Common iliac artery and vein Transparent: 12. Liver, 13. Large intestine, 14. Pelvis

24.3 The Human Urinary System

The urinary system, also known as the renal system or urinary tract, consists of the kidneys, ureters, bladder, and the urethra. The purpose of the urinary system is to eliminate waste from the body, regulate blood volume and blood pressure, control levels of electrolytes and metabolites, and regulate blood pH. The urinary tract is the body's drainage system for the eventual removal of urine. The kidneys have an extensive blood supply via the renal arteries which leave the kidneys via the renal vein. Each kidney consists of functional units called nephrons. Following filtration of blood and further processing, wastes (in the form of urine) exit the kidney via the ureters, tubes made of smooth muscle fibres that propel urine towards the urinary bladder, where it is stored and subsequently expelled from the body by urination (voiding). The female and male urinary system are very similar, differing only in the length of the urethra.

Urine is formed in the kidneys through a filtration of blood. The urine is then passed through the ureters to the bladder, where it is stored. During urination, the urine is passed from the bladder through the urethra to the outside of the body.

800–2,000 milliliters (mL) of urine are normally produced every day in a healthy human. This amount varies

according to fluid intake and kidney function.

The urinary system refers to the structures that produce and transport urine to the point of excretion. In the human urinary system there are two kidneys that are located between the dorsal body wall and parietal peritoneum on both the left and right sides.

The formation of urine begins within the functional unit of the kidney, the nephrons. Urine then flows through the nephrons, through a system of converging tubules called collecting ducts. These collecting ducts then join together to form the minor calyces, followed by the major calyces that ultimately join the renal pelvis. From here, urine continues its flow from the renal pelvis into the ureter, transporting urine into the urinary bladder. The anatomy of the human urinary system differs between males and females at the level of the urinary bladder. In males, the urethra begins at the internal urethral orifice in the trigone of the bladder, continues through the external urethral orifice, and then becomes the prostatic, membranous, bulbar, and penile urethra. Urine exits through the external urethral meatus. The female urethra is much shorter, beginning at the bladder neck and terminating in the vaginal vestibule.

The main functions of the urinary system and its components are to

- Regulate blood volume and composition (e.g. sodium, potassium and calcium)
- Regulate blood pressure.
- Regulate pH homeostasis of the blood.
- Contributes to the production of red blood cells by the kidney.
- Helps synthesize calcitriol the (active form of Vitamin D).
- Stores waste product (mainly urea and uric acid) before it and other products are removed from the body.

Average urine production in adult humans is about 1–2 litres (L) per day, depending on state of hydration, activity level, environmental factors, weight, and the individual's health. Producing too much or too little urine requires medical attention. Polyuria is a condition of excessive urine production (> 2.5 L/day). Oliguria when < 400 mL (millilitres) are produced, and anuria one of < 100 mL per day.

The first step in urine formation is the filtration of blood in the kidneys. In a healthy human the kidney receives between 12 and 30% of cardiac output, but it averages about 20% or about 1.25 L/min.

The kidneys are two bean-shaped organs found in vertebrates. They are located on the left and right in the retroperitoneal space, and in adult humans are about 12

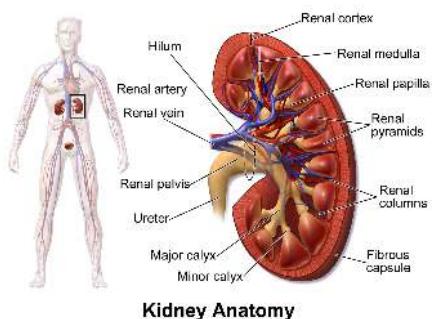


Figure 24.4: The kidneys lie in the retroperitoneal space behind the abdomen, and act to filter blood to create urine.⁴

centimetres (4 1/2 inches) in length. They receive blood from the paired renal arteries; blood exits into the paired renal veins. Each kidney is attached to a ureter, a tube that carries excreted urine to the bladder.

The nephron is the structural and functional unit of the kidney. Each human adult kidney contains around 1 million nephrons, while a mouse kidney contains only about 12,500 nephrons. The kidney participates in the control of the volume of various body fluids, fluid osmolality, acid-base balance, various electrolyte concentrations, and removal of toxins. Filtration occurs in the glomerulus: one-fifth of the blood volume that enters the kidneys is filtered. Examples of substances reabsorbed are solute-free water, sodium, bicarbonate, glucose, and amino acids. Examples of substances secreted are hydrogen, ammonium, potassium and uric acid. The kidneys also carry out functions independent of the nephron. For example, they convert a precursor of vitamin D to its active form, calcitriol; and synthesize the hormones erythropoietin and renin.

In humans, the kidneys are located high in the abdominal cavity, one on each side of the spine, and lie in a retroperitoneal position at a slightly oblique angle. The asymmetry within the abdominal cavity, caused by the position of the liver, typically results in the right kidney being slightly lower and smaller than the left, and being placed slightly more to the middle than the left kidney. The left kidney is approximately at the vertebral level T12 to L3, and the right is slightly lower. The right kidney sits just below the diaphragm and posterior to the liver. The left kidney sits below the diaphragm and posterior to the spleen. On top of each kidney is an adrenal gland. The upper parts of the kidneys are partially protected by the 11th and 12th ribs. Each kidney, with its adrenal gland is surrounded by two layers of fat: the perirenal fat present between renal fascia and renal capsule and pararenal fat superior to the renal fascia.

The kidney is a bean-shaped structure with a convex

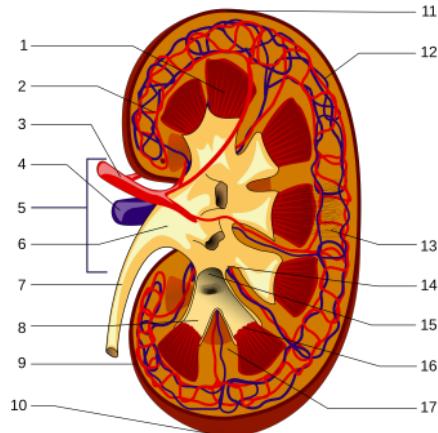


Figure 24.5: The structure of the kidney.⁵ 1. Renal pyramid • 2. Interlobular artery • 3. Renal artery • 4. Renal vein 5. Renal hilum • 6. Renal pelvis • 7. Ureter • 8. Minor calyx • 9. Renal capsule • 10. Inferior renal capsule • 11. Superior renal capsule • 12. Interlobular vein • 13. Nephron • 14. Renal sinus • 15. Major calyx • 16. Renal papilla • 17. Renal column

and a concave border. A recessed area on the concave border is the renal hilum, where the renal artery enters the kidney and the renal vein and ureter leave. The kidney is surrounded by tough fibrous tissue, the renal capsule, which is itself surrounded by perirenal fat, renal fascia, and pararenal fat. The anterior (front) surface of these tissues is the peritoneum, while the posterior (rear) surface is the transversalis fascia.

The superior pole of the right kidney is adjacent to the liver. For the left kidney, it is next to the spleen. Both, therefore, move down upon inhalation.

The basic structural and functional unit of the kidney is the nephron. Its chief function is to regulate the concentration of water and soluble substances like sodium by filtering the blood, reabsorbing what is needed and excreting the rest as urine.

In the first part of the nephron, Bowman's capsule filters blood from the circulatory system into the tubules. Hydrostatic and osmotic pressure gradients facilitate filtration across a semipermeable membrane. The filtrate includes water, small molecules, and ions that easily pass through the filtration membrane. However larger molecules such as proteins and blood cells are prevented from passing through the filtration membrane. The amount of filtrate produced every minute is called the glomerular filtration rate or GFR and amounts to 180 litres per day. About 99% of this filtrate is reabsorbed as it passes through the nephron and the remaining 1% becomes urine.

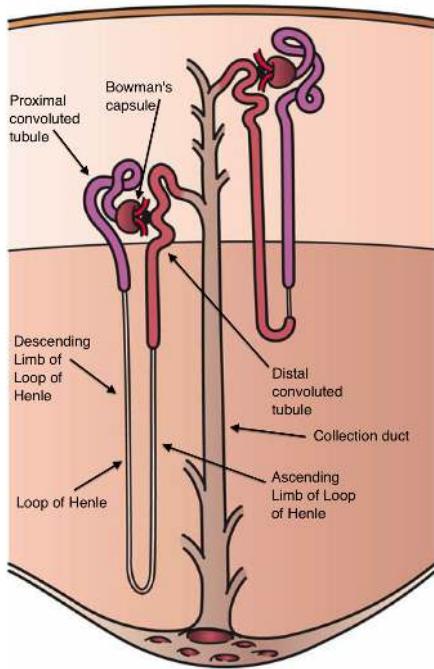


Figure 24.6: The nephron, shown here, is the functional unit of the kidneys. Its parts are labelled except the (gray) connecting tubule located after the (dark red) distal convoluted tubule and before the large (gray) collecting duct (mislabeled collection duct).⁶

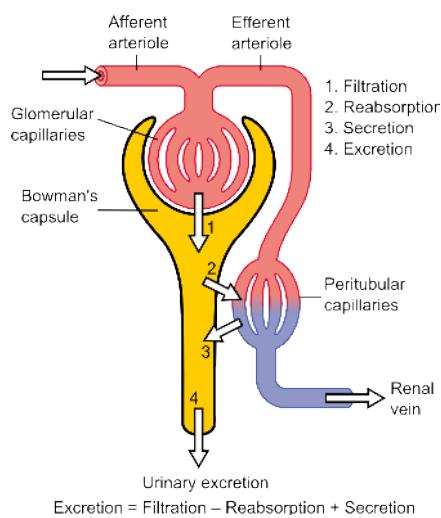


Figure 24.7: (ref.)

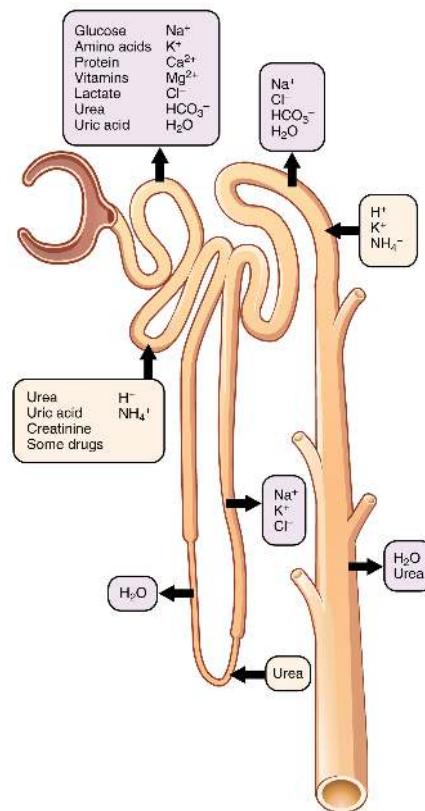


Figure 24.8: Secretion and reabsorption of water, ions and various substances throughout the nephron.⁷

The urinary system is regulated by the endocrine system by hormones such as antidiuretic hormone, aldosterone, and parathyroid hormone.

The urinary system is under influence of the circulatory system, nervous system, and endocrine system.

Aldosterone plays a central role in regulating blood pressure through its effects on the kidney. It acts on the distal tubules and collecting ducts of the nephron and increases reabsorption of sodium from the glomerular filtrate. Reabsorption of sodium results in retention of water, which increases blood pressure and blood volume. Antidiuretic hormone (ADH), is a neurohypophyseal hormone found in most mammals. Its two primary functions are to retain water in the body and vasoconstriction. Vasopressin regulates the body's retention of water by increasing water reabsorption in the collecting ducts of the kidney nephron. Vasopressin increases water permeability of the kidney's collecting duct and distal convoluted tubule by inducing translocation of aquaporin-CD water channels in the kidney nephron collecting duct plasma membrane.

In the majority of vertebrates, the mesonephros persists into the adult, albeit usually fused with the more advanced metanephros; only in amniotes is the mesonephros restricted to the embryo. The kidneys of fish and amphibians are typically narrow, elongated organs, occupying a significant portion of the trunk. The collecting ducts from each cluster of nephrons usually drain into an archinephric duct, which is homologous with the vas deferens of amniotes. However, the situation is not always so simple; in cartilaginous fish and some amphibians, there is also a shorter duct, similar to the amniote ureter, which drains the posterior (metanephric) parts of the kidney, and joins with the archinephric duct at the bladder or cloaca. Indeed, in many cartilaginous fish, the anterior portion of the kidney may degenerate or cease to function altogether in the adult.

In the most primitive vertebrates, the hagfish and lampreys, the kidney is unusually simple: it consists of a row of nephrons, each emptying directly into the archinephric duct. Invertebrates may possess excretory organs that are sometimes referred to as "kidneys", but, even in *Amphioxus*, these are never homologous with the kidneys of vertebrates, and are more accurately referred to by other names, such as nephridia. In amphibians, kidneys and the urinary bladder harbour specialized parasites, monogeneans of the family Polystomatidae.

The kidneys of reptiles consist of a number of lobules arranged in a broadly linear pattern. Each lobule contains a single branch of the ureter in its centre, into which the collecting ducts empty. Reptiles have relatively few nephrons compared with other amniotes of a similar size, possibly

because of their lower metabolic rate.

Birds have relatively large, elongated kidneys, each of which is divided into three or more distinct lobes. The lobes consist of several small, irregularly arranged, lobules, each centred on a branch of the ureter. Birds have small glomeruli, but about twice as many nephrons as similarly sized mammals.

The human kidney is fairly typical of that of mammals. Distinctive features of the mammalian kidney, in comparison with that of other vertebrates, include the presence of the renal pelvis and renal pyramids and a clearly distinguishable cortex and medulla. The latter feature is due to the presence of elongated loops of Henle; these are much shorter in birds, and not truly present in other vertebrates (although the nephron often has a short intermediate segment between the convoluted tubules). It is only in mammals that the kidney takes on its classical "kidney" shape, although there are some exceptions, such as the multi-lobed reniculate kidneys of pinnipeds and cetaceans.

Chapter 25

Nervous Systems

The nervous system is a highly complex part of an animal that coordinates its actions and sensory information by transmitting signals to and from different parts of its body. The nervous system detects environmental changes that impact the body, then works in tandem with the endocrine system to respond to such events.

The nervous system derives its name from nerves, which are cylindrical bundles of fibers (the axons of neurons), that emanate from the brain and spinal cord, and branch repeatedly to innervate every part of the body. Nerves are large enough to have been recognized by the ancient Egyptians, Greeks, and Romans, but their internal structure was not understood until it became possible to examine them using a microscope.

Neuroscience (or neurobiology) is a multidisciplinary branch of biology that combines physiology, anatomy, molecular biology, developmental biology, cytology, mathematical modeling, and psychology to understand the fundamental and emergent properties of neurons and neural circuits. The understanding of the biological basis of learning, memory, behavior, perception, and consciousness has been described as the “ultimate challenge” of the biological sciences. The human brain is often referred to as the most complicated structure in the universe. The scope of neuroscience has broadened over time to include different approaches used to study the nervous system at different scales and the techniques used by neuroscientists have expanded enormously, from molecular and cellular studies of individual neurons to imaging of sensory, motor and cognitive tasks in the brain. Malfunction of the nervous system can occur as a result of genetic defects, physical damage due to trauma or toxicity, infection or simply of ageing. The medical specialty of neurology studies disorders of the nervous system and looks for interventions that can prevent or treat them. Although mental illnesses are believed by many to be neurological disorders affecting the central nervous system, traditionally they are classified separately, and treated by psychiatrists.

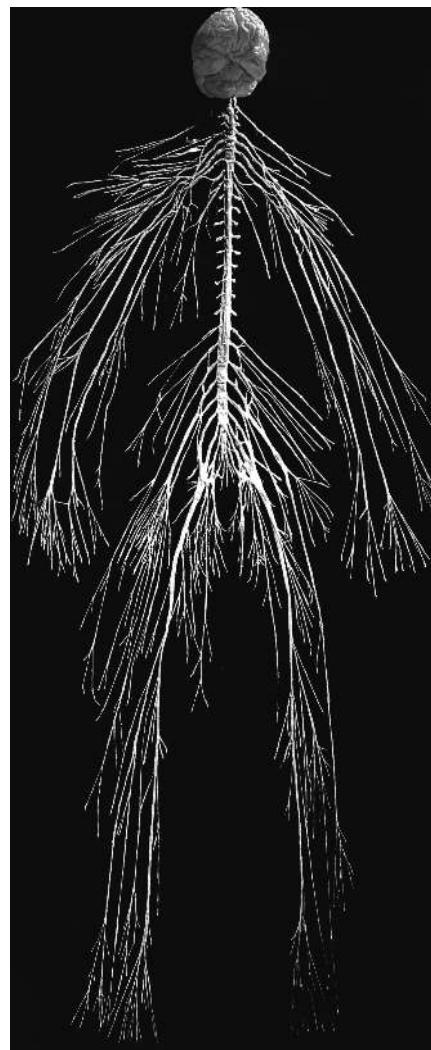


Figure 25.1: The human nervous system.

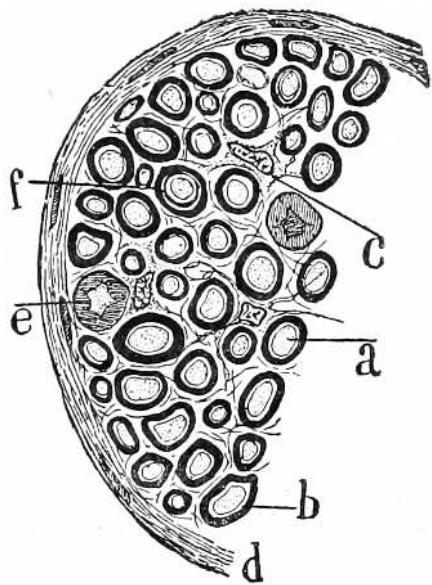


Figure 25.2: Transverse section of a nerve. a) a single nerve fibre (axon) surrounded by a thick layer of myelin. c) an interstitial cell. Histologie du système nerveux de l'homme & des vertébrés, Tome Premier¹ (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay.

Nervous systems are found in most multicellular animals, but vary greatly in complexity. The only multicellular animals that have no nervous system at all are sponges, placozoans, and mesozoans, which have very simple body plans. However, even sponges, unicellular animals, and even protists such as slime molds have cell-to-cell signalling mechanisms that are precursors to those of neurons. The nervous systems of the radially symmetric organisms ctenophores (comb jellies) and cnidarians (which include anemones, hydras, corals and jellyfish) consist of a diffuse nerve net. All other animal species, with the exception of a few types of worm, have a nervous system containing a brain, a central cord (or two cords running in parallel), and nerves radiating from the brain and central cord. The size of the nervous system ranges from a few hundred cells in the simplest worms, to around 300 billion cells in African elephants.

Nervous tissue first arose in wormlike organisms about 550 to 600 million years ago. In humans and other vertebrates it consists of two main parts, the central nervous system (CNS) and the peripheral nervous system (PNS). The CNS consists of the brain and spinal cord. The PNS consists mainly of nerves, which are enclosed bundles of the long fibers or axons, that connect the CNS to every other part of the body. Nerves that transmit signals from the brain are called motor or efferent nerves, while those nerves that transmit information from the body to the CNS are called sensory or afferent. Spinal nerves serve both functions and are called mixed nerves. The PNS is divided into three separate subsystems, the somatic, autonomic, and enteric nervous systems. Somatic nerves carry sensory information from the periphery to the CNS and signals for voluntary movement from the CNS to the muscles. The autonomic nervous system is further subdivided into the sympathetic and the parasympathetic nervous systems. The sympathetic nervous system is activated in cases of emergencies to mobilize energy, while the parasympathetic nervous system is activated when organisms are in a relaxed state. The enteric nervous system functions to control the gastrointestinal system. Both autonomic and enteric nervous systems function involuntarily. Nerves that exit from the cranium are called cranial nerves while those exiting from the spinal cord are called spinal nerves.

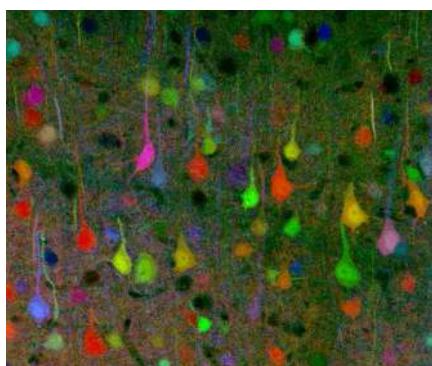


Figure 25.3: A rainbow of mouse neurons.²

25.1 The Cells Of The Nervous System

At the cellular level, the nervous system is defined by the presence of a special type of cell, called the neuron, also known as a nerve cell. Neurons have special structures that allow them to receive and send signals from and to other cells. They send these signals in the form of elec-

trochemical waves traveling along thin fibers called axons, which cause chemicals called neurotransmitters to be released at junctions called synapses. Neurons usually receive signals at tree-like processes called dendrites. A cell that receives a synaptic signal from another neuron may be excited, inhibited, or otherwise modulated. The connections between neurons can form neural pathways, neural circuits, and larger networks that generate an organism's perception of the world and determine its behavior. Along with neurons, the nervous system contains other specialized cells called glial cells (or simply glia), which provide structural and metabolic support.

Even in the nervous system of a single species such as humans, hundreds of different types of neurons exist, with a wide variety of morphologies and functions. These include sensory neurons that convert physical stimuli such as light and sound into neural signals, and motor neurons that activate muscles or glands; however in many species the great majority of neurons participate in the formation of centralized structures (the brain and ganglia) and they receive all of their input from other neurons and send their output to other neurons.

Glial cells (named from the Greek for “glue”) are non-neuronal cells that provide support and nutrition, maintain homeostasis, form myelin, and participate in signal transmission in the nervous system. In the human brain, it is estimated that the total number of glia roughly equals the number of neurons, although the proportions vary in different brain areas. Among the most important functions of glial cells are to support neurons and hold them in place; to supply nutrients to neurons; to insulate neurons electrically; to destroy pathogens and remove dead neurons; and to provide guidance cues directing the axons of neurons to their targets. A very important type of glial cell (oligodendrocytes in the central nervous system, and Schwann cells in the peripheral nervous system) generates layers of a fatty substance called myelin that wraps around axons and provides electrical insulation which allows them to transmit action potentials much more rapidly and efficiently. Microglia serve as important resident immune cells within the central nervous system.

25.2 Comparative Anatomy And Evolution Of Nervous Systems

Porifera (sponges) have no cells connected to each other by synaptic junctions, that is, no neurons, and therefore no nervous system. They do, however, have homologs of many genes that play key roles in synaptic function. Recent studies have shown that sponge cells express a group of proteins that cluster together to form a structure resembling a postsynaptic density (the signal-receiving part of a

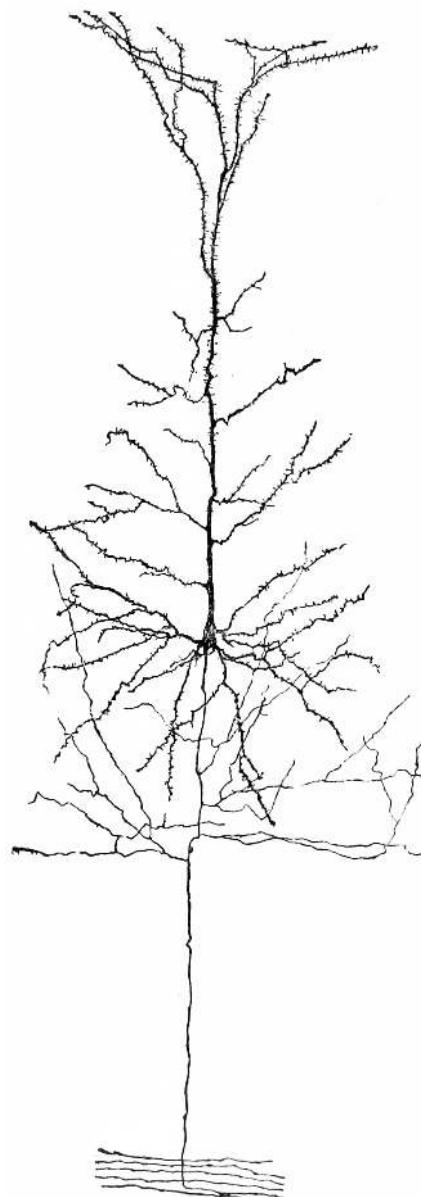


Figure 25.4: A nerve cell from the cerebral cortex of a rabbit. Notice the extensive tree of dendrites at the top, the long axon at the bottom. Because of the pyramid-like shape of the cell body, this type of neuron is referred to as a pyramidal cell.

synapse). However, the function of this structure is currently unclear. Although sponge cells do not show synaptic transmission, they do communicate with each other via calcium waves and other impulses, which mediate some simple actions such as whole-body contraction.

Radiata such as the cnidaria (jellyfish) and ctenophora (comb jellies) have diffuse nerve nets rather than a central nervous system. In most jellyfish the nerve net is spread more or less evenly across the body; in comb jellies it is concentrated near the mouth. The nerve nets consist of sensory neurons, which pick up chemical, tactile, and visual signals; motor neurons, which can activate contractions of the body wall; and intermediate neurons, which detect patterns of activity in the sensory neurons and, in response, send signals to groups of motor neurons. In some cases groups of intermediate neurons are clustered into discrete ganglia.

The development of the nervous system in radiata is relatively unstructured. Unlike bilaterians, radiata only have two primordial cell layers, endoderm and ectoderm. Neurons are generated from a special set of ectodermal precursor cells, which also serve as precursors for every other ectodermal cell type.

The vast majority of existing animals are bilaterians, meaning animals with left and right sides that are approximate mirror images of each other. All bilateria are thought to have descended from a common wormlike ancestor that appeared in the Ediacaran period, 550–600 million years ago. The fundamental bilaterian body form is a tube with a hollow gut cavity running from mouth to anus, and a nerve cord with an enlargement (a “ganglion”) for each body segment, with an especially large ganglion at the front, called the “brain”.

Even mammals, including humans, show the segmented bilaterian body plan at the level of the nervous system. The spinal cord contains a series of segmental ganglia, each giving rise to motor and sensory nerves that innervate a portion of the body surface and underlying musculature. On the limbs, the layout of the innervation pattern is complex, but on the trunk it gives rise to a series of narrow bands. The top three segments belong to the brain, giving rise to the forebrain, midbrain, and hindbrain.

Bilaterians can be divided, based on events that occur very early in embryonic development, into two groups (superphyla) called protostomes and deuterostomes. Deuterostomes include vertebrates as well as echinoderms, hemichordates (mainly acorn worms), and Xenoturbellidans. Protostomes, the more diverse group, include arthropods, molluscs, and numerous types of worms. There is a basic difference between the two groups

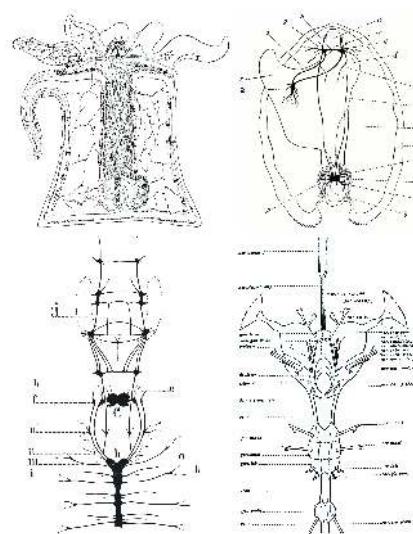


Figure 25.5: Comparison of nervous systems of invertebrates. Top left: A diffuse nerve net in *Actinia* (a genus of sea anemones in the family *Actiniidae* in the phylum *Cnidaria*); top right: The nervous system of *Anadonta anatina*, a freshwater mussel in the family *Unionidae* in the phylum *Mollusca*. c, foot; k, pedal ganglion; i, cerebro-pedal connective; g, cerebral ganglion; h, cerebral connective; a, anterior adductor muscle; r, q, anterior pallial nerves; d, liver; s, visceral nerve; l, cerebro-visceral connective; e, gill; f, edge of mantle; n, branchial nerves; m, visceral ganglion; o, posterior pallial nerves; b, posterior adductor muscle; p, lateral pallial nerves; bottom left: the nervous system of *Alitta virens*, a polychaete worm in the phylum *Annelida*. J, jaws; b, antennal nerves; c, palpal nerves; f, ganglia for the dorsal peristomial cirri; n¹, ganglion; n, nerves for the dissepimenta; m, parapodial nerves; i, parapodial branch; h, ventral chain of ganglia; C, cerebral ganglion; o, nerve passing through dissepiment to preceding segment; k, parapodial ganglion. Bottom right: the nervous system of an insect (*Arthropoda*). From Morphology of invertebrate types, by Alexander Petrunkevitch. New York, Macmillan company, 1916.³

in the placement of the nervous system within the body: protostomes possess a nerve cord on the ventral (usually bottom) side of the body, whereas in deuterostomes the nerve cord is on the dorsal (usually top) side. In fact, numerous aspects of the body are inverted between the two groups, including the expression patterns of several genes that show dorsal-to-ventral gradients. Most anatomists now consider that the bodies of protostomes and deuterostomes are “flipped over” with respect to each other, a hypothesis that was first proposed by Geoffroy Saint-Hilaire for insects in comparison to vertebrates. Thus insects, for example, have nerve cords that run along the ventral midline of the body, while all vertebrates have spinal cords that run along the dorsal midline.

There are a few types of existing bilaterians that lack a recognizable brain, including echinoderms and tunicates. It has not been definitively established whether the existence of these brainless species indicates that the earliest bilaterians lacked a brain, or whether their ancestors evolved in a way that led to the disappearance of a previously existing brain structure.

The diversity of invertebrate body plans is matched by an equal diversity in brain structures. Two groups of invertebrates have notably complex brains: arthropods (insects, crustaceans, arachnids, and others), and cephalopods (octopuses, squids, and similar molluscs). The brains of arthropods and cephalopods arise from twin parallel nerve cords that extend through the body of the animal. Arthropods have a central brain, the supraesophageal ganglion, with three divisions and large optical lobes behind each eye for visual processing. Cephalopods such as the octopus and squid have the largest brains of any invertebrates.

There are several invertebrate species whose brains have been studied intensively because they have properties that make them convenient for experimental work:

- Fruit flies (*Drosophila melanogaster*), because of the large array of techniques available for studying their genetics, have been a natural subject for studying the role of genes in brain development. In spite of the large evolutionary distance between insects and mammals, many aspects of *Drosophila* neurogenetics have been shown to be relevant to humans. The first biological clock genes, for example, were identified by examining *Drosophila* mutants that showed disrupted daily activity cycles. A search in the genomes of vertebrates revealed a set of analogous genes, which were found to play similar roles in the mouse biological clock—and therefore almost certainly in the human biological clock as well. Studies done on *Drosophila*, also show that

most neuropil regions of the brain are continuously reorganized throughout life in response to specific living conditions.

- The nematode worm *Caenorhabditis elegans*, like *Drosophila*, has been studied largely because of its importance in genetics. In the early 1970s, Sydney Brenner⁴ chose it as a model organism for studying the way that genes control development. One of the advantages of working with this worm is that the body plan is very stereotyped: the nervous system of the hermaphrodite contains exactly 302 neurons, always in the same places, making identical synaptic connections in every worm. Brenner’s team sliced worms into thousands of ultrathin sections and photographed each one under an electron microscope, then visually matched fibers from section to section, to map out every neuron and synapse in the entire body. The complete neuronal wiring diagram of *C. elegans* – its connectome was achieved. Nothing approaching this level of detail is available for any other organism, and the information gained has enabled a multitude of studies that would otherwise have not been possible.
- The sea slug *Aplysia californica* was chosen by Nobel Prize-winning neurophysiologist Eric Kandel⁵ as a model for studying the cellular basis of learning and memory, because of the simplicity and accessibility of its nervous system, and it has been examined in hundreds of experiments.

Worms are the simplest bilaterian animals, and reveal the basic structure of the bilaterian nervous system in the most straightforward way. As an example, earthworms have dual nerve cords running along the length of the body and merging at the tail and the mouth. These nerve cords are connected by transverse nerves like the rungs of a ladder. These transverse nerves help coordinate the two sides of the animal. Two ganglia at the head (the “nerve ring”) end function similar to a simple brain. Photoreceptors on the animal’s eyespots provide sensory information on light and dark.

Arthropods, such as insects and crustaceans, have a nervous system made up of a series of ganglia, connected by a ventral nerve cord made up of two parallel connectives running along the length of the belly. Typically, each body segment has one ganglion on each side, though some ganglia are fused to form the brain and other large ganglia. The head segment contains the brain, also known as the supraesophageal ganglion. In the insect nervous system, the brain is anatomically divided into the protocerebrum, deutocerebrum, and tritocerebrum. Immediately behind the brain is the subesophageal gan-

⁴https://en.wikipedia.org/wiki/Sydney_Brenner

⁵https://en.wikipedia.org/wiki/Eric_Kandel

glion, which is composed of three pairs of fused ganglia. It controls the mouthparts, the salivary glands and certain muscles. Many arthropods have well-developed sensory organs, including compound eyes for vision and antennae for olfaction and pheromone sensation. The sensory information from these organs is processed by the brain.

In insects, many neurons have cell bodies that are positioned at the edge of the brain and are electrically passive—the cell bodies serve only to provide metabolic support and do not participate in signalling. A protoplasmic fiber runs from the cell body and branches profusely, with some parts transmitting signals and other parts receiving signals. Thus, most parts of the insect brain have passive cell bodies arranged around the periphery, while the neural signal processing takes place in a tangle of protoplasmic fibers called neuropil, in the interior.

Brains are most simply compared in terms of their size. The relationship between brain size, body size and other variables has been studied across a wide range of vertebrate species. As a rule, brain size increases with body size, but not in a simple linear proportion. In general, smaller animals tend to have larger brains, measured as a fraction of body size. For mammals, the relationship between brain volume and body mass essentially follows a power law with an exponent of about 0.75. This formula describes the central tendency, but every family of mammals departs from it to some degree, in a way that reflects in part the complexity of their behavior. For example, primates have brains 5 to 10 times larger than the formula predicts. Predators tend to have larger brains than their prey, relative to body size.

All vertebrate brains share a common underlying form, which appears most clearly during early stages of embryonic development. In its earliest form, the brain appears as three swellings at the front end of the neural tube; these swellings eventually become the forebrain, midbrain, and hindbrain (the prosencephalon, mesencephalon, and rhombencephalon, respectively). At the earliest stages of brain development, the three areas are roughly equal in size. In many classes of vertebrates, such as fish and amphibians, the three parts remain similar in size in the adult, but in mammals the forebrain becomes much larger than the other parts, and the midbrain becomes very small.

The brains of vertebrates are made of very soft tissue. Living brain tissue is pinkish on the outside and mostly white on the inside, with subtle variations in color. Vertebrate brains are surrounded by a system of connective tissue membranes called meninges that separate the skull from the brain. Blood vessels enter the central nervous system through holes in the meningeal layers. The cells in the blood vessel walls are joined tightly to one another,

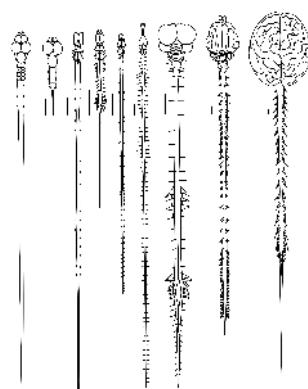


Figure 25.6: Dorsal views of the central nervous systems of the teleosts (from left to right) *Trigla hirundo* (a) and *Mola mola*, the urodele *Ambystoma tigrinum*, the anuran *Xenopus laevis*, the tortoise *Testudo hermanni*, the tegu lizard *Tupinambis teguixin*, the pigeon, the cat and human. In a, c, d, e, f, g and j the full length of the spinalcord, including the filum terminale (where present) is shown; in *Mola mola* and the cat most of the filum terminale is cut. Vertical black bars correspond to 1 cm in length. Modified from Nieuwenhuys, R., ten Donkelaar, H. J., & Nicholson, C. (1998). The Meaning of It All. The Central Nervous System of Vertebrates, 2135-2195⁶

forming the blood-brain barrier, which blocks the passage of many toxins and pathogens (though at the same time blocking antibodies and some drugs, thereby presenting special challenges in treatment of diseases of the brain).

Neuroanatomists usually divide the vertebrate brain into six main regions: the telencephalon (cerebral hemispheres), diencephalon (thalamus and hypothalamus), mesencephalon (midbrain), cerebellum, pons, and medulla oblongata. Each of these areas has a complex internal structure. Some parts, such as the cerebral cortex and the cerebellar cortex, consist of layers that are folded or convoluted to fit within the available space. Other parts, such as the thalamus and hypothalamus, consist of clusters of many small nuclei. Thousands of distinguishable areas can be identified within the vertebrate brain based on fine distinctions of neural structure, chemistry, and connectivity.

There is an anatomical convention that a cluster of neurons in the brain or spinal cord is called a nucleus, whereas a cluster of neurons in the periphery is called a ganglion. There are, however, a few exceptions to this rule, notably including the part of the forebrain called the basal ganglia.

Although the same basic components are present in all vertebrate brains, some branches of vertebrate evolution have led to substantial distortions of brain geometry, especially in the forebrain area. The brain of a shark shows the basic components in a straightforward way, but in teleost fishes (the great majority of existing fish species), the forebrain has become “everted”, like a sock turned inside out. In birds, there are also major changes in forebrain structure. These distortions can make it difficult to match brain components from one species with those of another species.

Here is a list of some of the most important vertebrate brain components, along with a brief description of their functions as currently understood:

- The medulla, along with the spinal cord, contains many small nuclei involved in a wide variety of sensory and involuntary motor functions such as vomiting, heart rate and digestive processes.
- The pons lies in the brainstem directly above the medulla. Among other things, it contains nuclei that control often voluntary but simple acts such as sleep, respiration, swallowing, bladder function, equilibrium, eye movement, facial expressions, and posture.
- The hypothalamus is a small region at the base of the forebrain, whose complexity and importance belies its size. It is composed of numerous small nuclei, each with distinct connections and neurochemistry. The hypothalamus is engaged in additional

involuntary or partially voluntary acts such as sleep and wake cycles, eating and drinking, and the release of some hormones.

- The thalamus is a collection of nuclei with diverse functions: some are involved in relaying information to and from the cerebral hemispheres, while others are involved in motivation. The subthalamic area (*zona incerta*) seems to contain action-generating systems for several types of “consummatory” behaviors such as eating, drinking, defecation, and copulation.
- The cerebellum modulates the outputs of other brain systems, whether motor related or thought related, to make them certain and precise. Removal of the cerebellum does not prevent an animal from doing anything in particular, but it makes actions hesitant and clumsy. This precision is not built-in, but learned by trial and error. The muscle coordination learned while riding a bicycle is an example of a type of neural plasticity that may take place largely within the cerebellum. 10% of the brain’s total volume consists of the cerebellum and 50% of all neurons are held within its structure.
- The optic tectum allows actions to be directed toward points in space, most commonly in response to visual input. In mammals it is usually referred to as the superior colliculus, and its best-studied function is to direct eye movements. It also directs reaching movements and other object-directed actions. It receives strong visual inputs, but also inputs from other senses that are useful in directing actions, such as auditory input in owls and input from the thermosensitive pit organs in snakes. In some primitive fishes, such as lampreys, this region is the largest part of the brain. The superior colliculus is part of the midbrain.
- The pallium is a layer of gray matter that lies on the surface of the forebrain and is the most complex and most recent evolutionary development of the brain as an organ. In reptiles and mammals, it is called the cerebral cortex. Multiple functions involve the pallium, including smell and spatial memory. In mammals, where it becomes so large as to dominate the brain, it takes over functions from many other brain areas. In many mammals, the cerebral cortex consists of folded bulges called gyri that create deep furrows or fissures called sulci. The folds increase the surface area of the cortex and therefore increase the amount of gray matter and the amount of information that can be stored and processed.
- The hippocampus, strictly speaking, is found only in mammals. However, the area it derives from, the medial pallium, has counterparts in all vertebrates. There is evidence that this part of the brain is in-

volved in complex events such as spatial memory and navigation in fishes, birds, reptiles, and mammals.

- The basal ganglia are a group of interconnected structures in the forebrain. The primary function of the basal ganglia appears to be action selection: they send inhibitory signals to all parts of the brain that can generate motor behaviors, and in the right circumstances can release the inhibition, so that the action-generating systems are able to execute their actions. Reward and punishment exert their most important neural effects by altering connections within the basal ganglia.
- The olfactory bulb is a special structure that processes olfactory sensory signals and sends its output to the olfactory part of the pallium. It is a major brain component in many vertebrates, but is greatly reduced in humans and other primates (whose senses are dominated by information acquired by sight rather than smell).

The most obvious difference between the brains of mammals and other vertebrates is in terms of size. On average, a mammal has a brain roughly twice as large as that of a bird of the same body size, and ten times as large as that of a reptile of the same body size.

Size, however, is not the only difference: there are also substantial differences in shape. The hindbrain and midbrain of mammals are generally similar to those of other vertebrates, but dramatic differences appear in the forebrain, which is greatly enlarged and also altered in structure. The cerebral cortex is the part of the brain that most strongly distinguishes mammals. In non-mammalian vertebrates, the surface of the cerebrum is lined with a comparatively simple three-layered structure called the pallium. In mammals, the pallium evolves into a complex six-layered structure called neocortex or isocortex. Several areas at the edge of the neocortex, including the hippocampus and amygdala, are also much more extensively developed in mammals than in other vertebrates.

The elaboration of the cerebral cortex carries with it changes to other brain areas. The superior colliculus, which plays a major role in visual control of behavior in most vertebrates, shrinks to a small size in mammals, and many of its functions are taken over by visual areas of the cerebral cortex. The cerebellum of mammals contains a large portion (the neocerebellum) dedicated to supporting the cerebral cortex, which has no counterpart in other vertebrates.

The brains of humans and other primates contain the same structures as the brains of other mammals, but are generally larger in proportion to body size. The encephal-

ization quotient (EQ) is used to compare brain sizes across species. It takes into account the nonlinearity of the brain-to-body relationship. Humans have an average EQ in the 7-to-8 range, while most other primates have an EQ in the 2-to-3 range. Dolphins have values higher than those of primates other than humans, but nearly all other mammals have EQ values that are substantially lower.

Most of the enlargement of the primate brain comes from a massive expansion of the cerebral cortex, especially the prefrontal cortex and the parts of the cortex involved in vision. The visual processing network of primates includes at least 30 distinguishable brain areas, with a complex web of interconnections. It has been estimated that visual processing areas occupy more than half of the total surface of the primate neocortex. The prefrontal cortex carries out functions that include planning, working memory, motivation, attention, and executive control. It takes up a much larger proportion of the brain for primates than for other species, and an especially large fraction of the human brain.

The peripheral nervous system (PNS) is a collective term for the nervous system structures that do not lie within the CNS. The large majority of the axon bundles called nerves are considered to belong to the PNS, even when the cell bodies of the neurons to which they belong reside within the brain or spinal cord. The PNS is divided into somatic and visceral parts. The somatic part consists of the nerves that innervate the skin, joints, and muscles. The cell bodies of somatic sensory neurons lie in dorsal root ganglia of the spinal cord. The visceral part, also known as the autonomic nervous system, contains neurons that innervate the internal organs, blood vessels, and glands. The autonomic nervous system itself consists of two parts: the sympathetic nervous system and the parasympathetic nervous system. Some authors also include sensory neurons whose cell bodies lie in the periphery (for senses such as hearing) as part of the PNS; others, however, omit them.

The vertebrate nervous system can also be divided into areas called gray matter and white matter. Gray matter (which is only gray in preserved tissue, and is better described as pink or light brown in living tissue) contains a high proportion of cell bodies of neurons. White matter is composed mainly of myelinated axons, and takes its color from the myelin. White matter includes all of the nerves, and much of the interior of the brain and spinal cord. Gray matter is found in clusters of neurons in the brain and spinal cord, and in cortical layers that line their surfaces.

25.3 The Human Brain

The adult human brain weighs on average about 1.2–1.4 kg (2.6–3.1 lb) which is about 2% of the total body weight,

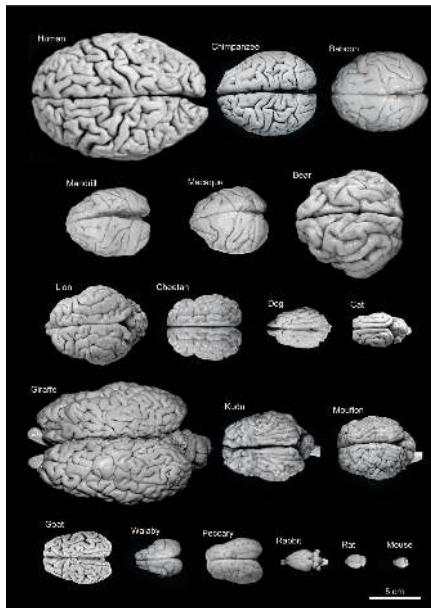


Figure 25.7: Variability of brain size and external topography. Photographs and weights of the brains of different species. Primates: human (*Homo sapiens*, 1.176 kg), chimpanzee (*Pan troglodytes*, 273 g), baboon (*Papio cynocephalus*, 151 g), mandrill (*Mandrillus sphinx*, 123 g), macaque (*Macaca tonkeana*, 110 g). Carnivores: bear (*Ursus arctos*, 289 g), lion (*Panthera leo*, 165 g), cheetah (*Acinonyx jubatus*, 119 g), dog (*Canis familiaris*, 95 g), cat (*Felis catus*, 32 g). Artiodactyls: giraffe (*Giraffa camelopardalis*, 700 g), kudu (*Tragelaphus strepsiceros*, 166 g), mouflon (*Ovis musimon*, 118 g), ibex (*Capra pyrenaica*, 115 g); peccary (*Tayassu pecari*, 41 g). Marsupials: wallaby (*Protemnodon rufogrisea*, 28 g). Lagomorphs: rabbit (*Oryctolagus cuniculus*, 5.2 g). Rodents: rat (*Rattus rattus*, 2.6 g), mouse (*Mus musculus*, 0.5 g). The chimpanzee brain was kindly supplied by Dr. Dean Falk. The rest of non-human brains were from material used in Ballesteros-Yáñez et al., 2005). Scale bar: 5 cm. From DeFelipe J (2011) The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. *Front. Neuroanat.* 5:29⁷

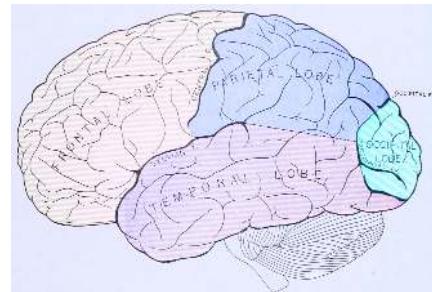


Figure 25.8: Principal lobes and fissures of the cerebrum viewed laterally. From Gray Henry, Anatomy of the Human Body. 20th Edition, Lea & Febiger, Philadelphia & New York, 1918⁸



Figure 25.9: Diagram showing a lateral view of the ridges (gyri) and grooves (sulci) of the left hemisphere of the brain. From Gray Henry, Anatomy of the Human Body. 20th Edition, Lea & Febiger, Philadelphia & New York, 1918⁹

with a volume of around 1260 cm^3 in men and 1130 cm^3 in women. There is substantial individual variation, with the standard reference range for men being 1,180–1,620 g (2.60–3.57 lb) and for women 1,030–1,400 g (2.27–3.09 lb).

The human brain is divided into nearly symmetrical left and right hemispheres by a deep groove, the longitudinal fissure. Each hemisphere is conventionally divided into four main lobes; the frontal lobe, parietal lobe, temporal lobe, and occipital lobe, named according to the skull bones that overlie them. The surface of the brain is folded into ridges (gyri) and grooves (sulci), many of which are named, usually according to their position, such as the frontal gyrus of the frontal lobe or the central sulcus separating the central regions of the hemispheres. There are many small variations in the secondary and tertiary folds.

Although the human brain represents only 2% of the body weight, it receives 15% of the cardiac output, 20% of total body oxygen consumption, and 25% of total body glucose utilization. The brain mostly uses glucose for energy, and deprivation of glucose, as can happen in hypoglycemia, can result in loss of consciousness. The energy consumption of the brain does not vary greatly over time, but active regions of the cortex consume somewhat more energy

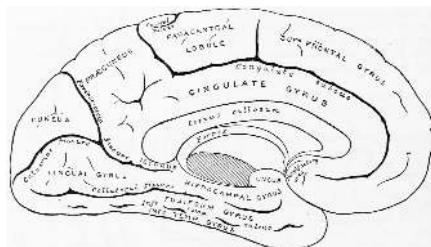


Figure 25.10: Diagram showing a medial view of the ridges (gyri) and grooves (sulci) of the left hemisphere of the brain. From Gray Henry, Anatomy of the Human Body. 20th Edition, Lea & Febiger, Philadelphia & New York, 1918¹⁰

than inactive regions: this fact forms the basis for the functional brain imaging methods PET and fMRI. These functional imaging techniques provide a three-dimensional image of metabolic activity.

The simplest way to gain information about brain anatomy is by visual inspection, but many more sophisticated techniques have been developed. Brain tissue in its natural state is too soft to work with, but it can be hardened by immersion in alcohol or other fixatives, and then sliced apart for examination of the interior. Visually, the interior of the brain consists of areas of so-called grey matter, with a dark color, separated by areas of white matter, with a lighter color. Further information can be gained by staining slices of brain tissue with a variety of chemicals that bring out areas where specific types of molecules are present in high concentrations. It is also possible to examine the microstructure of brain tissue using a microscope, and to trace the pattern of connections from one brain area to another.

25.4 Development Of The Nervous System

All bilaterian animals at an early stage of development form a gastrula, which is polarized, with one end called the animal pole and the other the vegetal pole. The gastrula has the shape of a disk with three layers of cells, an inner layer called the endoderm, which gives rise to the lining of most internal organs, a middle layer called the mesoderm, which gives rise to the bones and muscles, and an outer layer called the ectoderm, which gives rise to the skin and nervous system.

In vertebrates, the first sign of the nervous system is the appearance of a thin strip of cells along the center of the back, called the neural plate. The inner portion of the neural plate (along the midline) is destined to become the central nervous system (CNS), the outer portion the peripheral nervous system (PNS). As development proceeds, a fold called the neural groove appears along the midline. This fold deepens, and then closes up at the top. At this point the future CNS appears as a cylindrical structure called the neural tube, whereas the future PNS appears as two strips of tissue called the neural crest, running lengthwise above the neural tube. The sequence of stages from neural plate to neural tube and neural crest is known as neurulation.

In the early 20th century, a set of famous experiments by Hans Spemann¹¹ and Hilde Mangold¹² showed that the formation of nervous tissue is “induced” by signals from

¹¹https://en.wikipedia.org/wiki/Hans_Spemann

¹²https://en.wikipedia.org/wiki/Hilde_Mangold

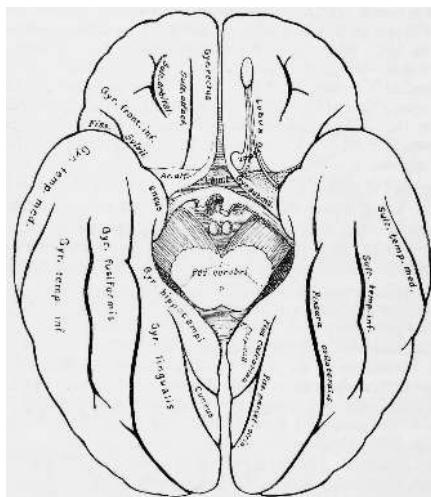


Figure 25.11: Diagram showing a view from the bottom of the ridges (gyri) and grooves (sulci) of the left hemisphere of the brain.

a group of mesodermal cells called the organizer region. For decades, though, the nature of neural induction defeated every attempt to figure it out, until finally it was resolved by genetic approaches in the 1990s. Induction of neural tissue requires inhibition of the gene for a so-called bone morphogenetic protein, or BMP. Specifically the protein BMP4 appears to be involved. Two proteins called Noggin and Chordin, both secreted by the mesoderm, are capable of inhibiting BMP4 and thereby inducing ectoderm to turn into neural tissue. It appears that a similar molecular mechanism is involved for widely disparate types of animals, including arthropods as well as vertebrates. In some animals, however, another type of molecule called Fibroblast Growth Factor or FGF may also play an important role in induction.

Induction of neural tissues causes formation of neural precursor cells, called neuroblasts. In *drosophila*, neuroblasts divide asymmetrically, so that one product is a “ganglion mother cell” (GMC), and the other is a neuroblast. A GMC divides once, to give rise to either a pair of neurons or a pair of glial cells. In all, a neuroblast is capable of generating an indefinite number of neurons or glia.

One factor common to all bilateral organisms (including humans) is a family of secreted signaling molecules called neurotrophins which regulate the growth and survival of neurons. Because neurotrophins have now been identified in both vertebrate and invertebrates, this evidence suggests that neurotrophins were present in an ancestor common to bilateral organisms and may represent a common mechanism for nervous system formation.

25.5 The Function Of The Nervous System

Organisms need information to solve at least three kinds of problems: (a) to maintain an appropriate environment, i.e., homeostasis; (b) to time activities (e.g., seasonal changes in behavior) or synchronize activities with those of conspecifics; and (c) to locate and respond to resources or threats (e.g., by moving towards resources or evading or attacking threats). Organisms also need to transmit information in order to influence another's behavior: to identify themselves, warn conspecifics of danger, coordinate activities, or deceive.

At the most basic level, the function of the nervous system is to send signals from one cell to others, or from one part of the body to others. There are multiple ways that a cell can send signals to other cells. One is by releasing chemicals called hormones into the internal circulation, so that they can diffuse to distant sites. In contrast to this “broadcast” mode of signaling, the nervous system pro-

vides “point-to-point” signals—neurons project their axons to specific target areas and make synaptic connections with specific target cells. Thus, neural signaling is capable of a much higher level of specificity than hormonal signaling. It is also much faster: the fastest nerve signals travel at speeds that exceed 100 meters per second.

At a more integrative level, the primary function of the nervous system is to control the body. It does this by extracting information from the environment using sensory receptors, sending signals that encode this information into the central nervous system, processing the information to determine an appropriate response, and sending output signals to muscles or glands to activate the response. The evolution of a complex nervous system has made it possible for various animal species to have advanced perception abilities such as vision, complex social interactions, rapid coordination of organ systems, and integrated processing of concurrent signals. In humans, the sophistication of the nervous system makes it possible to have language, abstract representation of concepts, transmission of culture, and many other features of human society that would not exist without the human brain.

25.6 The Sensory System

The sensory nervous system is a part of the nervous system responsible for processing sensory information. A sensory system consists of sensory neurons (including the sensory receptor cells), neural pathways, and parts of the brain involved in sensory perception. Commonly recognized sensory systems are those for vision, hearing, touch, taste, smell, and balance. In short, senses are transducers from the physical world to the realm of the mind where we interpret the information, creating our perception of the world around us.

Sensory systems code for four aspects of a stimulus: type (modality), intensity, location, and duration. Arrival time of a sound pulse and phase differences of continuous sound are used for sound localization. Certain receptors are sensitive to certain types of stimuli (for example, different mechanoreceptors respond best to different kinds of touch stimuli, like sharp or blunt objects). Receptors send impulses in certain patterns to send information about the intensity of a stimulus (for example, how loud a sound is). The location of the receptor that is stimulated gives the brain information about the location of the stimulus (for example, stimulating a mechanoreceptor in a finger will send information to the brain about that finger). The duration of the stimulus (how long it lasts) is conveyed by firing patterns of receptors. These impulses are transmitted to the brain through afferent neurons.

While debate exists among neurologists as to the spe-

cific number of senses due to differing definitions of what constitutes a sense, Gautama Buddha and Aristotle classified five 'traditional' human senses which have become universally accepted: touch, taste, smell, sight, and hearing. Other senses that have been well-accepted in most mammals, including humans, include nociception, equilibrioception, kinaesthesia, and thermoception. Furthermore, some nonhuman animals have been shown to possess alternate senses, including magnetoreception and electroreception.

The human sensory system consists of the following subsystems:

- Somatosensory system consists of the receptors, transmitters (pathways) leading to area S1, and area S1 in the cortex that is involved in creating the conscious experience of the sensations labelled as touch or pressure, temperature (warm or cold), pain (including itch and tickle), and the sensations of muscle movement and joint position including posture, movement, and facial expression (collectively also called proprioception)
- Visual system
- Auditory system
- Vestibular system
- Olfactory system
- Gustatory system

The receptive field is the area of the body or environment to which a receptor organ and receptor cells respond. For instance, the part of the world an eye can see, is its receptive field; the light that each rod or cone can see, is its receptive field. Receptive fields have been identified for the visual system, auditory system and somatosensory system.

25.7 The Motor System

The motor system is the set of central and peripheral structures in the nervous system that support motor functions, i.e. movement. Peripheral structures may include skeletal muscles and neural connections with muscle tissues. Central structures include cerebral cortex, brainstem, spinal cord, pyramidal system including the upper motor neurons, extrapyramidal system, cerebellum, and the lower motor neurons in the brainstem and the spinal cord.

The pyramidal motor system, also called the pyramidal tract or the corticospinal tract, start in the motor center of the cerebral cortex. There are upper and lower motor neurons in the corticospinal tract. The motor impulses originate in the giant pyramidal cells or Betz cells of the motor area; i.e., precentral gyrus of cerebral cortex. These are the upper motor neurons (UMN) of the corticospinal tract. The axons of these cells pass in the depth of the cerebral cor-

tex to the corona radiata and then to the internal capsule passing through the posterior branch of internal capsule and continue to descend in the midbrain and the medulla oblongata. In the lower part of medulla oblongata 80 to 85% of these fibers decussate (pass to the opposite side) and descend in the white matter of the lateral funiculus of the spinal cord on the opposite side. The remaining 15 to 20% pass to the same side. Fibers for the extremities (limbs) pass 100% to the opposite side. The fibers of the corticospinal tract terminate at different levels in the anterior horn of the grey matter of the spinal cord. Here the lower motor neurons (LMN) of the spinal cord are located. Peripheral motor nerves carry the motor impulses from the anterior horn to the voluntary muscles.

The extrapyramidal system is called extrapyramidal to distinguish it from the tracts of the motor cortex that reach their targets by traveling through the pyramids of the medulla. The pyramidal tracts (corticospinal tract and corticobulbar tracts) may directly innervate motor neurons of the spinal cord or brainstem (anterior (ventral) horn cells or certain cranial nerve nuclei), whereas the extrapyramidal system centers on the modulation and regulation (indirect control) of anterior (ventral) horn cells.

Extrapyramidal tracts are chiefly found in the reticular formation of the pons and medulla, and target lower motor neurons in the spinal cord that are involved in reflexes, locomotion, complex movements, and postural control. These tracts are in turn modulated by various parts of the central nervous system, including the nigrostriatal pathway, the basal ganglia, the cerebellum, the vestibular nuclei, and different sensory areas of the cerebral cortex. All of these regulatory components can be considered part of the extrapyramidal system, in that they modulate motor activity without directly innervating motor neurons.

25.8 Neuronal Signalling

Most neurons send signals via their axons, although some types are capable of dendrite-to-dendrite communication. (In fact, the types of neurons in the retina of the eye called amacrine cells have no axons, and communicate only via their dendrites.) Neural signals propagate along an axon in the form of electrochemical waves called action potentials, which produce cell-to-cell signals at points where axon terminals make synaptic contact with other cells.

Synapses may be electrical or chemical. Electrical synapses make direct electrical connections between neurons, but chemical synapses are much more common, and much more diverse in function. At a chemical synapse, the cell that sends signals is called presynaptic, and the cell that receives signals is called postsynaptic. Both the presynaptic and postsynaptic areas are full of molecular

machinery that carries out the signalling process. The presynaptic area contains large numbers of tiny spherical vessels called synaptic vesicles, packed with neurotransmitter chemicals. When the presynaptic terminal is electrically stimulated, an array of molecules embedded in the membrane are activated, and cause the contents of the vesicles to be released into the narrow space between the presynaptic and postsynaptic membranes, called the synaptic cleft. The neurotransmitter then binds to receptors embedded in the postsynaptic membrane, causing them to enter an activated state. Depending on the type of receptor, the resulting effect on the postsynaptic cell may be excitatory, inhibitory, or modulatory in more complex ways. For example, release of the neurotransmitter acetylcholine at a synaptic contact between a motor neuron and a muscle cell induces rapid contraction of the muscle cell. The entire synaptic transmission process takes only a fraction of a millisecond, although the effects on the postsynaptic cell may last much longer (even indefinitely, in cases where the synaptic signal leads to the formation of a memory trace).

There are literally hundreds of different types of synapses. In fact, there are over a hundred known neurotransmitters, and many of them have multiple types of receptors. Many synapses use more than one neurotransmitter—a common arrangement is for a synapse to use one fast-acting small-molecule neurotransmitter such as glutamate or GABA, along with one or more peptide neurotransmitters that play slower-acting modulatory roles. Molecular neuroscientists generally divide receptors into two broad groups: chemically gated ion channels and second messenger systems. When a chemically gated ion channel is activated, it forms a passage that allows specific types of ions to flow across the membrane. Depending on the type of ion, the effect on the target cell may be excitatory or inhibitory. When a second messenger system is activated, it starts a cascade of molecular interactions inside the target cell, which may ultimately produce a wide variety of complex effects, such as increasing or decreasing the sensitivity of the cell to stimuli, or even altering gene transcription.

According to a rule called Dale's principle, which has only a few known exceptions, a neuron releases the same neurotransmitters at all of its synapses. This does not mean, though, that a neuron exerts the same effect on all of its targets, because the effect of a synapse depends not on the neurotransmitter, but on the receptors that it activates. Because different targets can (and frequently do) use different types of receptors, it is possible for a neuron to have excitatory effects on one set of target cells, inhibitory effects on others, and complex modulatory effects on others still. Nevertheless, it happens that the two most widely used neurotransmitters, glutamate and GABA, each have

largely consistent effects. Glutamate has several widely occurring types of receptors, but all of them are excitatory or modulatory. Similarly, GABA has several widely occurring receptor types, but all of them are inhibitory. Because of this consistency, glutamatergic cells are frequently referred to as “excitatory neurons”, and GABAergic cells as “inhibitory neurons”. Strictly speaking, this is an abuse of terminology—it is the receptors that are excitatory and inhibitory, not the neurons—but it is commonly seen even in scholarly publications.

One very important subset of synapses are capable of forming memory traces by means of long-lasting activity-dependent changes in synaptic strength. The best-known form of neural memory is a process called long-term potentiation (abbreviated LTP), which operates at synapses that use the neurotransmitter glutamate acting on a special type of receptor known as the NMDA receptor. The NMDA receptor functions as a molecular “coincidence detector”: although the NMDA-receptor associated ion-channel opens upon binding of glutamate, extracellular Mg^{2+} ions will enter and block the channel immediately. Only concomitant membrane depolarization (e.g. induced by Na^+ influx via concomitantly stimulated non-NMDA (AMPA) type glutamate receptors in the same cell), will overcome the Mg^{2+} block and allow Na^+ and Ca^{2+} ions to enter the cell through the NMDA-receptor. Calcium entering the postsynaptic cell via NMDA receptors then initiates a second messenger cascade that ultimately leads to an increase in the number of AMPA-type glutamate receptors in the target cell, thereby increasing the effective strength of the synapse. This change in strength can last for weeks or longer. Besides the NMDA-receptor based processes, further cellular mechanisms allow of the association between two different input signals converging on the same neuron, in a defined timeframe. Upon a simultaneous increase in the intracellular concentrations of cAMP and Ca^{2+} , a transcriptional coactivator called TORC1 (CRTC1) becomes activated, that converts the temporal coincidence of the two second messengers into long term changes such as LTP. This cellular mechanism, through calcium-dependent adenylate cyclase activation, might also account for the detection of the repetitive stimulation of a given synapse.

In 1949, Donald Hebb¹³ postulated that synaptic efficiency will increase through repeated and persistent stimulation of a postsynaptic cell by a presynaptic cell. This is often informally summarized as “cells that fire together, wire together”. The theory was validated in part by the discovery of long-term potentiation. Studies of LTP on multiple presynaptic cells stimulating a postsynaptic cell uncovered the property of associativity. A weak neuronal stimulation

¹³https://en.wikipedia.org/wiki/Donald_O._Hebb

onto a pyramidal neuron may not induce long-term potentiation. However, this same stimulation paired with a simultaneous strong stimulation from another neuron will strengthen both synapses. This process suggests that two neuronal pathways converging on the same cell may both strengthen if stimulated coincidentally.

Since the discovery of LTP in 1973, many other types of synaptic memory traces have been found, involving increases or decreases in synaptic strength that are induced by varying conditions, and last for variable periods of time. The reward system, that reinforces desired behaviour for example, depends on a variant form of LTP that is conditioned on an extra input coming from a reward-signalling pathway that uses dopamine as neurotransmitter. All these forms of synaptic modifiability, taken collectively, give rise to neural plasticity, that is, to a capability for the nervous system to adapt itself to variations in the environment.

25.9 Neurons And Glial Cells

25.9.1 Neurons

The neuron doctrine is the now fundamental idea that neurons are the basic structural and functional units of the nervous system. The theory was put forward by Santiago Ramón y Cajal in the late 19th century. It held that neurons are discrete cells (not connected in a meshwork), acting as metabolically distinct units.

Later discoveries yielded refinements to the doctrine. For example, glial cells, which are not considered neurons, play an essential role in information processing. Also, electrical synapses are more common than previously thought, comprising direct, cytoplasmic connections between neurons. In fact, neurons can form even tighter couplings: the squid giant axon arises from the fusion of multiple axons.

Ramón y Cajal also postulated the Law of Dynamic Polarization, which states that a neuron receives signals at its dendrites and cell body and transmits them, as action potentials, along the axon in one direction: away from the cell body. The Law of Dynamic Polarization has important exceptions; dendrites can serve as synaptic output sites of neurons and axons can receive synaptic inputs.

The number of neurons in the brain varies dramatically from species to species. In a human, there are an estimated 10–20 billion neurons in the cerebral cortex and 55–70 billion neurons in the cerebellum. By contrast, the nematode worm *Caenorhabditis elegans* has just 302 neurons, making it an ideal model organism as scientists have been able to map all of its neurons. The fruit fly *Drosophila melanogaster*, a common subject in biological experiments, has around 100,000 neurons and exhibits many complex behaviors. Many properties of neurons,

from the type of neurotransmitters used to ion channel composition, are maintained across species, allowing scientists to study processes occurring in more complex organisms in much simpler experimental systems.

A neuron, neurone (old British spelling) or nerve cell, is an electrically excitable cell that communicates with other cells via specialized connections called synapses. It is the main component of nervous tissue. All animals except sponges and placozoans have neurons, but other multicellular organisms such as plants do not.

Neurons are typically classified into three types based on their function. Sensory neurons respond to stimuli such as touch, sound, or light that affect the cells of the sensory organs, and they send signals to the spinal cord or brain. Motor neurons receive signals from the brain and spinal cord to control everything from muscle contractions to glandular output. Interneurons connect neurons to other neurons within the same region of the brain or spinal cord. A group of connected neurons is called a neural circuit.

A typical neuron consists of a cell body (soma), dendrites, and a single axon. The soma is usually compact. The axon and dendrites are filaments that extrude from it. Dendrites typically branch profusely and extend a few hundred micrometers from the soma. The axon leaves the soma at a swelling called the axon hillock, and travels for as far as 1 meter in humans or more in other species. It branches but usually maintains a constant diameter. At the farthest tip of the axon's branches are axon terminals, where the neuron can transmit a signal across the synapse to another cell. Neurons may lack dendrites or have no axon. The term neurite is used to describe either a dendrite or an axon, particularly when the cell is undifferentiated.

Most neurons receive signals via the dendrites and soma and send out signals down the axon. At the majority of synapses, signals cross from the axon of one neuron to a dendrite of another. However, synapses can connect an axon to another axon or a dendrite to another dendrite.

The signaling process is partly electrical and partly chemical. Neurons are electrically excitable, due to maintenance of voltage gradients across their membranes. If the voltage changes by a large enough amount over a short interval, the neuron generates an all-or-nothing electrochemical pulse called an action potential. This potential travels rapidly along the axon, and activates synaptic connections as it reaches them. Synaptic signals may be excitatory or inhibitory, increasing or reducing the net voltage that reaches the soma.

In most cases, neurons are generated by neural stem cells during brain development and childhood. Neurogenesis largely ceases during adulthood in most areas of the

brain. However, strong evidence supports generation of substantial numbers of new neurons in the hippocampus and olfactory bulb.

Neurons are highly specialized for the processing and transmission of cellular signals. Given their diversity of functions performed in different parts of the nervous system, there is a wide variety in their shape, size, and electrochemical properties. For instance, the soma of a neuron can vary from 4 to 100 micrometers in diameter.

The soma is the body of the neuron. As it contains the nucleus, most protein synthesis occurs here. The nucleus can range from 3 to 18 micrometers in diameter. The dendrites of a neuron are cellular extensions with many branches. This overall shape and structure is referred to metaphorically as a dendritic tree. This is where the majority of input to the neuron occurs via the dendritic spine.

The axon is a finer, cable-like projection that can extend tens, hundreds, or even tens of thousands of times the diameter of the soma in length. The axon primarily carries nerve signals away from the soma, and carries some types of information back to it. Many neurons have only one axon, but this axon may—and usually will—undergo extensive branching, enabling communication with many target cells. The part of the axon where it emerges from the soma is called the axon hillock. Besides being an anatomical structure, the axon hillock also has the greatest density of voltage-dependent sodium channels. This makes it the most easily excited part of the neuron and the spike initiation zone for the axon. In electrophysiological terms, it has the most negative threshold potential. While the axon and axon hillock are generally involved in information outflow, this region can also receive input from other neurons.

The axon terminal is found at the end of the axon farthest from the soma and contains synapses. Synaptic boutons are specialized structures where neurotransmitter chemicals are released to communicate with target neurons. In addition to synaptic boutons at the axon terminal, a neuron may have en passant boutons, which are located along the length of the axon.

The accepted view of the neuron attributes dedicated functions to its various anatomical components; however, dendrites and axons often act in ways contrary to their so-called main function.

Axons and dendrites in the central nervous system are typically only about one micrometer thick, while some in the peripheral nervous system are much thicker. The soma is usually about 10–25 micrometers in diameter and often is not much larger than the cell nucleus it contains. The longest axon of a human motor neuron can be over a meter long, reaching from the base of the spine to the toes.

Sensory neurons can have axons that run from the toes to the posterior column of the spinal cord, over 1.5 meters in adults. Giraffes have single axons several meters in length running along the entire length of their necks. Much of what is known about axonal function comes from studying the squid giant axon, an ideal experimental preparation because of its relatively immense size (0.5–1 millimeters thick, several centimeters long).

Fully differentiated neurons are permanently postmitotic however, stem cells present in the adult brain may regenerate functional neurons throughout the life of an organism. Astrocytes are star-shaped glial cells. They have been observed to turn into neurons by virtue of their stem cell-like characteristic of pluripotency.

Like all animal cells, the cell body of every neuron is enclosed by a plasma membrane, a bilayer of lipid molecules with many types of protein structures embedded in it. A lipid bilayer is a powerful electrical insulator, but in neurons, many of the protein structures embedded in the membrane are electrically active. These include ion channels that permit electrically charged ions to flow across the membrane and ion pumps that transport ions from one side of the membrane to the other. Most ion channels are permeable only to specific types of ions. Some ion channels are voltage gated, meaning that they can be switched between open and closed states by altering the voltage difference across the membrane. Others are chemically gated, meaning that they can be switched between open and closed states by interactions with chemicals that diffuse through the extracellular fluid. The ions include sodium, potassium, chloride, and calcium. The interactions between ion channels and ion pumps produce a voltage difference across the membrane, typically a bit less than 1/10 of a volt at baseline. This voltage has two functions: first, it provides a power source for an assortment of voltage-dependent protein machinery that is embedded in the membrane; second, it provides a basis for electrical signal transmission between different parts of the membrane.

Numerous microscopic clumps called Nissl bodies (or Nissl substance) are seen when nerve cell bodies are stained with a basophilic (“base-loving”) dye. These structures consist of rough endoplasmic reticulum and associated ribosomal RNA. Named after German psychiatrist and neuropathologist Franz Nissl¹⁴ (1860–1919), they are involved in protein synthesis and their prominence can be explained by the fact that nerve cells are very metabolically active. Basophilic dyes such as aniline or (weakly) haematoxylin highlight negatively charged components, and so bind to the phosphate backbone of the ribosomal RNA.

¹⁴https://en.wikipedia.org/wiki/Franz_Nissl

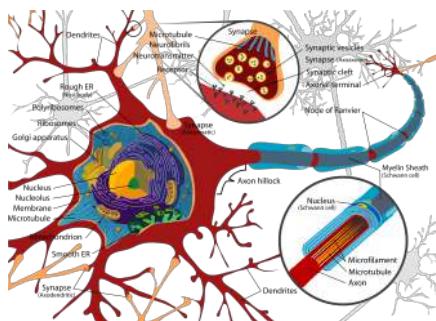


Figure 25.12: Diagram of a myelinated vertebrate motor neuron.¹⁵

The cell body of a neuron is supported by a complex mesh of structural proteins called neurofilaments, which together with neurotubules (neuronal microtubules) are assembled into larger neurofibrils. Some neurons also contain pigment granules, such as neuromelanin (a brownish-black pigment that is byproduct of synthesis of catecholamines), and lipofuscin (a yellowish-brown pigment), both of which accumulate with age. Other structural proteins that are important for neuronal function are actin and the tubulin of microtubules. Class III β -tubulin is found almost exclusively in neurons. Actin is predominately found at the tips of axons and dendrites during neuronal development. There the actin dynamics can be modulated via an interplay with microtubule.

There are different internal structural characteristics between axons and dendrites. Typical axons almost never contain ribosomes, except some in the initial segment. Dendrites contain granular endoplasmic reticulum or ribosomes, in diminishing amounts as the distance from the cell body increases.

Neurons vary in shape and size and can be classified by their morphology and function. The anatomist Camillo Golgi¹⁶ grouped neurons into two types; type I with long axons used to move signals over long distances and type II with short axons, which can often be confused with dendrites. Type I cells can be further classified by the location of the soma. The basic morphology of type I neurons, represented by spinal motor neurons, consists of a cell body called the soma and a long thin axon covered by a myelin sheath.

The dendritic tree wraps around the cell body and receives signals from other neurons. The end of the axon has branching terminals (axon terminal) that release neurotransmitters into a gap called the synaptic cleft between the terminals and the dendrites of the next neuron.

Most neurons can be anatomically characterized as:

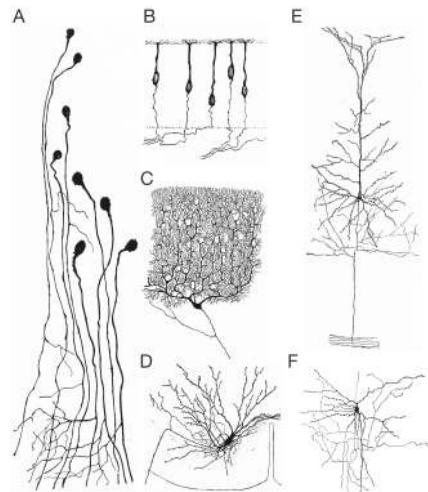


Figure 25.13: Morphologically distinct types of neurons after Cajal. A) Unipolar neurons; B) bipolar neurons; Golgi I neurons: C) a Purkinje cell; D) spinal motor neuron E) a pyramidal cell; F) Golgi II neuron. Histologie du système nerveux de l'homme & des vertébrés, Tome Premier¹⁷ (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay.

- Unipolar: single process
- Bipolar: 1 axon and 1 dendrite
- Multipolar: 1 axon and 2 or more dendrites
- Golgi I: neurons with projecting axonal processes; examples are pyramidal cells, Purkinje cells, and anterior horn cells
- Golgi II: neurons whose axonal process projects locally; the best example is the granule cell
- Anaxonic: where the axon cannot be distinguished from the dendrite(s)
- Pseudounipolar: 1 process which then serves as both an axon and a dendrite
- Other

Neurons can also be characterized based on various aspects of their function:

- Afferent neurons convey information from tissues and organs into the central nervous system and are also called sensory neurons.
- Efferent neurons (motor neurons) transmit signals from the central nervous system to the effector cells.
- Interneurons connect neurons within specific regions of the central nervous system. Afferent and efferent also refer generally to neurons that, respectively, bring information to or send information from the brain.

The axons of neurons in the human peripheral nervous system can be classified based on their physical features and signal conduction properties. Axons were known to

¹⁶https://en.wikipedia.org/wiki/Camillo_Golgi

have different thicknesses (from 0.1 to 20 μm) and these differences were thought to relate to the speed at which an action potential could travel along the axon – its conduction velocity. Erlanger and Gasser proved this hypothesis, and identified several types of nerve fiber, establishing a relationship between the diameter of an axon and its nerve conduction velocity. They published their findings in 1941 giving the first classification of axons.

Neurons communicate with each other via synapses, where either the axon terminal of one cell contacts another neuron's dendrite, soma or, less commonly, axon. Neurons such as Purkinje cells in the cerebellum can have over 1000 dendritic branches, making connections with tens of thousands of other cells; other neurons, such as the magnocellular neurons of the supraoptic nucleus, have only one or two dendrites, each of which receives thousands of synapses.

Synapses can be excitatory or inhibitory, either increasing or decreasing activity in the target neuron, respectively. Some neurons also communicate via electrical synapses, which are direct, electrically conductive junctions between cells.

When an action potential reaches the axon terminal, it opens voltage-gated calcium channels, allowing calcium ions to enter the terminal. Calcium causes synaptic vesicles filled with neurotransmitter molecules to fuse with the membrane, releasing their contents into the synaptic cleft. The neurotransmitters diffuse across the synaptic cleft and activate receptors on the postsynaptic neuron. High cytosolic calcium in the axon terminal triggers mitochondrial calcium uptake, which, in turn, activates mitochondrial energy metabolism to produce ATP to support continuous neurotransmission.

An autapse is a synapse in which a neuron's axon connects to its own dendrites.

The human brain has some 8.6×10^{10} (eighty six billion) neurons. Each neuron has on average 7,000 synaptic connections to other neurons. It has been estimated that the brain of a three-year-old child has about 10^{15} synapses (1 quadrillion). This number declines with age, stabilizing by adulthood. Estimates vary for an adult, ranging from 10^{14} to 5×10^{14} synapses (100 to 500 trillion).

The two most common neurotransmitters in the brain, glutamate and GABA, have largely consistent actions. Glutamate acts on several types of receptors, and has effects that are excitatory at ionotropic receptors and a modulatory effect at metabotropic receptors. Similarly, GABA acts on several types of receptors, but all of them have inhibitory effects (in adult animals, at least). Because of this consistency, it is common for neuroscientists to refer to cells that release glutamate as "excitatory neurons",

and cells that release GABA as "inhibitory neurons". Some other types of neurons have consistent effects, for example, "excitatory" motor neurons in the spinal cord that release acetylcholine, and "inhibitory" spinal neurons that release glycine.

The distinction between excitatory and inhibitory neurotransmitters is not absolute. Rather, it depends on the class of chemical receptors present on the postsynaptic neuron. In principle, a single neuron, releasing a single neurotransmitter, can have excitatory effects on some targets, inhibitory effects on others, and modulatory effects on others still. For example, photoreceptor cells in the retina constantly release the neurotransmitter glutamate in the absence of light. So-called OFF bipolar cells are, like most neurons, excited by the released glutamate. However, neighboring target neurons called ON bipolar cells are instead inhibited by glutamate, because they lack typical ionotropic glutamate receptors and instead express a class of inhibitory metabotropic glutamate receptors. When light is present, the photoreceptors cease releasing glutamate, which relieves the ON bipolar cells from inhibition, activating them; this simultaneously removes the excitation from the OFF bipolar cells, silencing them.

Neurons can also be classified based on the neurotransmitter they release at synapses:

- Cholinergic neurons—acetylcholine. Acetylcholine is released from presynaptic neurons into the synaptic cleft. It acts as a ligand for both ligand-gated ion channels and metabotropic (GPCRs) muscarinic receptors. Nicotinic receptors are pentameric ligand-gated ion channels composed of alpha and beta subunits that bind nicotine. Ligand binding opens the channel causing influx of Na^+ depolarization and increases the probability of presynaptic neurotransmitter release. Acetylcholine is synthesized from choline and acetyl coenzyme A.
- GABAergic neurons—gamma aminobutyric acid. GABA is one of two neuroinhibitors in the central nervous system (CNS), along with glycine. GABA has a homologous function to ACh, gating anion channels that allow Cl^- ions to enter the post synaptic neuron. Cl^- causes hyperpolarization within the neuron, decreasing the probability of an action potential firing as the voltage becomes more negative (for an action potential to fire, a positive voltage threshold must be reached). GABA is synthesized from glutamate by the enzyme glutamate decarboxylase.
- Glutamatergic neurons—glutamate. Glutamate is one of two primary excitatory amino acid neurotransmitters, along with aspartate. Glutamate can cause excitotoxicity when blood flow to the brain is

interrupted, resulting in brain damage. When blood flow is suppressed, glutamate is released from presynaptic neurons, causing abnormal NMDA and AMPA receptor activation, leading to elevated Ca^{2+} and Na^+ entering the post synaptic neuron and cell damage. Glutamate is synthesized from the amino acid glutamine by the enzyme glutamate synthase. There are three main types of ionotropic glutamate receptors

- AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptors
- Kainate receptors
- NMDA receptors and three groups of metabotropic (G-protein coupled) receptors.
- Dopaminergic neurons—dopamine. Dopamine is a neurotransmitter that acts on D1 type (D1 and D5) G_s-coupled receptors, which stimulate the production of cAMP which stimulates protein kinase A (PKA), and D2 type (D2, D3, and D4) receptors, which activate Gi-coupled receptors that decrease cAMP and PKA. Dopamine is connected to mood and behavior and modulates both pre- and post-synaptic neurotransmission. Loss of dopamine neurons in the substantia nigra has been linked to Parkinson's disease. Dopamine is synthesized from the amino acid tyrosine. Tyrosine is converted into levadopa (or L-DOPA) by tyrosine hydroxylase, and levadopa is then converted into dopamine by the aromatic amino acid decarboxylase.
- Serotonergic neurons—serotonin. Serotonin (5-hydroxytryptamine, 5-HT) can act as excitatory or inhibitory. Of its four 5-HT receptor classes, 3 are GPCR and 1 is a ligand-gated cation channel. Serotonin is synthesized from tryptophan by tryptophan hydroxylase, and then further by decarboxylase. A lack of 5-HT at postsynaptic neurons has been linked to depression. Drugs that block the presynaptic serotonin transporter are used for treatment, such as Prozac and Zoloft.

25.9.2 Glia

Glia, also called glial cells or neuroglia, are non-neuronal cells in the central nervous system (brain and spinal cord) and the peripheral nervous system that do not produce electrical impulses. They maintain homeostasis, form myelin, and provide support and protection for neurons. In the central nervous system, glial cells include oligodendrocytes, astrocytes, ependymal cells, and microglia, and in the peripheral nervous system glial cells include Schwann cells and satellite cells. They have four main functions: (1) to surround neurons and hold them in place; (2) to supply nutrients and oxygen to neurons;

(3) to insulate one neuron from another; (4) to destroy pathogens and remove dead neurons.

Glia cells exhibit great cellular and functional diversity. Glial cells can respond to and manipulate neurotransmission in many ways.

Glia were discovered in 1856, by the pathologist Rudolf Virchow in his search for a “connective tissue” in the brain. The term derives from Greek γλία and γλοία “glue”, and suggests the original impression that they were the glue of the nervous system.

In general, neuroglial cells are smaller than neurons. There are approximately 85 billion glia cells in the human brain, about the same number as neurons. Glial cells make up about half the total volume of the brain and spinal cord. The glia to neuron-ratio varies from one part of the brain to another. The glia to neuron-ratio in the cerebral cortex is 3.72 (60.84 billion glia (72%); 16.34 billion neurons), while that of the cerebellum is only 0.23 (16.04 billion glia; 69.03 billion neurons). The ratio in the cerebral cortex gray matter is 1.48, with 3.76 for the gray and white matter combined. The ratio of the basal ganglia, diencephalon and brainstem combined is 11.35.

The total number of glia cells in the human brain is distributed into the different types with oligodendrocytes being the most frequent (45–75%), followed by astrocytes (19–40%) and microglia (about 10% or less).

Most glia are derived from ectodermal tissue of the developing embryo, in particular the neural tube and crest. The exception is microglia, which are derived from hemopoietic stem cells. In the adult, microglia are largely a self-renewing population and are distinct from macrophages and monocytes, which infiltrate an injured and diseased CNS.

In the central nervous system, glia develop from the ventricular zone of the neural tube. These glia include the oligodendrocytes, ependymal cells, and astrocytes. In the peripheral nervous system, glia derive from the neural crest. These PNS glia include Schwann cells in nerves and satellite glial cells in ganglia.

Glia retain the ability to undergo cell division in adulthood, whereas most neurons cannot. The view is based on the general inability of the mature nervous system to replace neurons after an injury, such as a stroke or trauma, where very often there is a substantial proliferation of glia, or gliosis, near or at the site of damage. However, detailed studies have found no evidence that ‘mature’ glia, such as astrocytes or oligodendrocytes, retain mitotic capacity. Only the resident oligodendrocyte precursor cells seem to keep this ability once the nervous system matures.

Some glial cells function primarily as the physical sup-

port for neurons. Others provide nutrients to neurons and regulate the extracellular fluid of the brain, especially surrounding neurons and their synapses. During early embryogenesis, glial cells direct the migration of neurons and produce molecules that modify the growth of axons and dendrites.

Glia are crucial in the development of the nervous system and in processes such as synaptic plasticity and synaptogenesis. Glia have a role in the regulation of repair of neurons after injury. In the central nervous system (CNS), glia suppress repair. Glial cells known as astrocytes enlarge and proliferate to form a scar and produce inhibitory molecules that inhibit regrowth of a damaged or severed axon. In the peripheral nervous system (PNS), glial cells known as Schwann cells promote repair. After axonal injury, Schwann cells regress to an earlier developmental state to encourage regrowth of the axon. This difference between the CNS and the PNS, raises hopes for the regeneration of nervous tissue in the CNS. For example, a spinal cord may be able to be repaired following injury or severance. Schwann cells are also known as neuri-lemmocytes. These cells envelop nerve fibers of the PNS by winding repeatedly around a nerve fiber with the nucleus inside of it. This process creates a myelin sheath, which not only aids in conductivity but also assists in the regeneration of damaged fibers.

Astrocytes are crucial participants in the tripartite synapse. They have several crucial functions, including clearance of neurotransmitters from within the synaptic cleft, which aids in distinguishing between separate action potentials and prevents toxic build-up of certain neurotransmitters such as glutamate, which would otherwise lead to excitotoxicity. Furthermore, astrocytes release gliotransmitters such as glutamate, ATP, and D-serine in response to stimulation.

Oligodendrocytes are found in the CNS and resemble an octopus: they have bulbous cell bodies with up to fifteen arm-like processes. Each “arm” reaches out to a nerve fiber and spirals around it, creating a myelin sheath. The myelin sheath insulates the nerve fiber from the extracellular fluid and speeds up signal conduction along the nerve fiber.

Microglia are specialized macrophages capable of phagocytosis that protect neurons of the central nervous system. They are derived from the earliest wave of mononuclear cells that originate in yolk sac blood islands early in development, and colonize the brain shortly after the neural precursors begin to differentiate.

These cells are found in all regions of the brain and spinal cord. Microglial cells are small relative to macroglial cells, with changing shapes and oblong nuclei. They are mobile within the brain and multiply when the brain is dam-

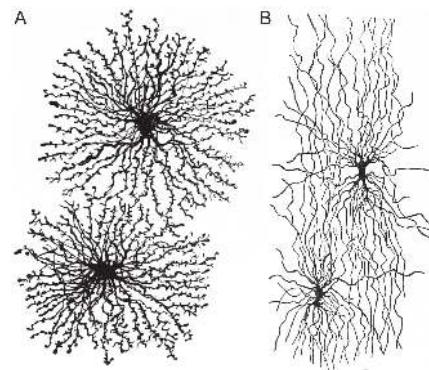


Figure 25.14: Astrocytes (A) and oligodendrocytes (B) are the major types of macroglia in the grey and white matter of the brain, respectively. *Histologie du système nerveux de l'homme & des vertébrés, Tome Premier*¹⁸ (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay.

aged. In the healthy central nervous system, microglia processes constantly sample all aspects of their environment (neurons, macroglia and blood vessels). In a healthy brain, microglia direct the immune response to brain damage and play an important role in the inflammation that accompanies the damage. Many diseases and disorders are associated with deficient microglia, such as Alzheimer's disease, Parkinson's disease, and ALS. During developmental wiring of the brain, microglial cells play a large role regulating numbers of neural precursor cells and removing apoptotic neurons. There is also evidence that microglia can refine synaptic circuitry by engulfing and eliminating synapses. Post development, the majority of dead or apoptotic cells are found in the cerebral cortex and the subcortical white matter. This may explain why the majority of ameboid microglial cells are found within the “fountains of microglia” in the cerebral cortex.

25.10 Electrical Basis Of Neuronal Function

25.10.1 The Membrane Potential

Membrane potential (also transmembrane potential or membrane voltage) is the difference in electric potential between the interior and the exterior of a biological cell. With respect to the exterior of the cell, typical values of membrane potential, normally given in units of millivolts and denoted as mV, ranges from -40 mV to -80 mV.

All animal cells are surrounded by a membrane composed of a lipid bilayer with proteins embedded in it. The membrane serves as both an insulator and a diffusion barrier to the movement of ions. Transmembrane proteins,

also known as ion transporter or ion pump proteins, actively push ions across the membrane and establish concentration gradients across the membrane, and ion channels allow ions to move across the membrane down those concentration gradients. Ion pumps and ion channels are electrically equivalent to a set of batteries and resistors inserted in the membrane, and therefore create a voltage between the two sides of the membrane.

Almost all plasma membranes have an electrical potential across them, with the inside usually negative with respect to the outside. The membrane potential has two basic functions. First, it allows a cell to function as a battery, providing power to operate a variety of “molecular devices” embedded in the membrane. Second, in electrically excitable cells such as neurons and muscle cells, it is used for transmitting signals between different parts of a cell. Signals are generated by opening or closing of ion channels at one point in the membrane, producing a local change in the membrane potential. This change in the electric field can quickly affect adjacent and more distant ion channels in the membrane. Those ion channels can then open or close as a result of the potential change, reproducing the signal.

In non-excitable cells, and in excitable cells in their baseline states, the membrane potential is held at a relatively stable value, called the resting potential. For neurons, typical values of the resting potential range from -70 to -80 millivolts; that is, the interior of a cell has a negative baseline voltage of a bit less than one-tenth of a volt. The opening and closing of ion channels can induce a departure from the resting potential. This is called a depolarization if the interior voltage becomes less negative (say from -70 mV to -60 mV), or a hyperpolarization if the interior voltage becomes more negative (say from -70 mV to -80 mV). In excitable cells, a sufficiently large depolarization can evoke an action potential, in which the membrane potential changes rapidly and significantly for a short time (on the order of 1 to 100 milliseconds), often reversing its polarity. Action potentials are generated by the activation of certain voltage-gated ion channels.

In neurons, the factors that influence the membrane potential are diverse. They include numerous types of ion channels, some of which are chemically gated and some of which are voltage-gated. Because voltage-gated ion channels are controlled by the membrane potential, while the membrane potential itself is influenced by these same ion channels, feedback loops that allow for complex temporal dynamics arise, including oscillations and regenerative events such as action potentials.

The membrane potential in a cell derives ultimately from two factors: electrical force and diffusion. Electrical force arises from the mutual attraction between particles

with opposite electrical charges (positive and negative) and the mutual repulsion between particles with the same type of charge (both positive or both negative). Diffusion arises from the statistical tendency of particles to redistribute from regions where they are highly concentrated to regions where the concentration is low.

Voltage, which is synonymous with difference in electrical potential, is the ability to drive an electric current across a resistance. Indeed, the simplest definition of a voltage is given by Ohm's law: $V=IR$, where V is voltage, I is current and R is resistance. If a voltage source such as a battery is placed in an electrical circuit, the higher the voltage of the source the greater the amount of current that it will drive across the available resistance. The functional significance of voltage lies only in potential differences between two points in a circuit. The idea of a voltage at a single point is meaningless. It is conventional in electronics to assign a voltage of zero to some arbitrarily chosen element of the circuit, and then assign voltages for other elements measured relative to that zero point. There is no significance in which element is chosen as the zero point—the function of a circuit depends only on the differences not on voltages per se. However, in most cases and by convention, the zero level is most often assigned to the portion of a circuit that is in contact with ground.

The same principle applies to voltage in cell biology. In electrically active tissue, the potential difference between any two points can be measured by inserting an electrode at each point, for example one inside and one outside the cell, and connecting both electrodes to the leads of what is in essence a specialized voltmeter. By convention, the zero potential value is assigned to the outside of the cell and the sign of the potential difference between the outside and the inside is determined by the potential of the inside relative to the outside zero.

25.10.2 Ions And The Forces Driving Their Motion

Electrical signals within biological organisms are, in general, driven by ions. An ion is an atom or molecule that has a net electrical charge. Since the charge of the electron (considered negative by convention) is equal and opposite to that of the proton (considered positive by convention), the net charge of an ion is non-zero due to its total number of electrons being unequal to its total number of protons. A cation is a positively charged ion, with fewer electrons than protons, while an anion is negatively charged, with more electrons than protons. Because of their opposite electric charges, cations and anions attract each other and readily form ionic compounds.

The word ion comes from the Greek word *iōv*, ion,

"going", the present participle of *ἰέναι*, *ienai*, "to go". This term was introduced (after a suggestion by William Whewell¹⁹) by English physicist and chemist Michael Faraday²⁰ in 1834 for the then-unknown species that goes from one electrode to the other through an aqueous medium. Faraday did not know the nature of these species, but he knew that since metals dissolved into and entered a solution at one electrode and new metal came forth from a solution at the other electrode; that some kind of substance has moved through the solution in a current. This conveys matter from one place to the other. In correspondence with Faraday, Whewell also coined the words anode and cathode, as well as anion (from the Greek word ἄνω (ánō), meaning "up") and cation (from the Greek word κάτω (káto), meaning "down") as ions that are attracted to the respective electrodes.

Svante Arrhenius²¹ put forth, in his 1884 dissertation, his explanation of the fact that solid crystalline salts dissociate into paired charged particles when dissolved, for which he would win the 1903 Nobel Prize in Chemistry. Arrhenius' explanation was that in forming a solution, the salt dissociates into Faraday's ions. Arrhenius proposed that ions formed even in the absence of an electric current.

The most important cations for the action potential are sodium (Na^+) and potassium (K^+). Both of these are monovalent cations that carry a single positive charge. Action potentials can also involve calcium (Ca^{2+}), which is a divalent cation that carries a double positive charge. The chloride anion (Cl^-) plays a major role in the action potentials of some algae, but plays a negligible role in the action potentials of most animals.

Ions cross the cell membrane under two influences: diffusion and electric fields. A simple example wherein two solutions—A and B—are separated by a porous barrier illustrates that diffusion will ensure that they will eventually mix into equal solutions. This mixing occurs because of the difference in their concentrations. The region with high concentration will diffuse out toward the region with low concentration. To extend the example, let solution A have 30 sodium ions and 30 chloride ions. Also, let solution B have only 20 sodium ions and 20 chloride ions. Assuming the barrier allows both types of ions to travel through it, then a steady state will be reached whereby both solutions have 25 sodium ions and 25 chloride ions. If, however, the porous barrier is selective to which ions are let through, then diffusion alone will not determine the resulting solution. Returning to the previous example, let's now construct a barrier that is permeable only to sodium ions. Now, only sodium is allowed to diffuse cross the barrier from its

higher concentration in solution A to the lower concentration in solution B. This will result in a greater accumulation of sodium ions than chloride ions in solution B and a lesser number of sodium ions than chloride ions in solution A.

This means that there is a net positive charge in solution B from the higher concentration of positively charged sodium ions than negatively charged chloride ions. Likewise, there is a net negative charge in solution A from the greater concentration of negative chloride ions than positive sodium ions. Since opposite charges attract and like charges repel, the ions are now also influenced by electrical fields as well as forces of diffusion. Therefore, positive sodium ions will be less likely to travel to the now-more-positive B solution and remain in the now-more-negative A solution. The point at which the forces of the electric fields completely counteract the force due to diffusion is called the equilibrium potential. At this point, the net flow of the specific ion (in this case sodium) is zero.

25.10.3 The Plasma Membrane

Because it is made of lipid molecules, the plasma membrane intrinsically has a high electrical resistivity, in other words a low intrinsic permeability to ions. However, some of the molecules embedded in the membrane are capable either of actively transporting ions from one side of the membrane to the other or of providing channels through which they can move.

In electrical terminology, the plasma membrane functions as a combined resistor and capacitor. Resistance arises from the fact that the membrane impedes the movement of charges across it. Capacitance arises from the fact that the lipid bilayer is so thin that an accumulation of charged particles on one side gives rise to an electrical force that pulls oppositely charged particles toward the other side. The capacitance of the membrane is relatively unaffected by the molecules that are embedded in it, so it has a more or less invariant value estimated at about $2 \mu\text{F}/\text{cm}^2$ (the total capacitance of a patch of membrane is proportional to its area). The conductance of a pure lipid bilayer is so low, on the other hand, that in biological situations it is always dominated by the conductance of alternative pathways provided by embedded molecules. Thus, the capacitance of the membrane is more or less fixed, but the resistance is highly variable.

The thickness of a plasma membrane is estimated to be about 7–8 nanometers. Because the membrane is so thin, it does not take a very large transmembrane voltage to create a strong electric field within it. Typical membrane potentials in animal cells are on the order of 100 millivolts (that is, one tenth of a volt), but calculations show that this generates an electric field close to the maximum that the membrane can sustain—it has been calculated that a volt-

¹⁹https://en.wikipedia.org/wiki/William_Whewell

²⁰https://en.wikipedia.org/wiki/Michael_Faraday

²¹https://en.wikipedia.org/wiki/Svante_Arrhenius

age difference much larger than 200 millivolts could cause dielectric breakdown, that is, arcing across the membrane.

The resistance of a pure lipid bilayer to the passage of ions across it is very high, but structures embedded in the membrane can greatly enhance ion movement, either actively or passively, via mechanisms called facilitated transport and facilitated diffusion. The two types of structure that play the largest roles are ion channels and ion pumps, both usually formed from assemblages of protein molecules. Ion channels provide passageways through which ions can move. In most cases, an ion channel is permeable only to specific types of ions (for example, sodium and potassium but not chloride or calcium), and sometimes the permeability varies depending on the direction of ion movement. Ion pumps, also known as ion transporters or carrier proteins, actively transport specific types of ions from one side of the membrane to the other, using energy derived from metabolic processes to do so.

25.10.4 Ion Pumps

Ion pumps are integral membrane proteins that carry out active transport, i.e., use cellular energy (ATP) to “pump” the ions against their concentration gradient. Such ion pumps take in ions from one side of the membrane (decreasing its concentration there) and release them on the other side (increasing its concentration there).

The ion pump most relevant to the action potential is the sodium-potassium pump, which transports three sodium ions out of the cell and two potassium ions in. As a consequence, the concentration of potassium ions K^+ inside the neuron is roughly 20-fold larger than the outside concentration, whereas the sodium concentration outside is roughly ninefold larger than inside. In a similar manner, other ions have different concentrations inside and outside the neuron, such as calcium, chloride and magnesium.

If the numbers of each type of ion were equal, the sodium-potassium pump would be electrically neutral, but, because of the three-for-two exchange, it gives a net movement of one positive charge from intracellular to extracellular for each cycle, thereby contributing to a positive voltage difference. The pump has three effects: (1) it makes the sodium concentration high in the extracellular space and low in the intracellular space; (2) it makes the potassium concentration high in the intracellular space and low in the extracellular space; (3) it gives the intracellular space a negative voltage with respect to the extracellular space.

The sodium-potassium pump is relatively slow in operation. If a cell were initialized with equal concentrations of sodium and potassium everywhere, it would take hours

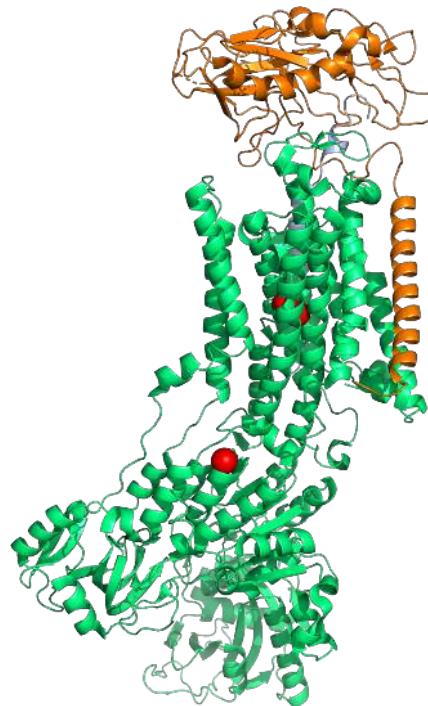


Figure 25.15: A cartoon representation of the molecular structure of the sodium – potassium pump based on atomic coordinates of PDB 2ZXE²², rendered with open source molecular visualization tool PyMol.

for the pump to establish equilibrium. The pump operates constantly, but becomes progressively less efficient as the concentrations of sodium and potassium available for pumping are reduced.

Ion pumps influence the action potential only by establishing the relative ratio of intracellular and extracellular ion concentrations. The action potential involves mainly the opening and closing of ion channels not ion pumps. If the ion pumps are turned off by removing their energy source, or by adding an inhibitor such as ouabain, the axon can still fire hundreds of thousands of action potentials before their amplitudes begin to decay significantly. In particular, ion pumps play no significant role in the repolarization of the membrane after an action potential.

Another functionally important ion pump is the sodium-calcium exchanger. This pump operates in a conceptually similar way to the sodium-potassium pump, except that in each cycle it exchanges three Na^+ from the extracellular space for one Ca^{2+} from the intracellular space. Because the net flow of charge is inward, this pump runs “downhill”, in effect, and therefore does not require any energy source except the membrane voltage. Its most important effect is to pump calcium outward—it also allows an inward flow of sodium, thereby counteracting

the sodium-potassium pump, but, because overall sodium and potassium concentrations are much higher than calcium concentrations, this effect is relatively unimportant. The net result of the sodium-calcium exchanger is that in the resting state, intracellular calcium concentrations become very low.

25.10.5 Ion Channels

Ion channels are integral membrane proteins with a pore through which ions can travel between extracellular space and cell interior. Most channels are specific (selective) for one ion; for example, most potassium channels are characterized by 1000:1 selectivity ratio for potassium over sodium, though potassium and sodium ions have the same charge and differ only slightly in their radius. The channel pore is typically so small that ions must pass through it in single-file order. Channel pores can be either open or closed for ion passage, although a number of channels demonstrate various sub-conductance levels. When a channel is open, ions permeate through the channel pore down the transmembrane concentration gradient for that particular ion. Rate of ionic flow through the channel, i.e. single-channel current amplitude, is determined by the maximum channel conductance and electrochemical driving force for that ion, which is the difference between the instantaneous value of the membrane potential and the value of the reversal potential.

The fundamental properties of currents mediated by ion channels were analyzed by the British biophysicists Alan Hodgkin²³ and Andrew Huxley²⁴ as part of their Nobel Prize-winning research on the action potential, published in 1952. They built on the work of other physiologists, such as Kenneth Stewart Cole's²⁵ research into voltage-gated membrane pores from 1941. The existence of ion channels was confirmed in the 1970s by Bernard Katz²⁶ and Ricardo Miledi²⁷ using noise analysis. It was then shown more directly with an electrical recording technique known as the "patch clamp", which led to a Nobel Prize to Erwin Neher²⁸ and Bert Sakmann²⁹, the technique's inventors. Hundreds if not thousands of researchers continue to pursue a more detailed understanding of how these proteins work.

Channels differ with respect to the ion they let pass (for example, Na^+ , K^+ , Cl^-), the ways in which they may be regulated, the number of subunits of which they are composed and other aspects of structure. Channels belonging to the

largest class, which includes the voltage-gated channels that underlie the nerve impulse, consists of four subunits with six transmembrane helices each. On activation, these helices move about and open the pore. Two of these six helices are separated by a loop that lines the pore and is the primary determinant of ion selectivity and conductance in this channel class and some others. The existence and mechanism for ion selectivity was first postulated in the late 1960s by Bertil Hille³⁰ and Clay Armstrong³¹. The idea of the ionic selectivity for potassium channels was that the carbonyl oxygens of the protein backbones of the "selectivity filter" (named by Bertil Hille) could efficiently replace the water molecules that normally shield potassium ions, but that sodium ions were smaller and cannot be completely dehydrated to allow such shielding, and therefore could not pass through. This mechanism was finally confirmed when the first structure of an ion channel was elucidated. A bacterial potassium channel KcsA, consisting of just the selectivity filter, "P" loop and two transmembrane helices was used as a model to study the permeability and the selectivity of ion channels. The determination of the molecular structure of KcsA by Roderick MacKinnon³² using X-ray crystallography won a share of the 2003 Nobel Prize in Chemistry.

Because of their small size and the difficulty of crystallizing integral membrane proteins for X-ray analysis, it is only very recently that scientists have been able to directly examine what channels "look like." Most of what researchers have deduced about channel operation so far they have established through electrophysiology, biochemistry, gene sequence comparison and mutagenesis and structural studies (X-ray crystallography and cryoelectron microscopy).

Channels can have single (e.g. members of the Chloride Intracellular Ion Channel family) or multiple transmembrane (K^+ channels, P2X receptors, Na^+ channels) domains which span the plasma membrane to form pores.

Ion channels can be classified by how they respond to their environment. For example, the ion channels involved in the action potential are voltage-sensitive channels; they open and close in response to the voltage across the membrane. Ligand-gated channels form another important class; these ion channels open and close in response to the binding of a ligand molecule, such as a neurotransmitter. Other ion channels open and close with mechanical forces. Still other ion channels—such as those of sensory neurons—open and close in response to other stimuli, such as light, temperature or pressure.

Leakage channels are the simplest type of ion channel,

²³https://en.wikipedia.org/wiki/Alan_Hodgkin

²⁴https://en.wikipedia.org/wiki/Andrew_Huxley

²⁵https://en.wikipedia.org/wiki/Kenneth_Stewart_Cole

²⁶https://en.wikipedia.org/wiki/Bernard_Katz

²⁷https://en.wikipedia.org/wiki/Ricardo_Miledi

²⁸https://en.wikipedia.org/wiki/Erwin_Neher

²⁹https://en.wikipedia.org/wiki/Bert_Sakmann

³⁰https://en.wikipedia.org/wiki/Bertil_Hille

³¹https://en.wikipedia.org/wiki/Clay_Armstrong

³²https://en.wikipedia.org/wiki/Roderick_MacKinnon

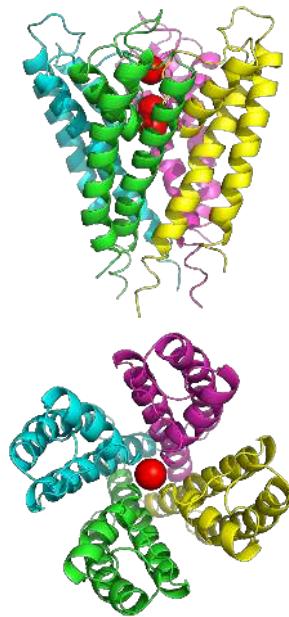


Figure 25.16: The structure of the potassium channel KcsA from *Streptomyces lividans* determined by X-ray crystallography. Top: side view; bottom: top view. KcsA shares sequence similarity with all known K^+ channels and was the first ion channel to have its structure solved at atomic resolution. It consists of four identical subunits that together form a cone shaped structure (top) with a ion selectivity filter at its outer end. Three K^+ ions (red spheres) are present in the channel, two of which are held 7.5 angstroms apart by the selectivity filter, a third K^+ ion shown in the channel pore below. Data from PDB 1BL8³³, rendered with open source molecular visualization tool PyMol³⁴.

in that their permeability is more or less constant. The types of leakage channels that have the greatest significance in neurons are potassium and chloride channels. Even these are not perfectly constant in their properties: First, most of them are voltage-dependent in the sense that they conduct better in one direction than the other (in other words, they are rectifiers); second, some of them are capable of being shut off by chemical ligands even though they do not require ligands in order to operate.

Voltage-gated ion channels³⁵ are a class of transmembrane proteins that form ion channels that are activated by changes in the electrical membrane potential near the channel. The membrane potential alters the conformation of the channel proteins, regulating their opening and closing. Cell membranes are generally impermeable to ions, thus they must diffuse through the membrane through transmembrane protein channels. They have a crucial role in excitable cells such as neuronal and muscle tissues, allowing a rapid and co-ordinated depolarization in response to triggering voltage change. Found along the axon and at the synapse, voltage-gated ion channels directionally propagate electrical signals. Voltage-gated ion-channels are usually ion-specific, and channels specific to sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), and chloride (Cl^-) ions have been identified. The opening and closing of the channels are triggered by changing ion concentration, and hence charge gradient, between the sides of the cell membrane.

Voltage-gated ion channels are generally composed of several subunits arranged in such a way that there is a central pore through which ions can travel down their electrochemical gradients. The channels tend to be ion-specific, although similarly sized and charged ions may sometimes travel through them. The functionality of voltage-gated ion channels is attributed to its three main discrete units: the voltage sensor, the pore or conducting pathway, and the gate. Na^+ , K^+ , and Ca^{2+} channels are composed of four transmembrane domains arranged around a central pore; these four domains are part of a single α -subunit in the case of most Na^+ and Ca^{2+} channels, whereas there are four α -subunits, each contributing one transmembrane domain, in most K^+ channels. The membrane-spanning segments, designated S1–S6, all take the form of alpha helices with specialized functions. The fifth and sixth transmembrane segments (S5 and S6) and pore loop serve the principal role of ion conduction, comprising the gate and pore of the channel, while S1–S4 serve as the voltage-sensing region. The four subunits may be identical, or different from one another. In addition to the four central α -subunits, there are also regulatory β -subunits, with oxidoreductase activity, which

³⁵ https://en.wikipedia.org/wiki/Voltage-gated_ion_channel

are located on the inner surface of the cell membrane and do not cross the membrane, and which are coassembled with the α -subunits in the endoplasmic reticulum.

Crystallographic structural studies of a potassium channel have shown that, when a potential difference is introduced over the membrane, the associated electric field induces a conformational change in the potassium channel. The conformational change distorts the shape of the channel proteins sufficiently such that the cavity, or channel, opens to allow influx or efflux to occur across the membrane. This movement of ions down their concentration gradients subsequently generates an electric current sufficient to depolarize the cell membrane.

Voltage-gated sodium channels and calcium channels are made up of a single polypeptide with four homologous domains. Each domain contains 6 membrane spanning alpha helices. One of these helices, S4, is the voltage sensing helix. The S4 segment contains many positive charges such that a high positive charge outside the cell repels the helix, keeping the channel in its closed state.

In general, the voltage sensing portion of the ion channel is responsible for the detection of changes in transmembrane potential that trigger the opening or closing of the channel. The S1-4 alpha helices are generally thought to serve this role. In potassium and sodium channels, voltage-sensing S4 helices contain positively-charged lysine or arginine residues in repeated motifs. In its resting state, half of each S4 helix is in contact with the cell cytosol. Upon depolarization, the positively-charged residues on the S4 domains move toward the exoplasmic surface of the membrane. It is thought that the first 4 arginines account for the gating current, moving toward the extracellular solvent upon channel activation in response to membrane depolarization. The movement of 10-12 of these protein-bound positive charges triggers a conformational change that opens the channel. The exact mechanism by which this movement occurs is not currently agreed upon, however the canonical, transporter, paddle, and twisted models are examples of current theories.

Movement of the voltage-sensor triggers a conformational change of the gate of the conducting pathway, controlling the flow of ions through the channel.

The main functional part of the voltage-sensitive protein domain of these channels generally contains a region composed of S3b and S4 helices, known as the "paddle" due to its shape, which appears to be a conserved sequence, interchangeable across a wide variety of cells and species. A similar voltage sensor paddle has also been found in a family of voltage sensitive phosphatases in various species. Genetic engineering of the paddle region

from a species of volcano-dwelling archaeabacteria into rat brain potassium channels results in a fully functional ion channel, as long as the whole intact paddle is replaced. This "modularity" allows use of simple and inexpensive model systems to study the function of this region, its role in disease, and pharmaceutical control of its behavior rather than being limited to poorly characterized, expensive, and/or difficult to study preparations.

Although voltage-gated ion channels are typically activated by membrane depolarization, some channels, such as inward-rectifier potassium ion channels, are activated instead by hyperpolarization.

The gate is thought to be coupled to the voltage sensing regions of the channels and appears to contain a mechanical obstruction to ion flow. While the S6 domain has been agreed upon as the segment acting as this obstruction, its exact mechanism is unknown. Possible explanations include: the S6 segment makes a scissor-like movement allowing ions to flow through, the S6 segment breaks into two segments allowing of passing of ions through the channel, or the S6 channel serving as the gate itself. The mechanism by which the movement of the S4 segment affects that of S6 is still unknown, however it is theorized that there is a S4-S5 linker whose movement allows the opening of S6.

Inactivation of ion channels occurs within milliseconds after opening. Inactivation is thought to be mediated by an intracellular gate that controls the opening of the pore on the inside of the cell. This gate is modeled as a ball tethered to a flexible chain. During inactivation, the chain folds in on itself and the ball blocks the flow of ions through the channel. Fast inactivation is directly linked to the activation caused by intramembrane movements of the S4 segments, though the mechanism linking movement of S4 and the engagement of the inactivation gate is unknown.

Important representative families of voltage-gated ion channels are the:

- Voltage-gated sodium channels: This family contains at least 9 members and is largely responsible for action potential creation and propagation. The pore-forming α subunits are very large (up to 4,000 amino acids) and consist of four homologous repeat domains (I-IV) each comprising six transmembrane segments (S1-S6) for a total of 24 transmembrane segments. The members of this family also coassemble with auxiliary β subunits, each spanning the membrane once. Both α and β subunits are extensively glycosylated.
- Voltage-gated calcium channels: This family consists of channels that are formed as a complex of several different subunits: α_1 , $\alpha_2\delta$, β_1-4 , and γ .

The $\alpha 1$ subunit forms the Ca^{2+} selective ion conducting pore while the associated subunits have several functions including modulation of gating. These channels play an important role in both linking muscle excitation with contraction as well as neuronal excitation with transmitter release. The α subunits have an overall structural resemblance to those of the sodium channels and are equally large.

- Cation channels of sperm: This small family of channels, normally referred to as CatSper channels, is related to the two-pore channels and distantly related to TRP channels.
- Voltage-gated potassium channels (KV): This family contains almost 40 members, which are further divided into 12 subfamilies. These channels are known mainly for their role in repolarizing the cell membrane following action potentials. The α subunits have six transmembrane segments, homologous to a single domain of the sodium channels. Correspondingly, they assemble as tetramers to produce a functioning channel.
- Some transient receptor potential channels: This group of channels, normally referred to simply as TRP channels, is named after their role in *Drosophila* phototransduction. This family, containing at least 28 members, is incredibly diverse in its method of activation. Some TRP channels seem to be constitutively open, while others are gated by voltage, intracellular Ca^{2+} , pH, redox state, osmolarity, and mechanical stretch. These channels also vary according to the ion(s) they pass, some being selective for Ca^{2+} while others are less selective, acting as cation channels. This family is subdivided into 6 subfamilies based on homology: classical (TRPC), vanilloid receptors (TRPV), melastatin (TRPM), polycystins (TRPP), mucolipins (TRPML), and ankyrin transmembrane protein 1 (TRPA).
- Hyperpolarization-activated cyclic nucleotide-gated channels: The opening of these channels is due to hyperpolarization rather than the depolarization required for other cyclic nucleotide-gated channels. These channels are also sensitive to the cyclic nucleotides cAMP and cGMP, which alter the voltage sensitivity of the channel's opening. These channels are permeable to the monovalent cations K^+ and Na^+ . There are 4 members of this family, all of which form tetramers of six-transmembrane α subunits. As these channels open under hyperpolarizing conditions, they function as pacemaking channels in the heart, particularly the SA node.
- Voltage-gated proton channels: Voltage-gated proton channels open with depolarization, but in a strongly pH-sensitive manner. The result is that these channels open only when the electrochemical

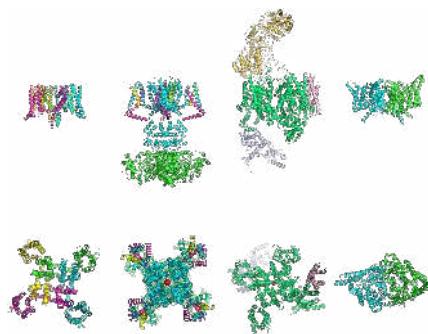


Figure 25.17: The structures (from left to right) of voltage-gated sodium, potassium, calcium and chloride channels. Top: side view; bottom: top view. Data from PDB 5EK0³⁶, PDB 2A79³⁷, PDB 3JBR³⁸ and PDB 1KPL³⁹ rendered with open source molecular visualization tool PyMol⁴⁰.

gradient is outward, such that their opening will only allow protons to leave cells. Their function thus appears to be acid extrusion from cells. Another important function occurs in phagocytes (e.g. eosinophils, neutrophils, macrophages) during the “respiratory burst.” When bacteria or other microbes are engulfed by phagocytes, the enzyme NADPH oxidase assembles in the membrane and begins to produce reactive oxygen species (ROS) that help kill bacteria. NADPH oxidase is electrogenic, moving electrons across the membrane, and proton channels open to allow proton flux to balance the electron movement electrically.

Ligand-gated ion channels⁴¹, also commonly referred to as ionotropic receptors, are a group of transmembrane ion-channel proteins which open to allow ions such as Na^+ , K^+ , Ca^{2+} , and Cl^- to pass through the membrane in response to the binding of a chemical messenger (i.e. a ligand), such as a neurotransmitter.

When a presynaptic neuron is excited, it releases a neurotransmitter from vesicles into the synaptic cleft. The neurotransmitter then binds to receptors located on the postsynaptic neuron. If these receptors are ligand-gated ion channels, a resulting conformational change opens the ion channels, which leads to a flow of ions across the cell membrane. This, in turn, results in either a depolarization, for an excitatory receptor response, or a hyperpolarization, for an inhibitory response.

These receptor proteins are typically composed of at least two different domains: a transmembrane domain which includes the ion pore, and an extracellular domain which includes the ligand binding location (an allosteric binding site). This modularity has enabled a ‘divide and

⁴¹https://en.wikipedia.org/wiki/Ligand-gated_ion_channel

'conquer' approach to finding the structure of the proteins (crystallising each domain separately). The function of such receptors located at synapses is to convert the chemical signal of presynaptically released neurotransmitter directly and very quickly into a postsynaptic electrical signal. Many ligand-gated ion channels are additionally modulated by allosteric ligands, by channel blockers, ions, or the membrane potential. Ligand-gated ion channels are classified into three superfamilies which lack evolutionary relationship: cys-loop receptors, ionotropic glutamate receptors and ATP-gated channels.

The cys-loop receptors are named after a characteristic loop formed by a disulfide bond between two cysteine residues in the N terminal extracellular domain. They are part of a larger family of pentameric ligand-gated ion channels that usually lack this disulfide bond, hence the tentative name "Pro-loop receptors". A binding site in the extracellular N-terminal ligand-binding domain gives them receptor specificity for (1) acetylcholine (AcCh), (2) serotonin, (3) glycine, (4) glutamate and (5) γ -aminobutyric acid (GABA) in vertebrates. The receptors are subdivided with respect to the type of ion that they conduct (anionic or cationic) and further into families defined by the endogenous ligand. They are usually pentameric with each subunit containing 4 transmembrane helices constituting the transmembrane domain, and a beta sheet sandwich type, extracellular, N terminal, ligand binding domain. Some also contain an intracellular domain like shown in the image.

The prototypic ligand-gated ion channel is the nicotinic acetylcholine receptor. It consists of a pentamer of protein subunits (typically $\alpha\beta\gamma\delta$), with two binding sites for acetylcholine (one at the interface of each alpha sub-unit). When the acetylcholine binds it alters the receptor's configuration (twists the T2 helices which moves the leucine residues, which block the pore, out of the channel pathway) and causes the constriction in the pore of approximately 3 angstroms to widen to approximately 8 angstroms so that ions can pass through. This pore allows Na^+ ions to flow down their electrochemical gradient into the cell. With a sufficient number of channels opening at once, the inward flow of positive charges carried by Na^+ ions depolarizes the postsynaptic membrane sufficiently to initiate an action potential.

While single-cell organisms like bacteria would have little apparent need for the transmission of an action potential, a bacterial homologue to an LIC has been identified, hypothesized to act nonetheless as a chemoreceptor. This prokaryotic nAChR variant is known as the GLIC receptor, after the species in which it was identified; *Gloeobacter* Ligand-gated Ion C channel.

The ionotropic glutamate receptors bind the neuro-

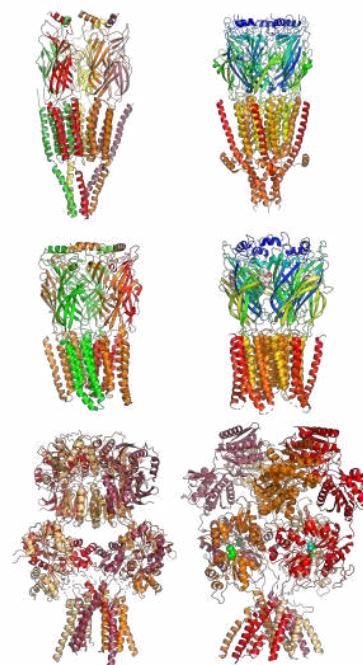


Figure 25.18: Structures of ligand-gated ion channels. Top row: The nicotinic acetylcholine receptor (left) and the serotonin 5-HT3 receptor. Middle row: The GABA_A receptor (left) and glycine receptor (right). Bottom row: The AMPA receptor (left) and NMDA receptor (right). Data from PDB 2BG9⁴², PDB 6HIO⁴³, PDB 6D6U⁴⁴, PDB 3JAD⁴⁵, PDB 5IDE⁴⁶, and PDB 4PE5⁴⁷, rendered with open source molecular visualization tool PyMol⁴⁸.

transmitter glutamate. They form tetramers with each subunit consisting of an extracellular amino terminal domain (ATD, which is involved tetramer assembly), an extracellular ligand binding domain (LBD, which binds glutamate), and a transmembrane domain (TMD, which forms the ion channel). The transmembrane domain of each subunit contains three transmembrane helices as well as a half membrane helix with a reentrant loop. The structure of the protein starts with the ATD at the N terminus followed by the first half of the LBD which is interrupted by helices 1,2 and 3 of the TMD before continuing with the final half of the LBD and then finishing with helix 4 of the TMD at the C terminus. This means there are three links between the TMD and the extracellular domains. Each subunit of the tetramer has a binding site for glutamate formed by the two LBD sections forming a clamshell like shape. Only two of these sites in the tetramer need to be occupied to open the ion channel. The pore is mainly formed by the half helix 2 in a way which resembles an inverted potassium channel.

Also called G protein-coupled receptor, seven-

transmembrane domain receptor, 7 TM receptor, constitute a large protein family of receptors that sense molecules outside the cell and activate inside signal transduction pathways and, ultimately, cellular responses. They pass through the cell membrane 7 times. G-protein-Linked receptors are a huge family that have hundreds of members identified. Ion-channel-linked receptors (e.g. GABAB, NMDA, etc.) are only a part of them.

25.10.6 The Reversal Potential

The reversal potential (or equilibrium potential) of an ion is the value of transmembrane voltage at which diffusive and electrical forces counterbalance, so that there is no net ion flow across the membrane. This means that the transmembrane voltage exactly opposes the force of diffusion of the ion, such that the net current of the ion across the membrane is zero and unchanging. The reversal potential is important because it corresponds to the voltage that acts on channels permeable to that ion—in other words, it gives the voltage that the ion concentration gradient generates when it acts as a battery.

The equilibrium potential of a particular ion is usually designated by the notation E_{ion} . The equilibrium potential for any ion can be calculated using the Nernst equation. For example, reversal potential for potassium ions will be as follows:

$$E_{eq,K^+} = \frac{RT}{zF} \ln \frac{[K^+]_o}{[K^+]_i}$$

where

- E_{eq,K^+} is the equilibrium potential for potassium, measured in volts
- R is the universal gas constant, equal to $8.314 \text{ Joule}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$
- T is the absolute temperature, measured in Kelvin
- z is the number of elementary charges of the ion in question involved in the reaction
- F is the Faraday constant, equal to $96,485 \text{ Coulomb}\cdot\text{mol}^{-1}$ or $\text{J}\cdot\text{V}^{-1}\cdot\text{mol}^{-1}$
- $[K^+]_o$ is the extracellular concentration of potassium, measured in $\text{mol}\cdot\text{m}^{-3}$ or $\text{mmol}\cdot\text{l}^{-1}$
- $[K^+]_i$ is the intracellular concentration of potassium

Even if two different ions have the same charge (i.e., K^+ and Na^+), they can still have very different equilibrium potentials, provided their outside and/or inside concentrations differ. Take, for example, the equilibrium potentials of potassium and sodium in neurons. The potassium equilibrium potential E_K is -84 mV with 5 mM potassium outside and 140 mM inside. On the other hand, the sodium equilibrium potential, E_{Na} , is approximately $+66 \text{ mV}$ with approximately 12 mM sodium inside and 140 mM outside.

A neuron's resting membrane potential actually changes during the development of an organism. In order for a neuron to eventually adopt its full adult function, its potential must be tightly regulated during development. As an organism progresses through development the resting membrane potential becomes more negative. Glial cells are also differentiating and proliferating as development progresses in the brain. The addition of these glial cells increases the organism's ability to regulate extracellular potassium. The drop in extracellular potassium can lead to a decrease in membrane potential of 35 mV.

Cell excitability is the change in membrane potential that is necessary for cellular responses in various tissues. Cell excitability is a property that is induced during early embryogenesis. Excitability of a cell has also been defined as the ease with which a response may be triggered. The resting potential forms the basis of cell excitability and these processes are fundamental for the generation of graded and action potentials.

The most important regulators of cell excitability are the extracellular calcium concentration and the calcium-sensing receptor. Calcium is also the most important second messenger in excitable cell signaling. Other important proteins that regulate cell excitability are voltage-gated ion channels, ion transporters, membrane receptors and hyperpolarization-activated cyclic-nucleotide-gated channels. For example, potassium channels are important regulators of excitability in neurons, cardiac myocytes and many other excitable cells like astrocytes. Activation of synaptic receptors initiates long-lasting changes in neuronal excitability.

Many cell types are considered to have an excitable membrane. Excitable cells are neurons, cardiac myocytes, skeletal myocytes, smooth muscle cells, many types of endothelial cells (e.g. beta cells), glial cells (e.g. astrocytes), mechanoreceptor cells (e.g. hair cells and Merkel cells), chemoreceptor cells (e.g. glomus cells, taste receptors), some plant cells and possibly immune cells. Astrocytes display a form of non-electrical excitability based on intracellular calcium variations related to the expression of several receptors through which they can detect the synaptic signal. In neurons, there are different membrane properties in some portions of the cell, for example, dendritic excitability endows neurons with the capacity for coincidence detection of spatially separated inputs.

25.10.7 The Resting Potential

When the membrane potential of a cell goes for a long period of time without changing significantly, it is referred to as a resting potential or resting voltage. This term is used for the membrane potential of non-excitable cells,

but also for the membrane potential of excitable cells in the absence of excitation. In excitable cells, the other possible states are graded membrane potentials (of variable amplitude), and action potentials, which are large, all-or-nothing rises in membrane potential that usually follow a fixed time course. Excitable cells include neurons, muscle cells, and some secretory cells in glands. Even in other types of cells, however, the membrane voltage can undergo changes in response to environmental or intracellular stimuli. For example, depolarization of the plasma membrane appears to be an important step in programmed cell death.

The interactions that generate the resting potential are modeled by the Goldman equation. This is similar in form to the Nernst equation shown above, in that it is based on the charges of the ions in question, as well as the difference between their inside and outside concentrations. However, it also takes into consideration the relative permeability of the plasma membrane to each ion in question.

$$E_m = \frac{RT}{F} \ln \left(\frac{P_K[K^+]_{out} + P_{Na}[Na^+]_{out} + P_{Cl}[Cl^-]_{in}}{P_K[K^+]_{in} + P_{Na}[Na^+]_{in} + P_{Cl}[Cl^-]_{out}} \right)$$

The three ions that appear in this equation are potassium (K^+), sodium (Na^+), and chloride (Cl^-). Calcium is omitted, but can be added to deal with situations in which it plays a significant role. Being an anion, the chloride terms are treated differently from the cation terms; the intracellular concentration is in the numerator, and the extracellular concentration in the denominator, which is reversed from the cation terms. P_i stands for the relative permeability of the ion type i .

In essence, the Goldman formula expresses the membrane potential as a weighted average of the reversal potentials for the individual ion types, weighted by permeability. (Although the membrane potential changes about 100 mV during an action potential, the concentrations of ions inside and outside the cell do not change significantly. They remain close to their respective concentrations when the membrane is at resting potential.) In most animal cells, the permeability to potassium is much higher in the resting state than the permeability to sodium. As a consequence, the resting potential is usually close to the potassium reversal potential. The permeability to chloride can be high enough to be significant, but, unlike the other ions, chloride is not actively pumped, and therefore equilibrates at a reversal potential very close to the resting potential determined by the other ions.

Values of resting membrane potential in most animal cells usually vary between the potassium reversal potential (usually around -80 mV) and around -40 mV. The resting potential in excitable cells (capable of producing action

potentials) is usually near -60 mV—more depolarized voltages would lead to spontaneous generation of action potentials. Immature or undifferentiated cells show highly variable values of resting voltage, usually significantly more positive than in differentiated cells. In such cells, the resting potential value correlates with the degree of differentiation: undifferentiated cells in some cases may not show any transmembrane voltage difference at all.

Maintenance of the resting potential can be metabolically costly for a cell because of its requirement for active pumping of ions to counteract losses due to leakage channels. The cost is highest when the cell function requires an especially depolarized value of membrane voltage. For example, the resting potential in daylight-adapted blowfly (*Calliphora vicina*) photoreceptors can be as high as -30 mV. This elevated membrane potential allows the cells to respond very rapidly to visual inputs; the cost is that maintenance of the resting potential may consume more than 20% of overall cellular ATP.

On the other hand, the high resting potential in undifferentiated cells can be a metabolic advantage. This apparent paradox is resolved by examination of the origin of that resting potential. Little-differentiated cells are characterized by extremely high input resistance, which implies that few leakage channels are present at this stage of cell life. As an apparent result, potassium permeability becomes similar to that for sodium ions, which places resting potential in-between the reversal potentials for sodium and potassium as discussed above. The reduced leakage currents also mean there is little need for active pumping in order to compensate, therefore low metabolic cost.

25.10.8 The Action Potential

An action potential occurs when the membrane potential of a specific cell location rapidly rises and falls: this depolarisation then causes adjacent locations to similarly depolarise. Action potentials occur in several types of animal cells, called excitable cells, which include neurons, muscle cells, endocrine cells, glomus cells (peripheral chemoreceptor cells mainly located in the carotid and aortic bodies), and in some plant cells.

In neurons, action potentials play a central role in cell-to-cell communication by providing for—or with regard to saltatory conduction, assisting—the propagation of signals along the neuron's axon toward synaptic boutons situated at the ends of an axon; these signals can then connect with other neurons at synapses, or to motor cells or glands. In other types of cells, their main function is to activate intracellular processes. In muscle cells, for example, an action potential is the first step in the chain of events leading to contraction. In beta cells of the pancreas, they provoke release of insulin. Action potentials in neurons are

also known as “nerve impulses” or “spikes”, and the temporal sequence of action potentials generated by a neuron is called its “spike train”. A neuron that emits an action potential, or nerve impulse, is often said to “fire”.

Action potentials are generated by special types of voltage-gated ion channels embedded in a cell’s plasma membrane. These channels are shut when the membrane potential is near the (negative) resting potential of the cell, but they rapidly begin to open if the membrane potential increases to a precisely defined threshold voltage, depolarising the transmembrane potential. When the channels open, they allow an inward flow of sodium ions, which changes the electrochemical gradient, which in turn produces a further rise in the membrane potential. This then causes more channels to open, producing a greater electric current across the cell membrane and so on. The process proceeds explosively until all of the available ion channels are open, resulting in a large upswing in the membrane potential. The rapid influx of sodium ions causes the polarity of the plasma membrane to reverse, and the ion channels then rapidly inactivate. As the sodium channels close, sodium ions can no longer enter the neuron, and they are then actively transported back out of the plasma membrane. Potassium channels are then activated, and there is an outward current of potassium ions, returning the electrochemical gradient to the resting state. After an action potential has occurred, there is a transient negative shift, called the afterhyperpolarization.

In animal cells, there are two primary types of action potentials. One type is generated by voltage-gated sodium channels, the other by voltage-gated calcium channels. Sodium-based action potentials usually last for under one millisecond, but calcium-based action potentials may last for 100 milliseconds or longer. In some types of neurons, slow calcium spikes provide the driving force for a long burst of rapidly emitted sodium spikes. In cardiac muscle cells, on the other hand, an initial fast sodium spike provides a “primer” to provoke the rapid onset of a calcium spike, which then produces muscle contraction

25.10.9 Graded Potentials

As explained above, the potential at any point in a cell’s membrane is determined by the ion concentration differences between the intracellular and extracellular areas, and by the permeability of the membrane to each type of ion. The ion concentrations do not normally change very quickly (with the exception of Ca^{2+} , where the baseline intracellular concentration is so low that even a small influx may increase it by orders of magnitude), but the permeabilities of the ions can change in a fraction of a millisecond, as a result of activation of ligand-gated ion channels. The

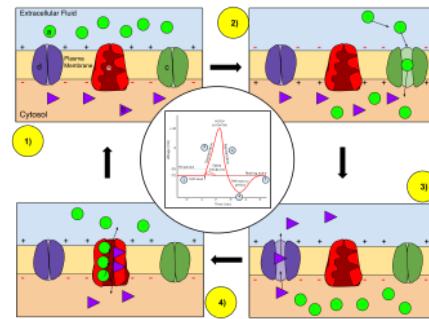


Figure 25.19: Ion movement during an action potential.⁴⁹

Key: a) Sodium (Na^+) ion. b) Potassium (K^+) ion. c) Sodium channel. d) Potassium channel. e) Sodium-potassium pump. In the stages of an action potential, the permeability of the membrane of the neuron changes. At the resting state (1), sodium and potassium ions have limited ability to pass through the membrane, and the neuron has a net negative charge inside. Once the action potential is triggered, the depolarization (2) of the neuron activates sodium channels, allowing sodium ions to pass through the cell membrane into the cell, resulting in a net positive charge in the neuron relative to the extracellular fluid. After the action potential peak is reached, the neuron begins repolarization (3), where the sodium channels close and potassium channels open, allowing potassium ions to cross the membrane into the extracellular fluid, returning the membrane potential to a negative value. Finally, there is a refractory period (4), during which the voltage-dependent ion channels are inactivated while the Na^+ and K^+ ions return to their resting state distributions across the membrane (1), and the neuron is ready to repeat the process for the next action potential.

change in membrane potential can be either large or small, depending on how many ion channels are activated and what type they are, and can be either long or short, depending on the lengths of time that the channels remain open. Changes of this type are referred to as graded potentials, in contrast to action potentials, which have a fixed amplitude and time course.

As can be derived from the Goldman equation shown above, the effect of increasing the permeability of a membrane to a particular type of ion shifts the membrane potential toward the reversal potential for that ion. Thus, opening Na^+ channels shifts the membrane potential toward the Na^+ reversal potential, which is usually around +100 mV. Likewise, opening K^+ channels shifts the membrane potential toward about -90 mV, and opening Cl^- channels shifts it toward about -70 mV (resting potential of most membranes). Thus, Na^+ channels shift the membrane potential in a positive direction, K^+ channels shift it in a negative direction (except when the membrane is hyperpolarized to a value more negative than the K^+ reversal potential), and Cl^- channels tend to shift it towards the resting potential.

Graded membrane potentials are particularly important in neurons, where they are produced by synapses—a temporary change in membrane potential produced by activation of a synapse by a single graded or action potential is called a postsynaptic potential. Neurotransmitters that act to open Na^+ channels typically cause the membrane potential to become more positive, while neurotransmitters that activate K^+ channels typically cause it to become more negative; those that inhibit these channels tend to have the opposite effect.

Whether a postsynaptic potential is considered excitatory or inhibitory depends on the reversal potential for the ions of that current, and the threshold for the cell to fire an action potential (around -50mV). A postsynaptic current with a reversal potential above threshold, such as a typical Na^+ current, is considered excitatory. A current with a reversal potential below threshold, such as a typical K^+ current, is considered inhibitory. A current with a reversal potential above the resting potential, but below threshold, will not by itself elicit action potentials, but will produce subthreshold membrane potential oscillations. Thus, neurotransmitters that act to open Na^+ channels produce excitatory postsynaptic potentials, or EPSPs, whereas neurotransmitters that act to open K^+ or Cl^- channels typically produce inhibitory postsynaptic potentials, or IPSPs. When multiple types of channels are open within the same time period, their postsynaptic potentials summate (are added together).

From the viewpoint of biophysics, the resting membrane potential is merely the membrane potential that results from the membrane permeabilities that predominate

when the cell is resting. The Goldman equation of weighted averages always applies, but the following approach may be more easily visualized. At any given moment, there are two factors for an ion that determine how much influence that ion will have over the membrane potential of a cell:

1. That ion's driving force
2. That ion's permeability

If the driving force is high, then the ion is being "pushed" across the membrane. If the permeability is high, it will be easier for the ion to diffuse across the membrane.

- Driving force is the net electrical force available to move that ion across the membrane. It is calculated as the difference between the voltage that the ion "wants" to be at (its equilibrium potential) and the actual membrane potential (E_m). So, in formal terms, the driving force for an ion = $E_m - E_{\text{ion}}$
- For example, at our earlier calculated resting potential of -73 mV, the driving force on potassium is 7 mV : (-73 mV) - (-80 mV) = 7 mV. The driving force on sodium would be (-73 mV) - (60 mV) = -133 mV.
- Permeability is a measure of how easily an ion can cross the membrane. It is normally measured as the (electrical) conductance and the unit, siemens (S), corresponds to $1 \text{ C}\cdot\text{s}^{-1}\cdot\text{V}^{-1}$, that is one coulomb per second per volt of potential.

So, in a resting membrane, while the driving force for potassium is low, its permeability is very high. Sodium has a huge driving force but almost no resting permeability. In this case, potassium carries about 20 times more current than sodium, and thus has 20 times more influence over E_m than does sodium.

However, consider another case—the peak of the action potential. Here, permeability to Na is high and K permeability is relatively low. Thus, the membrane moves to near E_{Na} and far from E_{K} .

25.11 Neurotransmission

Neurotransmission (Latin: *transmissio* “passage, crossing” from *transmittere* “send, let through”) is the process by which signaling molecules called neurotransmitters are released by the axon terminal of a neuron (the presynaptic neuron), and bind to and react with the receptors on the dendrites of another neuron (the postsynaptic neuron) a short distance away.

Neurotransmission is regulated by several different factors: the availability and rate-of-synthesis of the neurotransmitter, the release of that neurotransmitter, the baseline activity of the postsynaptic cell, the number of

available postsynaptic receptors for the neurotransmitter to bind to, and the subsequent removal or deactivation of the neurotransmitter by enzymes or presynaptic reuptake.

In response to a threshold action potential or graded electrical potential, a neurotransmitter is released at the presynaptic terminal. The released neurotransmitter may then move across the synaptic cleft to bind to receptors in the postsynaptic neuron. Binding of neurotransmitters may influence the postsynaptic neuron in either an inhibitory or excitatory way. The binding of neurotransmitters to receptors in the postsynaptic neuron can trigger either short term changes, such as changes in the membrane potential called postsynaptic potentials, or longer term changes by the activation of signaling cascades.

Neurons form complex biological neural networks through which nerve impulses (action potentials) travel. Neurons do not touch each other (except in the case of an electrical synapse through a gap junction); instead, neurons interact at close contact points called synapses. When the nerve impulse arrives at the synapse, it may cause the release of neurotransmitters, which influence another (postsynaptic) neuron. The postsynaptic neuron may receive inputs from many additional neurons, both excitatory and inhibitory. The excitatory and inhibitory influences are summed, and if the net effect is inhibitory, the neuron will be less likely to "fire" (i.e., generate an action potential), and if the net effect is excitatory, the neuron will be more likely to fire. How likely a neuron is to fire depends on how far its membrane potential is from the threshold potential, the voltage at which an action potential is triggered because enough voltage-dependent sodium channels are activated so that the net inward sodium current exceeds all outward currents. Excitatory inputs bring a neuron closer to threshold, while inhibitory inputs bring the neuron farther from threshold. An action potential is an "all-or-none" event; neurons whose membranes have not reached threshold will not fire, while those that do must fire. Once the action potential is initiated (traditionally at the axon hillock), it will propagate along the axon, leading to release of neurotransmitters at the synaptic bouton to pass along information to yet another adjacent neuron.

Stages in neurotransmission at the synapse

- Synthesis of the neurotransmitter. This can take place in the cell body, in the axon, or in the axon terminal.
- Storage of the neurotransmitter in storage granules or vesicles in the axon terminal.
- Calcium enters the axon terminal during an action potential, causing release of the neurotransmitter into the synaptic cleft.

- After its release, the transmitter binds to and activates a receptor in the postsynaptic membrane.
- Deactivation of the neurotransmitter. The neurotransmitter is either destroyed enzymatically, or taken back into the terminal from which it came, where it can be reused, or degraded and removed.

25.11.1 The Synapse

In the nervous system, a synapse is a structure that permits a neuron (or nerve cell) to pass an electrical or chemical signal to another neuron or to the target effector cell.

Santiago Ramón y Cajal proposed that neurons are not continuous throughout the body, yet still communicate with each other, an idea known as the neuron doctrine. The word "synapse" – from the Greek *synapsis* (συνάψις), meaning "conjunction", in turn from συνάπτειν (συν ("together") and ἄπτειν ("to fasten")) – was introduced in 1897 by the English neurophysiologist Charles Sherrington⁵⁰ in Michael Foster's Textbook of Physiology. Sherrington struggled to find a good term that emphasized a union between two separate elements, and the actual term "synapse" was suggested by the English classical scholar Arthur Woollgar Verrall, a friend of Foster. Some authors generalize the concept of the synapse to include the communication from a neuron to any other cell type, such as to a motor cell, although such non-neuronal contacts may be referred to as junctions (a historically older term). A landmark electronmicroscopy study by Sanford Palay⁵¹ demonstrated the existence of synapses. Palay examined thin sections of the abducens nucleus, and on the surfaces of dendrites and cell bodies he encountered clublike profiles that were filled with mitochondria and contained vesicles that were concentrated close to the presynaptic membrane. He also noticed that the pre- and postsynaptic membranes were thickened and appeared denser, and most importantly although these membranes appeared to adhere together, they were in fact separated by a thin intercellular space, the synaptic cleft. This observation directly confirmed Cajal's idea about the synaptic junctions between nerve cells.

Synapses are essential to neuronal function: neurons are cells that are specialized to pass signals to individual target cells, and synapses are the means by which they do so. At a synapse, the plasma membrane of the signal-passing neuron (the presynaptic neuron) comes into close apposition with the membrane of the target (postsynaptic) cell. Both the presynaptic and postsynaptic sites contain extensive arrays of molecular machinery that link the two membranes together and carry out the signaling process. In many synapses, the presynaptic part is located

⁵⁰https://en.wikipedia.org/wiki/Charles_Scott_Sherrington

⁵¹https://en.wikipedia.org/wiki/Sanford_Palay

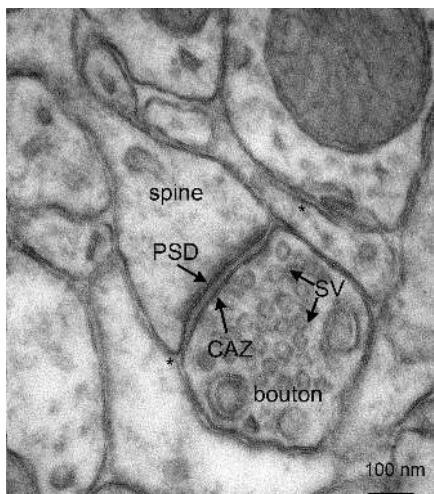


Figure 25.20: Electron micrograph of rat cortex showing multiple pre- and postsynaptic structures, as well as astrocytic endfeet (*) in close contact with synapses. Note the presence of numerous synaptic vesicles in the presynaptic boutons. CAZ, cytomatrix at the active zone; PSD, postsynaptic density; SV, synaptic vesicles. Scalebar: 100 nm. From Proteomics of the Synapse – A Quantitative Approach to Neuronal Plasticity Daniela C. Dieterich, Michael R. Kreutz Molecular & Cellular Proteomics February 1, 2016, First published on August 25, 2015, 15 (2) 368–381; DOI: 10.1074/mcp.R115.051482⁵²

on an axon and the postsynaptic part is located on a dendrite or soma. Astrocytes also exchange information with the synaptic neurons, responding to synaptic activity and, in turn, regulating neurotransmission. Synapses (at least chemical synapses) are stabilized in position by synaptic adhesion molecules (SAMs) projecting from both the pre- and post-synaptic neuron and sticking together where they overlap; SAMs may also assist in the generation and functioning of synapses.

There are two fundamentally different types of synapses:

- In a chemical synapse, electrical activity in the presynaptic neuron is converted (via the activation of voltage-gated calcium channels) into the release of a chemical called a neurotransmitter that binds to receptors located in the plasma membrane of the postsynaptic cell. The neurotransmitter may initiate an electrical response or a secondary messenger pathway that may either excite or inhibit the postsynaptic neuron. Chemical synapses can be classified according to the neurotransmitter released: glutamatergic (often excitatory), GABAergic (often inhibitory), cholinergic (e.g. vertebrate neuromuscular junction), and adrenergic (releasing

norepinephrine). Because of the complexity of receptor signal transduction, chemical synapses can have complex effects on the postsynaptic cell.

- In an electrical synapse, the presynaptic and postsynaptic cell membranes are connected by special channels called gap junctions that are capable of passing an electric current, causing voltage changes in the presynaptic cell to induce voltage changes in the postsynaptic cell. The main advantage of an electrical synapse is the rapid transfer of signals from one cell to the next.

Synapses can be classified by the type of cellular structures serving as the pre- and post-synaptic components. The vast majority of synapses in the mammalian nervous system are classical axo-dendritic synapses (axon synapsing upon a dendrite), however, a variety of other arrangements exist. These include but are not limited to axo-axonic, dendro-dendritic, axo-secretory, somato-dendritic, dendro-somatic, and somato-somatic synapses.

The axon can synapse onto a dendrite, onto a cell body, or onto another axon or axon terminal, as well as into the bloodstream or diffusely into the adjacent nervous tissue.

The postsynaptic density (PSD) is a protein dense specialization attached to the postsynaptic membrane. PSDs were originally identified by electron microscopy as an electron-dense region at the membrane of a postsynaptic neuron. The PSD is in close apposition to the presynaptic active zone and ensures that receptors are in close proximity to presynaptic neurotransmitter release sites. PSDs vary in size and composition among brain regions and have been studied in great detail at glutamatergic synapses. Hundreds of proteins have been identified in the postsynaptic density including glutamate receptors, scaffold proteins, and many signaling molecules.

PSDs are sized on the order of 250 to 500 nanometres in diameter and 25 to 50 nanometres in thickness, depending on the activity state of the synapse. During synaptic plasticity, the total size of the PSD is increasing along with an increase in synaptic size and strength after inducing long-term potentiation at single synapses.

Many proteins in the PSD are involved in the regulation of synaptic function. Key among these, are postsynaptic density-95 (PSD95), neuroligin (a cellular adhesion molecule), NMDA receptors, AMPA receptors, calcium/calmodulin-dependent protein kinase II and actin. As protein detection technologies have increased in sensitivity, such as with improvements in mass spectrometry techniques, many more proteins have been found to be part of the PSD. Current estimates are that several hundred proteins are found at PSDs in different

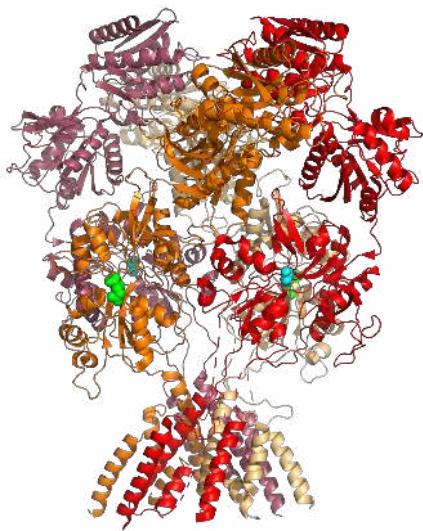


Figure 25.21: A cartoon representation of the atomic structure of the GluN1a/GluN2B N-Methyl-D-aspartate (NMDA) receptor subtype of the family of ionotropic glutamate receptors. The agonist glutamate (green spheres) is bound to the GluN2B subunit, the co-agonist glycine is bound to the GluN1A subunit. Data from PDB 4PE5⁵³, rendered with open source molecular visualization tool PyMol⁵⁴.

brain regions and during different states of development and synaptic activity. PSDs also contain cell adhesion molecules and a diverse set of other signaling proteins. Many of the PSD proteins contain PDZ domains.

The PSD has been proposed to concentrate and organize neurotransmitter receptors in the synaptic cleft. The PSD also serves as a signaling apparatus. For instance kinases and phosphatases in the PSD are activated and released from the PSD to change the activity of proteins located in the spine or are transported to the nucleus to affect protein synthesis. Some of the features of the PSD are similar to the neuromuscular junction and other cellular junctions, as the PSD has been modeled as a specialized cellular junction that allows for rapid, asymmetrical signaling.

The adult human brain is estimated to contain from 10^{14} to 5×10^{14} (100–500 trillion) synapses. Every cubic millimeter of cerebral cortex contains roughly a billion (10^9) of them. The number of synapses in the human cerebral cortex has separately been estimated at 0.15 quadrillion (150 trillion)

It is widely accepted that the synapse plays a role in the formation of memory. As neurotransmitters activate receptors across the synaptic cleft, the connection between the two neurons is strengthened when both neurons are active at the same time, as a result of the receptor's signaling

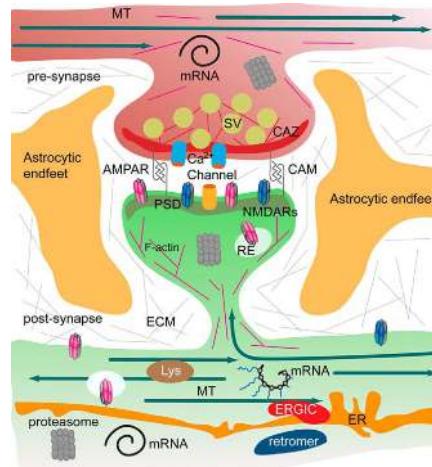


Figure 25.22: The tetrapartite synapse of principal neurons in the forebrain, consisting of the pre- and postsynaptic compartment, astrocytic endfeet, and the extracellular matrix has a tightly regulated protein composition. A microsecretory system is present in synapses and dendrites that allows for translation of mRNA, local synthesis of, processing and insertion of transmembrane proteins. Hence the turnover of the synaptic protein machinery is controlled by local and somatic de novo protein synthesis, protein degradation by the ubiquitin proteasome system, lysosomes and autophagosomes. In addition, the association of proteins with pre- and postsynaptic compartments is highly dynamic. Molecular machineries and organelles for proteostasis are shared between synapses in dendritic segments. Proteins are transported in and out of the synapse as well as by diffusion of transmembrane proteins. These processes govern the activity-dependent assembly of the pre- and postsynaptic scaffold and the synaptic surface expression of receptors, calcium channels and cell adhesion molecules. Abbreviations: CAM, cell adhesion molecules; CAZ, cytomatrix at the active zone; ECM, extracellular matrix; ER, endoplasmatic reticulum; ERGIC, endoplasmatic reticulum Golgi intermediate compartment; MT, microtubules; PSD, postsynaptic density; RE, recycling endosomes; Lys, lysomes; SV, synaptic vesicle. From Proteomics of the Synapse – A Quantitative Approach to Neuronal Plasticity Daniela C. Dieterich, Michael R. Kreutz Molecular & Cellular Proteomics February 1, 2016, First published on August 25, 2015, 15 (2) 368–381; DOI: 10.1074/mcp.R115.051482⁵⁵

mechanisms. The strength of two connected neural pathways is thought to result in the storage of information, resulting in memory. This process of synaptic strengthening is known as long-term potentiation.

Synaptic transmission can be changed by previous activity. These changes are called synaptic plasticity and may result in either a decrease in the efficacy of the synapse, called depression, or an increase in efficacy, called potentiation. These changes can either be long-term or short-term. Forms of short-term plasticity include synaptic fatigue or depression and synaptic augmentation. Forms of long-term plasticity include long-term depression and long-term potentiation. Synaptic plasticity can be either homosynaptic (occurring at a single synapse) or heterosynaptic (occurring at multiple synapses).

By altering the release of neurotransmitters, the plasticity of synapses can be controlled in the presynaptic cell. The postsynaptic cell can be regulated by altering the function and number of its receptors. Changes in postsynaptic signaling are most commonly associated with a N-methyl-D-aspartic acid receptor (NMDAR)-dependent long-term potentiation (LTP) and long-term depression (LTD) due to the influx of calcium into the post-synaptic cell, which are the most analyzed forms of plasticity at excitatory synapses.

A neurotransmitter can influence the function of a neuron through a remarkable number of mechanisms. In its direct actions in influencing a neuron's electrical excitability, however, a neurotransmitter acts in only one of two ways: excitatory or inhibitory. A neurotransmitter influences trans-membrane ion flow either to increase (excitatory) or to decrease (inhibitory) the probability that the cell with which it comes in contact will produce an action potential. Thus, despite the wide variety of synapses, they all convey messages of only these two types, and they are labeled as such. Type I synapses are excitatory in their actions, whereas type II synapses are inhibitory. Each type has a different appearance and is located on different parts of the neurons under its influence.

Type I (excitatory) synapses are typically located on the shafts or the spines of dendrites, whereas type II (inhibitory) synapses are typically located on a cell body. In addition, Type I synapses have round synaptic vesicles, whereas the vesicles of type II synapses are flattened. The material on the presynaptic and post-synaptic membranes is denser in a Type I synapse than it is in a type II, and the type I synaptic cleft is wider. Finally, the active zone on a Type I synapse is larger than that on a Type II synapse.

The different locations of type I and type II synapses divide a neuron into two zones: an excitatory dendritic tree and an inhibitory cell body. From an inhibitory perspective,

excitation comes in over the dendrites and spreads to the axon hillock to trigger an action potential. If the message is to be stopped, it is best stopped by applying inhibition on the cell body, close to the axon hillock where the action potential originates.

Here is a summary of the sequence of events that take place in synaptic transmission from a presynaptic neuron to a postsynaptic cell. Each step is explained in more detail below. Note that with the exception of the final step, the entire process may run only a few hundred microseconds, in the fastest synapses.

- The process begins with an action potential traveling along the membrane of the presynaptic cell, until it reaches the synapse.
- The electrical depolarization of the membrane at the synapse causes channels to open that are permeable to calcium ions.
- Calcium ions flow through the presynaptic membrane, rapidly increasing the calcium concentration in the interior.
- The high calcium concentration activates a set of calcium-sensitive proteins attached to vesicles that contain a neurotransmitter chemical.
- These proteins change shape, causing the membranes of some "docked" vesicles to fuse with the membrane of the presynaptic cell, thereby opening the vesicles and dumping their neurotransmitter contents into the synaptic cleft, the narrow space between the membranes of the pre- and postsynaptic cells.
- The neurotransmitter diffuses within the cleft. Some of it escapes, but some of it binds to chemical receptor molecules located on the membrane of the postsynaptic cell.
- The binding of neurotransmitter causes the receptor molecule to be activated.
- Due to thermal vibration, the motion of atoms, vibrating about their equilibrium positions in a crystalline solid, neurotransmitter molecules eventually break loose from the receptors and drift away.
- The neurotransmitter is either reabsorbed by the presynaptic cell, and then repackaged for future release, or else it is broken down metabolically.

In general, if an excitatory synapse is strong enough, an action potential in the presynaptic neuron will trigger an action potential in the postsynaptic cell. In many cases the excitatory postsynaptic potential (EPSP) will not reach the threshold for eliciting an action potential. When action potentials from multiple presynaptic neurons fire simultaneously, or if a single presynaptic neuron fires at a high enough frequency, the EPSPs can overlap and summate. If enough EPSPs overlap, the summated EPSP can reach the

threshold for initiating an action potential. This process is known as summation.

On the other hand, a presynaptic neuron releasing an inhibitory neurotransmitter, such as GABA, can cause an inhibitory postsynaptic potential (IPSP) in the postsynaptic neuron, moving the membrane potential farther away from the threshold, decreasing its excitability and making it more difficult for the neuron to initiate an action potential. If an IPSP overlaps with an EPSP, the IPSP can in many cases prevent the neuron from firing an action potential. In this way, the output of a neuron may depend on the input of many different neurons, each of which may have a different degree of influence, depending on the strength and type of synapse with that neuron. John Carew Eccles⁵⁶ performed some of the important early experiments on synaptic integration, for which he received the Nobel Prize for Physiology or Medicine in 1963.

Understanding the effects of drugs on neurotransmitters comprises a significant portion of research initiatives in the field of neuroscience. Most neuroscientists involved in this field of research believe that such efforts may further advance our understanding of the circuits responsible for various neurological diseases and disorders, as well as ways to effectively treat and someday possibly prevent or cure such illnesses.

25.11.2 Neurotransmitters

Neurotransmitters are endogenous chemicals that enable neurotransmission. It is a type of chemical messenger which transmits signals across a chemical synapse, such as a neuromuscular junction, from one neuron (nerve cell) to another “target” neuron, muscle cell, or gland cell. Neurotransmitters are released from synaptic vesicles in synapses into the synaptic cleft, where they are received by neurotransmitter receptors on the target cells. Many neurotransmitters are synthesized from simple and plentiful precursors such as amino acids, which are readily available from the diet and only require a small number of biosynthetic steps for conversion. Neurotransmitters play a major role in shaping everyday life and functions. Their exact numbers are unknown, but more than 200 unique chemical messengers have been identified.

Neurotransmitters are stored in synaptic vesicles, clustered close to the cell membrane at the axon terminal of the presynaptic neuron. Neurotransmitters are released into and diffuse across the synaptic cleft, where they bind to specific receptors on the membrane of the postsynaptic neuron.

Neurotransmitter action is terminated in three different ways:

- Diffusion – the neurotransmitter detaches from receptor, drifting out of the synaptic cleft, here it becomes absorbed by glial cells.
- Enzyme degradation – special chemicals called enzymes break it down. Usually, astrocytes absorb the excess neurotransmitters and pass them on to enzymes or pump them directly into the presynaptic neuron.
- Reuptake – re-absorption of a neurotransmitter into the neuron. Transporters, or membrane transport proteins, pump neurotransmitters from the synaptic cleft back into axon terminals (the presynaptic neuron) where they are stored.

For example, choline is taken up and recycled by the pre-synaptic neuron to synthesize more ACh. Other neurotransmitters such as dopamine are able to diffuse away from their targeted synaptic junctions and are eliminated from the body via the kidneys, or destroyed in the liver. Each neurotransmitter has very specific degradation pathways at regulatory points, which may be targeted by the body's regulatory system or by recreational drugs.

Until the early 20th century, scientists assumed that the majority of synaptic communication in the brain was electrical. But in 1921 German pharmacologist Otto Loewi⁵⁷ (1873–1961) demonstrated that neurons can communicate by releasing chemicals. Some neurons do, however, communicate via electrical synapses through the use of gap junctions, which allow specific ions to pass directly from one cell to another.

The anatomical localization of neurotransmitters is typically determined using immunocytochemical techniques, which identify the location of either the transmitter substances themselves, or of the enzymes that are involved in their synthesis. Immunocytochemical techniques have also revealed that many transmitters, particularly the neuropeptides, are co-localized, that is, one neuron may release more than one transmitter from its synaptic terminal. Various techniques have been used to identify neurotransmitters throughout the central nervous system.

Single ions (such as synaptically released zinc) are also considered neurotransmitters by some, as well as some gaseous molecules such as nitric oxide (NO), carbon monoxide (CO), and hydrogen sulfide (H₂S).

The most prevalent transmitter is glutamate, which is excitatory at well over 90% of the synapses in the human brain. The next most prevalent is Gamma-Aminobutyric Acid, or GABA, which is inhibitory at more than 90% of the synapses that do not use glutamate. Although other transmitters are used in fewer synapses, they may be very important functionally: the great majority of psychoactive

⁵⁶ [https://en.wikipedia.org/wiki/John_Eccles_\(neurophysiologist\)](https://en.wikipedia.org/wiki/John_Eccles_(neurophysiologist))

⁵⁷ https://en.wikipedia.org/wiki/Otto_Loewi

drugs exert their effects by altering the actions of some neurotransmitter systems, often acting through transmitters other than glutamate or GABA. Addictive drugs such as cocaine and amphetamines exert their effects primarily on the dopamine system. The addictive opiate drugs exert their effects primarily as functional analogs of opioid peptides, which, in turn, regulate dopamine levels.

25.11.3 Modulatory Neurotransmitter Systems

Neurons expressing certain types of neurotransmitters sometimes form distinct systems, where activation of the system acts in a modulatory fashion on a large number of neurons in large volumes of the brain. Such modulatory neurotransmitter systems include the noradrenaline (norepinephrine) system, the dopamine system, the serotonin system, and the cholinergic system, among others. Neuromodulatory neurotransmitters typically bind to metabotropic, G-protein coupled receptors to initiate a second messenger signaling cascade that induces a broad, long-lasting signal. This modulation can last for hundreds of milliseconds to several minutes. Some of the effects of neuromodulators include: altering the intrinsic firing activity, increasing or decreasing voltage-dependent currents, changing synaptic efficacy, increasing bursting activity and reconfiguring of synaptic connectivity.

25.11.4 Neurotransmitter Receptors

There are two major types of neurotransmitter receptors: ionotropic and metabotropic. Ionotropic means that ions can pass through the receptor, whereas metabotropic means that a second messenger inside the cell relays the message (i.e. metabotropic receptors do not have channels). Metabotropic receptors are G-protein-coupled receptors (GPCRs). Ionotropic receptors are also called ligand-gated ion channels. Conversely, GPCRs are neither excitatory nor inhibitory. Rather, they can have a broad number of functions such as modulating the actions of excitatory and inhibitory ion channels or triggering a signalling cascade that releases calcium from stores inside the cell.

25.11.5 Ionotropic Receptors: Neurotransmitter-Gated Ion Channels

Ligand-gated ion channels (LGICs) are one type of ionotropic receptor or channel-linked receptor. They are a group of transmembrane ion channels that are opened or closed in response to the binding of a chemical messenger (i.e., a ligand), such as a neurotransmitter.

The binding site of endogenous ligands on LGICs protein complexes are normally located on a different

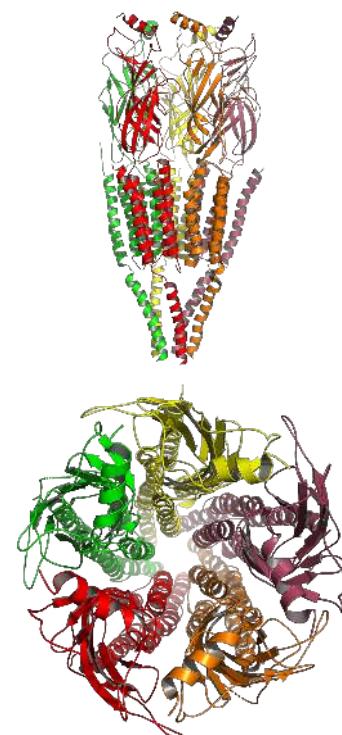


Figure 25.23: A cartoon representation of the atomic structure of the nicotinic acetylcholine receptor from the electric ray *Torpedo marmorata* at 4 Å resolution. Data from PDB 2BG9⁵⁸, rendered with open source molecular visualization tool PyMol⁵⁹.

portion of the protein (an allosteric binding site) compared to where the ion conduction pore is located. The direct link between ligand binding and opening or closing of the ion channel, which is characteristic of ligand-gated ion channels, is contrasted with the indirect function of metabotropic receptors, which use second messengers. LGICs are also different from voltage-gated ion channels (which open and close depending on membrane potential), and stretch-activated ion channels (which open and close depending on mechanical deformation of the cell membrane).

25.11.6 Metabotropic Receptors: G-Protein Coupled Receptors

GPCRs also known as seven-transmembrane domain receptors, 7TM receptors, heptahelical receptors, serpentine receptor comprise a large protein family of transmembrane receptors that sense molecules outside the cell and activate intracellular signal transduction pathways. GPCRs are found only in eukaryotes. The ligands that bind and activate these receptors include light-sensitive compounds, odors, pheromones, hormones, and neurotransmitters, and vary in size from small molecules to peptides to large proteins. G protein-coupled receptors are involved in many diseases, and are also the target of approximately 30% of all modern medicinal drugs.

There are two principal signal transduction pathways involving the G protein-coupled receptors: the cAMP signal pathway and the phosphatidylinositol signal pathway. When a ligand binds to the GPCR it causes a conformational change in the GPCR, which allows it to act as a guanine nucleotide exchange factor (GEF). The GPCR can then activate an associated G-protein by exchanging its bound GDP for a GTP. The G-protein's α subunit, together with the bound GTP, can then dissociate from the β and γ subunits to further affect intracellular signaling proteins or target functional proteins directly depending on the α subunit type ($G_{\alpha s}$, $G_{\alpha i/o}$, $G_{\alpha q/11}$, $G_{\alpha 12/13}$).

Neurotransmitter receptors are present on both post-synaptic neurons and presynaptic neurons with the former being used to receive neurotransmitters and the latter for the purpose of preventing further release of a given neurotransmitter. In addition to being found in neuron cells, neurotransmitter receptors are also found in various immune and muscle tissues. Many neurotransmitter receptors are categorized as a serpentine receptor or G protein-coupled receptor because they span the cell membrane not once, but seven times. Neurotransmitter receptors are known to become unresponsive to the type of neurotransmitter they receive when exposed for extended periods of time. This phenomenon is known as ligand-induced desensitization or downregulation.

The following are some major classes of neurotransmitter receptors:

- Adrenergic: $\alpha 1A$, $\alpha 1B$, $\alpha 1C$, $\alpha 1D$, $\alpha 2A$, $\alpha 2B$, $\alpha 2C$, $\alpha 2D$, $\beta 1$, $\beta 2$, $\beta 3$
- Dopaminergic: D1, D2, D3, D4, D5
- GABAergic: GABA_A, GABA_{B1a}, GABA_{B1b}, GABA_{B2}, GABA_C
- Glutamatergic: NMDA, AMPA, kainate, mGluR1, mGluR2, mGluR3, mGluR4, mGluR5, mGluR6, mGluR7
- Histaminergic: H1, H2, H3
- Cholinergic: Muscarinic: M1, M2, M3, M4, M5; Nicotinic: muscle, neuronal (α -bungarotoxin-insensitive), neuronal (α -bungarotoxin-sensitive)
- Opioid: μ , $\delta 1$, $\delta 2$, κ
- Serotonergic: 5-HT1A, 5-HT1B, 5-HT1D, 5-HT1E, 5-HT1F, 5-HT2A, 5-HT2B, 5-HT2C, 5-HT3, 5-HT4, 5-HT5, 5-HT6, 5-HT7
- Glycinergic: Glycine

Drugs can influence behavior by altering neurotransmitter activity. For instance, drugs can decrease the rate of synthesis of neurotransmitters by affecting the synthetic enzyme(s) for that neurotransmitter. When neurotransmitter synthesis is blocked, the amount of neurotransmitters available for release becomes substantially lower, resulting in a decrease in neurotransmitter activity. Some drugs block or stimulate the release of specific neurotransmitters. Alternatively, drugs can prevent neurotransmitter storage in synaptic vesicles by causing the synaptic vesicle membranes to leak. Drugs that prevent a neurotransmitter from binding to its receptor are called receptor antagonists. For example, drugs used to treat patients with schizophrenia such as haloperidol, chlorpromazine, and clozapine are antagonists at receptors in the brain for dopamine. Other drugs act by binding to a receptor and mimicking the normal neurotransmitter. Such drugs are called receptor agonists. An example of a receptor agonist is morphine, an opiate that mimics effects of the endogenous neurotransmitter β -endorphin to relieve pain. Other drugs interfere with the deactivation of a neurotransmitter after it has been released, thereby prolonging the action of a neurotransmitter. This can be accomplished by blocking re-uptake or inhibiting degradative enzymes. Lastly, drugs can also prevent an action potential from occurring, blocking neuronal activity throughout the central and peripheral nervous system.

Chapter 26

Sensation: Receptors, Organs And Systems

Sensation is the physical process during which sensory systems respond to stimuli and provide data for perception. A sense is any of the systems involved in sensation. During sensation, sense organs engage in stimulus collection and transduction. Sensation is often differentiated from the related and dependent concept of perception, which processes and integrates sensory information in order to give meaning to and understand detected stimuli, giving rise to subjective perceptual experience, or qualia. Sensation and perception are central to and precede almost all aspects of cognition, behavior and thought.

The sensory nervous system is a part of the nervous system responsible for processing sensory information. A sensory system consists of sensory neurons (including the sensory receptor cells), neural pathways, and parts of the brain involved in sensory perception. Commonly recognized sensory systems are those for vision, hearing, touch, taste, smell, and balance. In short, senses are transducers from the physical world to the realm of the mind where we interpret the information, creating our perception of the world around us.

Organisms need information to solve at least three kinds of problems: (a) to maintain an appropriate environment, i.e., homeostasis; (b) to time activities (e.g., seasonal changes in behavior) or synchronize activities with those of conspecifics; and (c) to locate and respond to resources or threats (e.g., by moving towards resources or evading or attacking threats). Organisms also need to transmit information in order to influence another's behavior: to identify themselves, warn conspecifics of danger, coordinate activities, or deceive.

The receptive field is the area of the body or environment to which a receptor organ and receptor cells respond. For instance, the part of the world an eye can see, is its receptive field; the light that each rod or cone can see, is its receptive field. Receptive fields have been identified for the visual system, auditory system and somatosensory system.

Sensory systems code for four aspects of a stimulus; type (modality), intensity, location, and duration. Arrival

time of a sound pulse and phase differences of continuous sound are used for sound localization. Certain receptors are sensitive to certain types of stimuli (for example, different mechanoreceptors respond best to different kinds of touch stimuli, like sharp or blunt objects). Receptors send impulses in certain patterns to send information about the intensity of a stimulus (for example, how loud a sound is). The location of the receptor that is stimulated gives the brain information about the location of the stimulus (for example, stimulating a mechanoreceptor in a finger will send information to the brain about that finger). The duration of the stimulus (how long it lasts) is conveyed by firing patterns of receptors. These impulses are transmitted to the brain through afferent neurons.

While debate exists among neurologists as to the specific number of senses due to differing definitions of what constitutes a sense, Gautama Buddha and Aristotle classified five 'traditional' human senses which have become universally accepted: touch, taste, smell, sight, and hearing. Other senses that have been well-accepted in most mammals, including humans, include nociception, equilibrioception, kinaesthesia, and thermoception. Furthermore, some nonhuman animals have been shown to possess alternate senses, including magnetoeception and electroreception.

In organisms, a sensory organ consists of a group of related sensory cells that respond to a specific type of physical stimulus. Via cranial and spinal nerves, the different types of sensory receptor cells (mechanoreceptors, photoreceptors, chemoreceptors, thermoreceptors) in sensory organs transduct sensory information from sensory organs towards the central nervous system, to the sensory cortices in the brain, where sensory signals are further processed and interpreted (perceived). Sensory systems, or senses, are often divided into external (exteroception) and internal (interoception) sensory systems. Sensory modalities or submodalities refer to the way sensory information is encoded or transduced. Multimodality integrates different senses into one unified perceptual experience. For example, information from one sense has the potential to influ-

ence how information from another is perceived. Sensation and perception are studied by a variety of related fields, most notably psychophysics, neurobiology, cognitive psychology, and cognitive science.

Humans have a multitude of sensory systems. Human external sensation is based on the sensory organs of the eyes, ears, skin, inner ear, nose, and mouth. The corresponding sensory systems of the visual system (sense of vision), auditory system (sense of hearing), somatosensory system (sense of touch), vestibular system (sense of balance), olfactory system (sense of smell), and gustatory system (sense of taste) contribute, respectively, to the perceptions of vision, hearing, touch, spatial orientation, smell, and taste (flavor). Internal sensation, or interoception, detects stimuli from internal organs and tissues. Many internal sensory and perceptual systems exist in humans, including proprioception (body position) and nociception (pain). Further internal chemoreception and osmoreception based sensory systems lead to various perceptions, such as hunger, thirst, suffocation, and nausea, or different involuntary behaviors, such as vomiting.

Nonhuman animals experience sensation and perception, with varying levels of similarity to and difference from humans and other animal species. For example, mammals, in general, have a stronger sense of smell than humans. Some animal species lack one or more human sensory system analogues, some have sensory systems that are not found in humans, while others process and interpret the same sensory information in very different ways. For example, some animals are able to detect electrical and magnetic fields, air moisture, or polarized light, while others sense and perceive through alternative systems, such as echolocation. Recently, it has been suggested that plants and artificial agents may be able to detect and interpret environmental information in an analogous manner to animals.

26.1 Sensory Receptors

Sensory receptors are the cells or structures that detect sensations. Stimuli in the environment activate specialized receptor cells in the peripheral nervous system. During transduction, physical stimulus is converted into action potential by receptors and transmitted towards the central nervous system for processing. Different types of stimuli are sensed by different types of receptor cells. Receptor cells can be classified into types on the basis of three different criteria: cell type, position, and function. Receptors can be classified structurally on the basis of cell type and their position in relation to stimuli they sense. Receptors can further be classified functionally on the basis of the transduction of stimuli, or how the mechanical stimulus,

light, or chemical changed the cell membrane potential.

One way to classify receptors is based on their location relative to the stimuli. An exteroceptor is a receptor that is located near a stimulus of the external environment, such as the somatosensory receptors that are located in the skin. An interoceptor is one that interprets stimuli from internal organs and tissues, such as the receptors that sense the increase in blood pressure in the aorta or carotid sinus.

The cells that interpret information about the environment can be either (1) a neuron that has a free nerve ending, with dendrites embedded in tissue that would receive a sensation; (2) a neuron that has an encapsulated ending in which the sensory nerve endings are encapsulated in connective tissue that enhances their sensitivity; or (3) a specialized receptor cell, which has distinct structural components that interpret a specific type of stimulus. The pain and temperature receptors in the dermis of the skin are examples of neurons that have free nerve endings (1). Also located in the dermis of the skin are lamellated corpuscles, neurons with encapsulated nerve endings that respond to pressure and touch (2). The cells in the retina that respond to light stimuli are an example of a specialized receptor (3), a photoreceptor.

A transmembrane protein receptor is a protein in the cell membrane that mediates a physiological change in a neuron, most often through the opening of ion channels or changes in the cell signaling processes. Transmembrane receptors are activated by chemicals called ligands. For example, a molecule in food can serve as a ligand for taste receptors. Other transmembrane proteins, which are not accurately called receptors, are sensitive to mechanical or thermal changes. Physical changes in these proteins increase ion flow across the membrane, and can generate an action potential or a graded potential in the sensory neurons.

A third classification of receptors is by how the receptor transduces stimuli into membrane potential changes. Stimuli are of three general types. Some stimuli are ions and macromolecules that affect transmembrane receptor proteins when these chemicals diffuse across the cell membrane. Some stimuli are physical variations in the environment that affect receptor cell membrane potentials. Other stimuli include the electromagnetic radiation from visible light. For humans, the only electromagnetic energy that is perceived by our eyes is visible light. Some other organisms have receptors that humans lack, such as the heat sensors of snakes, the ultraviolet light sensors of bees, or magnetic receptors in migratory birds.

Receptor cells can be further categorized on the basis of the type of stimuli they transduce. The different types of functional receptor cell types are mechanoreceptors,

photoreceptors, chemoreceptors (osmoreceptor), thermoreceptors, and nociceptors. Physical stimuli, such as pressure and vibration, as well as the sensation of sound and body position (balance), are interpreted through a mechanoreceptor. Photoreceptors convert light (visible electromagnetic radiation) into signals. Chemical stimuli can be interpreted by a chemoreceptor that interprets chemical stimuli, such as an object's taste or smell, while osmoreceptors respond to a chemical solute concentrations of body fluids. Nociception (pain) interprets the presence of tissue damage, from sensory information from mechano-, chemo-, and thermoreceptors. Another physical stimulus that has its own type of receptor is temperature, which is sensed through a thermoreceptor that is either sensitive to temperatures above (heat) or below (cold) normal body temperature.

Each sense organ (eyes or nose, for instance) requires a minimal amount of stimulation in order to detect a stimulus. This minimum amount of stimulus is called the absolute threshold. The absolute threshold is defined as the minimum amount of stimulation necessary for the detection of a stimulus 50% of the time. Absolute threshold is measured by using a method called signal detection. This process involves presenting stimuli of varying intensities to a subject in order to determine the level at which the subject can reliably detect stimulation in a given sense.

Differential threshold or just noticeable difference (JDS) is the smallest detectable difference between two stimuli, or the smallest difference in stimuli that can be judged to be different from each other. Weber's Law is an empirical law that states that the difference threshold is a constant fraction of the comparison stimulus. According to Weber's Law, bigger stimuli require larger differences to be noticed.

Perception occurs when nerves that lead from the sensory organs (e.g. eye) to the brain are stimulated, even if that stimulation is unrelated to the target signal of the sensory organ. For example, in the case of the eye, it does not matter whether light or something else stimulates the optic nerve, that stimulation will result in visual perception, even if there was no visual stimulus to begin with. (To prove this point to yourself (and if you are a human), close your eyes (preferably in a dark room) and press gently on the outside corner of one eye through the eyelid. You will see a visual spot toward the inside of your visual field, near your nose.

The initialization of sensation stems from the response of a specific receptor to a physical stimulus. The receptors which react to the stimulus and initiate the process of sensation are commonly characterized in four distinct categories: chemoreceptors, photoreceptors, mechanoreceptors, and thermoreceptors. All receptors receive distinct physical stimuli and transduce the signal into an electrical

action potential. This action potential then travels along afferent neurons to specific brain regions where it is processed and interpreted.

26.1.1 Chemoreceptors

Chemoreceptors, or chemosensors, detect certain chemical stimuli and transduce that signal into an electrical action potential. The two primary types of chemoreceptors are:

Distance chemoreceptors are integral to receiving stimuli in gases in the olfactory system through both olfactory receptor neurons and neurons in the vomeronasal organ. Direct chemoreceptors that detect stimuli in liquids include the taste buds in the gustatory system as well as receptors in the aortic bodies which detect changes in oxygen concentration.

26.1.2 Photoreceptors

Photoreceptors are capable of phototransduction, a process which converts light (electromagnetic radiation) into, among other types of energy, a membrane potential. The three primary types of photoreceptors are: Cones are photoreceptors which respond significantly to color. In humans the three different types of cones correspond with a primary response to short wavelength (blue), medium wavelength (green), and long wavelength (yellow/red). Rods are photoreceptors which are very sensitive to the intensity of light, allowing for vision in dim lighting. The concentrations and ratio of rods to cones is strongly correlated with whether an animal is diurnal or nocturnal. In humans rods outnumber cones by approximately 20:1, while in nocturnal animals, such as the tawny owl, the ratio is closer to 1000:1. Ganglion Cells reside in the adrenal medulla and retina where they are involved in the sympathetic response. Of the ~1.3 million ganglion cells present in the retina, 1–2% are believed to be photosensitive ganglia. These photosensitive ganglia play a role in conscious vision for some animals, and are believed to do the same in humans.

26.1.3 Mechanoreceptors

Mechanoreceptors are sensory receptors which respond to mechanical forces, such as pressure or distortion. While mechanoreceptors are present in hair cells and play an integral role in the vestibular and auditory systems, the majority of mechanoreceptors are cutaneous and are grouped into four categories:

- Slowly adapting type 1 receptors have small receptive fields and respond to static stimulation. These

receptors are primarily used in the sensations of form and roughness.

- Slowly adapting type 2 receptors have large receptive fields and respond to stretch. Similarly to type 1, they produce sustained responses to a continued stimuli.
- Rapidly adapting receptors have small receptive fields and underlie the perception of slip. Pacinian receptors have large receptive fields and are the predominant receptors for high-frequency vibration.

26.1.4 Thermoreceptors

Thermoreceptors are sensory receptors which respond to varying temperatures. While the mechanisms through which these receptors operate is unclear, recent discoveries have shown that mammals have at least two distinct types of thermoreceptors:[permanent dead link] See TfM>[failed verification]

- The end-bulb of Krause, or bulboid corpuscle, detects temperatures above body temperature.
- Ruffini's end organ detects temperatures below body temperature.

TRPV1 is a heat-activated channel that acts as a small heat detecting thermometer in the membrane which begins the polarization of the neural fiber when exposed to changes in temperature. Ultimately, this allows us to detect ambient temperature in the warm/hot range. Similarly, the molecular cousin to TRPV1, TRPM8, is a cold-activated ion channel that responds to cold. Both cold and hot receptors are segregated by distinct subpopulations of sensory nerve fibers, which shows us that the information coming into the spinal cord is originally separate. Each sensory receptor has its own "labeled line" to convey a simple sensation experienced by the recipient. Ultimately, TRP channels act as thermosensors, channels that help us to detect changes in ambient temperatures.

26.1.5 Nociceptors

Nociceptors respond to potentially damaging stimuli by sending signals to the spinal cord and brain. This process, called nociception, usually causes the perception of pain. They are found in internal organs, as well as on the surface of the body. Nociceptors detect different kinds of damaging stimuli or actual damage. Those that only respond when tissues are damaged are known as "sleeping" or "silent" nociceptors.

Thermal nociceptors are activated by noxious heat or cold at various temperatures. Mechanical nociceptors respond to excess pressure or mechanical deformation.

Chemical nociceptors respond to a wide variety of chemicals, some of which are signs of tissue damage. They are involved in the detection of some spices in food.

26.2 The Visual System

Visual perception is the ability to interpret the surrounding environment using light in the visible spectrum reflected by the objects in the environment. This is different from visual acuity, which refers to how clearly a person sees (for example "20/20 vision"). A person can have problems with visual perceptual processing even if they have 20/20 vision.

The resulting perception is also known as visual perception, eyesight, sight, or vision. The various physiological components involved in vision are referred to collectively as the visual system.

Different species are able to see different parts of the light spectrum; for example, bees can see into the ultraviolet, while pit vipers can accurately target prey with their pit organs, which are sensitive to infrared radiation. The mantis shrimp possesses arguably the most complex visual system in any species. The eye of the mantis shrimp holds 16 color receptive cones, whereas humans only have three. The variety of cones enables them to perceive an enhanced array of colors as a mechanism for mate selection, avoidance of predators, and detection of prey. Swordfish also possess an impressive visual system. The eye of a swordfish can generate heat to better cope with detecting their prey at depths of 2000 feet. Certain one-celled micro-organisms, the warnowiid dinoflagellates have eye-like ocelloids, with analogous structures for the lens and retina of the multi-cellular eye. The armored shell of the chiton *Acanthopleura granulata* is also covered with hundreds of aragonite crystalline eyes, named ocelli, which can form images.

Many fan worms, such as *Acromegalomma interruptum* which live in tubes on the sea floor of the Great Barrier Reef, have evolved compound eyes on their tentacles, which they use to detect encroaching movement. If movement is detected the fan worms will rapidly withdraw their tentacles.

Only higher primate Old World (African) monkeys and apes have the same kind of three-cone photoreceptor color vision humans have, while lower primate New World (South American) monkeys have a two-cone photoreceptor kind of color vision.

26.2.1 The Eye

Light entering the eye is refracted as it passes through the cornea. It then passes through the pupil (controlled by the

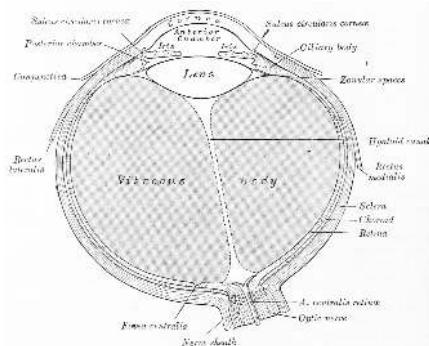


Figure 26.1: Horizontal section of the human eyeball. From Gray Henry, Anatomy of the Human Body. 20th Edition, Lea & Febiger, Philadelphia & New York, 1918¹

iris) and is further refracted by the lens. The cornea and lens act together as a compound lens to project an inverted image onto the retina.

26.2.2 The Retina

The retina is the light-sensitive layer of tissue of the eye of most vertebrates and some molluscs. The optics of the eye create a focused two-dimensional image of the visual world on the retina, which translates that image into electrical neural impulses to the brain to create visual perception.

The retina translates an optical image into neural impulses starting with the patterned excitation of the light-sensitive pigments of its rods and cones, the retina's photoreceptor cells. The excitation is processed by the neural system and various parts of the brain working in parallel to form a representation of the external environment in the brain.

Light striking the retina initiates a cascade of chemical and electrical events that ultimately trigger nerve impulses that are sent to various visual centres of the brain through the fibres of the optic nerve. Neural signals from the rods and cones undergo processing by other neurons, whose output takes the form of action potentials in retinal ganglion cells whose axons form the optic nerve. Several important features of visual perception can be traced to the retinal encoding and processing of light.

The cones respond to bright light and mediate high-resolution colour vision during daylight illumination (also called photopic vision). The rod responses are saturated at daylight levels and don't contribute to pattern vision. However, rods do respond to dim light and mediate lower-resolution, monochromatic vision under very low levels of illumination (called scotopic vision). The illumination in most office settings falls between these two levels and is called mesopic vision. At mesopic light levels, both the

rods and cones are actively contributing pattern information. What contribution the rod information makes to pattern vision under these circumstances is unclear.

The response of cones to various wavelengths of light is called their spectral sensitivity. In normal human vision, the spectral sensitivity of a cone falls into one of three subtypes, often called blue, green, and red, but more accurately known as short, medium, and long wavelength-sensitive cone subtypes. It is a lack of one or more of the cone subtypes that causes individuals to have deficiencies in colour vision or various kinds of colour blindness. These individuals are not blind to objects of a particular colour, but are unable to distinguish between colours that can be distinguished by people with normal vision. Humans have this trichromatic vision, while most other mammals lack cones with red sensitive pigment and therefore have poorer dichromatic colour vision. However, some animals have four spectral subtypes, e.g. the trout adds an ultraviolet subgroup to short, medium, and long subtypes that are similar to humans. Some fish are sensitive to the polarization of light as well.

When thus excited by light, the photoreceptor sends a proportional response synaptically to bipolar cells which in turn signal the retinal ganglion cells. The photoreceptors are also cross-linked by horizontal cells and amacrine cells, which modify the synaptic signal before it reaches the ganglion cells, the neural signals being intermixed and combined. Of the retina's nerve cells, only the retinal ganglion cells and few amacrine cells create action potentials.

In the retinal ganglion cells there are two types of response, depending on the receptive field of the cell. The receptive fields of retinal ganglion cells comprise a central, approximately circular area, where light has one effect on the firing of the cell, and an annular surround, where light has the opposite effect. In ON cells, an increment in light intensity in the centre of the receptive field causes the firing rate to increase. In OFF cells, it makes it decrease. Beyond this simple difference, ganglion cells are also differentiated by chromatic sensitivity and the type of spatial summation. Cells showing linear spatial summation are termed X cells (also called parvocellular, P, or midget ganglion cells), and those showing non-linear summation are Y cells (also called magnocellular, M, or parasol retinal ganglion cells), although the correspondence between X and Y cells (in the cat retina) and P and M cells (in the primate retina) is not as simple as it once seemed.

In the transfer of visual signals to the brain, the visual pathway, the retina is vertically divided in two, a temporal (nearer to the temple) half and a nasal (nearer to the nose) half. The axons from the nasal half cross the brain at the optic chiasma to join with axons from the temporal half of the other eye before passing into the lateral geniculate

body.

Although there are more than 130 million retinal receptors, there are only approximately 1.2 million fibres (axons) in the optic nerve. So, a large amount of pre-processing is performed within the retina. The fovea produces the most accurate information. Despite occupying about 0.01% of the visual field (less than 2° of visual angle), about 10% of axons in the optic nerve are devoted to the fovea.

The final result of all this processing is five different populations of ganglion cells that send visual (image-forming and non-image-forming) information to the brain:

- M cells, with large center-surround receptive fields that are sensitive to depth, indifferent to color, and rapidly adapt to a stimulus
- P cells, with smaller center-surround receptive fields that are sensitive to color and shape
- K cells, with very large center-only receptive fields that are sensitive to color and indifferent to shape or depth
- another population that is intrinsically photosensitive
- a final population that is involved in the control of eye movements. The neural retina consists of several layers of neurons interconnected by synapses, and is supported by an outer layer of pigmented epithelial cells. The primary light-sensing cells in the retina are the photoreceptor cells, which are of two types: rods and cones. Rods function mainly in dim light and provide black-and-white vision. Cones function in well-lit conditions and are responsible for the perception of colour, as well as high-acuity vision used for tasks such as reading. A third type of light-sensing cell, the photosensitive ganglion cell, is important for entrainment of circadian rhythms and reflexive responses such as the pupillary light reflex.

In vertebrate embryonic development, the retina and the optic nerve originate as outgrowths of the developing brain, specifically the embryonic diencephalon; thus, the retina is considered part of the central nervous system (CNS) and is actually brain tissue. It is the only part of the CNS that can be visualized non-invasively.

The vertebrate retina has ten distinct layers. From closest to farthest from the vitreous body:

- Inner limiting membrane – basement membrane elaborated by Müller cells.
- Nerve fibre layer – axons of the ganglion cell bodies (note that a thin layer of Müller cell footplates exists between this layer and the inner limiting membrane).
- Ganglion cell layer – contains nuclei of ganglion cells,

the axons of which become the optic nerve fibres, and some displaced amacrine cells.

- Inner plexiform layer – contains the synapse between the bipolar cell axons and the dendrites of the ganglion and amacrine cells.
- Inner nuclear layer – contains the nuclei and surrounding cell bodies (perikarya) of the amacrine cells, bipolar cells, and horizontal cells.
- Outer plexiform layer – projections of rods and cones ending in the rod spherule and cone pedicle, respectively. These make synapses with dendrites of bipolar cells and horizontal cells. In the macular region, this is known as the Fiber layer of Henle.
- Outer nuclear layer – cell bodies of rods and cones.
- External limiting membrane – layer that separates the inner segment portions of the photoreceptors from their cell nuclei.
- Inner segment / outer segment layer – inner segments and outer segments of rods and cones. The outer segments contain a highly specialized light-sensing apparatus.
- Retinal pigment epithelium – single layer of cuboidal epithelial cells. This layer is closest to the choroid, and provides nourishment and supportive functions to the neural retina. The black pigment melanin in the pigment layer prevents light reflection throughout the globe of the eyeball.

These layers can be grouped into 4 main processing stages: photoreception; transmission to bipolar cells; transmission to ganglion cells, which also contain photoreceptors, the photosensitive ganglion cells; and transmission along the optic nerve. At each synaptic stage there are also laterally connecting horizontal and amacrine cells.

The optic nerve is a central tract of many axons of ganglion cells connecting primarily to the lateral geniculate body, a visual relay station in the diencephalon (the rear of the forebrain). It also projects to the superior colliculus, the suprachiasmatic nucleus, and the nucleus of the optic tract. It passes through the other layers, creating the optic disc in primates.

Additional structures, not directly associated with vision, are found as outgrowths of the retina in some vertebrate groups. In birds, the pecten is a vascular structure of complex shape that projects from the retina into the vitreous humour; it supplies oxygen and nutrients to the eye, and may also aid in vision. Reptiles have a similar, but much simpler, structure.

In adult humans, the entire retina is approximately 72% of a sphere about 22 mm in diameter. The entire retina contains about 7 million cones and 75 to 150 million rods. The optic disc, a part of the retina sometimes called "the

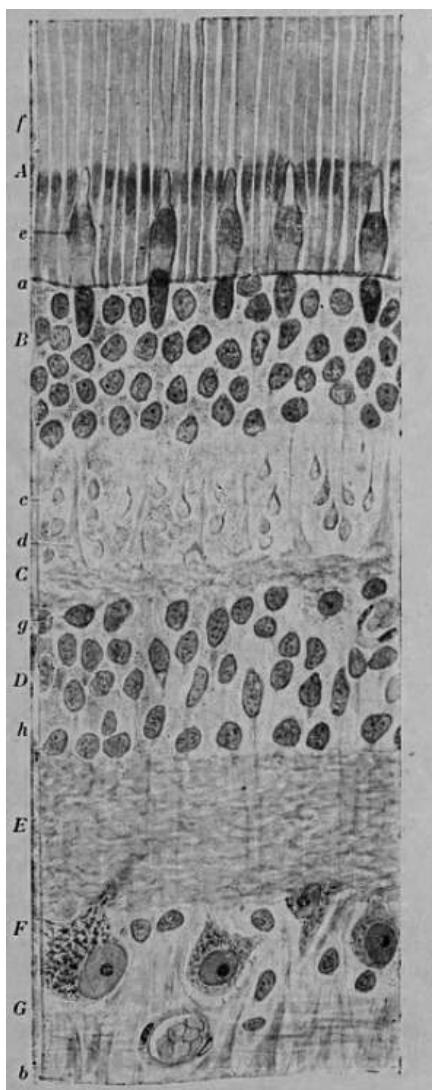


Figure 26.2: Vertical section of the adult human retina. Carmine and Nissl stain. *A*, Photoreceptor layer. *B*, Cell bodies of the photoreceptors. *C*, Outer plexiform layer. *D*, Internal granule layer. *E*, Internal plexiform layer. *F*, Ganglion cell layer. *G*, Ganglion cell axons. *a*, external limiting membrane. *b*, internal limiting membrane. *c*, Spherical endfeet of the rod photoreceptors. *d*, endfeet of the cones. *e*, *a*, cone. *f*, a rod *g*, horizontal cells. *h*, amacrine cells. Fig. 188 from *Histologie du système nerveux de l'homme & des vertébrés*² (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay.

"blind spot" because it lacks photoreceptors, is located at the optic papilla, where the optic-nerve fibres leave the eye. It appears as an oval white area of 3 mm^2 . Temporal (in the direction of the temples) to this disc is the macula, at whose centre is the fovea, a pit that is responsible for our sharp central vision but is actually less sensitive to light because of its lack of rods. Human and non-human primates possess one fovea, as opposed to certain bird species, such as hawks, who are bifoviate, and dogs and cats, who possess no fovea but a central band known as the visual streak. Around the fovea extends the central retina for about 6 mm and then the peripheral retina. The farthest edge of the retina is defined by the ora serrata.

In section, the retina is no more than 0.5 mm thick. It has three layers of nerve cells and two of synapses, including the unique ribbon synapse. The optic nerve carries the ganglion cell axons to the brain, and the blood vessels that supply the retina. The ganglion cells lie innermost in the eye while the photoreceptive cells lie beyond. Because of this counter-intuitive arrangement, light must first pass through and around the ganglion cells and through the thickness of the retina, before reaching the rods and cones. Light is absorbed by the retinal pigment epithelium or the choroid.

The white blood cells in the capillaries in front of the photoreceptors can be perceived as tiny bright moving dots when looking into blue light. This is known as the blue field entoptic phenomenon (or Scheerer's phenomenon).

Between the ganglion cell layer and the rods and cones there are two layers of neuropils where synaptic contacts are made. The neuropil layers are the outer plexiform layer and the inner plexiform layer. In the outer neuropil layer, the rods and cones connect to the vertically running bipolar cells, and the horizontally oriented horizontal cells connect to ganglion cells.

The central retina predominantly contains cones, while the peripheral retina predominantly contains rods. At the centre of the macula is the foveal pit where the cones are narrow and long, and, arranged in a hexagonal mosaic, the most dense, in contradistinction to the much fatter cones located more peripherally in the retina. At the foveal pit the other retinal layers are displaced, before building up along the foveal slope until the rim of the fovea, or parafovea, is reached, which is the thickest portion of the retina. The macula has a yellow pigmentation and is known as the macula lutea. The area directly surrounding the fovea has the highest density of rods converging on single bipolar cells. Since its cones have a much lesser convergence of signals, the fovea allows for the sharpest vision the eye can attain.

Though the rod and cones are a mosaic of sorts, trans-

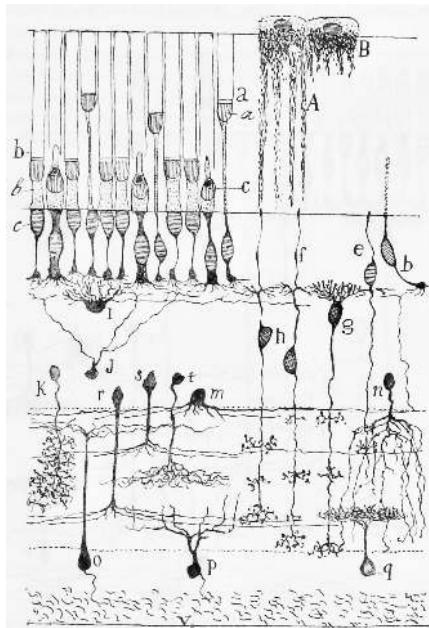


Figure 26.3: A semischematic diagram of the frog retina. a) green rods; b (left) red rods; c) cone; i) horizontal cell; h) bipolar cell; n,m,r,s,t) amacrine cells; o,p) ganglion cells; q) displaced amacrine cell. A) Pigment epithelial cell with extended process; B) Pigment epithelial cell with retracted process.

mission from receptors, to bipolars, to ganglion cells is not direct. Since there are about 150 million receptors and only 1 million optic nerve fibres, there must be convergence and thus mixing of signals. Moreover, the horizontal action of the horizontal and amacrine cells can allow one area of the retina to control another (e.g. one stimulus inhibiting another). This inhibition is key to lessening the sum of messages sent to the higher regions of the brain. In some lower vertebrates (e.g. the pigeon), there is a “centrifugal” control of messages – that is, one layer can control another, or higher regions of the brain can drive the retinal nerve cells, but in primates this does not occur.

26.2.3 The Photoreceptors

A photoreceptor cell is a specialized type of neuroepithelial cell found in the retina that is capable of visual phototransduction. The great biological importance of photoreceptors is that they convert light (visible electromagnetic radiation) into signals that can stimulate biological processes. To be more specific, photoreceptor proteins in the cell absorb photons, triggering a change in the cell's membrane potential.

There are currently three known types of photoreceptor cells in mammalian eyes: rods, cones, and intrinsically photosensitive retinal ganglion cells. The two classic pho-

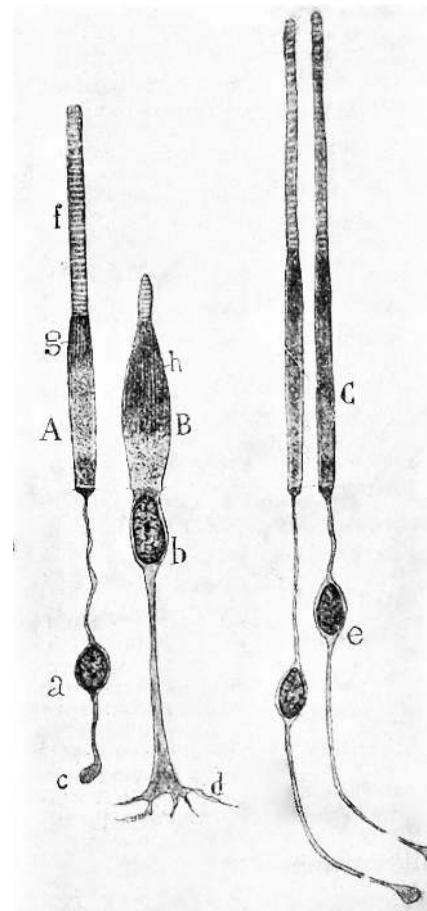


Figure 26.4: Rods and cones from the human retina. A) a rod from the peripheral retina; B) a cone from the peripheral retina; C) cones from the fovea.

toreceptor cells are rods and cones, each contributing information used by the visual system to form a representation of the visual world, sight. The rods are narrower than the cones and distributed differently across the retina, but the chemical process in each that supports phototransduction is similar. A third class of mammalian photoreceptor cell was discovered during the 1990s: the intrinsically photosensitive retinal ganglion cells. These cells do not contribute to sight directly, but are thought to support circadian rhythms and pupillary reflex.

There are major functional differences between the rods and cones. Rods are extremely sensitive, and can be triggered by a single photon. At very low light levels, visual experience is based solely on the rod signal.

Cones require significantly brighter light (that is, a larger number of photons) to produce a signal. In humans, there are three different types of cone cell, distinguished by their pattern of response to light of different wavelengths. Color experience is calculated from these three

distinct signals. This explains why colors cannot be seen at low light levels, when only the rod and not the cone photoreceptor cells are active. The three types of cone cell respond (roughly) to light of short, medium, and long wavelengths, so they may respectively be referred to as S-cones, M-cones, and L-cones. The different responses of the three types of cone cells are determined by the likelihoods that their respective photoreceptor proteins will absorb photons of different wavelengths. So, for example, an L cone cell contains a photoreceptor protein that more readily absorbs long wavelengths of light (that is, more "red"). Light of a shorter wavelength can also produce the same response, but it must be much brighter to do so.

The number and ratio of rods to cones varies among species, dependent on whether an animal is primarily diurnal or nocturnal.

26.2.4 Visual Phototransduction

Visual phototransduction is the sensory transduction of the visual system. It is a process by which light is converted into electrical signals in the rod cells, cone cells and photosensitive ganglion cells of the retina of the eye. This cycle was elucidated by George Wald³ (1906–1997) for which he received the Nobel Prize in 1967.

The visual cycle is the biological conversion of a photon into an electrical signal in the retina. This process occurs via G-protein coupled receptors called opsins which contain the chromophore 11-cis retinal. 11-cis retinal is covalently linked to the opsin receptor via Schiff base forming retinylidene protein. When struck by a photon, 11-cis retinal undergoes photoisomerization to all-trans retinal which changes the conformation of the opsin GPCR leading to signal transduction cascades which causes closure of cyclic GMP-gated cation channel, and hyperpolarization of the photoreceptor cell.

Following isomerization and release from the opsin protein, all-trans retinal is reduced to all-trans retinol and travels back to the retinal pigment epithelium to be "recharged". It is first esterified by lecithin retinol acyl-transferase (LRAT) and then converted to 11-cis retinol by the isomerohydrolase RPE65. Finally, it is oxidized to 11-cis retinal before traveling back to the rod outer segment where it is again conjugated to an opsin to form new, functional visual pigment (rhodopsin).

To understand the photoreceptor's behaviour to light intensities, it is necessary to understand the roles of different currents.

There is an ongoing outward potassium current

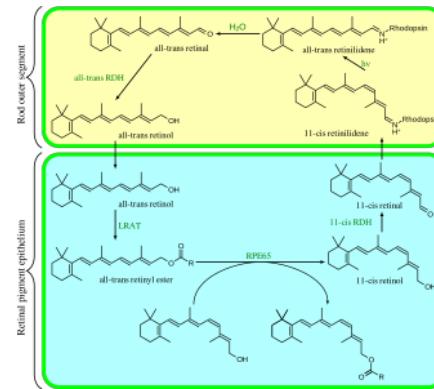


Figure 26.5: The chemical reactions involved in the photoreceptor visual cycle.⁴

through nongated K⁺-selective channels. This outward current tends to hyperpolarize the photoreceptor at around -70 mV (the equilibrium potential for K⁺).

There is also an inward sodium current carried by cGMP-gated sodium channels. This so-called 'dark current' depolarizes the cell to around -40 mV. Note that this is significantly more depolarized than most other neurons.

A high density of Na⁺-K⁺ pumps enables the photoreceptor to maintain a steady intracellular concentration of Na⁺ and K⁺.

Photoreceptor cells are unusual cells in that they are depolarized under scotopic conditions (darkness). In photopic conditions (light), photoreceptors are hyperpolarized to a potential of -60mV.

In the dark, cGMP levels are high and keep cGMP-gated sodium channels open allowing a steady inward current, called the dark current. This dark current keeps the cell depolarized at about -40 mV, leading to glutamate release.

The depolarization of the cell membrane in scotopic conditions opens voltage-gated calcium channels. An increased intracellular concentration of Ca²⁺ causes vesicles containing glutamate, the photoreceptor neurotransmitter, to merge with the cell membrane, therefore releasing glutamate.

In the cone pathway glutamate

- Hyperpolarizes on-center bipolar cells. Glutamate that is released from the photoreceptors in the dark binds to metabotropic glutamate receptors (mGluR6), which, through a G-protein coupling mechanism, causes non-specific cation channels in the cells to close, thus hyperpolarizing the bipolar cell.

³https://en.wikipedia.org/wiki/George_Wald

- Depolarizes off-center bipolar cells. Binding of glutamate to ionotropic glutamate receptors results in an inward cation current that depolarizes the bipolar cell.

Activation of the phototransduction cascade

- A light photon interacts with the retinal in a photoreceptor cell. The retinal undergoes isomerisation, changing from the 11-cis to all-trans configuration.
- Opsin therefore undergoes a conformational change to metarhodopsin II.
- Metarhodopsin II activates a G protein known as transducin. This causes transducin to dissociate from its bound GDP, and bind GTP, then the alpha subunit of transducin dissociates from the beta and gamma subunits, with the GTP still bound to the alpha subunit.
- The alpha subunit-GTP complex activates phosphodiesterase, also known as PDE6. It binds to one of two regulatory subunits of PDE (which itself is a tetramer) and inhibits its activity.
- PDE hydrolyzes cGMP, forming GMP. This lowers the intracellular concentration of cGMP and therefore the sodium channels close.
- Closure of the sodium channels causes hyperpolarization of the cell due to the ongoing efflux of potassium ions.
- Hyperpolarization of the cell causes voltage-gated calcium channels to close.
- As the calcium level in the photoreceptor cell drops, the amount of the neurotransmitter glutamate that is released by the cell also drops. This is because calcium is required for the glutamate-containing vesicles to fuse with cell membrane and release their contents (see SNARE proteins).
- A decrease in the amount of glutamate released by the photoreceptors causes depolarization of on-center bipolar cells (rod and cone On bipolar cells) and hyperpolarization of cone off-center bipolar cells.

Deactivation of the phototransduction cascade

In light, low cGMP levels close Na^+ and Ca^{2+} channels, reducing intracellular Na^+ and Ca^{2+} . During recovery (dark adaptation), the low Ca^{2+} levels induce recovery (termination of the phototransduction cascade), as follows:

- Low intracellular Ca^{2+} makes intracellular Ca-GCAP (Ca-Guanylate cyclase activating protein) dissociate into Ca^{2+} and GCAP. The liberated GCAP ultimately restores depleted cGMP levels, which re-opens the cGMP-gated cation channels (restoring dark current).
- Low intracellular Ca^{2+} makes intracellular Ca-GAP

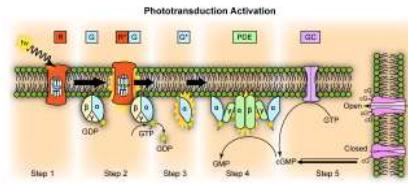


Figure 26.6: Representation⁵ of molecular steps in phototransduction activation (modified from Leskov et al., 2000). Depicted is an outer membrane disk in a rod. Step 1: Incident photon (hv) is absorbed and activates a rhodopsin by conformational change in the disk membrane to R. Step 2: Next, R makes repeated contacts with transducin molecules, catalyzing its activation to G^* by the release of bound GDP in exchange for cytoplasmic GTP, which expels its β and γ subunits. Step 3: G^* binds inhibitory γ subunits of the phosphodiesterase (PDE) activating its α and β subunits. Step 4: Activated PDE hydrolyzes cGMP. Step 5: Guanylyl cyclase (GC) synthesizes cGMP, the second messenger in the phototransduction cascade. Reduced levels of cytosolic cGMP cause cyclic nucleotide gated channels to close preventing further influx of Na^+ and Ca^{2+} .

(Ca-GTPase Accelerating Protein) dissociate into Ca^{2+} and GAP. The liberated GAP deactivates activated-transducin, terminating the phototransduction cascade (restoring dark current).

- Low intracellular Ca^{2+} makes intracellular Ca-recoverin-RK dissociate into Ca^{2+} and recoverin and RK. The liberated RK then phosphorylates metarhodopsin II, reducing its binding affinity for transducin. Arrestin then completely deactivates the phosphorylated-metarhodopsin II, terminating the phototransduction cascade (restoring dark current).
- Low intracellular Ca^{2+} make the Ca^{2+} /calmodulin complex within the cGMP-gated cation channels more sensitive to low cGMP levels (thereby, keeping the cGMP-gated cation channel open even at low cGMP levels, restoring dark current)

All-trans retinal cannot be synthesised by humans and must be supplied by vitamin A in the diet. Deficiency of all-trans retinal can lead to night blindness. This is part of the bleach and recycle process of retinoids in the photoreceptors and retinal pigment epithelium.

Photoreceptor cells are typically arranged in an irregular but approximately hexagonal grid, known as the retinal mosaic.

The opsin found in the intrinsically photosensitive ganglion cells of the retina is called melanopsin. These cells are involved in various reflexive responses of the brain and body to the presence of (day)light, such as the regulation

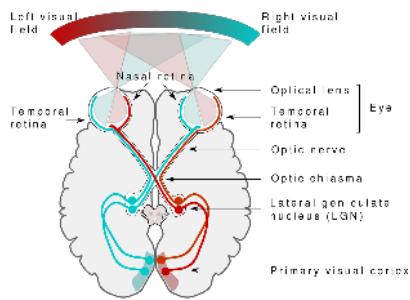


Figure 26.7: A simplified schema⁶ of the human visual pathway.

of circadian rhythms, pupillary reflex and other non-visual responses to light. Melanopsin functionally resembles invertebrate opsins.

26.2.5 The Visual Pathways

26.2.6 The Optic Nerve And Optic Tract

The optic nerve conducts the action potentials generated by the retinal ganglion cells through the optic canal to the subsequent processing centers in the brain. Upon reaching the optic chiasm the nerve fibers from the nasal part of the retina in each eye cross over to the other side (decussate). The fibers then branch and terminate in three places.

The optic nerve is composed of retinal ganglion cell axons and glial cells. Each human optic nerve contains between 770,000 and 1.7 million nerve fibers, which are axons of the retinal ganglion cells of one retina.

In humans, the optic nerve is derived from optic stalks during the seventh week of development. It extends from the optic disc to the optic chiasma and continues as the optic tract to the lateral geniculate nucleus, pretectal nuclei, and superior colliculus.

Most of the axons of the optic nerve terminate in the lateral geniculate nucleus from where information is relayed to the visual cortex, while other axons terminate in the pretectal nucleus and are involved in reflexive eye movements. Other axons terminate in the suprachiasmatic nucleus and are involved in regulating the sleep-wake cycle. Its diameter increases from about 1.6 mm within the eye to 3.5 mm in the orbit to 4.5 mm within the cranial space.

26.2.7 The Superior Colliculus

The superior colliculus (Latin, upper hill) is a structure lying on the roof of the mammalian midbrain. In non-mammalian vertebrates the homologous structure, is known as the optic tectum or optic lobe.

In mammals the superior colliculus forms a major component of the midbrain. It is a paired structure and together with the paired inferior colliculi form the corpora quadrigemina (from Latin quadruplet bodies). The superior colliculus is a layered structure, with a number of layers that varies by species. The layers can be grouped into the superficial layers (stratum opticum and above) and the deeper remaining layers. Neurons in the superficial layers receive direct input from the retina and respond almost exclusively to visual stimuli. Many neurons in the deeper layers also respond to other modalities, and some respond to stimuli in multiple modalities. The deeper layers also contain a population of motor-related neurons, capable of activating eye movements as well as other responses.

The general function of the tectal system is to direct behavioral responses toward specific points in egocentric ("body-centered") space. Each layer contains a topographic map of the surrounding world in retinotopic coordinates, and activation of neurons at a particular point in the map evokes a response directed toward the corresponding point in space. In primates, the superior colliculus has been studied mainly with respect to its role in directing eye movements. Visual input from the retina, or "command" input from the cerebral cortex, create a "bump" of activity in the tectal map, which, if strong enough, induces a saccadic eye movement. Even in primates, however, the superior colliculus is also involved in generating spatially directed head turns, arm-reaching movements, and shifts in attention that do not involve any overt movements. In mammals, and especially primates, the massive expansion of the cerebral cortex reduces the superior colliculus to a much smaller fraction of the whole brain. It remains nonetheless important in terms of function as the primary integrating center for eye movements.

Behavioral studies have shown that the SC is not needed for object recognition, but plays a critical role in the ability to direct behaviors toward specific objects, and can support this ability even in the absence of the cerebral cortex. Thus, cats with major damage to the visual cortex cannot recognize objects, but may still be able to follow and orient toward moving stimuli, although more slowly than usual. If one half of the SC is removed, however, the cats will circle constantly toward the side of the lesion, and orient compulsively toward objects located there, but fail to orient at all toward objects located in the opposite hemifield. These deficits diminish over time but never disappear.

In primates, eye movements can be divided into several types: fixation, in which the eyes are directed toward a motionless object, with eye movements only to compensate for movements of the head; smooth pursuit, in which

the eyes move steadily to track a moving object; saccades, in which the eyes move very rapidly from one location to another; and vergence, in which the eyes move simultaneously in opposite directions to obtain or maintain single binocular vision. The superior colliculus is involved in all of these, but its role in saccades has been studied most intensively.

The output from the motor sector of the SC goes to a set of midbrain and brainstem nuclei, which transform the "place" code used by the SC into the "rate" code used by oculomotor neurons. Eye movements are generated by six muscles, arranged in three orthogonally-aligned pairs. Thus, at the level of the final common path, eye movements are encoded in essentially a Cartesian coordinate system.

Although the SC receives a strong input directly from the retina, in primates it is largely under the control of the cerebral cortex, which contains several areas that are involved in determining eye movements. The frontal eye fields, a portion of the motor cortex, are involved in triggering intentional saccades, and an adjoining area, the supplementary eye fields, are involved in organizing groups of saccades into sequences. The parietal eye fields, farther back in the brain, are involved mainly in reflexive saccades, made in response to changes in the view.

The SC only receives visual inputs in its superficial layers, whereas the deeper layers of the colliculus receive also auditory and somatosensory inputs and are connected to many sensorimotor areas of the brain. The colliculus as a whole is thought to help orient the head and eyes toward something seen and heard.

The superior colliculus also receives auditory information from the inferior colliculus. This auditory information is integrated with the visual information already present to produce the ventriloquist effect.

26.2.8 The Lateral Geniculate Nucleus (LGN)

The lateral geniculate nucleus (LGN; also called the lateral geniculate body or lateral geniculate complex; named after its resemblance to a bent knee) is a relay center in the thalamus for the visual pathway. It receives a major sensory input from the retina. The LGN is the main central connection for the optic nerve to the occipital lobe, particularly the primary visual cortex. In humans, each LGN has six layers of neurons (grey matter) alternating with optic fibers (white matter).

The LGN is a small, ovoid, ventral projection at the termination of the optic tract on each side of the brain. The LGN and the medial geniculate nucleus which deals with auditory information are both thalamic nuclei and so are present in both hemispheres.

The LGN receives information directly from the ascending retinal ganglion cells via the optic tract and from the reticular activating system. Neurons of the LGN send their axons through the optic radiation, a direct pathway to the primary visual cortex. In addition, the LGN receives many strong feedback connections from the primary visual cortex. In humans as well as other mammals, the two strongest pathways linking the eye to the brain are those projecting to the dorsal part of the LGN in the thalamus, and to the superior colliculus.

In humans as well as in many other primates, the LGN has layers of magnocellular cells and parvocellular cells that are interleaved with layers of koniocellular cells. In humans the LGN is normally described as having six distinctive layers. The inner two layers, (1 and 2) are magnocellular layers, while the outer four layers, (3,4,5 and 6), are parvocellular layers. An additional set of neurons, known as the koniocellular layers, are found ventral to each of the magnocellular and parvocellular layers.

The magnocellular, parvocellular, and koniocellular layers of the LGN correspond with the similarly named types of retinal ganglion cells. Retinal P ganglion cells send axons to a parvocellular layer, M ganglion cells send axons to a magnocellular layer, and K ganglion cells send axons to a koniocellular layer.^{:269}

Koniocellular cells are functionally and neurochemically distinct from M and P cells and provide a third channel to the visual cortex. They project their axons between the layers of the lateral geniculate nucleus where M and P cells project. Their role in visual perception is presently unclear; however, the koniocellular system has been linked with the integration of somatosensory system-proprioceptive information with visual perception, and it may also be involved in color perception.

The other major retino-cortical visual pathway is the tectopulvinar pathway, routing primarily through the superior colliculus and thalamic pulvinar nucleus onto posterior parietal cortex and visual area MT.

Ipsilateral and contralateral layers

Layer 1, 2

- Large cells, called magnocellular pathways
- Input from M-ganglion cells
- Very rapid conduction
- Colour blind system

Layer 3–6

- Parvocellular
- Input from P-ganglion cells
- Colour vision
- Moderate velocity.

Both the LGN in the right hemisphere and the LGN in the left hemisphere receive input from each eye. However, each LGN only receives information from one half of the visual field. This occurs due to axons of the ganglion cells from the inner halves of the retina (the nasal sides) decussating (crossing to the other side of the brain) through the optic chiasma (khiasma means “cross-shaped”). The axons of the ganglion cells from the outer half of the retina (the temporal sides) remain on the same side of the brain. Therefore, the right hemisphere receives visual information from the left visual field, and the left hemisphere receives visual information from the right visual field. Within one LGN, the visual information is divided among the various layers as follows:

- the eye on the same side (the ipsilateral eye) sends information to layers 2, 3 and 5
- the eye on the opposite side (the contralateral eye) sends information to layers 1, 4 and 6.

This description applies to the LGN of many primates, but not all.

The principal neurons in the LGN receive strong inputs from the retina. However, the retina only accounts for a small percentage of LGN input. As much as 95% of input in the LGN comes from the visual cortex, superior colliculus, pretectum, thalamic reticular nuclei, and local LGN interneurons. Regions in the brainstem that are not involved in visual perception also project to the LGN, such as the mesencephalic reticular formation, dorsal raphe nucleus, periaqueductal grey matter, and the locus coeruleus. These non-retinal inputs can be excitatory, inhibitory, or modulatory.

Information leaving the LGN travels out on the optic radiations, which form part of the retro-lentiform portion of the internal capsule.

The axons that leave the LGN go to V1 visual cortex. Both the magnocellular layers 1–2 and the parvocellular layers 3–6 send their axons to layer 4 in V1. Within layer 4 of V1, layer 4c β receives parvocellular input, and layer 4c α receives magnocellular input. However, the koniocellular layers, intercalated between LGN layers 1–6 send their axons primarily to the cytochrome–oxidase rich blobs of layers 2 and 3 in V1. Axons from layer 6 of visual cortex send information back to the LGN.

Studies involving blindsight have suggested that projections from the LGN travel not only to the primary visual cortex but also to higher cortical areas V2 and V3. Patients with blindsight are phenomenally blind in certain areas of the visual field corresponding to a contralateral lesion in the primary visual cortex; however, these patients are able to perform certain motor tasks accurately in their blind field, such as grasping. This suggests that neurons travel

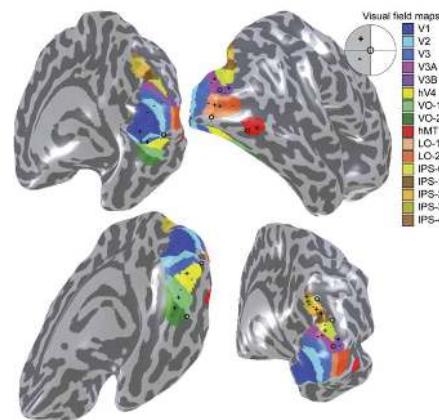


Figure 26.8: A visual field map⁷ of the primary visual cortex and the numerous extrastriate areas.

from the LGN to both the primary visual cortex and higher cortex regions.

26.2.9 The Visual Cortex

The visual cortex of the brain is that part of the cerebral cortex which processes visual information. It is located in the occipital lobe.

The visual cortex is the largest system in the human brain and is responsible for processing the visual information. The region that receives information directly from the LGN is called the primary visual cortex, (also called V1 and striate cortex). The primary visual cortex is the most studied visual area in the brain. Visual information then flows through a cortical hierarchy. These areas include V2, V3, V4 and area V5/MT (the exact connectivity depends on the species of the animal).

As visual information passes forward through the visual hierarchy, the complexity of the neural representations increases. Whereas a V1 neuron may respond selectively to a line segment of a particular orientation in a particular retinotopic location, neurons in the lateral occipital complex respond selectively to complete object (e.g., a figure drawing), and neurons in visual association cortex may respond selectively to human faces, or to a particular object.

Along with this increasing complexity of neural representation may come a level of specialization of processing into two distinct pathways: the dorsal stream and the ventral stream (the Two Streams hypothesis, first proposed by Ungerleider and Mishkin in 1982). The dorsal stream, commonly referred to as the “where” stream, is involved in spatial attention (covert and overt), and communicates with regions that control eye movements and hand movements. More recently, this area has been called the “how”

stream to emphasize its role in guiding behaviors to spatial locations. The ventral stream, commonly referred as the “what” stream, is involved in the recognition, identification and categorization of visual stimuli.

However, there is still much debate about the degree of specialization within these two pathways, since they are in fact heavily interconnected.

Visual information coming from the eye goes through the lateral geniculate nucleus in the thalamus and then reaches the visual cortex. The part of the visual cortex that receives the sensory inputs from the thalamus is the primary visual cortex, also known as visual area 1 (V1, Brodmann area 17), and the striate cortex. The extrastriate areas consist of visual areas 2 (V2, Brodmann area 18), 3, 4, and 5 (V3, V4, V5, all Brodmann area 19).

The primary visual cortex (V1) is located in and around the calcarine fissure in the occipital lobe. Each hemisphere's V1 receives information directly from its ipsilateral lateral geniculate nucleus that receives signals from the contralateral visual hemifield.

Neurons in the visual cortex fire action potentials when visual stimuli appear within their receptive field. By definition, the receptive field is the region within the entire visual field that elicits an action potential. But, for any given neuron, it may respond best to a subset of stimuli within its receptive field. This property is called neuronal tuning. In the earlier visual areas, neurons have simpler tuning. For example, a neuron in V1 may fire to any vertical stimulus in its receptive field. In the higher visual areas, neurons have complex tuning. For example, in the inferior temporal cortex (IT), a neuron may fire only when a certain face appears in its receptive field.

The visual cortex receives its blood supply primarily from the calcarine branch of the posterior cerebral artery.

V1 transmits information to two primary pathways, called the ventral stream and the dorsal stream. The ventral stream begins with V1, goes through visual area V2, then through visual area V4, and to the inferior temporal cortex (IT cortex). The ventral stream, sometimes called the “What Pathway”, is associated with form recognition and object representation. It is also associated with storage of long-term memory. The dorsal stream begins with V1, goes through Visual area V2, then to the dorsomedial area (DM/V6) and medial temporal area (MT/V5) and to the posterior parietal cortex. The dorsal stream, sometimes called the “Where Pathway” or “How Pathway”, is associated with motion, representation of object locations, and control of the eyes and arms, especially when visual information is used to guide saccades or reaching.

26.3 The Auditory And Vestibular Systems

The auditory system is the sensory system for the sense of hearing. It includes both the sensory organs (the ears) and the auditory parts of the sensory system. Hearing, or auditory perception, is the ability to perceive sounds by detecting vibrations, changes in the pressure of the surrounding medium over time, through the ear.

Providing balance, when moving or stationary, is also a central function of the ear. The ear facilitates two types of balance: static balance, which allows a person to feel the effects of gravity, and dynamic balance, which allows a person to sense acceleration.

26.3.1 The Ear

In mammals, the ear is usually described as having three parts—the outer ear, the middle ear and the inner ear. The outer ear consists of the pinna and the ear canal. The folds of cartilage surrounding the ear canal are called the pinna. Sound waves are reflected and attenuated when they hit the pinna, and these changes provide additional information that will help the brain determine the sound direction. Since the outer ear is the only visible portion of the ear in most animals, the word “ear” often refers to the external part alone. The middle ear includes the tympanic cavity and the three ossicles. The inner ear sits in the bony labyrinth, and contains structures which are key to several senses: the semicircular canals, which enable balance and eye tracking when moving; the utricle and saccule, which enable balance when stationary; and the cochlea, which enables hearing. The ears of vertebrates are placed somewhat symmetrically on either side of the head, an arrangement that aids sound localisation.

The ear develops from the first pharyngeal pouch and six small swellings that develop in the early embryo called otic placodes, which are derived from ectoderm.

The ear canal of the outer ear is separated from the air-filled tympanic cavity of the middle ear by the eardrum. The middle ear contains the three small bones—the ossicles—involved in the transmission of sound, and is connected to the throat at the nasopharynx, via the pharyngeal opening of the Eustachian tube. The inner ear contains the otolith organs—the utricle and saccule—and the semicircular canals belonging to the vestibular system, as well as the cochlea of the auditory system.

Sound waves travel through the ear canal and hit the tympanic membrane, or eardrum. This wave information travels across the air-filled middle ear cavity via a series of delicate bones: the malleus (hammer), incus (anvil) and stapes (stirrup). These ossicles act as a lever,

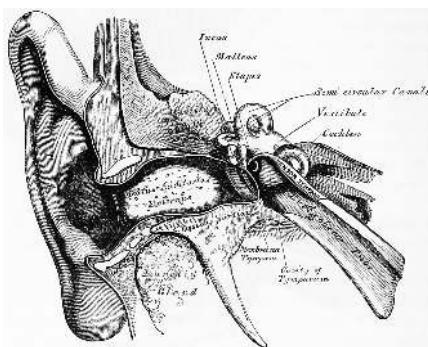


Figure 26.9: Front view of the right outer, middle and inner human ear. Gray, Henry, 1825–1861. Anatomy, descriptive and surgical; ed. by T. Pickering Pick and Robert Howden. A revised American, from the fifteenth English edition. Philadelphia, Lea, 1901⁸

converting the lower-pressure eardrum sound vibrations into higher-pressure sound vibrations at another, smaller membrane called the oval window or vestibular window. The manubrium (handle) of the malleus articulates with the tympanic membrane, while the footplate (base) of the stapes articulates with the oval window. Higher pressure is necessary at the oval window than at the tympanic membrane because the inner ear beyond the oval window contains liquid rather than air. The stapedius reflex of the middle ear muscles helps protect the inner ear from damage by reducing the transmission of sound energy when the stapedius muscle is activated in response to sound. The middle ear still contains the sound information in wave form; it is converted to nerve impulses in the cochlea. The middle-ear ossicles further amplify the vibration pressure roughly 20 times. The base of the stapes couples vibrations into the cochlea via the oval window, which vibrates the perilymph liquid (present throughout the inner ear) and causes the round window to bulge out as the oval window bulges in.

The inner ear consists of the cochlea and several non-auditory structures. The cochlea has three fluid-filled sections (i.e. the scala media, scala tympani and scala vestibuli), and supports a fluid wave driven by pressure across the basilar membrane separating two of the sections. Strikingly, one section, called the cochlear duct or scala media, contains endolymph. Endolymph is a fluid similar in composition to the intracellular fluid found inside cells. The organ of Corti is located in this duct on the basilar membrane, and transforms mechanical waves to electric signals in neurons. The other two sections are known as the scala tympani and the scala vestibuli. These are located within the bony labyrinth, which is filled with fluid called perilymph, similar in composition to cerebrospinal fluid. The chemical difference between

the fluids endolymph and perilymph fluids is important for the function of the inner ear.

26.4 The Auditory System

In humans and other vertebrates, hearing is performed primarily by the auditory system: mechanical waves, known as vibrations, are detected by the ear and transduced into nerve impulses that are perceived by the brain (primarily in the temporal lobe). Like touch, audition requires sensitivity to the movement of molecules in the world outside the organism. Both hearing and touch are types of mechanosensation. Sound may be heard through solid, liquid, or gaseous matter. It is one of the traditional five senses; partial or total inability to hear is called hearing loss.

26.4.1 Organ Of Corti

The organ of Corti, or spiral organ, is the receptor organ for hearing and is located in the mammalian cochlea. This highly varied strip of epithelial cells allows for transduction of auditory signals into nerve impulses. Transduction occurs through vibrations of structures in the inner ear causing displacement of cochlear fluid and movement of hair cells at the organ of Corti to produce electrochemical signals.

Italian anatomist Alfonso Giacomo Gaspare Corti⁹ (1822–1876) discovered the organ of corti in 1851.

The organ of corti is located in the scala media of the cochlea of the inner ear between the vestibular duct and the tympanic duct and is composed of mechanosensory cells, known as hair cells. Strategically positioned on the basilar membrane of the organ of corti are three rows of outer hair cells (ohcs) and one row of inner hair cells (ihcs). Separating these hair cells are supporting cells: Deiters cells, also called phalangeal cells, which separate and support both the ohcs and the ihcs.

Projecting from the tips of the hair cells are tiny finger like projections called stereocilia, which are arranged in a graduated fashion with the shortest stereocilia on the outer rows and the longest in the center.

If the cochlea were uncoiled it would roll out to be about 33 mm long in women and 34 mm in men, with about 2.28 mm of standard deviation for the population. The cochlea is also tonotopically organized, meaning that different frequencies of sound waves interact with different locations on the structure. The base of the cochlea, closest to the outer ear, is the most stiff and narrow and is where the high frequency sounds are transduced. The

⁹https://en.wikipedia.org/wiki/Alfonso_Giacomo_Gaspare_Corti

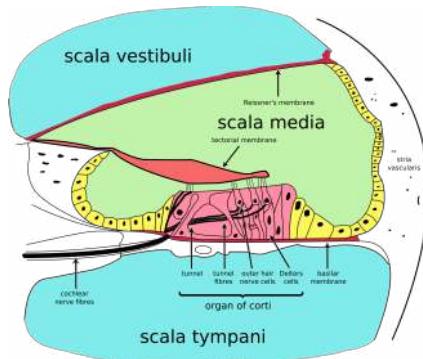


Figure 26.10: A cross section of the cochlea illustrating the organ of Corti.¹⁰

apex, or top, of the cochlea is wider and much more flexible and loose and functions as the transduction site for low frequency sounds.

26.4.2 Auditory Transduction

In normal hearing subjects, the majority of the auditory signals that reach the organ of Corti in the first place come from the outer ear. Sound waves enter through the auditory canal and vibrate the tympanic membrane, also known as the eardrum, which vibrates three small bones called the ossicles. As a result, the attached oval window moves and causes movement of the round window, which leads to displacement of the cochlear fluid. However, the stimulation can happen also via direct vibration of the cochlea from the skull. The latter is referred to as Bone Conduction (or BC) hearing, as complementary to the first one described, which is instead called Air Conduction (or AC) hearing. Both AC and BC stimulate the basilar membrane in the same way.

The basilar membrane on the tympanic duct presses against the hair cells of the organ as perilymphatic pressure waves pass. The stereocilia atop the IHCs move with this fluid displacement and in response their cation, or positive ion selective, channels are pulled open by cadherin structures called tip links that connect adjacent stereocilia. The organ of Corti, surrounded in potassium rich fluid endolymph, lies on the basilar membrane at the base of the scala media. Under the organ of Corti is the scala tympani and above it, the scala vestibuli. Both structures exist in a low potassium fluid called perilymph. Because those stereocilia are in the midst of a high concentration of potassium, once their cation channels are pulled open, potassium ions as well as calcium ions flow into the top of the hair cell. With this influx of positive ions the IHC becomes depolarized, opening voltage-gated calcium channels at the basolateral region of the hair cells and triggering the release of the neurotransmitter glutamate. An electrical signal is

then sent through the auditory nerve and into the auditory cortex of the brain as a neural message.

The organ of Corti is also capable of modulating the auditory signal. The outer hair cells (OHCs) can amplify the signal through a process called electromotility where they increase movement of the basilar and tectorial membranes and therefore increase deflection of stereocilia in the IHCs.

A crucial piece to this cochlear amplification is the motor protein prestin, which changes shape based on the voltage potential inside of the hair cell. When the cell is depolarized, prestin shortens, and because it is located on the membrane of OHCs it then pulls on the basilar membrane and increasing how much the membrane is deflected, creating a more intense effect on the inner hair cells (IHCs). When the cell hyperpolarizes prestin lengthens and eases tension on the IHCs, which decreases the neural impulses to the brain. In this way, the hair cell itself is able to modify the auditory signal before it even reaches the brain.

Hair cells are columnar cells, each with a bundle of 100–200 specialized cilia at the top, for which they are named. There are two types of hair cells; inner and outer hair cells. Inner hair cells are the mechanoreceptors for hearing: they transduce the vibration of sound into electrical activity in nerve fibers, which is transmitted to the brain. Outer hair cells are a motor structure. Sound energy causes changes in the shape of these cells, which serves to amplify sound vibrations in a frequency specific manner. Lightly resting atop the longest cilia of the inner hair cells is the tectorial membrane, which moves back and forth with each cycle of sound, tilting the cilia, which is what elicits the hair cells' electrical responses.

Inner hair cells, like the photoreceptor cells of the eye, show a graded response, instead of the spikes typical of other neurons.

26.4.3 Auditory Pathways

Afferent neurons innervate cochlear inner hair cells, at synapses where the neurotransmitter glutamate communicates signals from the hair cells to the dendrites of the primary auditory neurons.

There are far fewer inner hair cells in the cochlea than afferent nerve fibers – many auditory nerve fibers innervate each hair cell. The neural dendrites belong to neurons of the auditory nerve, which in turn joins the vestibular nerve to form the vestibulocochlear nerve, or cranial nerve number VIII. The region of the basilar membrane supplying the inputs to a particular afferent nerve fibre can be considered to be its receptive field.

Efferent projections from the brain to the cochlea also play a role in the perception of sound, although this is not

well understood. Efferent synapses occur on outer hair cells and on afferent (towards the brain) dendrites under inner hair cells

26.4.4 The Cochlear Nucleus

The cochlear nucleus is the first site of the neuronal processing of the newly converted “digital” data from the inner ear. In mammals, this region is anatomically and physiologically split into two regions, the dorsal cochlear nucleus (DCN), and ventral cochlear nucleus (VCN).

26.4.5 The Trapezoid Body

The trapezoid body is a bundle of decussating fibers in the ventral pons that carry information used for binaural computations in the brainstem. Some of these axons come from the cochlear nucleus and cross over to the other side before traveling on to the superior olivary nucleus. This is believed to help with localization of sound.

26.4.6 The superior olivary complex

The superior olivary complex is located in the pons, and receives projections predominantly from the ventral cochlear nucleus, although the dorsal cochlear nucleus projects there as well, via the ventral acoustic stria. Within the superior olivary complex lies the lateral superior olive (LSO) and the medial superior olive (MSO). The former is important in detecting interaural level differences while the latter is important in distinguishing interaural time difference.

26.4.7 The Lateral Lemniscus

The lateral lemniscus is a tract of axons in the brainstem that carries information about sound from the cochlear nucleus to various brainstem nuclei and ultimately the contralateral inferior colliculus of the midbrain.

26.4.8 The Inferior Colliculi

The inferior colliculi (IC) are located just below the visual processing centers known as the superior colliculi. The central nucleus of the IC is a nearly obligatory relay in the ascending auditory system, and most likely acts to integrate information (specifically regarding sound source localization from the superior olivary complex and dorsal cochlear nucleus) before sending it to the thalamus and cortex.

26.4.9 The Medial Geniculate Nucleus (MGN)

The medial geniculate nucleus is part of the thalamic relay system.

26.4.10 The Primary Auditory Cortex

The primary auditory cortex is the first region of cerebral cortex to receive auditory input.

Perception of sound is associated with the left posterior superior temporal gyrus (STG). The superior temporal gyrus contains several important structures of the brain, including Brodmann areas 41 and 42, marking the location of the primary auditory cortex, the cortical region responsible for the sensation of basic characteristics of sound such as pitch and rhythm. We know from research in non-human primates that the primary auditory cortex can probably be divided further into functionally differentiable sub-regions. The neurons of the primary auditory cortex can be considered to have receptive fields covering a range of auditory frequencies and have selective responses to harmonic pitches. Neurons integrating information from the two ears have receptive fields covering a particular region of auditory space.

The primary auditory cortex is surrounded by secondary auditory cortex, and interconnects with it. These secondary areas interconnect with further processing areas in the superior temporal gyrus, in the dorsal bank of the superior temporal sulcus, and in the frontal lobe. In humans, connections of these regions with the middle temporal gyrus are probably important for speech perception. The frontotemporal system underlying auditory perception allows us to distinguish sounds as speech, music, or noise.

26.4.11 The Auditory Ventral And Dorsal Streams

From the primary auditory cortex emerge two separate pathways: the auditory ventral stream and auditory dorsal stream. The auditory ventral stream includes the anterior superior temporal gyrus, anterior superior temporal sulcus, middle temporal gyrus and temporal pole. Neurons in these areas are responsible for sound recognition, and extraction of meaning from sentences. The auditory dorsal stream includes the posterior superior temporal gyrus and sulcus, inferior parietal lobule and intra-parietal sulcus. Both pathways project in humans to the inferior frontal gyrus. The most established role of the auditory dorsal stream in primates is sound localization. In humans, the auditory dorsal stream in the left hemisphere is also responsible for speech repetition and articulation, phonological long-term encoding of word names, and verbal working memory.

26.5 The Vestibular System

The vestibular system, in vertebrates, is part of the inner ear. In most mammals, the vestibular system is the sensory system that provides the leading contribution to the sense of balance and spatial orientation for the purpose of coordinating movement with balance. Together with the cochlea, a part of the auditory system, it constitutes the labyrinth of the inner ear in most mammals. As movements consist of rotations and translations, the vestibular system comprises two components: the semicircular canals which indicate rotational movements; and the otoliths which indicate linear accelerations. The vestibular system sends signals primarily to the neural structures that control eye movements, and to the muscles that keep an animal upright and in general control posture. The projections to the former provide the anatomical basis of the vestibulo-ocular reflex, which is required for vision; while the projections to the latter provide the anatomical means required to enable an animal to maintain its position in space.

The brain uses information from the vestibular system in the head and from proprioception throughout the body to enable the animal to understand its body's dynamics and kinematics (including its position and acceleration) from moment to moment. How these two perceptive sources are integrated to provide the underlying structure of the sensorium is unknown.

26.5.1 The Semicircular Canals

The semicircular canal system detects rotational movements.

The semicircular canals are a component of the bony labyrinth that are at right angles to each other. At one end of each of the semicircular canals is a dilated sac called an osseous ampulla which is more than twice the diameter of the canal. Each ampulla contains an ampulla crest, the crista ampullaris which consists of a thick gelatinous cap called a cupula and many hair cells. The superior and posterior semicircular canals are oriented vertically at right angles to each other. The lateral semicircular canal is about a 30-degree angle from the horizontal plane. The orientations of the canals cause a different canal to be stimulated by movement of the head in different planes, and more than one canal is stimulated at once if the movement is off those planes. The horizontal canal detects angular acceleration of the head when the head is turned and the superior and posterior canals detect vertical head movements when the head is moved up or down. When the head changes position, the endolymph in the canals lags behind due to inertia and this acts on the cupula which bends the cilia of the hair cells. The stimulation of the hair cells sends the message to the brain that acceleration is taking place. The

ampullae open into the vestibule by five orifices, one of the apertures being common to two of the canals.

Since the world is three-dimensional, the vestibular system contains three semicircular canals in each labyrinth. They are approximately orthogonal (at right angles) to each other, and are the horizontal (or lateral), the anterior semicircular canal (or superior), and the posterior (or inferior) semicircular canal. Anterior and posterior canals may collectively be called vertical semicircular canals.

The anterior and posterior semicircular canals detect rotations of the head in the sagittal plane (as when nodding), and in the frontal plane, as when cartwheeling. Both anterior and posterior canals are orientated at approximately 45° between frontal and sagittal planes. The movement of fluid pushes on a structure called the cupula which contains hair cells that transduce the mechanical movement to electrical signals.

The canals are arranged in such a way that each canal on the left side has an almost parallel counterpart on the right side. Each of these three pairs works in a push-pull fashion: when one canal is stimulated, its corresponding partner on the other side is inhibited, and vice versa.

26.5.2 The Otolithic Organs

While the semicircular canals respond to rotations, the otolithic organs sense linear accelerations. Humans have two otolithic organs on each side, one called the utricle, the other called the saccule. The utricle contains a patch of hair cells and supporting cells called a macula. Similarly, the saccule contains a patch of hair cells and a macula. Each hair cell of a macula has 40–70 stereocilia and one true cilium called a kinocilium. The tips of these cilia are embedded in an otolithic membrane. This membrane is weighted down with protein-calcium carbonate granules called otoconia. These otoconia add to the weight and inertia of the membrane and enhance the sense of gravity and motion. With the head erect, the otolithic membrane bears directly down on the hair cells and stimulation is minimal. When the head is tilted, however, the otolithic membrane sags and bends the stereocilia, stimulating the hair cells. Any orientation of the head causes a combination of stimulation to the utricles and saccules of the two ears. The brain interprets head orientation by comparing these inputs to each other and to other input from the eyes and stretch receptors in the neck, thereby detecting whether the head is tilted or the entire body is tipping. Essentially, these otolithic organs sense how quickly you are accelerating forward or backward, left or right, or up or down. Most of the utricular signals elicit eye movements, while the majority of the saccular signals project to muscles that control our posture.

While the interpretation of the rotation signals from the semicircular canals is straightforward, the interpretation of otolith signals is more difficult: since gravity is equivalent to a constant linear acceleration, one somehow has to distinguish otolith signals that are caused by linear movements from those caused by gravity. Humans can do that quite well, but the neural mechanisms underlying this separation are not yet fully understood. Humans can sense head tilting and linear acceleration even in dark environments because of the orientation of two groups of hair cell bundles on either side of the striola. Hair cells on opposite sides move with mirror symmetry, so when one side is moved, the other is inhibited. The opposing effects caused by a tilt of the head cause differential sensory inputs from the hair cell bundles allow humans to tell which way the head is tilting. Sensory information is then sent to the brain, which can respond with appropriate corrective actions to the nervous and muscular systems to ensure that balance and awareness are maintained.

Diseases of the vestibular system can take different forms, and usually induce vertigo and instability or loss of balance, often accompanied by nausea.

When the vestibular system and the visual system deliver incongruous results, nausea often occurs. When the vestibular system reports no movement but the visual system reports movement, the motion disorientation is often called motion sickness (or seasickness, car sickness, simulation sickness, or airsickness). In the opposite case, such as when a person is in a zero-gravity environment or during a virtual reality session, the disoriented sensation is often called space sickness or space adaptation syndrome. Either of these "sicknesses" usually ceases once the congruity between the two systems is restored.

Each canal is filled with a fluid called endolymph and contains motion sensors within the fluids. At the base of each canal, the bony region of the canal is enlarged which opens into the utricle and has a dilated sac at one end called the osseous ampullae. Within the ampulla is a mound of hair cells and supporting cells called crista ampullaris. These hair cells have many cytoplasmic projections on the apical surface called stereocilia which are embedded in a gelatinous structure called the cupula. As the head rotates the duct moves but the endolymph lags behind owing to inertia. This deflects the cupula and bends the stereocilia within. The bending of these stereocilia alters an electric signal that is transmitted to the brain. Within approximately 10 seconds of achieving constant motion, the endolymph catches up with the movement of the duct and the cupula is no longer affected, stopping the sensation of acceleration. The specific gravity of the cupula is comparable to that of the surrounding endolymph. Consequently, the cupula is not

displaced by gravity, unlike the otolithic membranes of the utricle and saccule. As with macular hair cells, hair cells of the crista ampullaris will depolarise when the stereocilia deflect towards the kinocilium. Deflection in the opposite direction results in hyperpolarisation and inhibition. In the horizontal canal, ampullopetal flow is necessary for hair-cell stimulation, whereas ampullofugal flow is necessary for the anterior and posterior canals.

26.5.3 Vestibular Pathways

The vestibular nerve is one of the two branches of the vestibulocochlear nerve (the cochlear nerve being the other). In humans the vestibular nerve transmits sensory information from vestibular hair cells located in the two otolith organs (the utricle and the saccule) and the three semicircular canals via the vestibular ganglion.

Axons of the vestibular nerve synapse in the vestibular nucleus are found on the lateral floor and wall of the fourth ventricle in the pons and medulla.

It arises from bipolar cells in the vestibular ganglion, ganglion of Scarpa, which is situated in the upper part of the outer end of the internal auditory meatus.

The fibers of the vestibular nerve enter the medulla oblongata on the medial side of those of the cochlear, and pass between the inferior peduncle and the spinal tract of the trigeminal nerve.

They then divide into ascending and descending fibers. The latter end by arborizing around the cells of the medial nucleus, which is situated in the area acustica of the rhomboid fossa. The ascending fibers either end in the same manner or in the lateral nucleus, which is situated lateral to the area acustica and farther from the ventricular floor.

Some of the axons of the cells of the lateral nucleus, and possibly also of the medial nucleus, are continued upward through the inferior peduncle to the roof nuclei of the opposite side of the cerebellum, to which also other fibers of the vestibular root are prolonged without interruption in the nuclei of the medulla oblongata.

A second set of fibers from the medial and lateral nuclei end partly in the tegmentum, while the remainder ascend in the medial longitudinal fasciculus to arborize around the cells of the nuclei of the oculomotor nerve.

Fibers from the lateral vestibular nucleus also pass via the vestibulospinal tract, to anterior horn cells at many levels in the spinal cord, in order to co-ordinate head and trunk movements.

The vestibular cortex is the portion of the cerebrum which responds to input from the vestibular system. In humans, it has not been completely delineated but is thought

to encompass regions in the parietal and temporal lobes.

26.6 The Somatic Sensory System

The somatic sensory system is a part of the sensory nervous system. The somatosensory system is a complex system of sensory neurons and neural pathways that responds to changes at the surface or inside the body. The axons (as afferent nerve fibers) of sensory neurons connect with, or respond to, various receptor cells. These sensory receptor cells are activated by different stimuli such as heat and nociception, giving a functional name to the responding sensory neuron, such as a thermoreceptor which carries information about temperature changes. Other types include mechanoreceptors, chemoreceptors, and nociceptors which send signals along a sensory nerve to the spinal cord where they may be processed by other sensory neurons and then relayed to the brain for further processing. Sensory receptors are found all over the body including the skin, epithelial tissues, muscles, bones and joints, internal organs, and the cardiovascular system.

Somatic senses are sometimes referred to as somesthetic senses, with the understanding that somesthesia includes the sense of touch and proprioception (sense of position and movement).

The mapping of the body surfaces in the brain is called somatotopy. In the cortex, it is also referred to as the cortical homunculus. This brain-surface ("cortical") map is not immutable, however. Dramatic shifts can occur in response to stroke or injury.

26.6.1 Touch

In contrast, the other sense, touch, is a somatic sense which does not have a specialized organ but comes from all over the body, most noticeably the skin but also the internal organs (viscera). Touch includes mechanoreception (pressure, vibration and proprioception), pain (nociception) and heat (thermoception), and such information is carried in general somatic afferents and general visceral afferents.

Skin is the soft outer tissue covering of vertebrates with three main functions: protection, regulation, and sensation.

Anatomy of human skin.¹¹ Diagrammatic section of hairless skin. Note tactile (Meissner) and Pacinian corpuscles.¹²

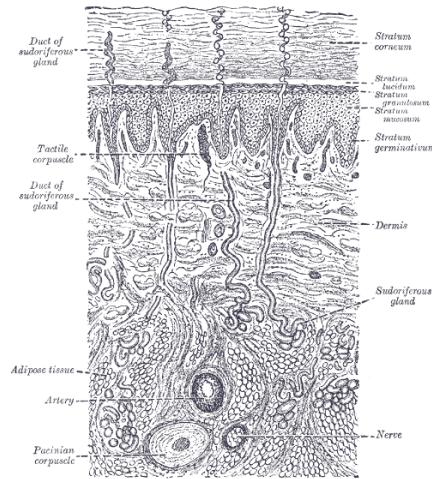


Figure 26.11: Anatomy of human skin.¹³

26.6.2 Cutaneous Mechanoreceptors

Cutaneous mechanoreceptors respond to mechanical stimuli that result from physical interaction, including pressure and vibration. They are located in the skin. They are all innervated by A β fibers, except the mechanoreceiving free nerve endings, which are A δ fibers. Cutaneous mechanoreceptors can be categorized by morphology, by what kind of sensation they perceive, and by the rate of adaptation. Furthermore, each has a different receptive field.

In the somatosensory system, receptive fields are regions of the skin or of internal organs. Some types of mechanoreceptors have large receptive fields, while others have smaller ones. Large receptive fields allow the cell to detect changes over a wider area, but lead to a less precise perception. Thus, the fingers, which require the ability to detect fine detail, have many, densely packed (up to 500 per cubic cm) mechanoreceptors with small receptive fields (around 10 square mm), while the back and legs, for example, have fewer receptors with large receptive fields.

Tactile-sense-related cortical neurons have receptive fields on the skin that can be modified by experience or by injury to sensory nerves resulting in changes in the field's size and position. In general these neurons have relatively large receptive fields (much larger than those of dorsal root ganglion cells). However, the neurons are able to discriminate fine detail due to patterns of excitation and inhibition relative to the field which leads to spatial resolution.

The term receptive field was first used by Sherrington (1906) to describe the area of skin from which a scratch reflex could be elicited in a dog. According to Alonso and Chen (2008) it was Hartline (1938) who applied the term to single neurons, in this case from the retina of a frog.

A sensory space can also map into a particular region

¹¹<https://commons.wikimedia.org/wiki/File:Skin.jpg>

¹²<https://commons.wikimedia.org/wiki/File:Gray940.png>

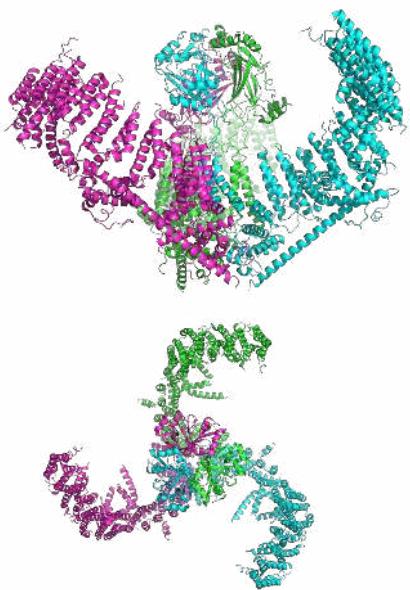


Figure 26.12: A cartoon representation of the mechanosensitive Piezo1 channel in side and top view. PDB 5Z10¹⁴, rendered with open source molecular visualization tool PyMol.

on an animal's body. For example, it could be a hair in the cochlea or a piece of skin, retina, or tongue or other part of an animal's body.

This concept of receptive fields can be extended further up the nervous system; if many sensory receptors all form synapses with a single cell further up, they collectively form the receptive field of that cell. For example, the receptive field of a ganglion cell in the retina of the eye is composed of input from all of the photoreceptors which synapse with it, and a group of ganglion cells in turn forms the receptive field for a cell in the brain. This process is called convergence.

Tactile corpuscles or Meissner's corpuscles are a type of mechanoreceptor discovered by anatomist Georg Meissner¹⁵ (1829–1905) and Rudolf Wagner¹⁶. They are a type of nerve ending in the skin that is responsible for sensitivity to light touch. In particular, they have their highest sensitivity (lowest threshold) when sensing vibrations between 10 and 50 hertz. They are rapidly adaptive receptors. They are most concentrated in thick hairless skin, especially at the finger pads.

Tactile corpuscles are encapsulated myelinated nerve endings, which consist of flattened supportive cells arranged as horizontal lamellae surrounded by a connective tissue capsule. The corpuscle is 30–140 µm in length and

40–60 µm in diameter. A single nerve fiber meanders between the lamellae and throughout the corpuscle.

Pacinian corpuscles (or lamellar corpuscles; discovered by Italian anatomist Filippo Pacini¹⁷) are one of the four major types of mechanoreceptor cell in glabrous (hairless) mammalian skin. They are nerve endings in the skin responsible for sensitivity to vibration and pressure. They respond only to sudden disturbances and are especially sensitive to vibration. The vibrational role may be used to detect surface texture, e.g., rough vs. smooth. Pacinian corpuscles are also found in the pancreas, where they detect vibration and possibly very low frequency sounds.[dubious – discuss] Pacinian corpuscles act as very rapidly adapting mechanoreceptors. Groups of corpuscles respond to pressure changes, e.g. on grasping or releasing an object.

Pacinian corpuscles are larger and fewer in number than Meissner's corpuscle, Merkel cells and Ruffini's corpuscles.

The Pacinian corpuscle is approximately oval-cylindrical-shaped and 1 mm in length. The entire corpuscle is wrapped by a layer of connective tissue. Its capsule consists of 20 to 60 concentric lamellae (hence the alternative lamellar corpuscle) including fibroblasts and fibrous connective tissue (mainly Type IV and Type II collagen network), separated by gelatinous material, more than 92% of which is water.

Pacinian corpuscles are rapidly adapting (phasic) receptors that detect gross pressure changes and vibrations in the skin. Any deformation in the corpuscle causes action potentials to be generated by opening pressure-sensitive sodium ion channels in the axon membrane. This allows sodium ions to influx, creating a receptor potential.

These corpuscles are especially susceptible to vibrations, which they can sense even centimeters away. Their optimal sensitivity is 250 Hz, and this is the frequency range generated upon fingertips by textures made of features smaller than 1 µm. Pacinian corpuscles cause action potentials when the skin is rapidly indented but not when the pressure is steady, due to the layers of connective tissue that cover the nerve ending. It is thought that they respond to high-velocity changes in joint position. They have also been implicated in detecting the location of touch sensations on handheld tools.

Pacinian corpuscles have a large receptive field on the skin's surface with an especially sensitive center.

Merkel cells, also known as Merkel–Ranvier cells or tactile epithelial cells, are oval-shaped mechanoreceptors essential for light touch sensation and found in the skin of vertebrates. They are abundant in highly sensitive skin like

¹⁵ https://en.wikipedia.org/wiki/Georg_Meissner

¹⁶ https://en.wikipedia.org/wiki/Rudolf_Wagner

¹⁷ https://en.wikipedia.org/wiki/Filippo_Pacini

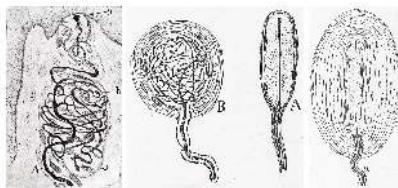


Figure 26.13: Tactile corpuscles (from left to right): Meissner, Krause, Pacini¹⁸

that of the fingertips in humans, and make synaptic contacts with somatosensory afferent nerve fibers. Although uncommon, these cells may become malignant and form a Merkel cell carcinoma—an aggressive and difficult to treat skin cancer.

Though it has been reported that Merkel cells are derived from neural crest cells, more recent experiments in mammals have indicated that they are in fact epithelial in origin.

Merkel cells are found in the skin and some parts of the mucosa of all vertebrates. In mammalian skin, they are clear cells found in the stratum basale (at the bottom of sweat duct ridges) of the epidermis approximately 10 µm in diameter. They also occur in epidermal invaginations of the plantar foot surface called rete ridges. Most often, they are associated with sensory nerve endings, when they are known as Merkel nerve endings (also called a Merkel cell-neurite complex). They are associated with slowly adapting (SA1) somatosensory nerve fibers. They react to low vibrations (5–15 Hz) and deep static touch such as shapes and edges. Due to a small receptive field (extremely detailed info) they are used in areas like fingertips the most; they are not covered (shelled) and thus respond to pressures over long periods.

The German anatomist Friedrich Sigmund Merkel¹⁹ referred to these cells as Tastzellen or “touch cells” but this proposed function has been controversial as it has been hard to prove. Though, genetic knockout mice have recently shown that Merkel cells are essential for the specialized coding by which afferent nerves resolve fine spatial details. Merkel cells are sometimes considered APUD cells (an older definition). More commonly classified as a part of dispersed neuroendocrine system because they contain dense core granules, and thus may also have a neuroendocrine function.

The bulboid corpuscles (end-bulbs of Krause) are cutaneous receptors in the human body. The end-bulbs of Krause were named after the German anatomist Wilhelm Krause²⁰ (1833–1910). End-bulbs are found in

the conjunctiva of the eye (where they are spheroidal in shape in humans, but cylindrical in most other animals), in the mucous membrane of the lips and tongue, and in the epineurium of nerve trunks. They are also found in the penis and the clitoris and have received the name of genital corpuscles. The end-bulbs of Krause were thought to be thermoreceptors, sensing cold temperatures, but their function is unknown.

The Bulbous corpuscle or Ruffini ending or Ruffini corpuscle is a slowly adapting mechanoreceptor located in the cutaneous tissue between the dermal papillae and the hypodermis. It is named after Angelo Ruffini.

Ruffini corpuscles are enlarged dendritic endings with elongated capsules.

This spindle-shaped receptor is sensitive to skin stretch, and contributes to the kinesthetic sense of and control of finger position and movement. They are at the highest density around the fingernails where they act in monitoring slippage of objects along the surface of the skin, allowing modulation of grip on an object.

Ruffini corpuscles respond to sustained pressure and show very little adaptation.

Ruffinian endings are located in the deep layers of the skin, and register mechanical deformation within joints, more specifically angle change, with a specificity of up to 2.75 degrees, as well as continuous pressure states. They also act as thermoreceptors that respond for a long time, so in case of deep burn there will be pain, as these receptors will be burned off.

26.6.3 Nociception

Nociception²¹ (also nociception or nociperception, from Latin nocere ‘to harm or hurt’) is the sensory nervous system’s response to certain harmful or potentially harmful stimuli. The term “nociception” was coined by Charles Scott Sherrington²² to distinguish the physiological process (nervous activity) from pain (a subjective experience). In nociception, intense chemical (e.g., cayenne powder), mechanical (e.g., cutting, crushing), or thermal (heat and cold) stimulation of sensory nerve cells called nociceptors produces a signal that travels along a chain of nerve fibers via the spinal cord to the brain. Nociception triggers a variety of physiological and behavioral responses and usually results in a subjective experience of pain in sentient beings. Nociception can also cause generalized autonomic responses before or without reaching consciousness to cause pallor, sweating,

¹⁹https://en.wikipedia.org/wiki/Friedrich_Sigmund_Merkel

²⁰https://en.wikipedia.org/wiki/Wilhelm_Krause

²¹<https://en.wikipedia.org/wiki/Nociception>

²²https://en.wikipedia.org/wiki/Charles_Scott_Sherrington

tachycardia, hypertension, lightheadedness, nausea and fainting.

Potentially damaging mechanical, thermal, and chemical stimuli are detected by nerve endings called nociceptors, which are found in the skin, on internal surfaces such as the periosteum, joint surfaces, and in some internal organs. Some nociceptors are unspecialized free nerve endings that have their cell bodies outside the spinal column in the dorsal root ganglia. Nociceptors are categorized according to the axons which travel from the receptors to the spinal cord or brain.

Nociceptors have a certain threshold; that is, they require a minimum intensity of stimulation before they trigger a signal. Once this threshold is reached a signal is passed along the axon of the neuron into the spinal cord.

Nociceptive threshold testing deliberately applies a noxious stimulus to a human or animal subject in order to study pain. In animals, the technique is often used to study the efficacy of analgesic drugs and to establish dosing levels and period of effect. After establishing a baseline, the drug under test is given and the elevation in threshold recorded at specified time points. When the drug wears off, the threshold should return to the baseline (pre-treatment) value.

In some conditions, excitation of pain fibers becomes greater as the pain stimulus continues, leading to a condition called hyperalgesia.

Thermoception refers to stimuli of moderate temperatures 24–28 °C (75–82 °F), as anything beyond that range is considered pain and moderated by nociceptors. TRP and potassium channels [TRPM (1–8), TRPV (1–6), TRAAK, and TREK] each respond to different temperatures (among other stimuli) which create action potentials in nerves which join the mechano (touch) system in the posterolateral tract. Thermoception, like proprioception, is then covered by the somatosensory system.

TRP channels that detect noxious stimuli (mechanical, thermal, and chemical pain) relay that info to nociceptors that generate an action potential. Mechanical TRP channels react to depression of their cells (like touch), thermal TRP change shape in different temperatures, and chemical TRP act like taste buds, signalling if their receptors bond to certain elements/chemicals.

26.6.4 Proprioception

Proprioception is the sense of self-movement and body position. It is sometimes described as the “sixth sense”.

Proprioception is mediated by mechanically sensitive proprioceptor neurons distributed throughout an animal's

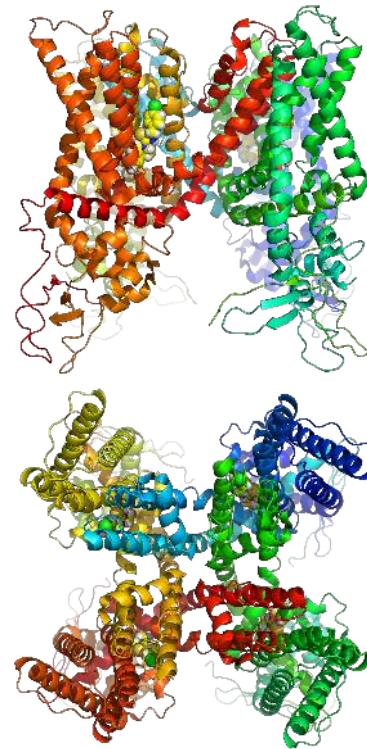


Figure 26.14: A cartoon representation of the transient receptor potential cation channel subfamily V member 1 (TRPV1), also known as the capsaicin receptor and the vanilloid receptor 1 viewed from the side and the top. The function of TRPV1 is detection and regulation of body temperature. In addition, TRPV1 provides a sensation of scalding heat and pain (nociception). In primary afferent sensory neurons, it cooperates with TRPA1 (a chemical irritant receptor) to mediate the detection of noxious environmental stimuli. PDB 5IS0²³, rendered with open source molecular visualization tool PyMol.

body. Most vertebrates possess three basic types of proprioceptors: muscle spindles, which are embedded in skeletal muscle fibers, Golgi tendon organs, which lie at the interface of muscles and tendons, and joint receptors, which are low-threshold mechanoreceptors embedded in joint capsules. Many invertebrates, such as insects, also possess three basic proprioceptor types with analogous functional properties: chordotonal neurons, campaniform sensilla, and hair plates.

The initiation of proprioception is the activation of a proprioceptor in the periphery. The proprioceptive sense is believed to be composed of information from sensory neurons located in the inner ear (motion and orientation) and in the stretch receptors located in the muscles and the joint-supporting ligaments (stance). There are specific nerve receptors for this form of perception termed "proprioceptors".

An important role for proprioception is to allow an animal to stabilize itself against perturbations. For instance, for a person to walk or stand upright, they must continuously monitor their posture and adjust muscle activity as needed to provide balance. Similarly, when walking on unfamiliar terrain or even tripping, the person must adjust the output of their muscles quickly based on estimated limb position and velocity. Proprioceptor reflex circuits are thought to play an important role to allow fast and unconscious execution of these behaviors. To make control of these behaviors efficient, proprioceptors are also thought to regulate reciprocal inhibition in muscles, leading to agonist-antagonist muscle pairs.

When planning complex movements such as reaching or grooming, animals must consider the current position and velocity of their limb and use it to adjust dynamics to target a final position. If the animal's estimate of their limb's initial position is wrong, this can lead to a deficiency in the movement. Furthermore, proprioception is crucial in refining the movement if it deviates from the trajectory.

26.6.5 Muscle Spindles

Muscle spindles are stretch receptors within the body of a muscle that primarily detect changes in the length of the muscle. They convey length information to the central nervous system via afferent nerve fibers. This information can be processed by the brain as proprioception. The responses of muscle spindles to changes in length also play an important role in regulating the contraction of muscles, for example, by activating motor neurons via the stretch reflex to resist muscle stretch.

Muscle spindles are found within the belly of muscles, between extrafusal muscle fibers.[b] The specialised fibers that constitute the muscle spindle are known as intrafusal

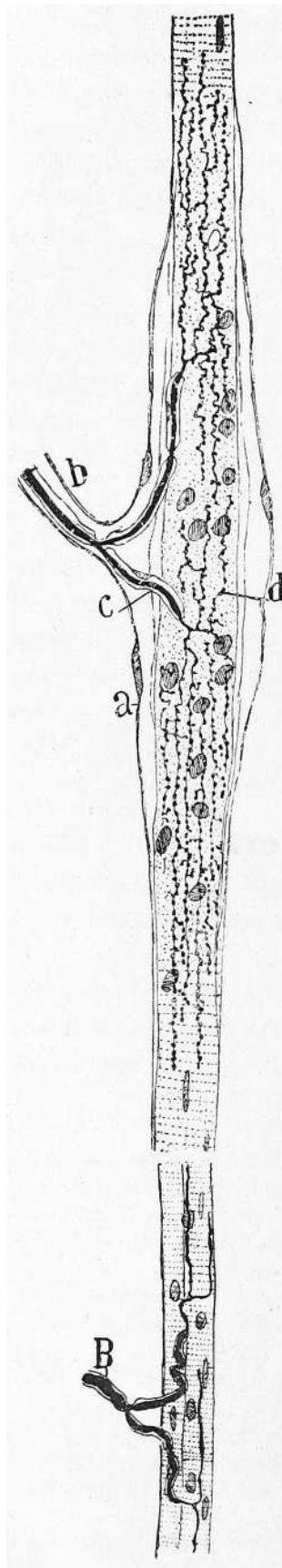


Figure 26.15: Muscle spindle from the pectoral muscle of a frog. Bottom part of figure: B) extrafusal muscle fibers with efferent motor nerve; top part of figure: myelinated afferent nerve fiber. *Histologie du système nerveux de l'homme & des vertébrés, Tome Premier*²⁴ (1909) by Georges W. B. Nissl.

fibers (as they are present within the spindle), to distinguish themselves from the fibres of the muscle itself which are called extrafusal fibers. Muscle spindles have a capsule of connective tissue, and run parallel to the extrafusal muscle fibers.

26.6.6 Golgi Tendon Organ

The Golgi tendon organ (GTO) (also called Golgi organ, tendon organ, neurotendinous organ or neurotendinous spindle) is a proprioceptive sensory receptor organ that senses changes in muscle tension. It lies at the origins and insertion of skeletal muscle fibers into the tendons of skeletal muscle. It provides the sensory component of the Golgi tendon reflex.

The body of the organ is made up of braided strands of collagen (intrafusal fasciculi) that are less compact than elsewhere in the tendon and are encapsulated. The capsule is connected in series with a group of muscle fibers (10–20 fibers) at one end, and merge into the tendon proper at the other. Each capsule is about 1 mm long, has a diameter of about 0.1 mm, and is perforated by one or more afferent type Ib sensory nerve fibers (A_a fiber), which are large (12–20 μm) myelinated axons that can conduct nerve impulses very rapidly. Inside the capsule, the afferent fibers lose their medullary sheaths, branch, intertwine with the collagen fibers, and terminate as flattened leaf-like endings between the collagen strands (see figure).

26.6.7 The Somatosensory Pathways

All afferent touch/vibration info ascends the spinal cord via the posterior (dorsal) column-medial lemniscus pathway via gracilis (T7 and below) or cuneatus (T6 and above). Cuneatus sends signals to the cochlear nucleus indirectly via spinal grey matter, this info is used in determining if a perceived sound is just villi noise/irritation. All fibers cross (left becomes right) in the medulla.

A somatosensory pathway will typically have three neurons: first-order, second-order, and third-order.

1. The first-order neuron is a type of pseudounipolar neuron and always has its cell body in the dorsal root ganglion of the spinal nerve with a peripheral axon innervating touch mechanoreceptors and a central axon synapsing on the second-order neuron. If the somatosensory pathway is in parts of the head or neck not covered by the cervical nerves, the first-order neuron will be the trigeminal nerve ganglia or the ganglia of other sensory cranial nerves).
2. The second-order neuron has its cell body either in the spinal cord or in the brainstem. This neuron's ascending axons will cross (decussate) to the opposite side either in the spinal cord or in the brainstem.

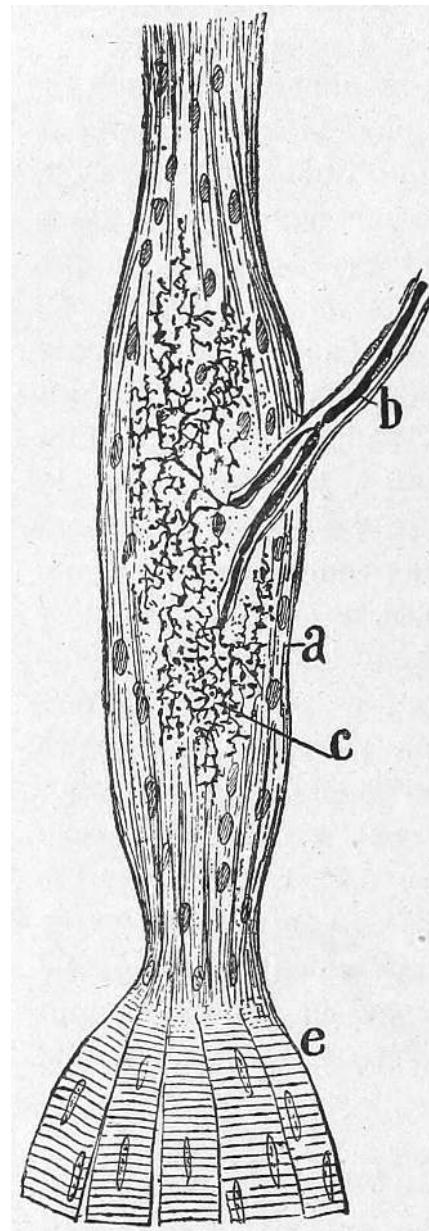


Figure 26.16: Golgi tendon organ from an adult cat. a) Terminal arborisation and c) fine ultimate branches of the afferent nerve; b) myelinated afferent nerve fiber; e) muscle fibers. *Histologie du système nerveux de l'homme & des vertébrés, Tome Premier*²⁵ (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay.

3. In the case of touch and certain types of pain, the third-order neuron has its cell body in the ventral posterior nucleus of the thalamus and ends in the postcentral gyrus of the parietal lobe in the primary somatosensory cortex (or S1).

Fine touch (or discriminative touch) is a sensory modality that allows a subject to sense and localize touch. The form of touch where localization is not possible is known as crude touch. The posterior column-medial lemniscus pathway is the pathway responsible for the sending of fine touch information to the cerebral cortex of the brain.

Crude touch (or non-discriminative touch) is a sensory modality that allows the subject to sense that something has touched them, without being able to localize where they were touched (contrasting “fine touch”). Its fibres are carried in the spinothalamic tract, unlike the fine touch, which is carried in the dorsal column. As fine touch normally works in parallel to crude touch, a person will be able to localize touch until fibres carrying fine touch (Posterior column-medial lemniscus pathway) have been disrupted. Then the subject will feel the touch, but be unable to identify where they were touched.

In humans, temperature sensation from thermoreceptors enters the spinal cord along the axons of Lissauer's tract that synapse on second order neurons in grey matter of the dorsal horn. The axons of these second order neurons then decussate, joining the spinothalamic tract as they ascend to neurons in the ventral posterolateral nucleus of the thalamus.

26.6.8 The Primary Somatosensory Cortex

The primary somatosensory cortex is located in the postcentral gyrus, and is part of the somatosensory system.

At the primary somatosensory cortex, tactile representation is orderly arranged (in an inverted fashion) from the toe (at the top of the cerebral hemisphere) to mouth (at the bottom). However, some body parts may be mapped to partially overlapping regions of cortex. Each cerebral hemisphere of the primary somatosensory cortex only contains a tactile representation of the opposite (contralateral) side of the body. The amount of primary somatosensory cortex devoted to a body part is not proportional to the absolute size of the body surface, but, instead, to the relative density of cutaneous tactile receptors on that body part. The density of cutaneous tactile receptors on a body part is generally indicative of the degree of sensitivity of tactile stimulation experienced at said body part. For example, the human lips and hands have a larger representation than other body parts.

Brodmann areas 3, 1, and 2 make up the primary somatosensory cortex of the human brain (or S1). Brodmann

area (BA) 3 is subdivided into areas 3a and 3b. Because Brodmann sliced the brain somewhat obliquely, he encountered area 1 first; however, from anterior to posterior, the Brodmann designations are 3, 1, and 2, respectively.

Areas 1 and 2 receive dense inputs from BA 3b. The projection from 3b to 1 primarily relays texture information; the projection to area 2 emphasizes size and shape. Lesions confined to these areas produce predictable dysfunction in texture, size, and shape discrimination.

Somatosensory cortex, like other neocortex, is layered. Like other sensory cortex (i.e., visual and auditory) the thalamic inputs project into layer IV, which in turn project into other layers. As in other sensory cortices, S1 neurons are grouped together with similar inputs and responses into vertical columns that extend across cortical layers.

26.7 The Olfactory System

The olfactory system, or sense of smell, is the sensory system used for smelling (olfaction). Olfaction is one of the special senses, that have directly associated specific organs. Most mammals and reptiles have a main olfactory system and an accessory olfactory system. The main olfactory system detects airborne substances, while the accessory system senses fluid-phase stimuli. Often, land organisms will have separate olfaction systems for smell and taste (orthonasal smell and retronasal smell), but water-dwelling organisms usually have only one system. The senses of smell and taste (gustatory system) are often referred to together as the chemosensory system, because they both give the brain information about the chemical composition of objects.

Olfaction is a chemoreception that forms the sense of smell. Olfaction has many purposes, such as the detection of hazards, pheromones, and food. It integrates with other senses to form the sense of flavor. Olfaction occurs when odorants bind to specific sites on olfactory receptors located in the nasal cavity. Glomeruli aggregate signals from these receptors and transmit them to the olfactory bulb, where the sensory input will start to interact with parts of the brain responsible for smell identification, memory, and emotion.

In insects, smells are sensed by olfactory sensory neurons in the chemosensory sensilla, which are present in insect antenna, palps, and tarsa, but also on other parts of the insect body. Odorants penetrate into the cuticle pores of chemosensory sensilla and get in contact with insect odorant-binding proteins (OBPs) or Chemosensory proteins (CSPs), before activating the sensory neurons.

In vertebrates, smells are sensed by olfactory sensory neurons in the olfactory epithelium. The olfactory epithe-

lum is made up of at least six morphologically and biochemically different cell types. The proportion of olfactory epithelium compared to respiratory epithelium (not innervated, or supplied with nerves) gives an indication of the animal's olfactory sensitivity. Humans have about 10 cm^2 of olfactory epithelium, whereas some dogs have 170 cm^2 . A dog's olfactory epithelium is also considerably more densely innervated, with a hundred times more receptors per square centimeter.

Molecules of odorants passing through the superior nasal concha of the nasal passages dissolve in the mucus that lines the superior portion of the cavity and are detected by olfactory receptors on the dendrites of the olfactory sensory neurons. This may occur by diffusion or by the binding of the odorant to odorant-binding proteins. The mucus overlying the epithelium contains mucopolysaccharides, salts, enzymes, and antibodies (these are highly important, as the olfactory neurons provide a direct passage for infection to pass to the brain). This mucus acts as a solvent for odor molecules, flows constantly, and is replaced approximately every ten minutes.

26.7.1 The Nose

The human nose is the most protruding part of the face. It bears the nostrils and is the first organ of the respiratory system. It is also the principal organ in the olfactory system. The shape of the nose is determined by the nasal bones and the nasal cartilages, including the nasal septum which separates the nostrils and divides the nasal cavity into two.

The main function of the nose is respiration, and the nasal mucosa lining the nasal cavity and the paranasal sinuses carries out the necessary conditioning of inhaled air by warming and moistening it. Nasal conchae, shell-like bones in the walls of the cavities, play a major part in this process. Filtering of the air by nasal hair in the nostrils prevents large particles from entering the lungs. Sneezing is a reflex to expel unwanted particles from the nose that irritate the mucosal lining. Sneezing can transmit infections, because aerosols are created in which the droplets can harbour pathogens.

Another major function of the nose is olfaction, the sense of smell. The area of olfactory epithelium, in the upper nasal cavity, contains specialised olfactory cells responsible for this function.

The peripheral olfactory system consists mainly of the nostrils, ethmoid bone, nasal cavity, and the olfactory epithelium (layers of thin tissue covered in mucus that line the nasal cavity). The primary components of the layers of epithelial tissue are the mucous membranes, olfactory glands, olfactory neurons, and nerve fibers of the olfactory

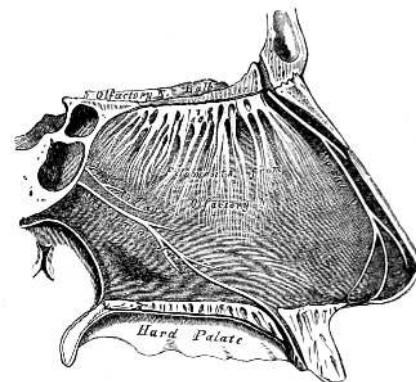


Figure 26.17: View of the right side of the nasal septum showing the olfactory bulb and the filaments of the olfactory nerve. From Gray, Henry: Anatomy Descriptive And Surgical. 7th Edition, Longmans, Green, And Co., London, 1875²⁶.

nerves.

Odor molecules can enter the peripheral pathway and reach the nasal cavity either through the nostrils when inhaling (olfaction) or through the throat when the tongue pushes air to the back of the nasal cavity while chewing or swallowing (retro-nasal olfaction). Inside the nasal cavity, mucus lining the walls of the cavity dissolves odor molecules. Mucus also covers the olfactory epithelium, which contains mucous membranes that produce and store mucus and olfactory glands that secrete metabolic enzymes found in the mucus.

26.7.2 Olfactory Sensory Neurons

Humans have between 10 and 20 million olfactory receptor neurons (ORNs). In vertebrates, ORNs are bipolar neurons with dendrites facing the external surface of the cribriform plate with axons that pass through the cribriform foramina with terminal end at olfactory bulbs. The ORNs are located in the olfactory epithelium in the nasal cavity. The cell bodies of the ORNs are distributed among all three of the stratified layers of the olfactory epithelium.

Many tiny hair-like cilia protrude from the olfactory receptor cell's dendrite into the mucus covering the surface of the olfactory epithelium. The surface of these cilia is covered with olfactory receptors, a type of G protein-coupled receptor. Each olfactory receptor cell expresses only one type of olfactory receptor (OR), but many separate olfactory receptor cells express ORs which bind the same set of odors. The axons of olfactory receptor cells which express the same OR converge to form glomeruli in the olfactory bulb. Olfactory sensory neurons in the epithelium detect odor molecules dissolved in the mucus and transmit infor-

mation about the odor to the brain in a process called sensory transduction. Olfactory neurons have cilia (tiny hairs) containing olfactory receptors that bind to odor molecules, causing an electrical response that spreads through the sensory neuron to the olfactory nerve fibers at the back of the nasal cavity.

Olfactory receptors (ORs), also known as odorant receptors, are expressed in the cell membranes of olfactory receptor neurons and are responsible for the detection of odorants (i.e., compounds that have an odor) which give rise to the sense of smell. Activated olfactory receptors trigger nerve impulses which transmit information about odor to the brain. These receptors are members of the class A rhodopsin-like family of G protein-coupled receptors (GPCRs). The olfactory receptors form a multigene family consisting of around 800 genes in humans and 1400 genes in mice

Rather than binding specific ligands, olfactory receptors display affinity for a range of odor molecules, and conversely a single odorant molecule may bind to a number of olfactory receptors with varying affinities, which depend on physio-chemical properties of molecules like their molecular volumes. Once the odorant has bound to the odor receptor, the receptor undergoes structural changes and it binds and activates the olfactory-type G protein on the inside of the olfactory receptor neuron. The G protein (G_{olf} and/or G_s) in turn activates adenylate cyclase – which converts ATP into cyclic AMP (cAMP). The cAMP opens cyclic nucleotide-gated ion channels which allow calcium and sodium ions to enter into the cell, depolarizing the olfactory receptor neuron and triggering the firing of action potentials which convey the information to the brain.

Olfactory nerves and fibers transmit information about odors from the peripheral olfactory system to the central olfactory system of the brain, which is separated from the epithelium by the cribriform plate of the ethmoid bone. Olfactory nerve fibers, which originate in the epithelium, pass through the cribriform plate, connecting the epithelium to the brain's limbic system at the olfactory bulbs.

26.7.3 The Olfactory Bulb

The main olfactory bulb transmits pulses to both mitral and tufted cells, which help determine odor concentration based off the time certain neuron clusters fire (called ‘timing code’). These cells also note differences between highly similar odors and use that data to aid in later recognition. The cells are different with mitral having low firing-rates and being easily inhibited by neighboring cells, while tufted have high rates of firing and are more difficult to inhibit.

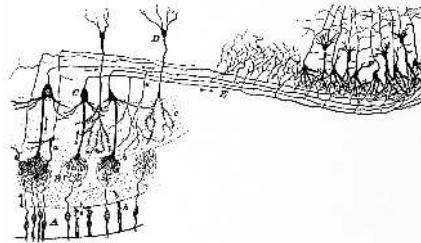


Figure 26.18: Diagram of the structure of the olfactory bulb and olfactory cortex. A) Olfactory mucosa; B) glomeruli in the olfactory bulb; C) mitral cells; D) granule cells; E) olfactory nerve; F) pyramidal cells in the olfactory cortex. Action potentials fired by the olfactory receptor cells and subsequently by the cells in the olfactory bulb, travel to the olfactory cortex (arrows). *Histologie du système nerveux de l'homme & des vertébrés, Tome Premier*²⁷ (1909) by Santiago Ramón y Cajal translated from Spanish by Dr. L. Azoulay.

26.7.4 The Olfactory Cortex

The uncus (an anterior extremity of the parahippocampal gyrus, a region that surrounds the hippocampus and is part of the limbic system) houses the olfactory cortex which includes the piriform cortex (posterior orbitofrontal cortex), amygdala, olfactory tubercle, and parahippocampal gyrus.

The olfactory tubercle connects to numerous areas of the amygdala, thalamus, hypothalamus, hippocampus, brain stem, retina, auditory cortex, and olfactory system.

The anterior olfactory nucleus distributes reciprocal signals between the olfactory bulb and piriform cortex. The anterior olfactory nucleus is the memory hub for smell.

26.7.5 Olfactory Pathways

Olfactory sensory neurons project axons to the brain within the olfactory nerve, (cranial nerve I). These nerve fibers, lacking myelin sheaths, pass to the olfactory bulb of the brain through perforations in the cribriform plate, which in turn projects olfactory information to the olfactory cortex and other areas. The axons from the olfactory receptors converge in the outer layer of the olfactory bulb within small (≈ 50 micrometers in diameter) structures called glomeruli. Mitral cells, located in the inner layer of the olfactory bulb, form synapses with the axons of the sensory neurons within glomeruli and send the information about the odor to other parts of the olfactory system, where multiple signals may be processed to form a synthesized olfactory perception. A large degree of convergence occurs, with 25,000 axons synapsing on 25 or so mitral cells, and with each of these mitral cells

projecting to multiple glomeruli. Mitral cells also project to periglomerular cells and granular cells that inhibit the mitral cells surrounding it (lateral inhibition). Granular cells also mediate inhibition and excitation of mitral cells through pathways from centrifugal fibers and the anterior olfactory nuclei. Neuromodulators like acetylcholine, serotonin and norepinephrine all send axons to the olfactory bulb and have been implicated in gain modulation, pattern separation, and memory functions, respectively.

The mitral cells leave the olfactory bulb in the lateral olfactory tract, which synapses on five major regions of the cerebrum: the anterior olfactory nucleus, the olfactory tubercle, the amygdala, the piriform cortex, and the entorhinal cortex. The anterior olfactory nucleus projects, via the anterior commissure, to the contralateral olfactory bulb, inhibiting it. The piriform cortex has two major divisions with anatomically distinct organizations and functions. The anterior piriform cortex (APC) appears to be better at determining the chemical structure of the odorant molecules, and the posterior piriform cortex (PPC) has a strong role in categorizing odors and assessing similarities between odors (e.g. minty, woody, and citrus are odors that can, despite being highly variant chemicals, be distinguished via the PPC in a concentration-independent manner). The piriform cortex projects to the medial dorsal nucleus of the thalamus, which then projects to the orbitofrontal cortex. The orbitofrontal cortex mediates conscious perception of the odor (citation needed). The three-layered piriform cortex projects to a number of thalamic and hypothalamic nuclei, the hippocampus and amygdala and the orbitofrontal cortex, but its function is largely unknown. The entorhinal cortex projects to the amygdala and is involved in emotional and autonomic responses to odor. It also projects to the hippocampus and is involved in motivation and memory. Odor information is stored in long-term memory and has strong connections to emotional memory. This is possibly due to the olfactory system's close anatomical ties to the limbic system and hippocampus, areas of the brain that have long been known to be involved in emotion and place memory, respectively.

Since any one receptor is responsive to various odorants, and there is a great deal of convergence at the level of the olfactory bulb, it may seem strange that human beings are able to distinguish so many different odors. It seems that a highly complex form of processing must be occurring; however, as it can be shown that, while many neurons in the olfactory bulb (and even the pyriform cortex and amygdala) are responsive to many different odors, half the neurons in the orbitofrontal cortex are responsive to only one odor, and the rest to only a few. It has been shown through microelectrode studies that each individual odor gives a particular spatial map of excitation in the olfactory bulb. It is possible that the brain is able to distin-

guish specific odors through spatial encoding, but temporal coding must also be taken into account. Over time, the spatial maps change, even for one particular odor, and the brain must be able to process these details as well.

Inputs from the two nostrils have separate inputs to the brain, with the result that, when each nostril takes up a different odorant, a person may experience perceptual rivalry in the olfactory sense akin to that of binocular rivalry.

In insects, smells are sensed by sensilla located on the antenna and maxillary palp and first processed by the antennal lobe (analogous to the olfactory bulb), and next by the mushroom bodies and lateral horn.

Many animals, including most mammals and reptiles, but not humans, have two distinct and segregated olfactory systems: a main olfactory system, which detects volatile stimuli, and an accessory olfactory system, which detects fluid-phase stimuli. Behavioral evidence suggests that these fluid-phase stimuli often function as pheromones, although pheromones can also be detected by the main olfactory system. In the accessory olfactory system, stimuli are detected by the vomeronasal organ, located in the vomer, between the nose and the mouth. Snakes use it to smell prey, sticking their tongue out and touching it to the organ.

The sensory receptors of the accessory olfactory system are located in the vomeronasal organ. As in the main olfactory system, the axons of these sensory neurons project from the vomeronasal organ to the accessory olfactory bulb, which in the mouse is located on the dorsal-posterior portion of the main olfactory bulb. Unlike in the main olfactory system, the axons that leave the accessory olfactory bulb do not project to the brain's cortex but rather to targets in the amygdala and bed nucleus of the stria terminalis, and from there to the hypothalamus, where they may influence aggression and mating behavior.

The process by which olfactory information is coded in the brain to allow for proper perception is still being researched, and is not completely understood. When an odorant is detected by receptors, they in a sense break the odorant down, and then the brain puts the odorant back together for identification and perception. The odorant binds to receptors that recognize only a specific functional group, or feature, of the odorant, which is why the chemical nature of the odorant is important.

After binding the odorant, the receptor is activated and will send a signal to the glomeruli. Each glomerulus receives signals from multiple receptors that detect similar odorant features. Because several receptor types are activated due to the different chemical features of the odorant, several glomeruli are activated as well. All of the signals

from the glomeruli are then sent to the brain, where the combination of glomeruli activation encodes the different chemical features of the odorant. The brain then essentially puts the pieces of the activation pattern back together in order to identify and perceive the odorant. This distributed code allows the brain to detect specific odors in mixtures of many background odors.

Different people smell different odors, and most of these differences are caused by genetic differences. Although odorant receptor genes make up one of the largest gene families in the human genome, only a handful of genes have been linked conclusively to particular smells. For instance, the odorant receptor OR5A1 and its genetic variants (alleles) are responsible for our ability (or failure) to smell β -ionone, a key aroma in foods and beverages. Similarly, the odorant receptor OR2J3 is associated with the ability to detect the "grassy" odor, cis-3-hexen-1-ol. The preference (or dislike) of cilantro (coriander) has been linked to the olfactory receptor OR6A2.

26.8 The Gustatory System

Taste, gustatory perception, or gustation is one of the five traditional senses that belongs to the gustatory system.

Chemicals that stimulate taste receptor cells are known as tastants. The tongue is equipped with many taste buds on its dorsal surface, and each taste bud is equipped with taste receptor cells that can sense particular classes of tastes. Distinct types of taste receptor cells respectively detect substances that are sweet, bitter, salty, sour, spicy, or taste of umami.

Taste, along with smell (olfaction) and trigeminal nerve stimulation (registering texture, pain, and temperature), determines flavors of food and/or other substances. Humans have taste receptors on taste buds (gustatory calyculi) and other areas including the upper surface of the tongue and the epiglottis. The gustatory cortex is responsible for the perception of taste.

26.8.1 The Tongue

The tongue is a muscular organ in the mouth of most vertebrates that manipulates food for mastication and is used in the act of swallowing. It has importance in the digestive system and is the primary organ of taste in the gustatory system. The tongue's upper surface (dorsum) is covered by taste buds housed in numerous lingual papillae. It is sensitive and kept moist by saliva and is richly supplied with nerves and blood vessels. The tongue also serves as a natural means of cleaning the teeth. A major function of the tongue is the enabling of speech in humans and vocalization in other animals.

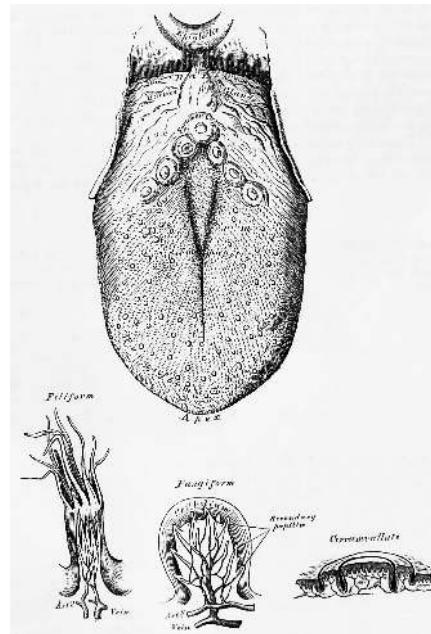


Figure 26.19: The human tongue and three kinds of papillae magnified. Bottom left to right: filiform, fungiform and circumvallate papillae. From Gray, Henry: Anatomy Descriptive And Surgical. 7th Edition, Longmans, Green, And Co., London, 1875²⁸

Innervation of the tongue consists of motor fibers, special sensory fibers for taste, and general sensory fibers for sensation.

The tongue is covered with thousands of small bumps called papillae, which are visible to the naked eye. Within each papilla are hundreds of taste buds. The exception to this is the filiform papillae that do not contain taste buds. Each taste bud contains 50 to 100 taste receptor cells.

26.8.2 The Five Basic Tastes

The sensation of taste includes five established basic tastes: sweetness, sourness, saltiness, bitterness, and savoriness (also known as savory or umami). Taste buds are able to distinguish between different tastes through detecting interaction with different molecules or ions. Sweet, savoriness, and bitter tastes are triggered by the binding of molecules to G protein-coupled receptors on the cell membranes of taste buds. Saltiness and sourness are perceived when alkali metal or hydrogen ions enter taste buds, respectively.

The basic tastes contribute only partially to the sensation and flavor of food in the mouth—other factors include smell, detected by the olfactory epithelium of the nose; texture, detected through a variety of mechanoreceptors; temperature, detected by thermoreceptors; and

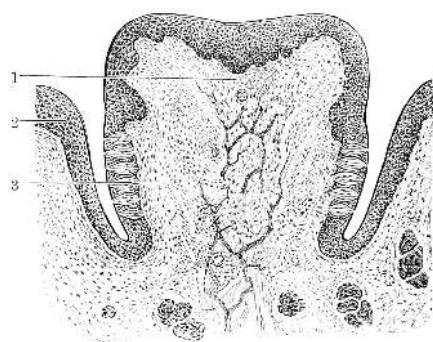


Figure 26.20: Section through a papilla vallata of the human tongue. 1) Papilla. 2) Vallum. 3) Taste buds. From Textbook of anatomy. Section 2. The muscular system: the nervous system: the organs of sense and integument edited by D. J. Cunningham²⁹

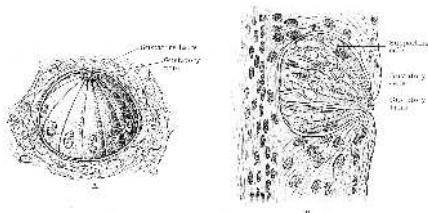


Figure 26.21: A. Three quarter surface view of a taste bud from the papilla foliata of a rabbit (highly magnified). B. Vertical section of taste bud from the papilla foliata of a rabbit (highly magnified). From Textbook of anatomy. Section 2. The muscular system: the nervous system: the organs of sense and integument edited by D. J. Cunningham³⁰

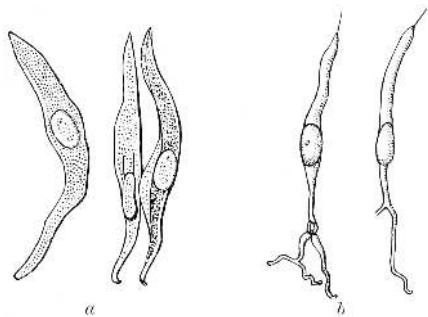


Figure 26.22: Isolated cells from taste bud of a rabbit. a, Supporting cells. b, Gustatory cells. From Textbook of anatomy. Section 2. The muscular system: the nervous system: the organs of sense and integument edited by D. J. Cunningham³¹

“coolness” (such as of menthol) and “hotness” (pungency), through chemesthesis³².

As taste senses both harmful and beneficial things, all basic tastes are classified as either aversive or appetitive, depending upon the effect the things they sense have on our bodies. Sweetness helps to identify energy-rich foods, while bitterness serves as a warning sign of poisons.

Among humans, taste perception begins to fade around 50 years of age because of loss of tongue papillae and a general decrease in saliva production. Humans can also have distortion of tastes through dysgeusia. Not all mammals share the same taste senses: some rodents can taste starch (which humans cannot), cats cannot taste sweetness, and several other carnivores including hyenas, dolphins, and sea lions, have lost the ability to sense up to four of their ancestral five taste senses.

26.8.3 Sweetness

Sweetness, usually regarded as a pleasurable sensation, is produced by the presence of sugars and substances that mimic sugar. Sweetness may be connected to aldehydes and ketones, which contain a carbonyl group. Sweetness is detected by a variety of G protein coupled receptors (GPCR) coupled to the G protein gustducin found on the taste buds. At least two different variants of the “sweetness receptors” must be activated for the brain to register sweetness. Compounds the brain senses as sweet are compounds that can bind with varying bond strength to two different sweetness receptors. These receptors are T1R2+3 (heterodimer) and T1R3 (homodimer), which account for all sweet sensing in humans and animals. Taste detection thresholds for sweet substances are rated relative to sucrose, which has an index of 1. The average human detection threshold for sucrose is 10 millimoles per liter. For lactose it is 30 millimoles per liter, with a sweetness index of 0.3, and 5-nitro-2-propoxyaniline 0.002 millimoles per liter. “Natural” sweeteners such as saccharides activate the GPCR, which releases gustducin. The gustducin then activates the molecule adenylate cyclase, which catalyzes the production of the molecule cAMP, or adenosine 3', 5'-cyclic monophosphate. This molecule closes potassium ion channels, leading to depolarization and neurotransmitter release. Synthetic sweeteners such as saccharin activate different GPCRs and induce taste receptor cell depolarization by an alternate pathway.

26.8.4 Sourness

Sourness is the taste that detects acidity. The sourness of substances is rated relative to dilute hydrochloric acid, which has a sourness index of 1. By comparison, tartaric

³²<https://en.wikipedia.org/wiki/Chemesthesis>

acid has a sourness index of 0.7, citric acid an index of 0.46, and carbonic acid an index of 0.06.

Sour taste is detected by a small subset of cells that are distributed across all taste buds in the tongue. Sour taste cells can be identified by expression of the protein PKD2L1, although this gene is not required for sour responses. There is evidence that the protons that are abundant in sour substances can directly enter the sour taste cells through apically located ion channels. In 2018, the proton-elective ion channel otopetrin 1 (Otop1) was implicated as the primary mediator of this proton influx. This transfer of positive charge into the cell can itself trigger an electrical response. It has also been proposed that weak acids such as acetic acid, which is not fully dissociated at physiological pH values, can penetrate taste cells and thereby elicit an electrical response. According to this mechanism, intracellular hydrogen ions inhibit potassium channels, which normally function to hyperpolarize the cell. By a combination of direct intake of hydrogen ions (which itself depolarizes the cell) and the inhibition of the hyperpolarizing channel, sourness causes the taste cell to fire action potentials and release neurotransmitter.

26.8.5 Saltiness

The simplest receptor found in the mouth is the sodium chloride (salt) receptor. Saltiness is a taste produced primarily by the presence of sodium ions. Other ions of the alkali metals group also taste salty, but the further from sodium, the less salty the sensation is. A sodium channel in the taste cell membrane allows sodium cations to enter the cell. This on its own depolarizes the cell, and opens voltage-dependent calcium channels, flooding the cell with positive calcium ions and leading to neurotransmitter release. This sodium channel is known as an epithelial sodium channel (ENaC) and is composed of three subunits. An ENaC can be blocked by the drug amiloride in many mammals, especially rats. The sensitivity of the salt taste to amiloride in humans, however, is much less pronounced, leading to conjecture that there may be additional receptor proteins besides ENaC to be discovered.

The size of lithium and potassium ions most closely resemble those of sodium, and thus the saltiness is most similar. In contrast, rubidium and caesium ions are far larger, so their salty taste differs accordingly. The saltiness of substances is rated relative to sodium chloride (NaCl), which has an index of 1. Potassium, as potassium chloride (KCl), is the principal ingredient in salt substitutes and has a saltiness index of 0.6.

Other monovalent cations, e.g. ammonium (NH_4^+), and divalent cations of the alkali earth metal group of the periodic table, e.g. calcium (Ca^{2+}), ions generally elicit a bitter

rather than a salty taste even though they, too, can pass directly through ion channels in the tongue, generating an action potential.

26.8.6 Bitterness

Bitterness is one of the most sensitive of the tastes, and many perceive it as unpleasant, sharp, or disagreeable, but it is sometimes desirable and intentionally added via various bittering agents. Common bitter foods and beverages include coffee, unsweetened cocoa, South American mate, coca tea, bitter gourd, uncured olives, citrus peel, many plants in the family *Brassicaceae*, dandelion greens, horehound, wild chicory, and escarole. The ethanol in alcoholic beverages tastes bitter, as do the additional bitter ingredients found in some alcoholic beverages including hops in beer and *Gentiana* in bitters. Quinine is also known for its bitter taste and is found in tonic water.

Bitterness is of interest to those who study evolution, as well as various health researchers since a large number of natural bitter compounds are known to be toxic. The ability to detect bitter-tasting, toxic compounds at low thresholds is considered to provide an important protective function. Plant leaves often contain toxic compounds, and among leaf-eating primates there is a tendency to prefer immature leaves, which tend to be higher in protein and lower in fiber and poisons than mature leaves. Amongst humans, various food processing techniques are used worldwide to detoxify otherwise inedible foods and make them palatable. Furthermore, the use of fire, changes in diet, and avoidance of toxins has led to neutral evolution in human bitter sensitivity. This has allowed several loss of function mutations that has led to a reduced sensory capacity towards bitterness in humans when compared to other species.

The threshold for stimulation of bitter taste by quinine averages a concentration of 8 μM (8 micromolar). The taste thresholds of other bitter substances are rated relative to quinine, which is thus given a reference index of 1. For example, brucine has an index of 11, is thus perceived as intensely more bitter than quinine, and is detected at a much lower solution threshold. The most bitter natural substance is amarogentin a compound present in the roots of the plant *Gentiana lutea* and the most bitter substance known is the synthetic chemical denatonium, which has an index of 1,000. It is used as an aversive agent (a bitterant) that is added to toxic substances to prevent accidental ingestion.

Research has shown that TAS2Rs (taste receptors, type 2, also known as T2Rs) such as TAS2R38 coupled to the G protein gustducin are responsible for the human ability to taste bitter substances. The TAS2R family in humans is thought to comprise about 25 different taste recep-

tors, some of which can recognize a wide variety of bitter-tasting compounds. Over 670 bitter-tasting compounds have been identified, on a bitter database, of which over 200 have been assigned to one or more specific receptors. Researchers use two synthetic substances, phenylthiocarbamide (PTC) and 6-n-propylthiouracil (PROP) to study the genetics of bitter perception. These two substances taste bitter to some people, but are virtually tasteless to others. Among the tasters, some are so-called "supertasters" to whom PTC and PROP are extremely bitter. The variation in sensitivity is determined by two common alleles at the TAS2R38 locus.

26.8.7 Savoriness (Umami)

Savory, or savoriness is an appetitive taste and is occasionally described by its Japanese name, umami or "meaty".

26.8.8 The Taste Receptors

There are four types taste receptors. When food or other substances enter the mouth, molecules interact with saliva and are bound to taste receptors in the oral cavity and other locations. Molecules which give a sensation of taste are considered "sapid".

Taste receptors are divided into two families:

- Type 1, sweet: TAS1R2+TAS1R3; umami: TAS1R1+TAS1R3
- Type 2, bitter: TAS2R

The standard bitter, sweet, or umami taste receptors are G protein-coupled receptors with seven transmembrane domains. Ligand binding at the taste receptors activate second messenger cascades to depolarize the taste cell. Gustducin is the most common taste $G\alpha$ subunit, having a major role in TAS2R bitter taste reception. Gustducin is a homologue for transducin, a G-protein involved in vision transduction. Additionally, taste receptors share the use of the TRPM5 ion channel, as well as a phospholipase PLC β 2.

The TAS1R1+TAS1R3 heterodimer receptor functions as an umami receptor, responding to L-amino acid binding, especially L-glutamate. The umami taste is most frequently associated with the food additive monosodium glutamate (MSG) and can be enhanced through the binding of inosine monophosphate (IMP) and guanosine monophosphate (GMP) molecules. TAS1R1+3 expressing cells are found mostly in the fungiform papillae at the tip and edges of the tongue and palate taste receptor cells in the roof of the mouth. These cells are shown to synapse upon the chorda tympani nerves to send their signals to the brain, although some activation of the glossopharyngeal nerve has been found.

Alternative candidate umami taste receptors include splice variants of metabotropic glutamate receptors, mGluR4 and mGluR1, and the N-methyl-D-aspartate type glutamate ion channel receptor.

The TAS1R2+TAS1R3 heterodimer receptor functions as the sweet receptor by binding to a wide variety of sugars and sugar substitutes. TAS1R2+3 expressing cells are found in circumvallate papillae and foliate papillae near the back of the tongue and palate taste receptor cells in the roof of the mouth. These cells are shown to synapse upon the chorda tympani and glossopharyngeal nerves to send their signals to the brain. The TAS1R3 homodimer also functions as a sweet receptor in much the same way as TAS1R2+3 but has decreased sensitivity to sweet substances. Natural sugars are more easily detected by the TAS1R3 receptor than sugar substitutes. This may help explain why sugar and artificial sweeteners have different tastes. Genetic polymorphisms in TAS1R3 partly explain the difference in sweet taste perception and sugar consumption between people of African American ancestry and people of European and Asian ancestries.

The TAS2R proteins function as bitter taste receptors. There are 43 human TAS2R genes, each of which (excluding the five pseudogenes) lacks introns and codes for a GPCR protein. These proteins, as opposed to TAS1R proteins, have short extracellular domains and are located in circumvallate papillae, palate, foliate papillae, and epiglottis taste buds, with reduced expression in fungiform papillae. Though it is certain that multiple TAS2Rs are expressed in one taste receptor cell, it is still debated whether mammals can distinguish between the tastes of different bitter ligands. Some overlap must occur, however, as there are far more bitter compounds than there are TAS2R genes. Common bitter ligands include cycloheximide, denatonium, PROP (6-n-propyl-2-thiouracil), PTC (phenylthiocarbamide), and β -glucopyranosides.

Signal transduction of bitter stimuli is accomplished via the α -subunit of gustducin. This G protein subunit activates a taste phosphodiesterase and decreases cyclic nucleotide levels. Further steps in the transduction pathway are still unknown. The $\beta\gamma$ -subunit of gustducin also mediates taste by activating IP3 (inositol triphosphate) and DAG (diglyceride). These second messengers may open gated ion channels or may cause release of internal calcium. Though all TAS2Rs are located in gustducin-containing cells, knockout of gustducin does not completely abolish sensitivity to bitter compounds, suggesting a redundant mechanism for bitter tasting (unsurprising given that a bitter taste generally signals the presence of a toxin). One proposed mechanism for gustducin-independent bitter tasting is via ion channel interaction by specific bitter ligands, similar to the ion

channel interaction which occurs in the tasting of sour and salty stimuli.

One of the best-researched TAS2R proteins is TAS2R38, which contributes to the tasting of both PROP and PTC. It is the first taste receptor whose polymorphisms are shown to be responsible for differences in taste perception. Current studies are focused on determining other such taste phenotype-determining polymorphisms. More recent studies show that genetic polymorphisms in other bitter taste receptor genes influence bitter taste perception of caffeine, quinine and denatonium benzoate.

Historically it was thought that the sour taste was produced solely when free hydrogen ions (H^+) directly depolarised taste receptors. However, specific receptors for sour taste with other methods of action are now being proposed. HCN1 and HCN4 (HCN channels) were two such proposals; both of these receptors are cyclic nucleotide-gated channels. The two ion channels suggested to contribute to sour taste are ACCN1 and TASK-1.

Various receptors have also been proposed for salty tastes, along with the possible taste detection of lipids, complex carbohydrates, and water. Evidence for these receptors is, however, shaky at best, and is often unconvincing in mammal studies. For example, the proposed ENaC receptor for sodium detection can only be shown to contribute to sodium taste in *Drosophila*.

Visual, olfactory, "sapictive" (the perception of tastes), trigeminal (hot, cool), mechanical, all contribute to the perception of taste. Of these, transient receptor potential cation channel subfamily V member 1 (TRPV1) vanilloid receptors are responsible for the perception of heat from some molecules such as capsaicin, and a CMR1 receptor is responsible for the perception of cold from molecules such as menthol, eucalyptol, and icilin.

26.8.9 The Gustatory Cortex

The primary gustatory cortex is a brain structure responsible for the perception of taste. It consists of two substructures: the anterior insula on the insular lobe and the frontal operculum on the inferior frontal gyrus of the frontal lobe. Because of its composition the primary gustatory cortex is sometimes referred to in literature as the AI/FO(Anterior Insula/Frontal Operculum). By using extracellular unit recording techniques, scientists have elucidated that neurons in the AI/FO respond to sweetness, saltiness, bitterness, and sourness, and they code the intensity of the taste stimulus.

Like the olfactory system, the taste system is defined by its specialized peripheral receptors and central pathways that relay and process taste information. Peripheral taste receptors are found on the upper surface of the tongue,

soft palate, pharynx, and the upper part of the esophagus. Taste cells synapse with primary sensory axons that run in the chorda tympani and greater superficial petrosal branches of the facial nerve (cranial nerve VII), the lingual branch of the glossopharyngeal nerve (cranial nerve IX), and the superior laryngeal branch of the vagus nerve (Cranial nerve X) to innervate the taste buds in the tongue, palate, epiglottis, and esophagus respectively. The central axons of these primary sensory neurons in the respective cranial nerve ganglia project to rostral and lateral regions of the nucleus of the solitary tract in the medulla, which is also known as the gustatory nucleus of the solitary tract complex. Axons from the rostral (gustatory) part of the solitary nucleus project to the ventral posterior complex of the thalamus, where they terminate in the medial half of the ventral posterior medial nucleus. This nucleus projects in turn to several regions of the neocortex which includes the gustatory cortex (the frontal operculum and the insula), which becomes activated when the subject is consuming and experiencing taste.

Chapter 27

Animal Locomotion And Support Systems

Animal locomotion is any of a variety of methods that animals use to move from one place to another. Some modes of locomotion are (initially) self-propelled, e.g., running, swimming, jumping, flying, hopping, soaring and gliding. There are also many animal species that depend on their environment for transportation, a type of mobility called passive locomotion, e.g., sailing (some jellyfish), kiting (spiders), rolling (some beetles and spiders) or riding other animals (phoresis).

The term “locomotion” is formed in English from Latin *loco* “from a place” (ablative of *locus* “place”) + *motio* “motion, a moving”.

Animals move for a variety of reasons, such as to find food, a mate, a suitable microhabitat, or to escape predators. For many animals, the ability to move is essential for survival and, as a result, natural selection has shaped the locomotion methods and mechanisms used by moving organisms. For example, migratory animals that travel vast distances (such as the Arctic tern) typically have a locomotion mechanism that costs very little energy per unit distance, whereas non-migratory animals that must frequently move quickly to escape predators are likely to have energetically costly, but very fast, locomotion.

The anatomical structures that animals use for movement, including cilia, legs, wings, arms, fins, or tails are sometimes referred to as locomotory organs or locomotory structures.

Animals move through, or on, four types of environment: aquatic (in or on water), terrestrial (on ground or other surface, including arboreal, or tree-dwelling), fossorial (underground), and aerial (in the air). Many animals—for example semi-aquatic animals, and diving birds—regularly move through more than one type of medium. In some cases, the surface they move on facilitates their method of locomotion.

27.1 The Skeleton

The skeleton is the body part that provides support, shape and protection to the soft tissues and delicate organs of animals. There are several different skeletal types: the exoskeleton, which is the stable outer shell of an organism, the endoskeleton, which forms the support structure inside the body, the hydroskeleton, a flexible skeleton supported by fluid pressure, and the cytoskeleton present in the cytoplasm of all cells, including bacteria, and archaea. The term comes from Greek σκελετός (*skeletós*), meaning ‘dried up’.

There are two major types of skeletons: solid and fluid. Solid skeletons can be internal, called an endoskeleton, or external, called an exoskeleton, and may be further classified as pliant (elastic/movable) or rigid (hard/non-movable). Fluid skeletons are always internal.

27.1.1 Exoskeleton

Exoskeletons are external, and are found in many invertebrates; they enclose and protect the soft tissues and organs of the body. Some kinds of exoskeletons undergo periodic moulting or ecdysis as the animal grows, as is the case in many arthropods including insects and crustaceans.

The exoskeleton of insects is not only a form of protection, but also serves as a surface for muscle attachment, as a watertight protection against drying, and as a sense organ to interact with the environment. The shell of mollusks also performs all of the same functions, except that in most cases it does not contain sense organs.

An external skeleton can be quite heavy in relation to the overall mass of an animal, so on land, organisms that have an exoskeleton are mostly relatively small. Somewhat larger aquatic animals can support an exoskeleton because weight is less of a consideration underwater. The southern giant clam, a species of extremely large saltwater clam in the Pacific Ocean, has a shell that is massive in both size

and weight. *Syrinx aruanus* is a species of sea snail with a very large shell.

27.1.2 Endoskeleton

The endoskeleton is the internal support structure of an animal, composed of mineralized tissue and is typical of vertebrates. Endoskeletons vary in complexity from functioning purely for support (as in the case of sponges), to serving as an attachment site for muscles and a mechanism for transmitting muscular forces. A true endoskeleton is derived from mesodermal tissue. Such a skeleton is present in echinoderms and chordates.

Pliant skeletons are capable of movement; thus, when stress is applied to the skeletal structure, it deforms and then reverts to its original shape. This skeletal structure is used in some invertebrates, for instance in the hinge of bivalve shells or the mesoglea of cnidarians such as jellyfish. Pliant skeletons are beneficial because only muscle contractions are needed to bend the skeleton; upon muscle relaxation, the skeleton will return to its original shape. Cartilage is one material that a pliant skeleton may be composed of, but most pliant skeletons are formed from a mixture of proteins, polysaccharides, and water. For additional structure or protection, pliant skeletons may be supported by rigid skeletons. Organisms that have pliant skeletons typically live in water, which supports body structure in the absence of a rigid skeleton.

Rigid skeletons are not capable of movement when stressed, creating a strong support system most common in terrestrial animals. Such a skeleton type used by animals that live in water are more for protection (such as barnacle and snail shells) or for fast-moving animals that require additional support of musculature needed for swimming through water. Rigid skeletons are formed from materials including chitin (in arthropods), calcium compounds such as calcium carbonate (in stony corals and mollusks) and silicate (for diatoms and radiolarians).

27.1.3 Hydrostatic Skeleton (Hydroskeleton)

A hydrostatic skeleton is a semi-rigid, soft tissue structure filled with liquid under pressure, surrounded by muscles. Longitudinal and circular muscles around their body sectors allow movement by alternate lengthening and contractions along their lengths. A common example of this is the earthworm.

27.1.4 Organisms With Skeletons

The endoskeletons of echinoderms and some other soft-bodied invertebrates such as jellyfish and earthworms are also termed hydrostatic; a body cavity the coelom is filled



Figure 27.1: Vertebrate skulls and skeletons¹: 1. *Homo sapiens*², 2. Human skull³, 3. *Australopithecus*⁴, 4. *Homo neanderthalensis*⁵, 5. *Pan troglodytes*⁶, 6. *Papio hamadryas*⁷, 7. *Rhinopithecus roxellana*⁸, 8. *Gorilla*⁹, 9. *Sus scrofa*¹⁰, 10. *Cattle*¹¹, 11. *Panthera leo*¹², 12. *Canis lupus*¹³, 13. *Equus caballus*¹⁴, 14. *Elephantidae*¹⁵, 15. *Capra*¹⁶, 16. *Hippopotamus*¹⁷, 17. *Camelus*¹⁸, 18. *Macropus*¹⁹, 19. *Damaliscus korrigum*²⁰, 20. *Odobenus rosmarus*²¹, 21. *Bat*²², 22. *Cetacea*²³, 23. *Accipitridae*²⁴, 24. *Psittacidae*²⁵, 25. *Gallus gallus*²⁶, 26. *Roosters*²⁷, 27. *Ramphastos sulfuratus*²⁸, 28. *Casuariidae*²⁹, 29. *Spheniscidae*³⁰, 30. *Gruidae*³¹, 32. *Sharovipteryx mirabilis*³², 32. *Natrix natrix*³³, 33. *Crotalinae*³⁴, 34. *Boa constrictor*³⁵, 35. *Crocodile*³⁶, 36. *Lizard*³⁷, 37. *Testudines*³⁸, 38. *Frog*³⁹, 39. *Salamandra salamandra*⁴⁰, 40. *Perca fluviatilis*⁴¹, 41. *Acipenser*⁴², 42. *Balistoides viridescens*⁴³, 43. *Rajidae*⁴⁴, 44. *Polypterus*⁴⁵

with coelomic fluid and the pressure from this fluid acts together with the surrounding muscles to change the organism's shape and produce movement.

The skeleton of sponges consists of microscopic calcareous or silicious spicules. The demosponges include 90% of all species of sponges. Their "skeletons" are made of spicules consisting of fibers of the protein spongin, the mineral silica, or both. Where spicules of silica are present, they have a different shape from those in the otherwise similar glass sponges.

The skeleton of the echinoderms, which include, among other things, the starfish, is composed of calcite and a small amount of magnesium oxide. It lies below the epidermis in the mesoderm and is within cell clusters of frame-forming cells. This structure formed is porous and therefore firm and at the same time light. It coalesces into small calcareous ossicles (bony plates), which can grow in all directions and thus can replace the loss of a body part. Connected by joints, the individual skeletal parts can be moved by the muscles.

In most vertebrates, the main skeletal component is referred to as bone. These bones compose a unique skeletal system for each type of animal. Another important component is cartilage which in mammals is found mainly in the joint areas. In other animals, such as the cartilaginous fishes, which include the sharks, the skeleton is composed entirely of cartilage. The segmental pattern of the skeleton is present in all vertebrates (mammals, birds, fish, reptiles and amphibians) with basic units being repeated. This segmental pattern is particularly evident in the vertebral column and the ribcage.

Bones in addition to supporting the body also serve, at the cellular level, as calcium and phosphate storage.

The skeleton, which forms the support structure inside the fish is either made of cartilage as in the (Chondrichthyes), or bones as in the (Osteichthyes). The main skeletal element is the vertebral column, composed of articulating vertebrae which are lightweight yet strong. The ribs attach to the spine and there are no limbs or limb girdles. They are supported only by the muscles. The main external features of the fish, the fins, are composed of either bony or soft spines called rays, which with the exception of the caudal fin (tail fin), have no direct connection with the spine. They are supported by the muscles which compose the main part of the trunk.

The bird skeleton is highly adapted for flight. It is extremely lightweight, yet still strong enough to withstand the stresses of taking off, flying, and landing. One key adaptation is the fusing of bones into single ossifications, such as the pygostyle. Because of this, birds usually have a smaller number of bones than other terrestrial vertebrates.

Birds also lack teeth or even a true jaw, instead having evolved a beak, which is far more lightweight. The beaks of many baby birds have a projection called an egg tooth, which facilitates their exit from the amniotic egg.

27.2 The Human Musculoskeletal System

The human musculoskeletal system (also known as the locomotor system, and previously the activity system) is an organ system that gives humans the ability to move using their muscular and skeletal systems. The musculoskeletal system provides form, support, stability, and movement to the body.

It is made up of the bones of the skeleton, muscles, cartilage, tendons, ligaments, joints, and other connective tissue that supports and binds tissues and organs together. The musculoskeletal system's primary functions include supporting the body, allowing motion, and protecting vital organs. The skeletal portion of the system serves as the main storage system for calcium and phosphorus and contains critical components of the hematopoietic system.

This system describes how bones are connected to other bones and muscle fibers via connective tissue such as tendons and ligaments. The bones provide stability to the body. Muscles keep bones in place and also play a role in the movement of bones. To allow motion, different bones are connected by joints. Cartilage prevents the bone ends from rubbing directly onto each other. Muscles contract to move the bone attached at the joint.

There are, however, diseases and disorders that may adversely affect the function and overall effectiveness of the system. These diseases can be difficult to diagnose due to the close relation of the musculoskeletal system to other internal systems. The musculoskeletal system refers to the system having its muscles attached to an internal skeletal system and is necessary for humans to move to a more favorable position. Complex issues and injuries involving the musculoskeletal system are usually handled by a physiatrist (specialist in physical medicine and rehabilitation) or an orthopaedic surgeon.

The skeletal system serves many important functions; it provides the shape and form for the body, support and protection, allows bodily movement, produces blood for the body, and stores minerals. The number of bones in the human skeletal system is a controversial topic. Humans are born with over 300 bones; however, many bones fuse together between birth and maturity. As a result, an average adult skeleton consists of 206 bones. The number of bones varies according to the method used to derive the count. While some consider certain structures to be a sin-

gle bone with multiple parts, others may see it as a single part with multiple bones. There are five general classifications of bones. These are long bones, short bones, flat bones, irregular bones, and sesamoid bones. The human skeleton is composed of both fused and individual bones supported by ligaments, tendons, muscles and cartilage. It is a complex structure with two distinct divisions; the axial skeleton, which includes the vertebral column, and the appendicular skeleton.

The skeletal system serves as a framework for tissues and organs to attach themselves to. This system acts as a protective structure for vital organs. Major examples of this are the brain being protected by the skull and the lungs being protected by the rib cage.

Located in long bones are two distinctions of bone marrow (yellow and red). The yellow marrow has fatty connective tissue and is found in the marrow cavity. During starvation, the body uses the fat in yellow marrow for energy. The red marrow of some bones is an important site for blood cell production, approximately 2.6 million red blood cells per second in order to replace existing cells that have been destroyed by the liver. Here all erythrocytes, platelets, and most leukocytes form in adults. From the red marrow, erythrocytes, platelets, and leukocytes migrate to the blood to do their special tasks.

Another function of bones is the storage of certain minerals. Calcium and phosphorus are among the main minerals being stored. The importance of this storage "device" helps to regulate mineral balance in the bloodstream. When the fluctuation of minerals is high, these minerals are stored in bone; when it is low it will be withdrawn from the bone.

There are three types of muscles—cardiac, skeletal, and smooth. Smooth muscles are used to control the flow of substances within the lumens of hollow organs, and are not consciously controlled. Skeletal and cardiac muscles have striations that are visible under a microscope due to the components within their cells. Only skeletal and smooth muscles are part of the musculoskeletal system and only the skeletal muscles can move the body. Cardiac muscles are found in the heart and are used only to circulate blood; like the smooth muscles, these muscles are not under conscious control. Skeletal muscles are attached to bones and arranged in opposing groups around joints. Muscles are innervated, to communicate nervous energy to, by nerves, which conduct electrical currents from the central nervous system and cause the muscles to contract.

In mammals, when a muscle contracts, a series of reactions occur. Muscle contraction is stimulated by the motor neuron sending a message to the muscles from the somatic

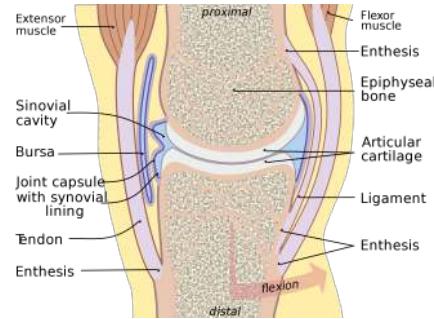


Figure 27.2: A synovial joint.⁴⁶

nervous system. Depolarization of the motor neuron results in neurotransmitters being released from the nerve terminal. The space between the nerve terminal and the muscle cell is called the neuromuscular junction. These neurotransmitters diffuse across the synapse and bind to specific receptor sites on the cell membrane of the muscle fiber. When enough receptors are stimulated, an action potential is generated and the permeability of the sarcolemma is altered. This process is known as initiation.

27.2.1 Tendons

A tendon is a tough, flexible band of fibrous connective tissue that connects muscles to bones. The extra-cellular connective tissue between muscle fibers binds to tendons at the distal and proximal ends, and the tendon binds to the periosteum of individual bones at the muscle's origin and insertion. As muscles contract, tendons transmit the forces to the relatively rigid bones, pulling on them and causing movement. Tendons can stretch substantially, allowing them to function as springs during locomotion, thereby saving energy.

27.2.2 Joints, Ligaments And Bursae

Joints are structures that connect individual bones and may allow bones to move against each other to cause movement. There are three divisions of joints, diarthroses which allow extensive mobility between two or more articular heads; amphiarthrosis, which is a joint that allows some movement, and false joints or synarthroses, joints that are immovable, that allow little or no movement and are predominantly fibrous. Synovial joints, joints that are not directly joined, are lubricated by a solution called synovial fluid that is produced by the synovial membranes. This fluid lowers the friction between the articular surfaces and is kept within an articular capsule, binding the joint with its taut tissue.

A ligament is a small band of dense, white, fibrous elastic tissue. Ligaments connect the ends of bones together in

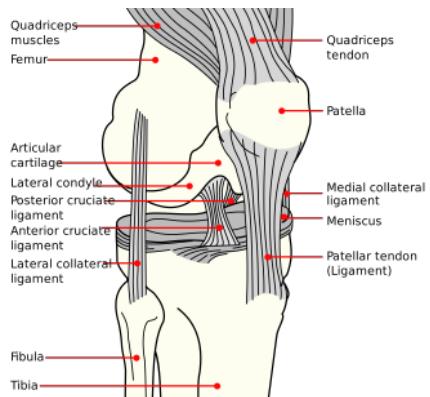


Figure 27.3: Diagram of the right knee. Anterior cruciate ligament labeled at center left.⁴⁷

order to form a joint. Most ligaments limit dislocation, or prevent certain movements that may cause breaks. Since they are only elastic they increasingly lengthen when under pressure. When this occurs the ligament may be susceptible to break resulting in an unstable joint.

Ligaments may also restrict some actions: movements such as hyper extension and hyper flexion are restricted by ligaments to an extent. Also ligaments prevent certain directional movement.

A bursa is a small fluid-filled sac made of white fibrous tissue and lined with synovial membrane. Bursa may also be formed by a synovial membrane that extends outside of the joint capsule. It provides a cushion between bones and tendons or muscles around a joint; bursa are filled with synovial fluid and are found around almost every major joint of the body.

27.2.3 The Human Skeleton

The human skeleton consists of both fused and individual bones supported and supplemented by ligaments, tendons, muscles and cartilage. It is composed of around 270 bones at birth – this total decreases to around 206 bones by adulthood after some bones get fused together. The biggest bone in the body is the femur in the upper leg, and the smallest is the stapes bone in the middle ear. In an adult, the skeleton comprises around 14% of the total body weight, and half of this weight is water. The bone mass in the skeleton reaches maximum density around age 21. The human skeleton can be divided into the axial skeleton and the appendicular skeleton. The axial skeleton is formed by the vertebral column, the rib cage, the skull and other associated bones. The appendicular skeleton, which is attached to the axial skeleton, is formed by the shoulder girdle, the pelvic girdle and the bones of the upper and lower limbs.

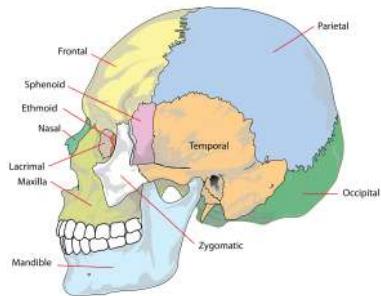


Figure 27.4: Major bones of the human skull viewed from the side.⁴⁸

The human skeleton performs six major functions; support, movement, protection, production of blood cells, storage of minerals, and endocrine regulation. It serves as a scaffold which supports organs, anchors muscles, and protects organs such as the brain, lungs, heart and spinal cord. Although the teeth do not consist of tissue commonly found in bones, the teeth are usually considered as members of the skeletal system.

Fused bones include those of the pelvis and the cranium. Not all bones are interconnected directly: There are three bones in each middle ear called the ossicles that articulate only with each other. The hyoid bone, which is located in the neck and serves as the point of attachment for the tongue, does not articulate with any other bones in the body, being supported by muscles and ligaments.

There are 206 bones in the adult human skeleton, although this number depends on whether the pelvic bones (the hip bones on each side) are counted as one or three bones on each side (ilium, ischium, and pubis), whether the coccyx or tail bone is counted as one or four separate bones, and does not count the variable wormian bones between skull sutures. Similarly, the sacrum is usually counted as a single bone, rather than five fused vertebrae. There is also a variable number of small sesamoid bones, commonly found in tendons. The patella or kneecap on each side is an example of a larger sesamoid bone. The patellae are counted in the total, as they are constant. The number of bones varies between individuals and with age – newborn babies have over 270 bones some of which fuse together. These bones are organized into a longitudinal axis, the axial skeleton, to which the appendicular skeleton is attached.

The human skeleton takes 20 years before it is fully developed, and the bones contain marrow, which produces blood cells.

There exist several general differences between the male and female skeletons. The male skeleton, for

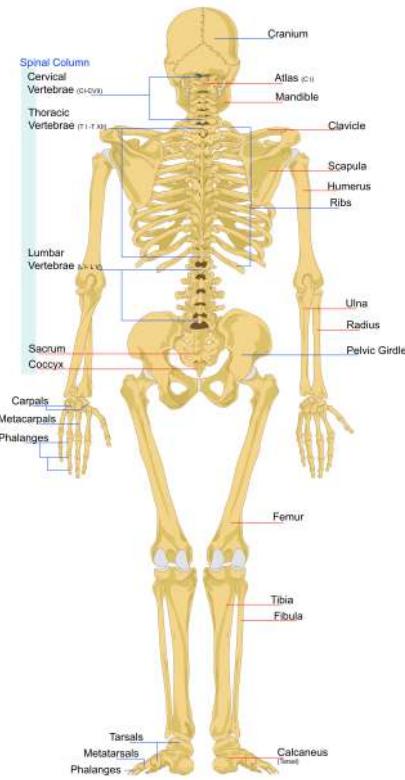
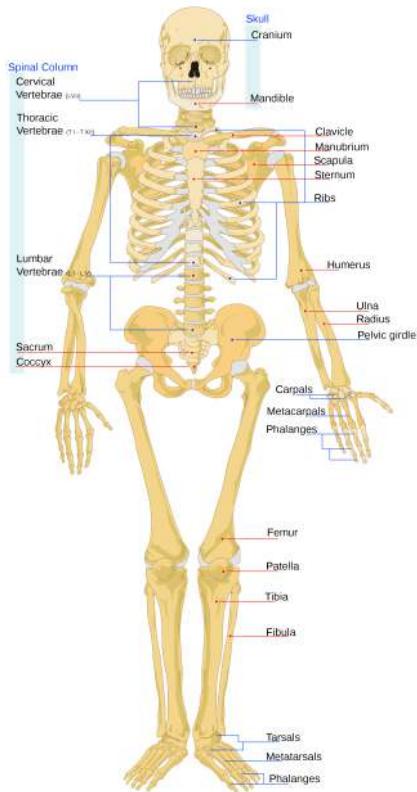


Figure 27.5: The human skeleton viewed from the front.⁴⁹

Figure 27.6: The human skeleton viewed from the back.⁵⁰

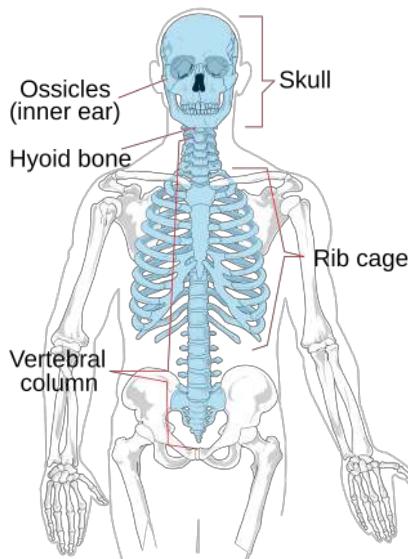


Figure 27.7: The Axial skeleton⁵¹ consists of the bones in the head and trunk of the human body. It is composed of five parts; the human skull, the ossicles of the inner ear, the hyoid bone of the throat, the chest, and the vertebral column. The axial skeleton and the appendicular skeleton together form the complete skeleton.

example, is generally larger and heavier than the female skeleton. In the female skeleton, the bones of the skull are generally less angular. The female skeleton also has wider and shorter breastbone and slimmer wrists. There exist significant differences between the male and female pelvis which are related to the female's pregnancy and childbirth capabilities. The female pelvis is wider and shallower than the male pelvis. Female pelvises also have an enlarged pelvic outlet and a wider and more circular pelvic inlet. The angle between the pubic bones is known to be sharper in males, which results in a more circular, narrower, and near heart-shaped pelvis.

27.2.4 The Axial Skeleton

The axial skeleton (80 bones) is formed by the vertebral column (32 bones), a part of the rib cage (12 pairs of ribs and the sternum), and the skull (22 bones and 7 associated bones).

The upright posture of humans is maintained by the axial skeleton, which transmits the weight from the head, the trunk, and the upper extremities down to the lower extremities at the hip joints. The bones of the spine are supported by many ligaments. The erector spinae muscles are also supporting and are useful for balance.

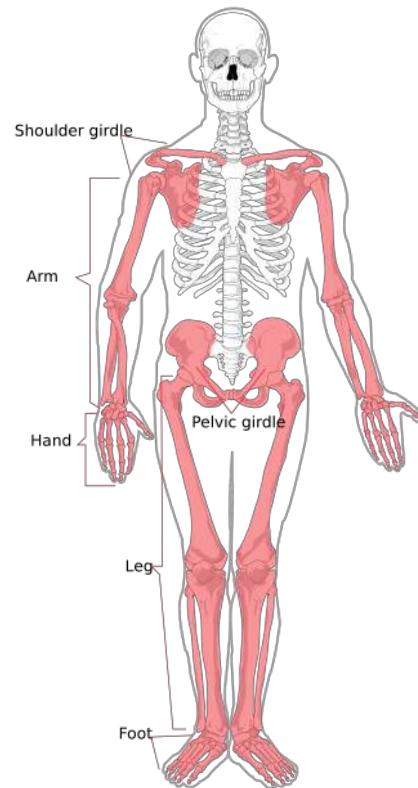


Figure 27.8: The appendicular skeleton (red) and the axial skeleton together form the complete skeleton.⁵²

27.2.5 The Appendicular Skeleton

The appendicular skeleton (126 bones) is formed by the pectoral girdles, the upper limbs, the pelvic girdle or pelvis, and the lower limbs. Their functions are to make locomotion possible and to protect the major organs of digestion, excretion and reproduction.

27.2.6 Bone

A bone is a rigid organ that constitutes part of the vertebrate skeleton in animals. Bones protect the various organs of the body, produce red and white blood cells, store minerals, provide structure and support for the body, and enable mobility. Bones come in a variety of shapes and sizes and have a complex internal and external structure. They are lightweight yet strong and hard, and serve multiple functions.

Bone tissue (osseous tissue) is a hard tissue, a type of dense connective tissue. It has a honeycomb-like matrix internally, which helps to give the bone rigidity. Bone tissue is made up of different types of bone cells. Osteoblasts and osteocytes are involved in the formation and mineralization of bone; osteoclasts are involved in the resorption of bone tissue. Modified (flattened) osteoblasts be-

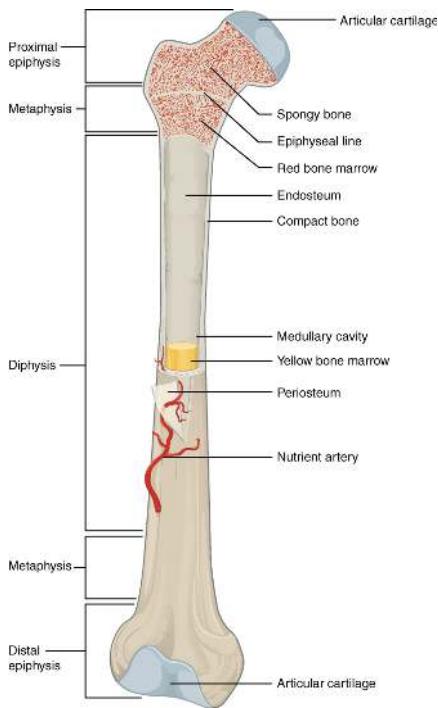


Figure 27.9: Structure of a long bone.⁵³

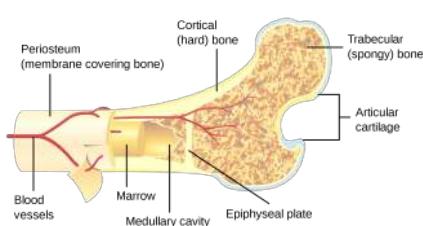


Figure 27.10: Cross-section of human bone.⁵⁴

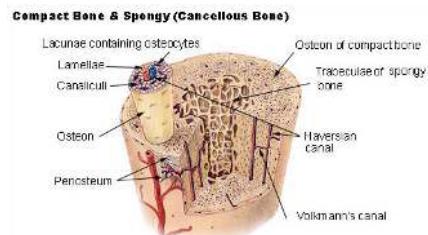


Figure 27.11: A cross section of a long bone, showing the internal structure.⁵⁵

come the lining cells that form a protective layer on the bone surface. The mineralised matrix of bone tissue has an organic component of mainly collagen called ossein and an inorganic component of bone mineral made up of various salts. Bone tissue is a mineralized tissue of two types, cortical bone and cancellous bone. Other types of tissue found in bones include bone marrow, endosteum, periosteum, nerves, blood vessels and cartilage.

The Greek word for bone is ὄστεον (“osteon”), hence the many terms that use it as a prefix—such as osteopathy.

Bone is not uniformly solid, but consists of a flexible matrix (about 30%) and bound minerals (about 70%) which are intricately woven and endlessly remodeled by a group of specialized bone cells. Their unique composition and design allows bones to be relatively hard and strong, while remaining lightweight.

Bone matrix is 90 to 95% composed of elastic collagen fibers, also known as ossein, and the remainder is ground substance. The elasticity of collagen improves fracture resistance. The matrix is hardened by the binding of inorganic mineral salt, calcium phosphate, in a chemical arrangement known as calcium hydroxyapatite. It is the bone mineralization that give bones rigidity.

Bone is actively constructed and remodeled throughout life by special bone cells known as osteoblasts and osteoclasts. Within any single bone, the tissue is woven into two main patterns, known as cortical and cancellous bone, and each with different appearance and characteristics.

27.2.7 Cortical Bone

The hard outer layer of bones is composed of cortical bone, which is also called compact bone as it is much denser than cancellous bone. It forms the hard exterior (cortex) of bones. The cortical bone gives bone its smooth, white, and solid appearance, and accounts for 80% of the total bone mass of an adult human skeleton. It facilitates bone's main functions – to support the whole body, to protect organs, to provide levers for movement, and to store and release chemical elements, mainly calcium. It consists of multi-

ple microscopic columns, each called an osteon or Haversian system. Each column is multiple layers of osteoblasts and osteocytes around a central canal called the haversian canal. Volkmann's canals at right angles connect the osteons together. The columns are metabolically active, and as bone is reabsorbed and created the nature and location of the cells within the osteon will change. Cortical bone is covered by a periosteum on its outer surface, and an endosteum on its inner surface. The endosteum is the boundary between the cortical bone and the cancellous bone. The primary anatomical and functional unit of cortical bone is the osteon.

27.2.8 Cancellous Bone

Cancellous bone, also called trabecular or spongy bone, is the internal tissue of the skeletal bone and is an open cell porous network. Cancellous bone has a higher surface-area-to-volume ratio than cortical bone and it is less dense. This makes it weaker and more flexible. The greater surface area also makes it suitable for metabolic activities such as the exchange of calcium ions. Cancellous bone is typically found at the ends of long bones, near joints and in the interior of vertebrae. Cancellous bone is highly vascular and often contains red bone marrow where hematopoiesis, the production of blood cells, occurs. The primary anatomical and functional unit of cancellous bone is the trabecula. The trabeculae are aligned towards the mechanical load distribution that a bone experiences within long bones such as the femur. As far as short bones are concerned, trabecular alignment has been studied in the vertebral pedicle. Thin formations of osteoblasts covered in endosteum create an irregular network of spaces, known as trabeculae. Within these spaces are bone marrow and hematopoietic stem cells that give rise to platelets, red blood cells and white blood cells. Trabecular marrow is composed of a network of rod- and plate-like elements that make the overall organ lighter and allow room for blood vessels and marrow. Trabecular bone accounts for the remaining 20% of total bone mass but has nearly ten times the surface area of compact bone.

The words cancellous and trabecular refer to the tiny lattice-shaped units (trabeculae) that form the tissue. It was first illustrated accurately in the engravings of Crisóstomo Martínez.

27.2.9 The Bone Marrow

Bone marrow, also known as myeloid tissue in red bone marrow, can be found in almost any bone that holds cancellous tissue. In newborns, all such bones are filled exclusively with red marrow or hematopoietic marrow, but as the child ages the hematopoietic fraction decreases in quantity and the fatty/ yellow fraction called marrow adip-

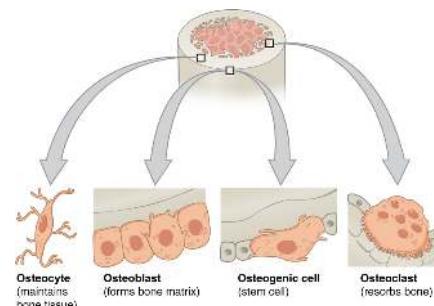


Figure 27.12: Bone cells.⁵⁶

pose tissue (MAT) increases in quantity. In adults, red marrow is mostly found in the bone marrow of the femur, the ribs, the vertebrae and pelvic bones.

27.2.10 Bone cells

Bone is a metabolically active tissue composed of several types of cells. These cells include osteoblasts, which are involved in the creation and mineralization of bone tissue, osteocytes, and osteoclasts, which are involved in the reabsorption of bone tissue. Osteoblasts and osteocytes are derived from osteoprogenitor cells, but osteoclasts are derived from the same cells that differentiate to form macrophages and monocytes. Within the marrow of the bone there are also hematopoietic stem cells. These cells give rise to other cells, including white blood cells, red blood cells, and platelets.

27.2.11 Osteoblast

Osteoblasts are mononucleate bone-forming cells. They are located on the surface of osteon seams and make a protein mixture known as osteoid, which mineralizes to become bone. The osteoid seam is a narrow region of newly formed organic matrix, not yet mineralized, located on the surface of a bone. Osteoid is primarily composed of Type I collagen. Osteoblasts also manufacture hormones, such as prostaglandins, to act on the bone itself. The osteoblast creates and repairs new bone by actually building around itself. First, the osteoblast puts up collagen fibers. These collagen fibers are used as a framework for the osteoblasts' work. The osteoblast then deposits calcium phosphate which is hardened by hydroxide and bicarbonate ions. The brand new bone created by the osteoblast is called osteoid. Once the osteoblast is finished working it is actually trapped inside the bone once it hardens. When the osteoblast becomes trapped, it becomes known as an osteocyte. Other osteoblasts remain on the top of the new bone and are used to protect the underlying bone, these become known as lining cells.

27.2.12 Osteocyte

Osteocytes are mostly inactive osteoblasts. Osteocytes originate from osteoblasts that have migrated into and become trapped and surrounded by bone matrix that they themselves produced. The spaces they occupy are known as lacunae. Osteocytes have many processes that reach out to meet osteoblasts and other osteocytes probably for the purposes of communication. Osteocytes remain in contact with other cells in the bone through gap junctions—coupled cell processes—which pass through small channels in the bone matrix called the canaliculi.

27.2.13 Osteoclast

Osteoclasts are very large multinucleate cells that are responsible for the breakdown of bones by the process of bone resorption. New bone is then formed by the osteoblasts. Bone is constantly remodelled by the resorption of osteoclasts and created by osteoblasts. Osteoclasts are large cells with multiple nuclei located on bone surfaces in what are called Howship's lacunae (or resorption pits). These lacunae are the result of surrounding bone tissue that has been reabsorbed. Because the osteoclasts are derived from a monocyte stem-cell lineage, they are equipped with phagocytic-like mechanisms similar to circulating macrophages. Osteoclasts mature and/or migrate to discrete bone surfaces. Upon arrival, active enzymes, such as tartrate resistant acid phosphatase, are secreted against the mineral substrate.[citation needed] The reabsorption of bone by osteoclasts also plays a role in calcium homeostasis.

27.2.14 Extracellular Matrix

Bones consist of living cells embedded in a mineralized organic matrix. This matrix consists of organic components, mainly type I collagen – “organic” referring to materials produced as a result of the human body – and inorganic components, primarily hydroxyapatite and other salts of calcium and phosphate. Above 30% of the acellular part of bone consists of the organic components, and 70% of salts. The collagen fibers give bone its tensile strength, and the interspersed crystals of hydroxyapatite give bone its compressive strength. These effects are synergistic.

The inorganic composition of bone (bone mineral) is primarily formed from salts of calcium and phosphate, the major salt being hydroxyapatite ($\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$). The exact composition of the matrix may be subject to change over time due to nutrition and biominerilization, with the ratio of calcium to phosphate varying between 1.3 and 2.0 (per weight), and trace minerals such as magnesium, sodium, potassium and carbonate also being found.

27.3 The Muscular System

The muscular system is an organ system consisting of skeletal, smooth and cardiac muscles. It permits movement of the body, maintains posture and circulates blood throughout the body. The muscular systems in vertebrates are controlled through the nervous system although some muscles (such as the cardiac muscle) can be completely autonomous. Together with the skeletal system, it forms the musculoskeletal system, which is responsible for movement of the human body.

There are three distinct types of muscles:

- Skeletal muscle or “voluntary muscle” is anchored by tendons (or by aponeuroses at a few places) to bone and is used to effect skeletal movement such as locomotion and in maintaining posture. Though this postural control is generally maintained as an unconscious reflex, the muscles responsible react to conscious control like non-postural muscles. An average adult male is made up of 42% of skeletal muscle and an average adult female is made up of 36% (as a percentage of body mass).
- Smooth muscle or “involuntary muscle” is found within the walls of organs and structures such as the esophagus, stomach, intestines, bronchi, uterus, urethra, bladder, blood vessels, and the arrector pili in the skin (in which it controls erection of body hair). Unlike skeletal muscle, smooth muscle is not under conscious control.
- Cardiac muscle (myocardium), is also an “involuntary muscle” but is more akin in structure to skeletal muscle, and is found only in the heart.

Skeletal muscle is arranged in discrete muscles, an example of which is the biceps brachii (biceps). The tough, fibrous epimysium of skeletal muscle is both connected to and continuous with the tendons. In turn, the tendons connect to the periosteum layer surrounding the bones, permitting the transfer of force from the muscles to the skeleton. Together, these fibrous layers, along with tendons and ligaments, constitute the deep fascia of the body.

Skeletal muscles, like other striated muscles, are composed of myocytes, or muscle fibers, which are in turn composed of myofibrils, which are composed of sarcomeres, the basic building block of striated muscle tissue. Upon stimulation by an action potential, skeletal muscles perform a coordinated contraction by shortening each sarcomere. The best proposed model for understanding contraction is the sliding filament model of muscle contraction. Within the sarcomere, actin and myosin fibers overlap in a contractile motion towards each other. Myosin filaments have club-shaped heads that project toward the actin filaments.

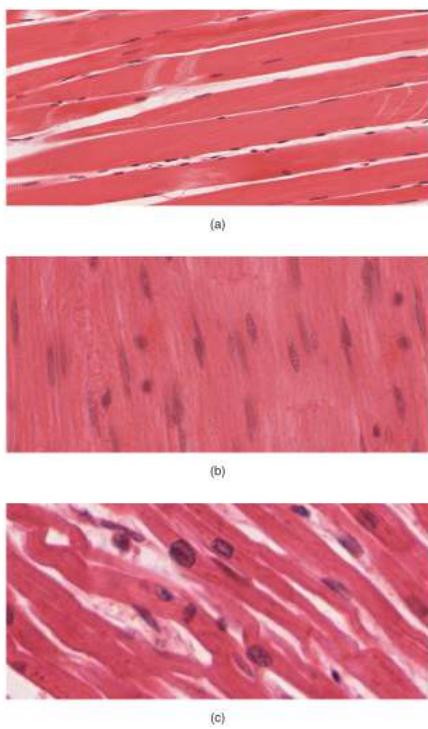


Figure 27.13: The body contains three types of muscle tissue⁵⁷: (a) skeletal muscle, (b) smooth muscle, and (c) cardiac muscle.

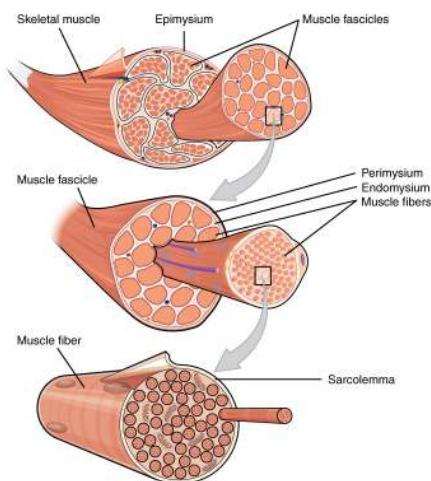


Figure 27.14: Bundles of muscle fibers, called fascicles, are covered by the perimysium. Muscle fibers are covered by the endomysium.⁵⁸

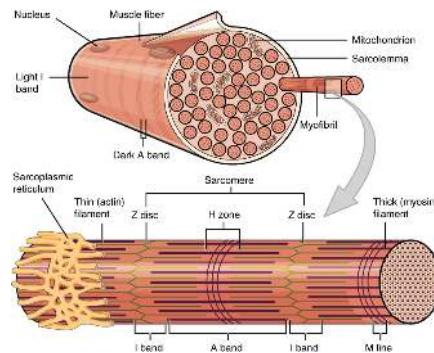


Figure 27.15: A skeletal muscle fiber is surrounded by a plasma membrane called the sarcolemma, which contains sarcoplasm, the cytoplasm of muscle cells. A muscle fiber is composed of many fibrils, which give the cell its striated appearance.⁵⁹

Larger structures along the myosin filament called myosin heads are used to provide attachment points on binding sites for the actin filaments. The myosin heads move in a coordinated style; they swivel toward the center of the sarcomere, detach and then reattach to the nearest active site of the actin filament. This is called a ratchet type drive system.

This process consumes large amounts of adenosine triphosphate (ATP), the energy source of the cell. ATP binds to the cross bridges between myosin heads and actin filaments. The release of energy powers the swiveling of the myosin head. When ATP is used, it becomes adenosine diphosphate (ADP), and since muscles store little ATP, they must continuously replace the discharged ADP with ATP. Muscle tissue also contains a stored supply of a fast acting recharge chemical, creatine phosphate, which when necessary can assist with the rapid regeneration of ADP into ATP.

Calcium ions are required for each cycle of the sarcomere. Calcium is released from the sarcoplasmic reticulum into the sarcomere when a muscle is stimulated to contract. This calcium uncovers the actin binding sites. When the muscle no longer needs to contract, the calcium ions are pumped from the sarcomere and back into storage in the sarcoplasmic reticulum.

There are approximately 639 skeletal muscles in the human body.

Heart muscles are distinct from skeletal muscles because the muscle fibers are laterally connected to each other. Furthermore, just as with smooth muscles, their movement is involuntary. Heart muscles are controlled by the sinus node influenced by the autonomic nervous system.

Smooth muscles are controlled directly by the autonomic nervous system and are involuntary, meaning that they are incapable of being moved by conscious thought. Functions such as heartbeat and lungs (which are capable of being willingly controlled, be it to a limited extent) are involuntary muscles but are not smooth muscles.

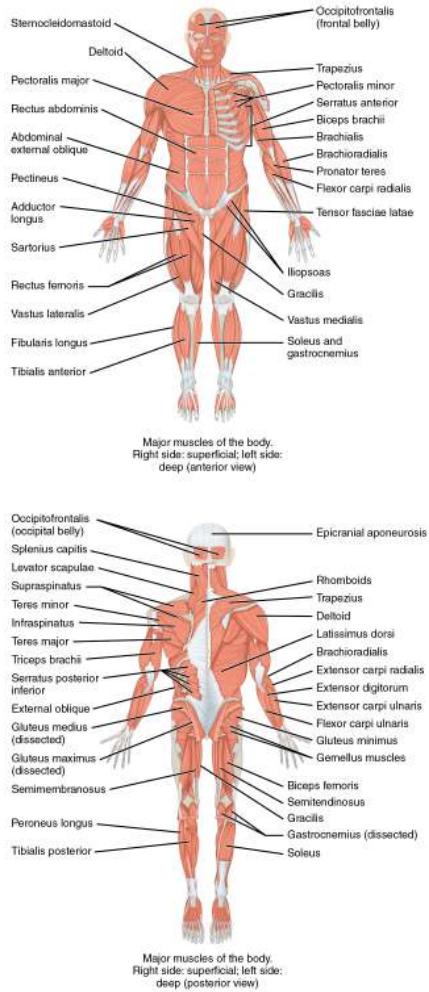


Figure 27.16: On the anterior and posterior views of the muscular system above, superficial muscles (those at the surface) are shown on the right side of the body while deep muscles (those underneath the superficial muscles) are shown on the left half of the body. For the legs, superficial muscles are shown in the anterior view while the posterior view shows both superficial and deep muscles.⁶⁰

27.3.1 Muscle Contraction

Muscle contraction is the activation of tension-generating sites within muscle fibers. In physiology, muscle contraction does not necessarily mean muscle shortening because muscle tension can be produced without changes in muscle length, such as when holding a heavy book or a dumbbell at the same position. The termination of muscle contraction is followed by muscle relaxation, which is a return of the muscle fibers to their low tension-generating state.

At rest, the body produces the majority of its ATP aerobically in the mitochondria without producing lactic acid or other fatiguing byproducts. During exercise, the method of ATP production varies depending on the fitness of the individual as well as the duration and intensity of exercise. At lower activity levels, when exercise continues for a long duration (several minutes or longer), energy is produced aerobically by combining oxygen with carbohydrates and fats stored in the body.

During activity that is higher in intensity, with possible duration decreasing as intensity increases, ATP production can switch to anaerobic pathways, such as the use of the creatine phosphate and the phosphagen system or anaerobic glycolysis. Aerobic ATP production is biochemically much slower and can only be used for long-duration, low-intensity exercise, but produces no fatiguing waste products that can not be removed immediately from the sarcomere and the body, and it results in a much greater number of ATP molecules per fat or carbohydrate molecule. Aerobic training allows the oxygen delivery system to be more efficient, allowing aerobic metabolism to begin quicker. Anaerobic ATP production produces ATP much faster and allows near-maximal intensity exercise, but also produces significant amounts of lactic acid which renders high-intensity exercise unsustainable for more than several minutes. The phosphagen system is also anaerobic. It allows for the highest levels of exercise intensity, but intramuscular stores of phosphocreatine are very limited and can only provide energy for exercises lasting up to ten seconds. Recovery is very quick, with full creatine stores regenerated within five minutes. Muscle contractions can be described based on two variables: length and tension. A muscle contraction is described as isometric if the muscle tension changes but the muscle length remains the same. In contrast, a muscle contraction is isotonic if muscle tension remains the same throughout

the contraction. If the muscle length shortens, the contraction is concentric; if the muscle length lengthens, the contraction is eccentric. In natural movements that underlie locomotor activity, muscle contractions are multifaceted as they are able to produce changes in length and tension in a time-varying manner. Therefore, neither length nor tension is likely to remain the same in muscles that contract during locomotor activity.

In vertebrates, skeletal muscle contractions are neurogenic as they require synaptic input from motor neurons to produce muscle contractions. A single motor neuron is able to innervate multiple muscle fibers, thereby causing the fibers to contract at the same time. Once innervated, the protein filaments within each skeletal muscle fiber slide past each other to produce a contraction, which is explained by the sliding filament theory. The contraction produced can be described as a twitch, summation, or tetanus, depending on the frequency of action potentials. In skeletal muscles, muscle tension is at its greatest when the muscle is stretched to an intermediate length as described by the length-tension relationship.

Unlike skeletal muscle, the contractions of smooth and cardiac muscles are myogenic (meaning that they are initiated by the smooth or heart muscle cells themselves instead of being stimulated by an outside event such as nerve stimulation), although they can be modulated by stimuli from the autonomic nervous system. The mechanisms of contraction in these muscle tissues are similar to those in skeletal muscle tissues.

Muscle contractions can be described based on two variables: force and length. Force itself can be differentiated as either tension or load. Muscle tension is the force exerted by the muscle on an object whereas a load is the force exerted by an object on the muscle. When muscle tension changes without any corresponding changes in muscle length, the muscle contraction is described as isometric. If the muscle length changes while muscle tension remains the same, then the muscle contraction is isotonic. In an isotonic contraction, the muscle length can either shorten to produce a concentric contraction or lengthen to produce an eccentric contraction. In natural movements that underlie locomotor activity, muscle contractions are multifaceted as they are able to produce changes in length and tension in a time-varying manner. Therefore, neither length nor tension is likely to remain constant when the muscle is active during locomotor activity.

An isometric contraction of a muscle generates tension without changing length. An example can be found when the muscles of the hand and forearm grip an object; the joints of the hand do not move, but muscles generate sufficient force to prevent the object from being dropped.

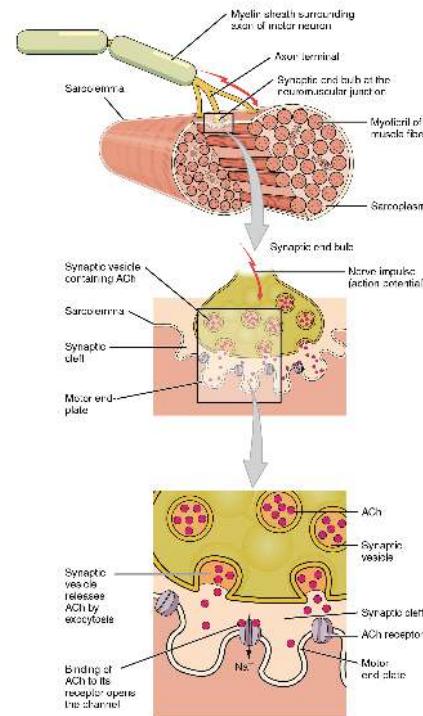


Figure 27.17: Structure of neuromuscular junction.⁶¹

In isotonic contraction, the tension in the muscle remains constant despite a change in muscle length. This occurs when a muscle's force of contraction matches the total load on the muscle.

27.3.2 Skeletal Muscle

Excluding reflexes, all skeletal muscle contractions occur as a result of conscious effort originating in the brain. The brain sends electrochemical signals through the nervous system to the motor neuron that innervates several muscle fibers. In the case of some reflexes, the signal to contract can originate in the spinal cord through a feedback loop with the grey matter. Other actions such as locomotion, breathing, and chewing have a reflex aspect to them: the contractions can be initiated both consciously or unconsciously.

A neuromuscular junction is a chemical synapse formed by the contact between a motor neuron and a muscle fiber. It is the site in which a motor neuron transmits a signal to a muscle fiber to initiate muscle contraction.

The sequence of events that results in the depolarization of the muscle fiber at the neuromuscular junction begins when an action potential is initiated in the cell body of a motor neuron, which is then propagated by saltatory conduction along its axon toward the neuromuscular

junction. Once it reaches the terminal bouton, the action potential causes a Ca^{2+} ion influx into the terminal by way of the voltage-gated calcium channels. The Ca^{2+} influx causes synaptic vesicles containing the neurotransmitter acetylcholine to fuse with the plasma membrane, releasing acetylcholine into the synaptic cleft between the motor neuron terminal and the neuromuscular junction of the skeletal muscle fiber. Acetylcholine diffuses across the synapse and binds to and activates nicotinic acetylcholine receptors on the neuromuscular junction. Activation of the nicotinic receptor opens its intrinsic sodium/potassium channel, causing sodium to rush in and potassium to trickle out. As a result, the sarcolemma reverses polarity and its voltage quickly jumps from the resting membrane potential of -90mV to as high as $+75\text{mV}$ as sodium enters. The membrane potential then becomes hyperpolarized when potassium exits and is then adjusted back to the resting membrane potential. This rapid fluctuation is called the end-plate potential. The voltage-gated ion channels of the sarcolemma next to the end plate open in response to the end plate potential. These voltage-gated channels are sodium and potassium specific and only allow one through. This wave of ion movements creates the action potential that spreads from the motor end plate in all directions. If action potentials stop arriving, then acetylcholine ceases to be released from the terminal bouton. The remaining acetylcholine in the synaptic cleft is either degraded by active acetylcholine esterase or reabsorbed by the synaptic knob and none is left to replace the degraded acetylcholine.

27.3.3 Excitation-contraction Coupling

Excitation-contraction coupling is the process by which a muscular action potential in the muscle fiber causes the myofibrils to contract. In skeletal muscle, excitation-contraction coupling relies on a direct coupling between key proteins, the sarcoplasmic reticulum (SR) calcium release channel (identified as the ryanodine receptor, RyR) and voltage-gated L-type calcium channels (identified as dihydropyridine receptors, DHPRs). DHPRs are located on the sarcolemma (which includes the surface sarcolemma and the transverse tubules), while the RyRs reside across the SR membrane. The close apposition of a transverse tubule and two SR regions containing RyRs is described as a triad and is predominantly where excitation-contraction coupling takes place. Excitation-contraction coupling occurs when depolarization of skeletal muscle cell results in a muscle action potential, which spreads across the cell surface and into the muscle fiber's network of T-tubules, thereby depolarizing the inner portion of the muscle fiber. Depolarization of the inner portions activates dihydropyridine receptors in the terminal cisternae, which are in close proximity to ryanodine receptors in the adjacent

sarcoplasmic reticulum. The activated dihydropyridine receptors physically interact with ryanodine receptors to activate them via foot processes (involving conformational changes that allosterically activates the ryanodine receptors). As the ryanodine receptors open, Ca^{2+} is released from the sarcoplasmic reticulum into the local junctional space, which then diffuses into the bulk cytoplasm to cause a calcium spark. Note that the sarcoplasmic reticulum has a large calcium buffering capacity partially due to a calcium-binding protein called calsequestrin. The near synchronous activation of thousands of calcium sparks by the action potential causes a cell-wide increase in calcium giving rise to the upstroke of the calcium transient. The Ca^{2+} released into the cytosol binds to Troponin C by the actin filaments, to allow crossbridge cycling, producing force and, in some situations, motion. The sarco/endoplasmic reticulum calcium-ATPase (SERCA) actively pumps Ca^{2+} back into the sarcoplasmic reticulum. As Ca^{2+} declines back to resting levels, the force declines and relaxation occurs.

27.3.4 The Sliding Filament Theory

The sliding filament theory describes a process used by muscles to contract. It is a cycle of repetitive events that cause a thin filament to slide over a thick filament and generate tension in the muscle. It was independently developed by Andrew Huxley and Rolf Niedergerke and by Hugh Huxley and Jean Hanson in 1954. Physiologically, this contraction is not uniform across the sarcomere; the central position of the thick filaments becomes unstable and can shift during contraction. However the actions of elastic proteins such as titin are hypothesised to maintain uniform tension across the sarcomere and pull the thick filament into a central position.

27.3.5 Crossbridge Cycling

Crossbridge cycling is a sequence of molecular events that underlies the sliding filament theory. A crossbridge is a myosin projection, consisting of two myosin heads, that extends from the thick filaments. Each myosin head has two binding sites: one for ATP and another for actin. The binding of ATP to a myosin head detaches myosin from actin, thereby allowing myosin to bind to another actin molecule. Once attached, the ATP is hydrolyzed by myosin, which uses the released energy to move into the "cocked position" whereby it binds weakly to a part of the actin binding site. The remainder of the actin binding site is blocked by tropomyosin. With the ATP hydrolyzed, the cocked myosin head now contains ADP + P_i . Two Ca^{2+} ions bind to troponin C on the actin filaments. The troponin- Ca^{2+} complex causes tropomyosin to slide over and unblock the remainder of the actin binding site.

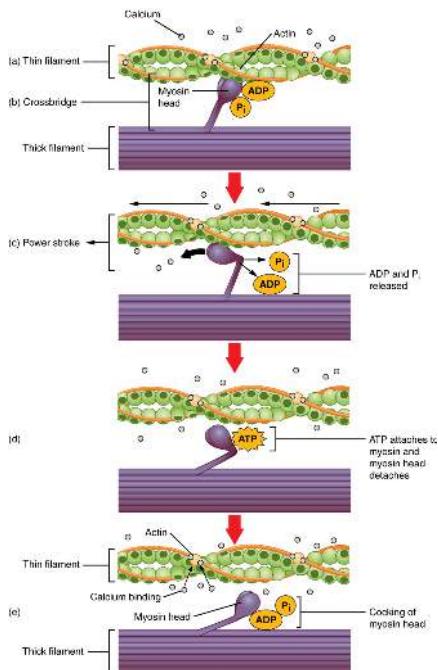


Figure 27.18: Crossbridge cycling.⁶²

Unblocking the rest of the actin binding sites allows the two myosin heads to close and myosin to bind strongly to actin. The myosin head then releases the inorganic phosphate and initiates a power stroke, which generates a force of 2 pN. The power stroke moves the actin filament inwards, thereby shortening the sarcomere. Myosin then releases ADP but still remains tightly bound to actin. At the end of the power stroke, ADP is released from the myosin head, leaving myosin attached to actin in a rigor state until another ATP binds to myosin. A lack of ATP would result in the rigor state characteristic of rigor mortis. Once another ATP binds to myosin, the myosin head will again detach from actin and another crossbridges cycle occurs.

Crossbridge cycling is able to continue as long as there are sufficient amounts of ATP and Ca^{2+} in the cytoplasm. Termination of crossbridge cycling can occur when Ca^{2+} is actively pumped back into the sarcoplasmic reticulum. When Ca^{2+} is no longer present on the thin filament, the tropomyosin changes conformation back to its previous state so as to block the binding sites again. The myosin ceases binding to the thin filament, and the muscle relaxes. The Ca^{2+} ions leave the troponin molecule in order to maintain the Ca^{2+} ion concentration in the sarcoplasm. The active pumping of Ca^{2+} ions into the sarcoplasmic reticulum creates a deficiency in the fluid around the myofibrils. This causes the removal of Ca^{2+} ions from the troponin. Thus, the tropomyosin-troponin complex again covers the binding sites on the actin filaments and contraction ceases.

The strength of skeletal muscle contractions can be broadly separated into twitch, summation, and tetanus. A twitch is a single contraction and relaxation cycle produced by an action potential within the muscle fiber itself. The time between a stimulus to the motor nerve and the subsequent contraction of the innervated muscle is called the latent period, which usually takes about 10 ms and is caused by the time taken for nerve action potential to propagate, the time for chemical transmission at the neuromuscular junction, then the subsequent steps in excitation-contraction coupling.

If another muscle action potential were to be produced before the complete relaxation of a muscle twitch, then the next twitch will simply sum onto the previous twitch, thereby producing a summation. Summation can be achieved in two ways: frequency summation and multiple fiber summation. In frequency summation, the force exerted by the skeletal muscle is controlled by varying the frequency at which action potentials are sent to muscle fibers. Action potentials do not arrive at muscles synchronously, and, during a contraction, some fraction of the fibers in the muscle will be firing at any given time. In a typical circumstance, when humans are exerting their muscles as hard as they are consciously able, roughly one-third of the fibers in each of those muscles will fire at once, though this ratio can be affected by various physiological and psychological factors (including Golgi tendon organs and Renshaw cells). This ‘low’ level of contraction is a protective mechanism to prevent avulsion of the tendon—the force generated by a 95% contraction of all fibers is sufficient to damage the body. In multiple fiber summation, if the central nervous system sends a weak signal to contract a muscle, the smaller motor units, being more excitable than the larger ones, are stimulated first. As the strength of the signal increases, more motor units are excited in addition to larger ones, with the largest motor units having as much as 50 times the contractile strength as the smaller ones. As more and larger motor units are activated, the force of muscle contraction becomes progressively stronger. A concept known as the size principle, allows for a gradation of muscle force during weak contraction to occur in small steps, which then become progressively larger when greater amounts of force are required.

Finally, if the frequency of muscle action potentials increases such that the muscle contraction reaches its peak force and plateaus at this level, then the contraction is a tetanus.

Length-tension relationship relates the strength of an isometric contraction to the length of the muscle at which the contraction occurs. Muscles operate with greatest active tension when close to an ideal length

(often their resting length). When stretched or shortened beyond this (whether due to the action of the muscle itself or by an outside force), the maximum active tension generated decreases. This decrease is minimal for small deviations, but the tension drops off rapidly as the length deviates further from the ideal. Due to the presence of elastic proteins within a muscle cell (such as titin) and extracellular matrix, as the muscle is stretched beyond a given length, there is an entirely passive tension, which opposes lengthening. Combined together, there is a strong resistance to lengthening an active muscle far beyond the peak of active tension.

Force-velocity relationship relates the speed at which a muscle changes its length (usually regulated by external forces, such as load or other muscles) to the amount of force that it generates. Force declines in a hyperbolic fashion relative to the isometric force as the shortening velocity increases, eventually reaching zero at some maximum velocity. The reverse holds true for when the muscle is stretched – force increases above isometric maximum, until finally reaching an absolute maximum. This intrinsic property of active muscle tissue plays a role in the active damping of joints that are actuated by simultaneously-active opposing muscles. In such cases, the force-velocity profile enhances the force produced by the lengthening muscle at the expense of the shortening muscle. This favoring of whichever muscle returns the joint to equilibrium effectively increases the damping of the joint. Moreover, the strength of the damping increases with muscle force. The motor system can thus actively control joint damping via the simultaneous contraction (co-contraction) of opposing muscle groups.

27.3.6 Smooth Muscle

Smooth muscles can be divided into two subgroups: single-unit (unitary) and multi-unit. Single-unit smooth muscle cells can be found in the gut and blood vessels. Because these cells are linked together by gap junctions, they are able to contract as a syncytium. Single-unit smooth muscle cells contract myogenically, which can be modulated by the autonomic nervous system.

Unlike single-unit smooth muscle cells, multi-unit smooth muscle cells are found in the muscle of the eye and in the base of hair follicles. Multi-unit smooth muscle cells contract by being separately stimulated by nerves of the autonomic nervous system. As such, they allow for fine control and gradual responses, much like motor unit recruitment in skeletal muscle.

The contractile activity of smooth muscle cells is influenced by multiple inputs such as spontaneous electrical activity, neural and hormonal inputs, local changes in chemical composition, and stretch. This is in contrast to

the contractile activity of skeletal muscle cells, which relies on a single neural input. Some types of smooth muscle cells are able to generate their own action potentials spontaneously, which usually occur following a pacemaker potential or a slow wave potential. These action potentials are generated by the influx of extracellular Ca^{2+} , and not Na^+ . Like skeletal muscles, cytosolic Ca^{2+} ions are also required for crossbridge cycling in smooth muscle cells.

The two sources for cytosolic Ca^{2+} in smooth muscle cells are the extracellular Ca^{2+} entering through calcium channels and the Ca^{2+} ions that are released from the sarcoplasmic reticulum. The elevation of cytosolic Ca^{2+} results in more Ca^{2+} binding to calmodulin, which then binds and activates myosin light-chain kinase. The calcium-calmodulin-myosin light-chain kinase complex phosphorylates myosin on the 20 kilodalton (kDa) myosin light chains on amino acid residue-serine 19, initiating contraction and activating the myosin ATPase. Unlike skeletal muscle cells, smooth muscle cells lack troponin, even though they contain the thin filament protein tropomyosin and other notable proteins – caldesmon and calponin. Thus, smooth muscle contractions are initiated by the Ca^{2+} -activated phosphorylation of myosin rather than Ca^{2+} binding to the troponin complex that regulates myosin binding sites on actin like in skeletal and cardiac muscles.

Termination of crossbridge cycling (and leaving the muscle in latch-state) occurs when myosin light chain phosphatase removes the phosphate groups from the myosin heads. Phosphorylation of the 20 kDa myosin light chains correlates well with the shortening velocity of smooth muscle. During this period, there is a rapid burst of energy utilization as measured by oxygen consumption. Within a few minutes of initiation, the calcium level markedly decreases, the 20 kDa myosin light chains' phosphorylation decreases, and energy utilization decreases; however, force in tonic smooth muscle is maintained. During contraction of muscle, rapidly cycling crossbridges form between activated actin and phosphorylated myosin, generating force. It is hypothesized that the maintenance of force results from dephosphorylated "latch-bridges" that slowly cycle and maintain force. A number of kinases such as rho kinase, ZIP kinase, and protein kinase C are believed to participate in the sustained phase of contraction, and Ca^{2+} flux may be significant.

Although smooth muscle contractions are myogenic, the rate and strength of their contractions can be modulated by the autonomic nervous system. Postganglionic nerve fibers of parasympathetic nervous system release the neurotransmitter acetylcholine, which binds to muscarinic acetylcholine receptors (mAChRs) on smooth muscle cells. These receptors are metabotropic, or G-

protein coupled receptors that initiate a second messenger cascade. Conversely, postganglionic nerve fibers of the sympathetic nervous system release the neurotransmitters epinephrine and norepinephrine, which bind to adrenergic receptors that are also metabotropic. The exact effects on the smooth muscle depend on the specific characteristics of the receptor activated—both parasympathetic input and sympathetic input can be either excitatory (contractile) or inhibitory (relaxing).

27.3.7 Cardiac Muscle

There are two types of cardiac muscle cells: autorhythmic and contractile. Autorhythmic cells do not contract, but instead set the pace of contraction for other cardiac muscle cells, which can be modulated by the autonomic nervous system. In contrast, contractile muscle cells (cardiomyocytes) constitute the majority of the heart muscle and are able to contract.

Unlike skeletal muscle, excitation-contraction coupling in cardiac muscle is thought to depend primarily on a mechanism called calcium-induced calcium release. Though the proteins involved are similar, the L-type calcium channels and ryanodine receptors (RyRs) are not physically coupled. Instead, RyRs are activated by a calcium trigger, which is brought about by the flow of Ca^{2+} through the L-type calcium channels. Furthermore, cardiac muscle tend to exhibit diad (or dyad) structures, rather than triads.

Excitation-contraction coupling in cardiac muscle cells occurs when an action potential is initiated by pacemaker cells in the sinoatrial node or Atrioventricular node and conducted to all cells in the heart via gap junctions. The action potential travels along the surface membrane into T-tubules (the latter are not seen in all cardiac cell types) and the depolarisation causes extracellular Ca^{2+} to enter the cell via L-type calcium channels and possibly sodium-calcium exchanger (NCX) during the early part of the plateau phase. This Ca^{2+} influx causes a small local increase in intracellular Ca^{2+} . The increase in Ca^{2+} is detected by ryanodine receptors in the membrane of the sarcoplasmic reticulum, which releases Ca^{2+} in a positive feedback physiological response. This positive feedback is known as calcium-induced calcium release and gives rise to calcium sparks (Ca^{2+} sparks). The spatial and temporal summation of ~30,000 Ca^{2+} sparks gives a cell-wide increase in cytosolic calcium concentration. The increase in cytosolic calcium following the flow of calcium through the cell membrane and sarcoplasmic reticulum is moderated by calcium buffers, which bind a large proportion of intracellular calcium. As a result, a large increase in total calcium leads to a relatively small rise in free Ca^{2+} .

Following systole, intracellular calcium is taken up by the sarco/endoplasmic reticulum ATPase (SERCA) pump back into the sarcoplasmic reticulum ready for the next cycle to begin. Calcium is also ejected from the cell mainly by the sodium-calcium exchanger (NCX) and, to a lesser extent, a plasma membrane calcium ATPase. Some calcium is also taken up by the mitochondria. An enzyme, phospholamban, serves as a brake for SERCA. At low heart rates, phospholamban is active and slows down the activity of the ATPase so that Ca^{2+} does not have to leave the cell entirely. At high heart rates, phospholamban is phosphorylated and deactivated thus taking most Ca^{2+} from the cytoplasm back into the sarcoplasmic reticulum. Once again, calcium buffers moderate this fall in Ca^{2+} concentration, permitting a relatively small decrease in free Ca^{2+} concentration in response to a large change in total calcium. The falling Ca^{2+} concentration allows the troponin complex to dissociate from the actin filament thereby ending contraction. The heart relaxes, allowing the ventricles to fill with blood and begin the cardiac cycle again.

27.3.8 The Muscular Systems of Invertebrates

In annelids such as earthworms and leeches, circular and longitudinal muscles cells form the body wall of these animals and are responsible for their movement. In an earthworm that is moving through a soil, for example, contractions of circular and longitudinal muscles occur reciprocally while the coelomic fluid serves as a hydro skeleton by maintaining turgidity of the earthworm. When the circular muscles in the anterior segments contract, the anterior portion of animal's body begins to constrict radially, which pushes the incompressible coelomic fluid forward and increasing the length of the animal. As a result, the front end of the animal moves forward. As the front end of the earthworm becomes anchored and the circular muscles in the anterior segments become relaxed, a wave of longitudinal muscle contractions passes backwards, which pulls the rest of animal's trailing body forward. These alternating waves of circular and longitudinal contractions is called peristalsis, which underlies the creeping movement of earthworms.

Invertebrates such as annelids, mollusks, and nematodes, possess obliquely striated muscles, which contain bands of thick and thin filaments that are arranged helically rather than transversely, like in vertebrate skeletal or cardiac muscles. In bivalves, the obliquely striated muscles can maintain tension over long periods without using too much energy. Bivalves use these muscles to keep their shells closed.

Advanced insects such as wasps, flies, bees, and

beetles possess asynchronous muscles that constitute the flight muscles in these animals. These flight muscles are often called fibrillar muscles because they contain myofibrils that are thick and conspicuous. A remarkable feature of these muscles is that they do not require stimulation for each muscle contraction. Hence, they are called asynchronous muscles because the number of contractions in these muscles do not correspond (or synchronize) with the number of action potentials. For example, a wing muscle of a tethered fly may receive action potentials at a frequency of 3 Hz but it is able to beat at a frequency of 120 Hz. The high frequency beating is made possible because the muscles are connected to a resonant system, which is driven to a natural frequency of vibration.

Chapter 28

Endocrine Systems

The endocrine system is a chemical messenger system comprising feedback loops of the hormones released by internal glands of an organism directly into the circulatory system, regulating distant target organs. In humans, the major endocrine glands are the thyroid gland and the adrenal glands. In vertebrates, the hypothalamus is the neural control center for all endocrine systems. The study of the endocrine system and its disorders is known as endocrinology. Endocrinology is a branch of internal medicine.

A neuroendocrine system has been observed in all animals with a nervous system and all vertebrates have a hypothalamus–pituitary axis. All vertebrates have a thyroid, which in amphibians is also crucial for transformation of larvae into adult form. All vertebrates have adrenal gland tissue, with mammals unique in having it organized into layers. All vertebrates have some form of a renin–angiotensin axis, and all tetrapods have aldosterone as a primary mineralocorticoid.

A number of glands that signal each other in sequence are usually referred to as an axis, such as the hypothalamic–pituitary–adrenal axis. In addition to the specialized endocrine organs mentioned above, many other organs that are part of other body systems have secondary endocrine functions, including bone, kidneys, liver, heart and gonads. For example, the kidneys secrete a variety of hormones, including erythropoietin, calcitriol, and renin. Erythropoietin is released in response to hypoxia (low levels of oxygen at tissue level) in the renal circulation. It stimulates erythropoiesis (production of red blood cells) in the bone marrow. Calcitriol, the activated form of vitamin D, promotes intestinal absorption of calcium and the renal reabsorption of phosphate. Renin is an enzyme which regulates angiotensin and aldosterone levels.

The endocrine system can be contrasted to both exocrine glands, which secrete hormones to the outside of the body, and paracrine signalling between cells over a relatively short distance. Endocrine glands have no ducts, are vascular, and commonly have intracellular vacuoles or

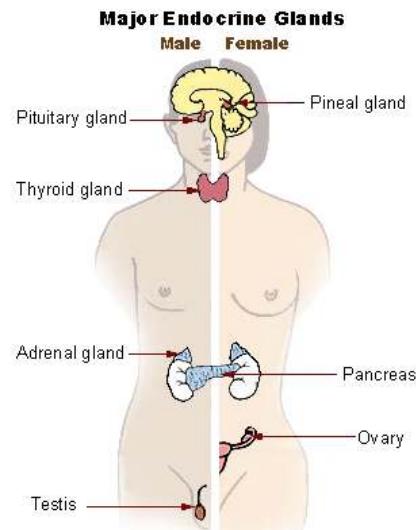


Figure 28.1: Main glands of the human endocrine system.¹

granules that store their hormones. In contrast, exocrine glands, such as salivary glands, sweat glands, and glands within the gastrointestinal tract, tend to be much less vascular and have ducts or a hollow lumen.

The word endocrine derives via New Latin from the Greek words ἔνδον, *endon*, “inside, within,” and “crine” from the κρίνω, *krīnō*, “to separate, distinguish”.

Endocrine glands are glands of the endocrine system that secrete their products, hormones, directly into interstitial spaces and then absorbed into blood rather than through a duct. The major glands of the endocrine system include the pineal gland, pituitary gland, pancreas, ovaries, testes, thyroid gland, parathyroid gland, hypothalamus and adrenal glands. The hypothalamus and pituitary gland are neuroendocrine organs.

A hormone (from the Greek participle ὀρμῶν, “setting in motion”) is any member of a class of signaling molecules, produced by glands in multicellular organisms, that are

Hormone Class	Components	Example(s)
Amine Hormones	Amino acids with hydroxyl groups, a ring which contains both a hydroxyl group & a nitrogen ring.	Norepinephrine
Peptide Hormones	Short chains of linked amino acids	Oxytocin
Protein Hormones	Long chains of linked amino acids	Human Growth Hormone
Steroid Hormones	Derived from the lipid backbone	Testosterone Progesterone

Figure 28.2: Representative examples of the three chemical classes of hormones in the human body.²

transported by the circulatory system to target distant organs to regulate physiology and behavior. Hormones have diverse chemical structures, mainly of three classes:

- eicosanoids
- steroids
- amino acid/protein derivatives (amines, peptides, and proteins)

The glands that secrete hormones comprise the endocrine system.

Hormonal effects are dependent on where they are released, as they can be released in different manners. Not all hormones are released from a cell and into the blood until it binds to a receptor on a target. The major types of hormone signaling are:

- Endocrine – Acts on the target cell after being released into the bloodstream.
- Paracrine – Acts on a nearby cell and does not have to enter general circulation.
- Autocrine – Affects the cell type that secreted it and causes a biological effect.
- Intracrine – Acts intracellularly on the cell that synthesized it.

Hormones are used to communicate between organs and tissues for physiological regulation and behavioral activities, such as digestion, metabolism, respiration, tissue function, sensory perception, sleep, excretion, lactation, stress, growth and development, movement, reproduction, and mood.

Hormones affect distant cells by binding to specific re-

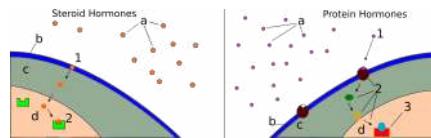


Figure 28.3: The left diagram shows a steroid (lipid) hormone (1) entering a cell and (2) binding to a receptor protein in the nucleus, causing (3) mRNA synthesis which is the first step of protein synthesis. The right side shows protein hormones (1) binding with receptors which (2) begins a transduction pathway. The transduction pathway ends (3) with transcription factors being activated in the nucleus, and protein synthesis beginning. In both diagrams, a is the hormone, b is the cell membrane, c is the cytoplasm, and d is the nucleus.³

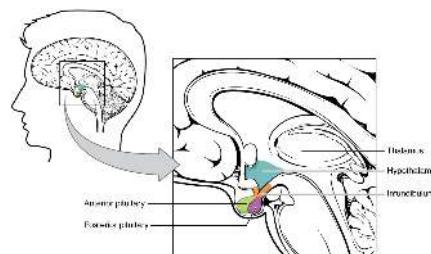


Figure 28.4: The Hypothalamus–Pituitary Complex.⁵

ceptor proteins in the target cell resulting in a change in cell function. This may lead to cell type-specific responses that include rapid changes to the activity of existing proteins, or slower changes in the expression of target genes. Amino acid-based hormones (amines and peptide or protein hormones) are water-soluble and act on the surface of target cells via signal transduction pathways; steroid hormones, being lipid-soluble, move through the plasma membranes of target cells to act within their nuclei.

28.1 The Hypothalamus

The hypothalamus⁴ (from Ancient Greek ὑπό, “under”, and θάλαμος, “chamber”) is a portion of the brain that has a central neuroendocrine function, most notably by its control of the anterior pituitary, which in turn regulates various endocrine glands and organs. Releasing hormones (also called releasing factors) are produced in hypothalamic nuclei then transported along axons to the posterior pituitary, where they are stored and released as needed.

In the hypothalamic–adenohypophyseal axis, releasing hormones, also known as hypophysiotropic or hypothalamic hormones, are released from the median eminence, a prolongation of the hypothalamus, into the hypophyseal

⁴<https://en.wikipedia.org/wiki/Hypothalamus>

portal system, which carries them to the anterior pituitary where they exert their regulatory functions on the secretion of adenohypophyseal hormones. These hypophysiotropic hormones are stimulated by parvocellular neurosecretory cells located in the periventricular area of the hypothalamus. After their release into the capillaries of the third ventricle, the hypophysiotropic hormones travel through what is known as the hypothalamo-pituitary portal circulation. Once they reach their destination in the anterior pituitary, these hormones bind to specific receptors located on the surface of pituitary cells. Depending on which cells are activated through this binding, the pituitary will either begin secreting or stop secreting hormones into the rest of the bloodstream.

Table 28.1: Hormones of the anterior pituitary gland.

Secreted hormone	Abbreviation	Produced by	Effect
Thyrotropin-releasing hormone (Prolactin-releasing hormone)	TRH, TRF, or PRH	Parvocellular neurosecretory cells of the paraventricular nucleus	Stimulate thyroid-stimulating hormone (TSH) release from anterior pituitary (primarily) Stimulate prolactin release from anterior pituitary
Corticotropin-releasing hormone	CRH or CRF	Parvocellular neurosecretory cells of the paraventricular nucleus	Stimulate adrenocorticotropic hormone (ACTH) release from anterior pituitary
Dopamine (Prolactin-inhibiting hormone)	DA or PIH	Dopamine neurons of the arcuate nucleus	Inhibit prolactin release from anterior pituitary
Growth-hormone-releasing hormone	GHRH	Neuroendocrine neurons of the Arcuate nucleus	Stimulate growth-hormone (GH) release from anterior pituitary
Gonadotropin-releasing hormone	GnRH or LHRH	Neuroendocrine cells of the Preoptic area	Stimulate follicle-stimulating hormone (FSH) release from anterior pituitary
Somatostatin (growth-hormone-inhibiting hormone)	SS, GHIH, or SRF	Neuroendocrine cells of the Periventricular nucleus	Stimulate luteinizing hormone (LH) release from anterior pituitary
			Inhibit growth-hormone (GH) release from anterior pituitary Inhibit (moderately) thyroid-stimulating hormone (TSH) release from anterior pituitary

In the hypothalamic–neurohypophyseal axis, neurohypophysial hormones are released from the posterior pituitary, which is actually a prolongation of the hypothalamus, into the circulation.

Table 28.2: Hormones of the posterior pituitary gland.

Secreted hormone	Abbreviation	Produced by	Effect
Oxytocin	OXY or OXT	Magnocellular neurosecretory cells of the paraventricular nucleus and supraoptic nucleus	Uterine contraction (lactation reflex)
Vasopressin (antidiuretic hormone)	ADH or AVP	Magnocellular and parvocellular neurosecretory cells of the paraventricular nucleus, magnocellular cells in supraoptic nucleus	Increase in the permeability to water of the cells of distal tubule and collecting duct in the kidney and thus allows water reabsorption and excretion of concentrated urine

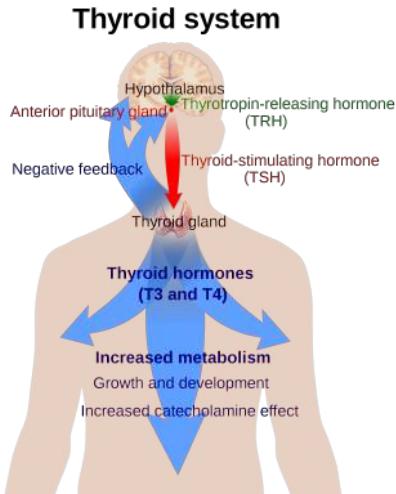


Figure 28.5: The hypothalamic–pituitary–thyroid axis (HPT axis for short, a.k.a. thyroid homeostasis or thyrotropic feedback control) is part of the neuroendocrine system responsible for the regulation of metabolism and also responds to stress.⁶

28.2 Hypothalamic-pituitary-thyroid axis

The hypothalamic–pituitary–thyroid axis (HPT axis for short, a.k.a. thyroid homeostasis or thyrotropic feedback control) is part of the neuroendocrine system responsible for the regulation of metabolism and also responds to stress.

As its name suggests, it depends upon the hypothalamus, the pituitary gland, and the thyroid gland.

The hypothalamus senses low circulating levels of thyroid hormone (Triiodothyronine (T3) and Thyroxine (T4)) and responds by releasing thyrotropin-releasing hormone (TRH). The TRH stimulates the anterior pituitary to produce thyroid-stimulating hormone (TSH). The TSH, in turn, stimulates the thyroid to produce thyroid hormone until levels in the blood return to normal. Thyroid hormone exerts negative feedback control over the hypothalamus as well as anterior pituitary, thus controlling the release of both TRH from hypothalamus and TSH from anterior pituitary gland.

Thyroid homeostasis results from a multi-loop feedback system that is found in virtually all higher vertebrates. Proper function of thyrotropic feedback control is indispensable for growth, differentiation, reproduction and intelligence. Very few animals (e.g. axolotls and sloths) have impaired thyroid homeostasis that exhibits a very low set-point that is assumed to underlie the metabolic and ontogenetic anomalies of these animals.

The pituitary gland secretes thyrotropin (TSH; Thyroid Stimulating Hormone) that stimulates the thyroid to secrete thyroxine (T4) and, to a lesser degree, triiodothyronine (T3). The major portion of T3, however, is produced in peripheral organs, e.g. liver, adipose tissue, glia and skeletal muscle by deiodination from circulating T4. Deiodination is controlled by numerous hormones and nerval signals including TSH, vasopressin and catecholamines.

Both peripheral thyroid hormones (iodothyronines) inhibit thyrotropin secretion from the pituitary (negative feedback). Consequently, equilibrium concentrations for all hormones are attained.

TSH secretion is also controlled by thyrotropin releasing hormone (thyroliberin, TRH), whose secretion itself is again suppressed by plasma T4 and T3 in CSF (long feedback, Fekete–Lechan loop). Additional feedback loops are ultrashort feedback control of TSH secretion (Brokken–Wiersinga–Prummel loop) and linear feedback loops controlling plasma protein binding.

Recent research suggested the existence of an additional feedforward motif linking TSH release to deiodinase activity in humans. The existence of this TSH-T3 shunt could explain why deiodinase activity is higher in hypothyroid patients and why a minor fraction of affected individuals may benefit from substitution therapy with T3.

Convergence of multiple afferent signals in the control of TSH release including but not limited to T3, cytokines and TSH receptor antibodies may be the reason for the observation that the relation between free T4 concentration and TSH levels deviates from a pure loglinear relation that has previously been proposed.

28.3 Hypothalamic–pituitary–gonadal axis

The hypothalamic–pituitary–gonadal axis (HPG axis) refers to the hypothalamus, pituitary gland, and gonadal glands as if these individual endocrine glands were a single entity. Because these glands often act in concert, physiologists and endocrinologists find it convenient and descriptive to speak of them as a single system.

The HPG axis plays a critical part in the development and regulation of a number of the body's systems, such as the reproductive and immune systems. Fluctuations in this axis cause changes in the hormones produced by each gland and have various local and systemic effects on the body.

The axis controls development, reproduction, and aging in animals. Gonadotropin-releasing hormone (GnRH) is

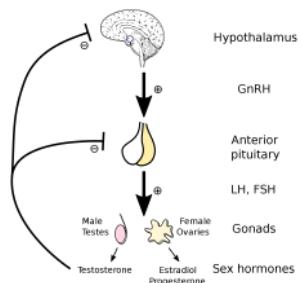


Figure 28.6: The hypothalamic-pituitary-gonadal axis (HPG axis) refers to the hypothalamus, pituitary gland, and gonadal glands as if these individual endocrine glands were a single entity. Because these glands often act in concert, physiologists and endocrinologists find it convenient and descriptive to speak of them as a single system.⁷

secreted from the hypothalamus by GnRH-expressing neurons. The anterior portion of the pituitary gland produces luteinizing hormone (LH) and follicle-stimulating hormone (FSH), and the gonads produce estrogen and testosterone.

In oviparous organisms (e.g. fish, reptiles, amphibians, birds), the HPG axis is commonly referred to as the hypothalamus-pituitary-gonadal-liver axis (HPGL-axis) in females. Many egg-yolk and chorionic proteins are synthesized heterologously in the liver, which are necessary for ovocyte growth and development. Examples of such necessary liver proteins are vitellogenin and choriogenin.

The hypothalamus is located in the brain and secretes GnRH. GnRH travels down the anterior portion of the pituitary via the hypophyseal portal system and binds to receptors on the secretory cells of the adenohypophysis. In response to GnRH stimulation these cells produce LH and FSH, which travel into the blood stream.

These two hormones play an important role in communicating to the gonads. In females FSH and LH act primarily to activate the ovaries to produce estrogen and inhibin and to regulate the menstrual cycle and ovarian cycle. Estrogen forms a negative feedback loop by inhibiting the production of GnRH in the hypothalamus. Inhibin acts to inhibit activin, which is a peripherally produced hormone that positively stimulates GnRH-producing cells. Folliculin, which is also produced in all body tissue, inhibits activin and gives the rest of the body more control over the axis. In males LH stimulates the interstitial cells located in the testes to produce testosterone, and FSH plays a role in spermatogenesis. Only small amounts of estrogen are secreted in males. Recent research has shown that a neurosteroid axis exists, which helps the cortex to regulate the hypothalamus's production of GnRH.

In addition, leptin and insulin have stimulatory effects

and ghrelin has inhibitory effects on gonadotropin-releasing hormone (GnRH) secretion from the hypothalamus. Kisspeptin also influences GnRH secretion.

One of the most important functions of the HPG axis is to regulate reproduction by controlling the uterine and ovarian cycles. In females, the positive feedback loop between estrogen and luteinizing hormone help to prepare the follicle in the ovary and the uterus for ovulation and implantation. When the egg is released, the empty follicle sac begins to produce progesterone to inhibit the hypothalamus and the anterior pituitary thus stopping the estrogen-LH positive feedback loop. If conception occurs, the placenta will take over the secretion of progesterone; therefore the mother cannot ovulate again. If conception does not occur, decreasing excretion of progesterone will allow the hypothalamus to restart secretion of GnRH. These hormone levels also control the uterine (menstrual) cycle causing the proliferation phase in preparation for ovulation, the secretory phase after ovulation, and menstruation when conception does not occur. The activation of the HPG axis in both males and females during puberty also causes individuals to acquire secondary sex characteristics.

In males, the production of GnRH, LH, and FSH are similar, but the effects of these hormones are different. FSH stimulates sustentacular cells to release androgen-binding protein, which promotes testosterone binding. LH binds to the interstitial cells, causing them to secrete testosterone. Testosterone is required for normal spermatogenesis and inhibits the hypothalamus. Inhibin is produced by the spermatogenic cells, which, also through inactivating activin, inhibits the hypothalamus. After puberty these hormone levels remain relatively constant.

The activation and deactivation of the HPG axis also helps to regulate life cycles. At birth FSH and LH levels are elevated, and females also have a lifetime supply of primary oocytes. These levels decrease and remain low through childhood. During puberty the HPG axis is activated by the secretions of estrogen from the ovaries or testosterone from the testes. This activation of estrogen and testosterone causes physiological and psychological changes. Once activated, the HPG axis continues to function in men for the rest of their life but becomes deregulated in women, leading to menopause. This deregulation is caused mainly by the lack of oocytes that normally produce estrogen to create the positive feedback loop. Over several years, the activity of the HPG axis decreases and women are no longer fertile.

Although males remain fertile until death, the activity of the HPG axis decreases. As males age, the testes begin to produce less testosterone, leading to a condition known as post-pubertal hypogonadism. The cause of the decreased testosterone is unclear and a current topic of

research. Post-pubertal hypogonadism results in progressive muscle mass decrease, increase in visceral fat mass, loss of libido, impotence, decreased attention, increased risk of fractures, and abnormal sperm production.

Sex steroids also affect behavior, because sex steroids affect the brain's structure and functioning. During development, hormones help determine how neurons synapse and migrate to result in sexual dimorphisms. These physical differences lead to differences in behavior. While GnRH has not been shown to have any direct influence on regulating brain structure and function, gonadotropins, sex steroids, and activin have been shown to have such effects. It is thought that FSH may have an important role in brain development and differentiation.

Testosterone levels have been shown to relate to prosocial behavior. This helps create synaptogenesis by promoting neurite development and migration. Activin promotes neural plasticity throughout the lifespan and regulates the neurotransmitters of peripheral neurons. Environment can also affect hormones and behavior interaction. Women have more connections between areas of language better enabling them to communicate than men. On average men outperform women on spatial reasoning tests, which is theorized to result from sexual differences. Testosterone has been linked to aggression and sex drive; therefore men tend to be more competitive or aggressive than women. There is also a large amount of individual diversity within all these traits and hormone levels.

28.4 Hypothalamic-pituitary-adrenal axis

The hypothalamic-pituitary-adrenal axis (HPA axis or HTPA axis) is a complex set of direct influences and feedback interactions among three components: the hypothalamus, the pituitary gland (a pea-shaped structure located below the thalamus), and the adrenal (also called "suprarenal" glands (small, conical organs on top of the kidneys).

These organs and their interactions constitute the HPA axis, a major neuroendocrine system that controls reactions to stress and regulates many body processes, including digestion, the immune system, mood and emotions, sexuality, and energy storage and expenditure. It is the common mechanism for interactions among glands, hormones, and parts of the midbrain that mediate the general adaptation syndrome (GAS). While steroid hormones are produced mainly in vertebrates, the physiological role of the HPA axis and corticosteroids in stress response is so fundamental that analogous systems can be found in invertebrates and monocellular organisms as well.

The key elements of the HPA axis are:

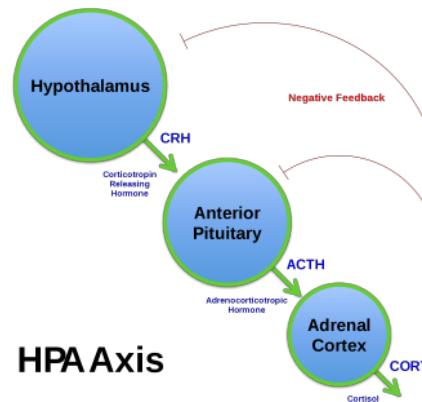


Figure 28.7: The hypothalamic-pituitary-adrenal axis (HPA axis or HTPA axis) is a complex set of direct influences and feedback interactions among three components: the hypothalamus, the pituitary gland (a pea-shaped structure located below the thalamus), and the adrenal (also called "suprarenal" glands (small, conical organs on top of the kidneys)).⁸

- The paraventricular nucleus of the hypothalamus, which contains neuroendocrine neurons which synthesize and secrete vasopressin and corticotropin-releasing hormone (CRH). These two peptides regulate:
 - The anterior lobe of the pituitary gland. In particular, CRH and vasopressin stimulate the secretion of adrenocorticotropic hormone (ACTH), once known as corticotropin. ACTH in turn acts on:
 - the adrenal cortex, which produces glucocorticoid hormones (mainly cortisol in humans) in response to stimulation by ACTH. Glucocorticoids in turn act back on the hypothalamus and pituitary (to suppress CRH and ACTH production) in a negative feedback cycle.

CRH and vasopressin are released from neurosecretory nerve terminals at the median eminence. CRH is transported to the anterior pituitary through the portal blood vessel system of the hypophyseal stalk and vasopressin is transported by axonal transport to the posterior pituitary gland. There, CRH and vasopressin act synergistically to stimulate the secretion of stored ACTH from corticotrope cells. ACTH is transported by the blood to the adrenal cortex of the adrenal gland, where it rapidly stimulates biosynthesis of corticosteroids such as cortisol from cholesterol. Cortisol is a major stress hormone and has effects on many tissues in the body, including the brain. In the brain, cortisol acts on two types of receptor – mineralocorticoid receptors and glucocorticoid receptors, and these are expressed by many different types of neurons. One important target of glucocorticoids is the hypothalamus, which is a major

controlling centre of the HPA axis.

Vasopressin can be thought of as “water conservation hormone” and is also known as “antidiuretic hormone.” It is released when the body is dehydrated and has potent water-conserving effects on the kidney. It is also a potent vasoconstrictor.

Important to the function of the HPA axis are some of the feedback loops:

- Cortisol produced in the adrenal cortex will negatively feedback to inhibit both the hypothalamus and the pituitary gland. This reduces the secretion of CRH and vasopressin, and also directly reduces the cleavage of proopiomelanocortin (POMC) into ACTH and β -endorphins.
- Epinephrine and norepinephrine (E/NE) are produced by the adrenal medulla through sympathetic stimulation and the local effects of cortisol (up-regulation enzymes to make E/NE). E/NE will positively feedback to the pituitary and increase the breakdown of POMCs into ACTH and β -endorphins.

Release of corticotropin-releasing hormone (CRH) from the hypothalamus is influenced by stress, physical activity, illness, by blood levels of cortisol and by the sleep/wake cycle (circadian rhythm). In healthy individuals, cortisol rises rapidly after wakening, reaching a peak within 30–45 minutes. It then gradually falls over the day, rising again in late afternoon. Cortisol levels then fall in late evening, reaching a trough during the middle of the night. This corresponds to the rest–activity cycle of the organism. An abnormally flattened circadian cortisol cycle has been linked with chronic fatigue syndrome, insomnia and burnout.

The HPA axis has a central role in regulating many homeostatic systems in the body, including the metabolic system, cardiovascular system, immune system, reproductive system and central nervous system. The HPA axis integrates physical and psychosocial influences in order to allow an organism to adapt effectively to its environment, use resources, and optimize survival.

Anatomical connections between brain areas such as the amygdala, hippocampus, prefrontal cortex and hypothalamus facilitate activation of the HPA axis. Sensory information arriving at the lateral aspect of the amygdala is processed and conveyed to the amygdala’s central nucleus, which then projects out to several parts of the brain involved in responses to fear. At the hypothalamus, fear-signaling impulses activate both the sympathetic nervous system and the modulating systems of the HPA axis.

Increased production of cortisol during stress results

in an increased availability of glucose in order to facilitate fighting or fleeing. As well as directly increasing glucose availability, cortisol also suppresses the highly demanding metabolic processes of the immune system, resulting in further availability of glucose.

Glucocorticoids have many important functions, including modulation of stress reactions, but in excess they can be damaging. Atrophy of the hippocampus in humans and animals exposed to severe stress is believed to be caused by prolonged exposure to high concentrations of glucocorticoids. Deficiencies of the hippocampus may reduce the memory resources available to help a body formulate appropriate reactions to stress.

28.5 The Renin-angiotensin System

The renin-angiotensin system⁹ (RAS), or renin-angiotensin-aldosterone system (RAAS), is a hormone system that regulates blood pressure and fluid and electrolyte balance, as well as systemic vascular resistance.

When renal blood flow is reduced, juxtaglomerular cells in the kidneys convert the precursor prorenin (already present in the blood) into renin and secrete it directly into circulation. Plasma renin then carries out the conversion of angiotensinogen, released by the liver, to angiotensin I. Angiotensin I is subsequently converted to angiotensin II by the angiotensin-converting enzyme (ACE) found on the surface of vascular endothelial cells, predominantly those of the lungs. Angiotensin II is a potent vasoconstrictive peptide that causes blood vessels to narrow, resulting in increased blood pressure. Angiotensin II also stimulates the secretion of the hormone aldosterone from the adrenal cortex. Aldosterone causes the renal tubules to increase the reabsorption of sodium which in consequence causes the reabsorption of water into the blood, while at the same time causing the excretion of potassium (to maintain electrolyte balance). This increases the volume of extracellular fluid in the body, which also increases blood pressure.

If the RAS is abnormally active, blood pressure will be too high. There are several types of drugs which includes ACE inhibitors, ARBs, and renin inhibitors that interrupt different steps in this system to improve blood pressure. These drugs are one of the primary ways to control high blood pressure, heart failure, kidney failure, and harmful effects of diabetes. Renin activates the renin-angiotensin system by cleaving angiotensinogen, produced by the liver, to yield angiotensin I, which is further converted into angiotensin II by ACE, the angiotensin-converting enzyme primarily within the capillaries of the lungs.

⁹https://en.wikipedia.org/wiki/Renin-angiotensin_system

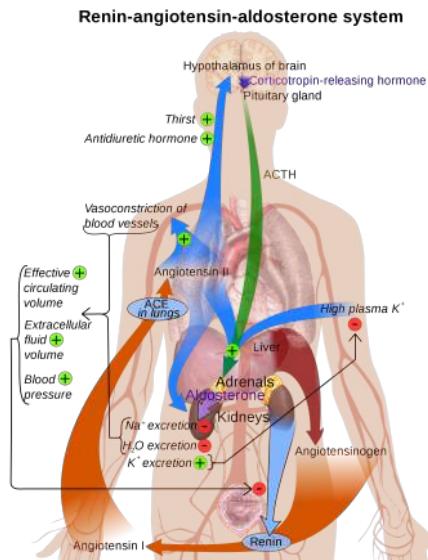


Figure 28.8: Diagram of the renin–angiotensin system.¹⁰

The system can be activated when there is a loss of blood volume or a drop in blood pressure (such as in hemorrhage or dehydration). This loss of pressure is interpreted by baroreceptors in the carotid sinus. It can also be activated by a decrease in the filtrate sodium chloride (NaCl) concentration or a decreased filtrate flow rate that will stimulate the macula densa to signal the juxtaglomerular cells to release renin.

1. If the perfusion of the juxtaglomerular apparatus in the kidney's macula densa decreases, then the juxtaglomerular cells (granular cells, modified pericytes in the glomerular capillary) release the enzyme renin.
2. Renin cleaves a decapeptide from angiotensinogen, a globular protein. The decapeptide is known as angiotensin I.
3. Angiotensin I is then converted to an octapeptide, angiotensin II by angiotensin-converting enzyme (ACE), which is thought to be found mainly in endothelial cells of the capillaries throughout the body, within the lungs and the epithelial cells of the kidneys. One study in 1992 found ACE in all blood vessel endothelial cells.
4. Angiotensin II is the major bioactive product of the renin-angiotensin system, binding to receptors on intraglomerular mesangial cells, causing these cells to contract along with the blood vessels surrounding them and causing the release of aldosterone from the zona glomerulosa in the adrenal cortex. Angiotensin II acts as an endocrine, autocrine/paracrine, and intracrine hormone.

Angiotensin I may have some minor activity, but an-

giotensin II is the major bio-active product. Angiotensin II has a variety of effects on the body:

- Throughout the body, angiotensin II is a potent vasoconstrictor of arterioles.
- In the kidneys, angiotensin II constricts glomerular arterioles, having a greater effect on efferent arterioles than afferent. As with most other capillary beds in the body, the constriction of afferent arterioles increases the arteriolar resistance, raising systemic arterial blood pressure and decreasing the blood flow. However, the kidneys must continue to filter enough blood despite this drop in blood flow, necessitating mechanisms to keep glomerular blood pressure up. To do this, angiotensin II constricts efferent arterioles, which forces blood to build up in the glomerulus, increasing glomerular pressure. The glomerular filtration rate (GFR) is thus maintained, and blood filtration can continue despite lowered overall kidney blood flow. Because the filtration fraction, which is the ratio of the glomerular filtration rate (GFR) to the renal plasma flow (RPF), has increased, there is less plasma fluid in the downstream peritubular capillaries. This in turn leads to a decreased hydrostatic pressure and increased oncotic pressure (due to unfiltered plasma proteins) in the peritubular capillaries. The effect of decreased hydrostatic pressure and increased oncotic pressure in the peritubular capillaries will facilitate increased reabsorption of tubular fluid.
- Angiotensin II decreases medullary blood flow through the vasa recta. This decreases the washout of NaCl and urea in the kidney medullary space. Thus, higher concentrations of NaCl and urea in the medulla facilitate increased absorption of tubular fluid. Furthermore, increased reabsorption of fluid into the medulla will increase passive reabsorption of sodium along the thick ascending limb of the Loop of Henle.
- Angiotensin II stimulates Na^+/H^+ exchangers located on the apical membranes (faces the tubular lumen) of cells in the proximal tubule and thick ascending limb of the loop of Henle in addition to Na^+ channels in the collecting ducts. This will ultimately lead to increased sodium reabsorption.
- Angiotensin II stimulates the hypertrophy of renal tubule cells, leading to further sodium reabsorption.
- In the adrenal cortex, angiotensin II acts to cause the release of aldosterone. Aldosterone acts on the tubules (e.g., the distal convoluted tubules and the cortical collecting ducts) in the kidneys, causing them to reabsorb more sodium and water from the urine. This increases blood volume and, therefore, increases blood pressure. In exchange

for the reabsorbing of sodium to blood, potassium is secreted into the tubules, becomes part of urine and is excreted.

- Angiotensin II causes the release of anti-diuretic hormone (ADH), also called vasopressin – ADH is made in the hypothalamus and released from the posterior pituitary gland. As its name suggests, it also exhibits vaso-constrictive properties, but its main course of action is to stimulate reabsorption of water in the kidneys. ADH also acts on the central nervous system to increase an individual's appetite for salt, and to stimulate the sensation of thirst.

These effects directly act together to increase blood pressure and are opposed by atrial natriuretic peptide (ANP).

Chapter 29

Reproductive Systems

The reproductive system of an organism, also known as the genital system, is the biological system made up of all the anatomical organs involved in sexual reproduction. Many non-living substances such as fluids, hormones, and pheromones are also important accessories to the reproductive system. Unlike most organ systems, the sexes of differentiated species often have significant differences. These differences allow for a combination of genetic material between two individuals, which allows for the possibility of greater genetic fitness of the offspring.

In mammals, the major organs of the reproductive system include the external genitalia (penis and vulva) as well as a number of internal organs, including the gamete-producing gonads (testicles and ovaries). Diseases of the human reproductive system are very common and widespread, particularly communicable sexually transmitted diseases.

Most other vertebrates have generally similar reproductive systems consisting of gonads, ducts, and openings. However, there is a great diversity of physical adaptations as well as reproductive strategies in every group of vertebrates.

Vertebrates share key elements of their reproductive systems. They all have gamete-producing organs known as gonads. In females, these gonads are then connected by oviducts to an opening to the outside of the body, typically the cloaca, but sometimes to a unique pore such as a vagina or intromittent organ.

The human reproductive system includes the male reproductive system which functions to produce and deposit sperm; and the female reproductive system which functions to produce egg cells, and to protect and nourish the fetus until birth. Humans have a high level of sexual differentiation. In addition to differences in nearly every reproductive organ, there are numerous differences in typical secondary sex characteristics.

Most mammal reproductive systems are similar, however, there are some notable differences between the

non-human mammals and humans. For instance, most male mammals have a penis which is stored internally until erect, and most have a penis bone or baculum. Additionally, males of most species do not remain continually sexually fertile as humans do. Like humans, most groups of mammals have descended testicles found within a scrotum, however, others have descended testicles that rest on the ventral body wall, and a few groups of mammals, such as elephants, have undescended testicles found deep within their body cavities near their kidneys.

The reproductive system of marsupials is unique in that the female has two vaginæ, both of which open externally through one orifice but lead to different compartments within the uterus; males usually have a two-pronged penis, which corresponds to the females' two vaginæ. Marsupials typically develop their offspring in an external pouch containing teats to which their newborn young (joeys) attach themselves for post uterine development. Also, marsupials have a unique prepenial scrotum. The 15mm (5/8 in) long newborn joey instinctively crawls and wriggles the several inches (15 cm), while clinging to fur, on the way to its mother's pouch.

The uterus and vagina are unique to mammals with no homologue in birds, reptiles, amphibians, or fish. In place of the uterus the other vertebrate groups have an unmodified oviduct leading directly to a cloaca, which is a shared exit-hole for gametes, urine, and feces. Monotremes (i.e. platypus and echidnas), a group of egg-laying mammals, also lack a uterus and vagina, and in that respect have a reproductive system resembling that of a reptile.

Male and female birds have a cloaca, an opening through which eggs, sperm, and wastes pass. Intercourse is performed by pressing the lips of the cloacæ together, which is sometimes known as intromittent organ which is known as a phallus that is analogous to the mammals' penis. The female lays amniotic eggs in which the young fetus continues to develop after it leaves the female's body. Unlike most vertebrates female birds typically have

only one functional ovary and oviduct. As a group, birds, like mammals, are noted for their high level of parental care.

Reptiles are almost all sexually dimorphic, and exhibit internal fertilization through the cloaca. Some reptiles lay eggs while others are ovoviparous (animals that deliver live young). Reproductive organs are found within the cloaca of reptiles. Most male reptiles have copulatory organs, which are usually retracted or inverted and stored inside the body. In turtles and crocodilians, the male has a single median penis-like organ, while male snakes and lizards each possess a pair of penis-like organs.

Most amphibians exhibit external fertilization of eggs, typically within the water, though some amphibians such as caecilians have internal fertilization. All have paired, internal gonads, connected by ducts to the cloaca.

Fish exhibit a wide range of different reproductive strategies. Most fish, however, are oviparous and exhibit external fertilization. In this process, females use their cloaca to release large quantities of their gametes, called spawn into the water and one or more males release "milt", a white fluid containing many sperm over the unfertilized eggs. Other species of fish are oviparous and have internal fertilization aided by pelvic or anal fins that are modified into an intromittent organ analogous to the human penis. A small portion of fish species are either viviparous or ovoviparous, and are collectively known as livebearers.

Fish gonads are typically pairs of either ovaries or testes. Most fish are sexually dimorphic but some species are hermaphroditic or unisexual.

Invertebrates have an extremely diverse array of reproductive systems, the only commonality may be that they all lay eggs. Also, aside from cephalopods and arthropods, nearly all other invertebrates are hermaphroditic and exhibit external fertilization.

All cephalopods are sexually dimorphic and reproduce by laying eggs. Most cephalopods have semi-internal fertilization, in which the male places his gametes inside the female's mantle cavity or pallial cavity to fertilize the ova found in the female's single ovary. Likewise, male cephalopods have only a single testicle. In the female of most cephalopods the nidamental glands aid in development of the egg.

The "penis" in most unshelled male cephalopods (*Coleoidea*) is a long and muscular end of the gonoduct used to transfer spermatophores to a modified arm called a hectocotylus. That in turn is used to transfer the spermatophores to the female. In species where the hectocotylus is missing, the "penis" is long and able to extend beyond the mantle cavity and transfer the

spermatophores directly to the female.

Most insects reproduce oviparously, i.e. by laying eggs. The eggs are produced by the female in a pair of ovaries. Sperm, produced by the male in one testis or more commonly two, is transmitted to the female during mating by means of external genitalia. The sperm is stored within the female in one or more spermathecae. At the time of fertilization, the eggs travel along oviducts to be fertilized by the sperm and are then expelled from the body ("laid"), in most cases via an ovipositor.

Arachnids may have one or two gonads, which are located in the abdomen. The genital opening is usually located on the underside of the second abdominal segment. In most species, the male transfers sperm to the female in a package, or spermatophore. Complex courtship rituals have evolved in many arachnids to ensure the safe delivery of the sperm to the female.

Arachnids usually lay yolk eggs, which hatch into immatures that resemble adults. Scorpions, however, are either ovoviparous or viviparous, depending on species, and bear live young.

Among all living organisms, flowers, which are the reproductive structures of angiosperms, are the most varied physically and show a correspondingly great diversity in methods of reproduction. Plants that are not flowering plants (green algae, mosses, liverworts, hornworts, ferns and gymnosperms such as conifers) also have complex interplays between morphological adaptation and environmental factors in their sexual reproduction. The breeding system, or how the sperm from one plant fertilizes the ovum of another, depends on the reproductive morphology, and is the single most important determinant of the genetic structure of nonclonal plant populations. Christian Konrad Sprengel (1793) studied the reproduction of flowering plants and for the first time it was understood that the pollination process involved both biotic and abiotic interactions.

Fungal reproduction is complex, reflecting the differences in lifestyles and genetic makeup within this diverse kingdom of organisms. It is estimated that a third of all fungi reproduce using more than one method of propagation; for example, reproduction may occur in two well-differentiated stages within the life cycle of a species, the teleomorph and the anamorph. Environmental conditions trigger genetically determined developmental states that lead to the creation of specialized structures for sexual or asexual reproduction. These structures aid reproduction by efficiently dispersing spores or spore-containing propagules.

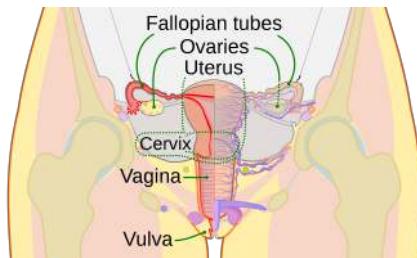


Figure 29.1: Diagram of the female reproductive system.¹

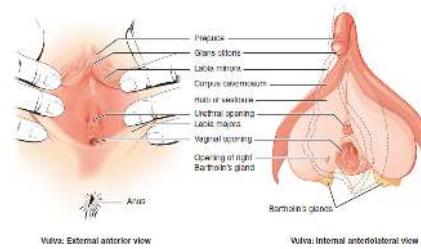


Figure 29.2: External and internal views of the vulva.²

29.1 The Human Reproductive System

The human reproductive system includes the male reproductive system which functions to produce and deposit sperm; and the female reproductive system which functions to produce egg cells, and to protect and nourish the fetus until birth. Humans have a high level of sexual differentiation. In addition to differences in nearly every reproductive organ, there are numerous differences in typical secondary sex characteristics.

29.2 The Female Reproductive System

The human female reproductive system is a series of organs primarily located inside the body and around the pelvic region of a female that contribute towards the reproductive process. The human female reproductive system contains three main parts: the vulva, which leads to the vagina, the vaginal opening, to the uterus; the uterus, which holds the developing fetus; and the ovaries, which produce the female's ova. The breasts are involved during the parenting stage of reproduction, but in most classifications they are not considered to be part of the female reproductive system.

The vagina meets the outside at the vulva, which also includes the labia, clitoris and urethra; during intercourse this area is lubricated by mucus secreted by the Bartholin's glands. The vagina is attached to the uterus through the cervix, while the uterus is attached to the ovaries via the Fallopian tubes. Each ovary contains hundreds of egg cells or ova (singular ovum).

Approximately every 28 days, the pituitary gland releases a hormone that stimulates some of the ova to develop and grow. One ovum is released and it passes through the Fallopian tube into the uterus. Hormones produced by the ovaries prepare the uterus to receive the ovum. The lining of the uterus, called the endometrium, and unfertilized ova are shed each cycle through the

process of menstruation. If the ovum is fertilized by sperm, it attaches to the endometrium and the fetus develops.

29.3 The Male Reproductive System

The male reproductive system is a series of organs located outside the body and around the pelvis region of a male that contribute towards the reproduction process. The primary direct function of the male reproductive system is to provide the male sperm for fertilization of the ovum.

The major reproductive organs of the male can be grouped into three categories. The first category produces and stores sperm (spermatozoa). These are produced in the testes, which are housed in the temperature-regulating scrotum; immature sperm then travel to the epididymis for development and storage. The second category are the ejaculatory fluid producing glands which include the Cowper's gland (also called bulbo-urethral gland), seminal vesicles, prostate, and vas deferens. The final category are those used for copulation and deposition of the sperm within the female; these include the penis, urethra, and vas deferens.

Major secondary sexual characteristics include: larger, more muscular stature, deepened voice, facial and body hair, broad shoulders, and development of an Adam's apple. An important sexual hormone of males is androgen, and particularly testosterone.

The testes release a hormone that controls the development of sperm. This hormone is also responsible for the development of physical characteristics in men such as facial hair and a deep voice.

The development of the reproductive system and the development of the urinary system are closely tied in with the development of the human fetus. Despite the differences between the adult female and male are derived from the intermediate mesoderm. The three main fetal precursors of the reproductive organs are the Wolffian duct, the Müllerian ducts, and the gonads. Endocrine hormones are a well-known and critical controlling factor in the normal

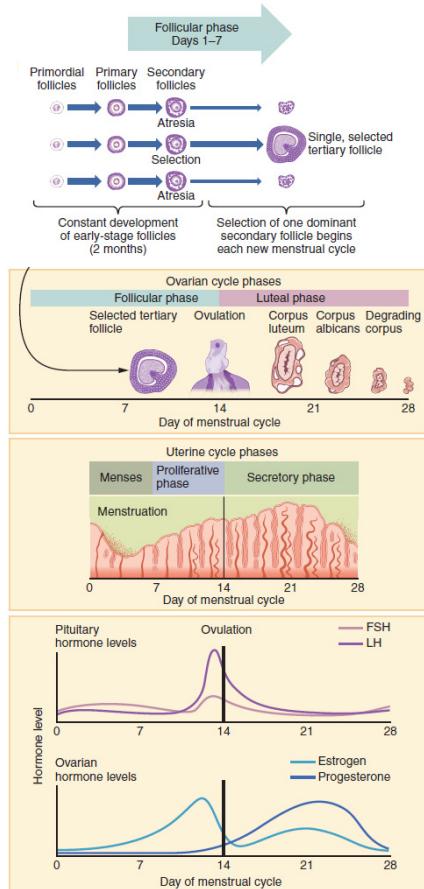


Figure 29.3: The progression of the menstrual cycle and the different hormones regulating it.³

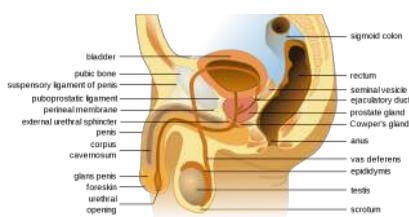


Figure 29.4: Diagram of the male reproductive system.⁴

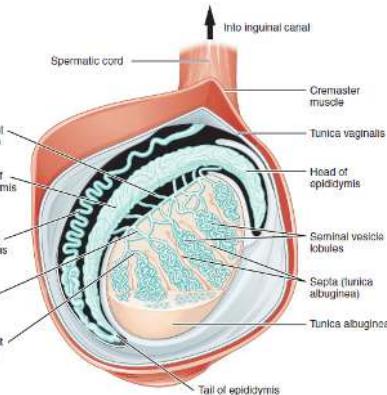


Figure 29.5: Diagram of inner structures of testes.⁵

differentiation of the reproductive system.

The Wolffian duct forms the epididymis, vas deferens, ductus deferens, ejaculatory duct, and seminal vesicle in the male reproductive system, but essentially disappears in the female reproductive system. The reverse is true for the Müllerian duct, as it essentially disappears in the male reproductive system and forms the Fallopian tubes, uterus, and vagina in the female system. In both sexes the gonads go on to form the testes and ovaries; because they are derived from the same undeveloped structure, they are considered homologous organs. There are a number of other homologous structures shared between male and female reproductive systems. However, despite the similarity in function of the female Fallopian tubes and the male epididymis and vas deferens, they are not homologous but rather analogous structures as they arise from different fetal structures.

Gametes are produced within the gonads through a process known as gametogenesis. This occurs when certain types of germ cells undergo meiosis to split the normal diploid number of chromosomes ($n=46$) into haploid cells containing only 23 chromosomes.

In males, this process is known as spermatogenesis, and takes place only after puberty in the seminiferous tubules of the testes. The immature spermatozoa or sperm are then sent to the epididymis, where they gain a tail, enabling motility. Each of the original diploid germ cells or primary spermatocytes forms four functional gametes. The production and survival of sperms require a temperature below the normal core body temperature. Since the scrotum, where the testes is present, is situated outside the body cavity, it provides a temperature about 3 °C below normal body temperature.

In females, gametogenesis is known as oogenesis; this occurs in the ovarian follicles of the ovaries. This process does not produce mature ovum until puberty. In contrast

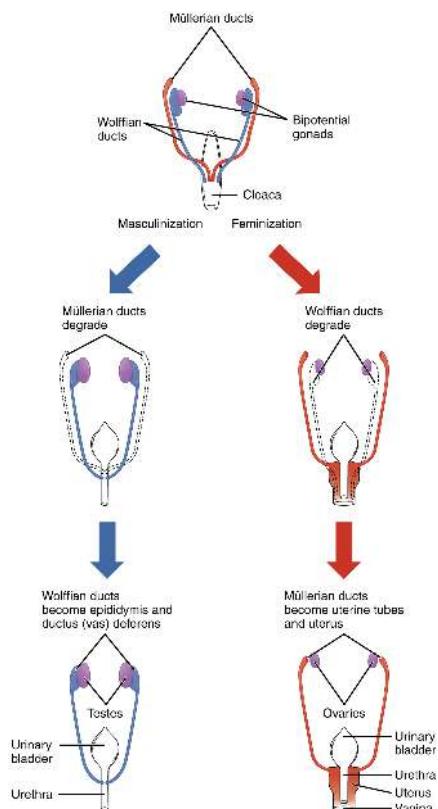


Figure 29.6: Differentiation of the male and female reproductive systems from a common origin.⁶

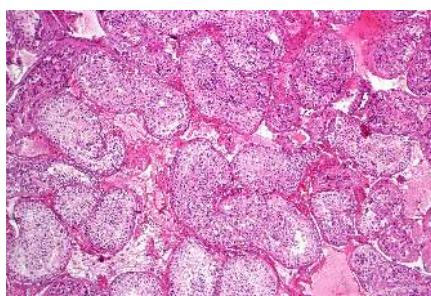


Figure 29.7: Normal spermatogenesis, testis biopsy.⁷

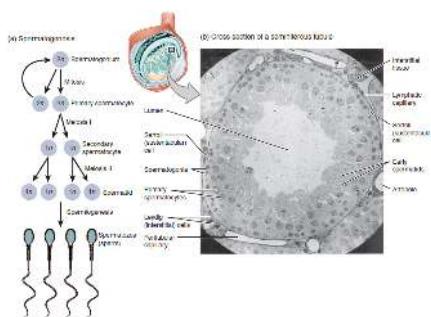


Figure 29.8: The process of spermatogenesis as the cells progress from primary spermatocytes, to secondary spermatocytes, to spermatids, to Sperm.⁸

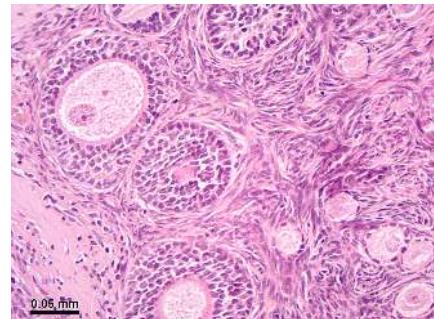


Figure 29.9: Micrograph of the ovarian cortex from a rhesus monkey showing several round follicles embedded in a matrix of stromal cells. A secondary follicle sectioned through the nucleus of an oocyte is at the upper left, and earlier stage follicles are at the lower right. The tissue was stained with the dyes hematoxylin and eosin.⁹

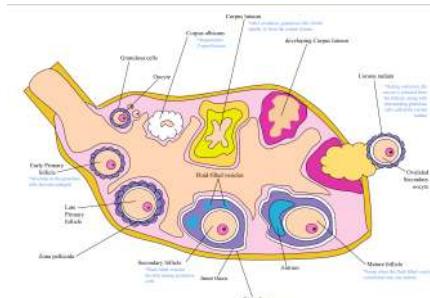


Figure 29.10: The process of ovulation and gamete production, oogenesis, in a human ovary.¹⁰

with males, each of the original diploid germ cells or primary oocytes will form only one mature ovum, and three polar bodies which are not capable of fertilization. It has long been understood that in females, unlike males, all of the primary oocytes ever found in a female will be created prior to birth, and that the final stages of ova production will then not resume until puberty. However, recent scientific research has challenged that hypothesis. This new research indicates that in at least some species of mammal, oocytes continue to be replenished in females well after birth.

Chapter 30

Development And Aging

Developmental biology is the study of the process by which animals and plants grow and develop.

Ageing or aging is the process of becoming older. The term refers especially to human beings, many animals, and fungi, whereas for example bacteria, perennial plants and some simple animals are potentially biologically immortal. In the broader sense, aging can refer to single cells within an organism which have ceased dividing (cellular senescence) or to the population of a species (population aging).

30.1 Animal Development

The main processes involved in the embryonic development of animals are: tissue patterning (via regional specification and patterned cell differentiation); tissue growth; and tissue morphogenesis.

- Regional specification refers to the processes that create spatial pattern in a ball or sheet of initially similar cells. This generally involves the action of cytoplasmic determinants, located within parts of the fertilized egg, and of inductive signals emitted from signaling centers in the embryo. The early stages of regional specification do not generate functional differentiated cells, but cell populations committed to develop to a specific region or part of the organism. These are defined by the expression of specific combinations of transcription factors.
- Cell differentiation relates specifically to the formation of functional cell types such as nerve, muscle, secretory epithelia etc. Differentiated cells contain large amounts of specific proteins associated with the cell function.
- Morphogenesis relates to the formation of three-dimensional shape. It mainly involves the orchestrated movements of cell sheets and of individual cells. Morphogenesis is important for creating the three germ layers of the early embryo (ectoderm, mesoderm and endoderm) and for building up

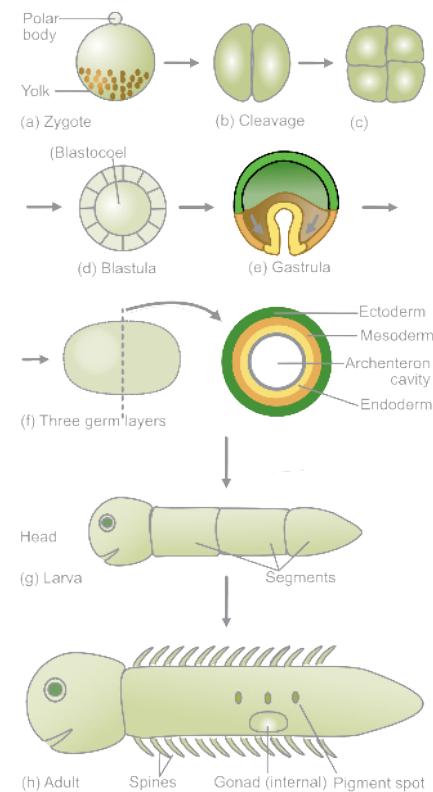


Figure 30.1: Generalized scheme of the embryonic development of animals.¹

complex structures during organ development.

- Tissue growth involves both an overall increase in tissue size, and also the differential growth of parts (allometry) which contributes to morphogenesis. Growth mostly occurs through cell proliferation but also through changes of cell size or the deposition of extracellular materials.

The development of plants involves similar processes to that of animals. However plant cells are mostly immobile so morphogenesis is achieved by differential growth, with-

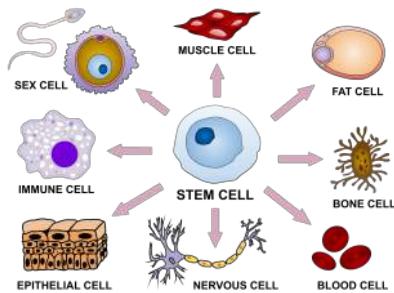


Figure 30.2: Stem cell differentiation into various tissue types.²

out cell movements. Also, the inductive signals and the genes involved are different from those that control animal development.

30.1.1 Cellular differentiation

Cellular differentiation is the process in which a cell changes from one cell type to another. Usually, the cell changes to a more specialized type. Differentiation occurs numerous times during the development of a multicellular organism as it changes from a simple zygote to a complex system of tissues and cell types. Differentiation continues in adulthood as adult stem cells divide and create fully differentiated daughter cells during tissue repair and during normal cell turnover. Some differentiation occurs in response to antigen exposure. Differentiation dramatically changes a cell's size, shape, membrane potential, metabolic activity, and responsiveness to signals. These changes are largely due to highly controlled modifications in gene expression and are the study of epigenetics. With a few exceptions, cellular differentiation almost never involves a change in the DNA sequence itself. Thus, different cells can have very different physical characteristics despite having the same genome.

A specialized type of differentiation, known as 'terminal differentiation', is of importance in some tissues, for example vertebrate nervous system, striated muscle, epidermis and gut. During terminal differentiation, a precursor cell formerly capable of cell division, permanently leaves the cell cycle, dismantles the cell cycle machinery and often expresses a range of genes characteristic of the cell's final function (e.g. myosin and actin for a muscle cell). Differentiation may continue to occur after terminal differentiation if the capacity and functions of the cell undergo further changes.

Among dividing cells, there are multiple levels of cell potency, the cell's ability to differentiate into other cell types. A greater potency indicates a larger number of cell types that can be derived. A cell that can differentiate into

all cell types, including the placental tissue, is known as totipotent. In mammals, only the zygote and subsequent blastomeres are totipotent, while in plants, many differentiated cells can become totipotent with simple laboratory techniques. A cell that can differentiate into all cell types of the adult organism is known as pluripotent. Such cells are called meristematic cells in higher plants and embryonic stem cells in animals, though some groups report the presence of adult pluripotent cells. Virally induced expression of four transcription factors Oct4, Sox2, c-Myc, and Klf4 (Yamanaka factors) is sufficient to create pluripotent (iPS) cells from adult fibroblasts. A multipotent cell is one that can differentiate into multiple different, but closely related cell types. Oligopotent cells are more restricted than multipotent, but can still differentiate into a few closely related cell types. Finally, unipotent cells can differentiate into only one cell type, but are capable of self-renewal. In cytopathology, the level of cellular differentiation is used as a measure of cancer progression. "Grade" is a marker of how differentiated a cell in a tumor is.

30.1.2 Morphogenesis

Morphogenesis (from the Greek morphē shape and genesis creation, literally "the generation of form") is the biological process that causes a cell, tissue or organism to develop its shape. It is one of three fundamental aspects of developmental biology along with the control of tissue growth and patterning of cellular differentiation.

The process controls the organized spatial distribution of cells during the embryonic development of an organism. Morphogenesis can take place also in a mature organism, such as in the normal maintenance of tissue homeostasis by stem cells or in regeneration of tissues after damage. Cancer is an example of highly abnormal and pathological tissue morphogenesis. Morphogenesis also describes the development of unicellular life forms that do not have an embryonic stage in their life cycle. Morphogenesis is essential for the evolution of a new form.

Morphogenesis is a mechanical process involving forces that generate mechanical stress, strain, and movement of cells, and can be induced by genetic programs according to the spatial patterning of cells within tissues.

Several types of molecules are important in morphogenesis. Morphogens are soluble molecules that can diffuse and carry signals that control cell differentiation via concentration gradients. Morphogens typically act through binding to specific protein receptors. An important class of molecules involved in morphogenesis are transcription factor proteins that determine the fate of cells by interacting with DNA. These can be coded for by master regulatory genes, and either activate or deactivate the transcription of other genes; in turn, these secondary gene products can

regulate the expression of still other genes in a regulatory cascade of gene regulatory networks. At the end of this cascade are classes of molecules that control cellular behaviors such as cell migration, or, more generally, their properties, such as cell adhesion or cell contractility. For example, during gastrulation, clumps of stem cells switch off their cell-to-cell adhesion, become migratory, and take up new positions within an embryo where they again activate specific cell adhesion proteins and form new tissues and organs. Developmental signaling pathways implicated in morphogenesis include Wnt, Hedgehog, and ephrins.

At a tissue level, ignoring the means of control, morphogenesis arises because of cellular proliferation and motility. Morphogenesis also involves changes in the cellular structure or how cells interact in tissues. These changes can result in tissue elongation, thinning, folding, invasion or separation of one tissue into distinct layers. The latter case is often referred as cell sorting. Cell "sorting out" consists of cells moving so as to sort into clusters that maximize contact between cells of the same type. The ability of cells to do this has been proposed to arise from differential cell adhesion by Malcolm Steinberg through his differential adhesion hypothesis. Tissue separation can also occur via more dramatic cellular differentiation events during which epithelial cells become mesenchymal (see Epithelial-mesenchymal transition). Mesenchymal cells typically leave the epithelial tissue as a consequence of changes in cell adhesive and contractile properties. Following epithelial-mesenchymal transition, cells can migrate away from an epithelium and then associate with other similar cells in a new location. In plants, cellular morphogenesis is tightly linked to the chemical composition and the mechanical properties of the cell wall.

The puzzle of how embryonic development was controlled began to be solved using the fruit fly *Drosophila melanogaster* as a model organism. The step-by-step control of its embryogenesis was visualized by attaching fluorescent dyes of different colours to specific types of protein made by genes expressed in the embryo. A dye such as green fluorescent protein, originally from a jellyfish, was typically attached to an antibody specific to a fruit fly protein, forming a precise indicator of where and when that protein appeared in the living embryo.

Using such a technique, in 1994 Walter Gehring found that the pax-6 gene, vital for forming the eyes of fruit flies, exactly matches an eye-forming gene in mice and humans. The same gene was quickly found in many other groups of animals, such as squid, a cephalopod mollusc. Biologists including Ernst Mayr had believed that eyes had arisen in the animal kingdom at least 40 times, as the anatomy of different types of eye varies widely. For example, the fruit fly's compound eye is made of hundreds of small lensed structures (ommatidia); the human eye has a blind spot where the optic nerve enters the eye, and the nerve fibres run over the surface of the retina, so light has to pass through a layer of nerve fibres before reaching the detector cells in the retina, so the structure is effectively "upside-down"; in contrast, the cephalopod eye has the retina, then a layer of nerve fibres, then the wall of the eye "the right way around". The evidence of pax-6, however, was that the same genes controlled the development of the eyes of all these animals, suggesting that they all evolved from a common ancestor. Ancient genes had been conserved through millions of years of evolution to create dissimilar structures for similar functions, demonstrating deep homology between structures once thought to be purely analogous. This has caused a radical revision of the meaning of homology in evolutionary biology.

30.2 Genetic Control of Development

Roughly spherical eggs of different animals give rise to extremely different bodies, from jellyfish to lobsters, butterflies to elephants. Many of these organisms share the same structural genes for body-building proteins like collagen and enzymes, but biologists had expected that each group of animals would have its own rules of development. The surprise of evolutionary developmental (evo-devo) biology is that the shaping of bodies is controlled by a rather small percentage of genes, and that these regulatory genes are ancient, shared by all animals. The giraffe does not have a gene for a long neck, any more than the elephant has a gene for a big body. Their bodies are patterned by a system of switching which causes development of different features to begin earlier or later, to occur in this or that part of the embryo, and to continue for more or less time.

A small fraction of the genes in an organism's genome control the organism's development. These genes are called the developmental-genetic toolkit. They are highly conserved among phyla, meaning that they are ancient and very similar in widely separated groups of animals. Differences in deployment of toolkit genes affect the body plan and the number, identity, and pattern of body parts. Most toolkit genes are parts of signalling pathways: they encode transcription factors, cell adhesion proteins, cell surface receptor proteins and signalling ligands that bind to them, and secreted morphogens that diffuse through the embryo. All of these help to define the fate of undifferentiated cells in the embryo. Together, they generate the patterns in time and space which shape the embryo, and ultimately form the body plan of the organism. Among the most important toolkit genes are the Hox genes. These transcription factors contain the homeobox protein-binding DNA motif, also found in other

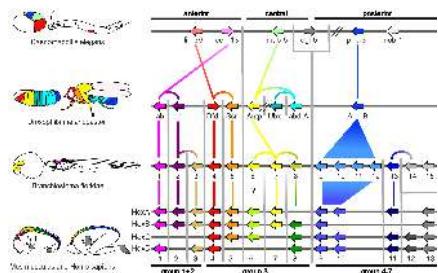


Figure 30.3: Homologous hox genes in such different animals as insects and vertebrates control embryonic development and hence the form of adult bodies. These genes have been highly conserved through hundreds of millions of years of evolution.³

toolkit genes, and create the basic pattern of the body along its front-to-back axis. Hox genes determine where repeating parts, such as the many vertebrae of snakes, will grow in a developing embryo or larva. Pax-6, already mentioned, is a classic toolkit gene. Homeobox genes are also found in plants, implying they are common to all eukaryotes

The protein products of the regulatory toolkit are reused not by duplication and modification, but by a complex mosaic of pleiotropy, being applied unchanged in many independent developmental processes, giving pattern to many dissimilar body structures. The loci of these pleiotropic toolkit genes have large, complicated and modular cis-regulatory elements. For example, while a non-pleiotropic rhodopsin gene in the fruit fly has a cis-regulatory element just a few hundred base pairs long, the pleiotropic eyeless cis-regulatory region contains 6 cis-regulatory elements in over 7000 base pairs. The regulatory networks involved are often very large. Each regulatory protein controls “scores to hundreds” of cis-regulatory elements. For instance, 67 fruit fly transcription factors controlled on average 124 target genes each. All this complexity enables genes involved in the development of the embryo to be switched on and off at exactly the right times and in exactly the right places. Some of these genes are structural, directly forming enzymes, tissues and organs of the embryo. But many others are themselves regulatory genes, so what is switched on is often a precisely-timed cascade of switching, involving turning on one developmental process after another in the developing embryo.

Such a cascading regulatory network has been studied in detail in the development of the fruit fly embryo. The young embryo is oval in shape, like a rugby ball. A small number of genes produce messenger RNAs that set up concentration gradients along the long axis of the embryo. In the early embryo, the bicoid and hunchback genes are at

high concentration near the anterior end, and give pattern to the future head and thorax; the caudal and nanos genes are at high concentration near the posterior end, and give pattern to the hindmost abdominal segments. The effects of these genes interact; for instance, the Bicoid protein blocks the translation of caudal’s messenger RNA, so the Caudal protein concentration becomes low at the anterior end. Caudal later switches on genes which create the fly’s hindmost segments, but only at the posterior end where it is most concentrated.

The Bicoid, Hunchback and Caudal proteins in turn regulate the transcription of gap genes such as giant, knirps, Krüppel, and tailless in a striped pattern, creating the first level of structures that will become segments. The proteins from these in turn control the pair-rule genes, which in the next stage set up 7 bands across the embryo’s long axis. Finally, the segment polarity genes such as engrailed split each of the 7 bands into two, creating 14 future segments.

This process explains the accurate conservation of toolkit gene sequences, which has resulted in deep homology and functional equivalence of toolkit proteins in dissimilar animals (seen, for example, when a mouse protein controls fruit fly development). The interactions of transcription factors and cis-regulatory elements, or of signalling proteins and receptors, become locked in through multiple usages, making almost any mutation deleterious and hence eliminated by natural selection.

30.3 Human Embryonic Development

Human embryonic development, or human embryogenesis, refers to the development and formation of the human embryo. It is characterised by the processes of cell division and cellular differentiation of the embryo that occurs during the early stages of development. In biological terms, the development of the human body entails growth from a one-celled zygote to an adult human being. Fertilisation occurs when the sperm cell successfully enters and fuses with an egg cell (ovum). The genetic material of the sperm and egg then combine to form a single cell called a zygote and the germinal stage of development commences. Embryonic development in the human, covers the first eight weeks of development; at the beginning of the ninth week the embryo is termed a fetus. Human embryology is the study of this development during the first eight weeks after fertilisation. The normal period of gestation (pregnancy) is about nine months or 40 weeks.

The germinal stage refers to the time from fertilization through the development of the early embryo until implantation is completed in the uterus. The germinal stage takes around 10 days. During this stage, the zygote begins to divide, in a process called cleavage. A blastocyst is then



Figure 30.4: Stages during pregnancy. Embryonic development is marked in green. Weeks and months are numbered by gestation.⁴

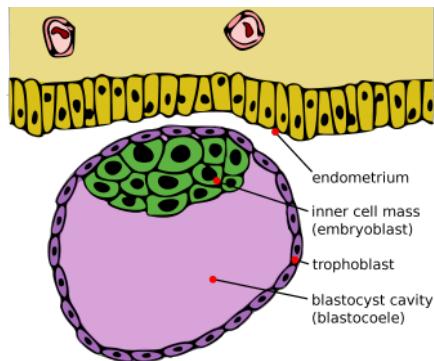


Figure 30.5: Blastocyst with an inner cell mass and trophoblast.⁵

formed and implanted in the uterus. Embryogenesis continues with the next stage of gastrulation, when the three germ layers of the embryo form in a process called histogenesis, and the processes of neurulation and organogenesis follow.

In comparison to the embryo, the fetus has more recognizable external features and a more complete set of developing organs. The entire process of embryogenesis involves coordinated spatial and temporal changes in gene expression, cell growth and cellular differentiation. A nearly identical process occurs in other species, especially among chordates.

Fertilization takes place when the spermatozoon has successfully entered the ovum and the two sets of genetic material carried by the gametes fuse together, resulting in

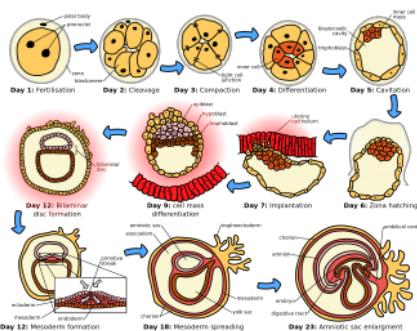


Figure 30.6: The initial stages of human embryonic development.⁶



Figure 30.7: An embryo at the 8-cell stage, at 3 days.⁷

the zygote (a single diploid cell). This usually takes place in the ampulla of one of the fallopian tubes. The zygote contains the combined genetic material carried by both the male and female gametes which consists of the 23 chromosomes from the nucleus of the ovum and the 23 chromosomes from the nucleus of the sperm. The 46 chromosomes undergo changes prior to the mitotic division which leads to the formation of the embryo having two cells.

Successful fertilization is enabled by three processes, which also act as controls to ensure species-specificity. The first is that of chemotaxis which directs the movement of the sperm towards the egg. Secondly there is an adhesive compatibility between the sperm and the egg. With the sperm adhered to the ovum, the third process of acrosomal reaction takes place; the front part of the spermatozoan head is capped by an acrosome which contains digestive enzymes to break down the zona pellucida and allow its entry. The entry of the sperm causes calcium to be released which blocks entry to other sperm cells. A parallel reaction takes place in the ovum called the zona reaction. This sees the release of cortical granules that release enzymes which digest sperm receptor proteins, thus preventing polyspermy. The granules also fuse with the plasma membrane and modify the zona pellucida in such a way as to prevent further sperm entry.

The beginning of the cleavage process is marked when the zygote divides through mitosis into two cells. This mitosis continues and the first two cells divide into four cells, then into eight cells and so on. Each division takes from 12 to 24 hours. The zygote is large compared to any other cell and undergoes cleavage without any overall increase in size. This means that with each successive subdivision, the ratio of nuclear to cytoplasmic material increases. Initially the dividing cells, called blastomeres (blastos Greek for sprout), are undifferentiated and aggregated into a sphere enclosed within the membrane of glycoproteins (termed the zona pellucida) of the ovum. When eight blastomeres have formed they begin to develop gap junctions, enabling them to develop in an integrated way and co-ordinate their

response to physiological signals and environmental cues.

When the cells number around sixteen the solid sphere of cells within the zona pellucida is referred to as a morula. At this stage the cells start to bind firmly together in a process called compaction, and cleavage continues as cellular differentiation.

Cleavage itself is the first stage in blastulation, the process of forming the blastocyst. Cells differentiate into an outer layer of cells (collectively called the trophoblast) and an inner cell mass. With further compaction the individual outer blastomeres, the trophoblasts, become indistinguishable. They are still enclosed within the zona pellucida. This compaction serves to make the structure watertight, containing the fluid that the cells will later secrete. The inner mass of cells differentiate to become embryoblasts and polarise at one end. They close together and form gap junctions, which facilitate cellular communication. This polarisation leaves a cavity, the blastocoel, creating a structure that is now termed the blastocyst. (In animals other than mammals, this is called the blastula.) The trophoblasts secrete fluid into the blastocoel. The resulting increase in size of the blastocyst causes it to hatch through the zona pellucida, which then disintegrates.

The inner cell mass will give rise to the pre-embryo, the amnion, yolk sac and allantois, while the fetal part of the placenta will form from the outer trophoblast layer. The embryo plus its membranes is called the conceptus, and by this stage the conceptus has reached the uterus. The zona pellucida ultimately disappears completely, and the now exposed cells of the trophoblast allow the blastocyst to attach itself to the endometrium, where it will implant. The formation of the hypoblast and epiblast, which are the two main layers of the bilaminar germ disc, occurs at the beginning of the second week. Either the embryoblast or the trophoblast will turn into two sub-layers. The inner cells will turn into the hypoblast layer, which will surround the other layer, called the epiblast, and these layers will form the embryonic disc that will develop into the embryo. The trophoblast will also develop two sub-layers: the cytotrophoblast, which is in front of the syncytiotrophoblast, which in turn lies within the endometrium. Next, another layer called the exocoelomic membrane or Heuser's membrane will appear and surround the cytotrophoblast, as well as the primitive yolk sac. The syncytiotrophoblast will grow and will enter a phase called lacunar stage, in which some vacuoles will appear and be filled by blood in the following days. The development of the yolk sac starts with the hypoblastic flat cells that form the exocoelomic membrane, which will coat the inner part of the cytotrophoblast to form the primitive yolk sac. An erosion of the endothelial lining of the maternal capillaries by the syncytiotrophoblastic cells of the sinusoids will form where the blood will be-

gin to penetrate and flow through the trophoblast to give rise to the uteroplacental circulation. Subsequently new cells derived from yolk sac will be established between trophoblast and exocoelomic membrane and will give rise to extra-embryonic mesoderm, which will form the chorionic cavity.

At the end of the second week of development, some cells of the trophoblast penetrate and form rounded columns into the syncytiotrophoblast. These columns are known as primary villi. At the same time, other migrating cells form into the exocoelomic cavity a new cavity named the secondary or definitive yolk sac, smaller than the primitive yolk sac.

After ovulation, the endometrial lining becomes transformed into a secretory lining in preparation of accepting the embryo. It becomes thickened, with its secretory glands becoming elongated, and is increasingly vascular. This lining of the uterine cavity (or womb) is now known as the decidua, and it produces a great number of large decidual cells in its increased interglandular tissue. The blastomeres in the blastocyst are arranged into an outer layer called the trophoblast. The trophoblast then differentiates into an inner layer, the cytotrophoblast, and an outer layer, the syncytiotrophoblast. The cytotrophoblast contains cuboidal epithelial cells and is the source of dividing cells, and the syncytiotrophoblast is a syncytial layer without cell boundaries.

The syncytiotrophoblast implants the blastocyst in the decidual epithelium by projections of chorionic villi, forming the embryonic part of the placenta. The placenta develops once the blastocyst is implanted, connecting the embryo to the uterine wall. The decidua here is termed the decidua basalis; it lies between the blastocyst and the myometrium and forms the maternal part of the placenta. The implantation is assisted by hydrolytic enzymes that erode the epithelium. The syncytiotrophoblast also produces human chorionic gonadotropin, a hormone that stimulates the release of progesterone from the corpus luteum. Progesterone enriches the uterus with a thick lining of blood vessels and capillaries so that it can oxygenate and sustain the developing embryo. The uterus liberates sugar from stored glycogen from its cells to nourish the embryo. The villi begin to branch and contain blood vessels of the embryo. Other villi, called terminal or free villi, exchange nutrients. The embryo is joined to the trophoblastic shell by a narrow connecting stalk that develops into the umbilical cord to attach the placenta to the embryo. Arteries in the decidua are remodelled to increase the maternal blood flow into the intervillous spaces of the placenta, allowing gas exchange and the transfer of nutrients to the embryo. Waste products from the embryo will diffuse across the placenta.

As the syncytiotrophoblast starts to penetrate the uterine wall, the inner cell mass (embryoblast) also develops. The inner cell mass is the source of embryonic stem cells, which are pluripotent and can develop into any one of the three germ layer cells, and which have the potency to give rise to all the tissues and organs.

The embryoblast forms an embryonic disc, which is a bilaminar disc of two layers, an upper layer called the epiblast (primitive ectoderm) and a lower layer called the hypoblast (primitive endoderm). The disc is stretched between what will become the amniotic cavity and the yolk sac. The epiblast is adjacent to the trophoblast and made of columnar cells; the hypoblast is closest to the blastocyst cavity and made of cuboidal cells. The epiblast migrates away from the trophoblast downwards, forming the amniotic cavity, the lining of which is formed from amnioblasts developed from the epiblast. The hypoblast is pushed down and forms the yolk sac (exocoelomic cavity) lining. Some hypoblast cells migrate along the inner cytotrophoblast lining of the blastocoel, secreting an extracellular matrix along the way. These hypoblast cells and extracellular matrix are called Heuser's membrane (or the exocoelomic membrane), and they cover the blastocoel to form the yolk sac (or exocoelomic cavity). Cells of the hypoblast migrate along the outer edges of this reticulum and form the extraembryonic mesoderm; this disrupts the extraembryonic reticulum. Soon pockets form in the reticulum, which ultimately coalesce to form the chorionic cavity (extraembryonic coelom). The primitive streak, a linear band of cells formed by the migrating epiblast, appears, and this marks the beginning of gastrulation, which takes place around the seventeenth day (week 3) after fertilisation. The process of gastrulation reorganises the two-layer embryo into a three-layer embryo, and also gives the embryo its specific head-to-tail, and front-to-back orientation, by way of the primitive streak which establishes bilateral symmetry. A primitive node (or primitive knot) forms in front of the primitive streak which is the organiser of neurulation. A primitive pit forms as a depression in the centre of the primitive node which connects to the notochord which lies directly underneath. The node has arisen from epiblasts of the amniotic cavity floor, and it is this node that induces the formation of the neural plate which serves as the basis for the nervous system. The neural plate will form opposite the primitive streak from ectodermal tissue which thickens and flattens into the neural plate. The epiblast in that region moves down into the streak at the location of the primitive pit where the process called ingression, which leads to the formation of the mesoderm takes place. This ingression sees the cells from the epiblast move into the primitive streak in an epithelial-mesenchymal transition; epithelial cells become mesenchymal stem cells, multipotent stromal cells

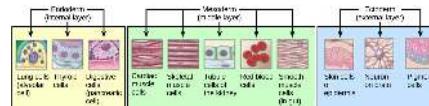


Figure 30.8: Histogenesis of the three germ layers.⁸

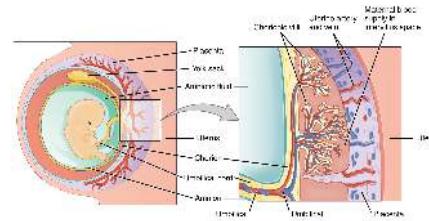


Figure 30.9: Embryo attached to the placenta in the amniotic cavity.⁹

that can differentiate into various cell types. The hypoblast is pushed out of the way and goes on to form the amnion. The epiblast keeps moving and forms a second layer, the mesoderm. The epiblast has now differentiated into the three germ layers of the embryo, so that the bilaminar disc is now a trilaminar disc, the gastrula.

The three germ layers are the ectoderm, mesoderm and endoderm, and are formed as three overlapping flat discs. It is from these three layers that all the structures and organs of the body will be derived through the processes of somitogenesis, histogenesis and organogenesis. The embryonic endoderm is formed by invagination of epiblastic cells that migrate to the hypoblast, while the mesoderm is formed by the cells that develop between the epiblast and endoderm. In general, all germ layers will derive from the epiblast. The upper layer of ectoderm will give rise to the outermost layer of skin, central and peripheral nervous systems, eyes, inner ear, and many connective tissues. The middle layer of mesoderm will give rise to the heart and the beginning of the circulatory system as well as the bones, muscles and kidneys. The inner layer of endoderm will serve as the starting point for the development of the lungs, intestine, thyroid, pancreas and bladder.

Following ingressation, a blastopore develops where the cells have ingressed, in one side of the embryo and it deepens to become the archenteron, the first formative stage of the gut. As in all deuterostomes, the blastopore becomes the anus whilst the gut tunnels through the embryo to the other side where the opening becomes the mouth. With a functioning digestive tube, gastrulation is now completed and the next stage of neurulation can begin.

Following gastrulation, the ectoderm gives rise to epithelial and neural tissue, and the gastrula is now referred to as the neurula. The neural plate that has formed as a thickened plate from the ectoderm, continues to broaden

and its ends start to fold upwards as neural folds. Neuralisation refers to this folding process whereby the neural plate is transformed into the neural tube, and this takes place during the fourth week. They fold, along a shallow neural groove which has formed as a dividing median line in the neural plate. This deepens as the folds continue to gain height, when they will meet and close together at the neural crest. The cells that migrate through the most cranial part of the primitive line form the paraxial mesoderm, which will give rise to the somitomeres that in the process of somitogenesis will differentiate into somites that will form the sclerotomes, the syndetomes, the myotomes and the dermatomes to form cartilage and bone, tendons, dermis (skin), and muscle. The intermediate mesoderm gives rise to the urogenital tract and consists of cells that migrate from the middle region of the primitive line. Other cells migrate through the caudal part of the primitive line and form the lateral mesoderm, and those cells migrating by the most caudal part contribute to the extraembryonic mesoderm.

The embryonic disc begins flat and round, but eventually elongates to have a wider cephalic part and narrow-shaped caudal end. At the beginning, the primitive line extends in cephalic direction and 18 days after fertilization returns caudally until it disappears. In the cephalic portion, the germ layer shows specific differentiation at the beginning of the 4th week, while in the caudal portion it occurs at the end of the 4th week. Cranial and caudal neuropores become progressively smaller until they close completely (by day 26) forming the neural tube.

30.3.1 Development of Organs And Organ Systems

Organogenesis is the development of the organs that begins during the third to eighth week, and continues until birth. Sometimes full development, as in the brain, continues after birth. Different organs take part in the development of the many organ systems of the body.

Haematopoietic stem cells that give rise to all the blood cells develop from the mesoderm. The development of blood formation takes place in clusters of blood cells, known as blood islands, in the yolk sac. Blood islands develop outside the embryo, on the umbilical vesicle, allantois, connecting stalk, and chorion, from mesodermal hemangioblasts.

In the centre of a blood island, hemangioblasts form the haematopoietic stem cells that are the precursor to all types of blood cell. In the periphery of a blood island the hemangioblasts differentiate into angioblasts the precursors to the blood vessels.

The heart is the first functional organ to develop

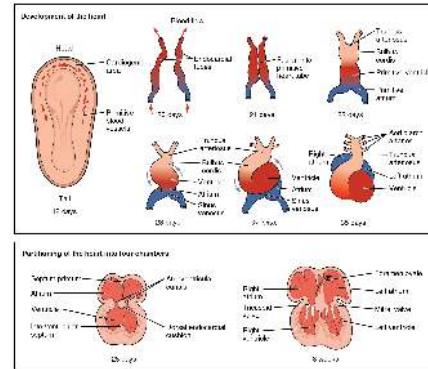


Figure 30.10: Embryonic development of the heart.¹⁰

and starts to beat and pump blood at around 21 or 22 days. Cardiac myoblasts and blood islands in the splanchnopleuric mesenchyme on each side of the neural plate, give rise to the cardiogenic region.¹⁶⁵ This is a horseshoe-shaped area near to the head of the embryo. By day 19, following cell signalling, two strands begin to form as tubes in this region, as a lumen develops within them. These two endocardial tubes grow and by day 21 have migrated towards each other and fused to form a single primitive heart tube, the tubular heart. This is enabled by the folding of the embryo which pushes the tubes into the thoracic cavity.

Also at the same time that the endocardial tubes are forming, vasculogenesis (the development of the circulatory system) has begun. This starts on day 18 with cells in the splanchnopleuric mesoderm differentiating into angioblasts that develop into flattened endothelial cells. These join to form small vesicles called angiocysts which join up to form long vessels called angioblastic cords. These cords develop into a pervasive network of plexuses in the formation of the vascular network. This network grows by the additional budding and sprouting of new vessels in the process of angiogenesis. Following vasculogenesis and the development of an early vasculature, a stage of vascular remodelling takes place.

The tubular heart quickly forms five distinct regions. From head to tail, these are the infundibulum, bulbus cordis, primitive ventricle, primitive atrium, and the sinus venosus. Initially, all venous blood flows into the sinus venosus, and is propelled from tail to head to the truncus arteriosus. This will divide to form the aorta and pulmonary artery; the bulbus cordis will develop into the right (primitive) ventricle; the primitive ventricle will form the left ventricle; the primitive atrium will become the front parts of the left and right atria and their appendages, and the sinus venosus will develop into the posterior part of the right atrium, the sinoatrial node and the coronary sinus.

Cardiac looping begins to shape the heart as one of the processes of morphogenesis, and this completes by the end of the fourth week. Programmed cell death (apoptosis) at the joining surfaces enables fusion to take place. In the middle of the fourth week, the sinus venosus receives blood from the three major veins: the vitelline, the umbilical and the common cardinal veins.

During the first two months of development, the interatrial septum begins to form. This septum divides the primitive atrium into a right and a left atrium. Firstly it starts as a crescent-shaped piece of tissue which grows downwards as the septum primum. The crescent shape prevents the complete closure of the atria allowing blood to be shunted from the right to the left atrium through the opening known as the ostium primum. This closes with further development of the system but before it does, a second opening (the ostium secundum) begins to form in the upper atrium enabling the continued shunting of blood.

A second septum (the septum secundum) begins to form to the right of the septum primum. This also leaves a small opening, the foramen ovale which is continuous with the previous opening of the ostium secundum. The septum primum is reduced to a small flap that acts as the valve of the foramen ovale and this remains until its closure at birth. Between the ventricles the septum inferius also forms which develops into the muscular interventricular septum.

The digestive system starts to develop from the third week and by the twelfth week, the organs have correctly positioned themselves.

The respiratory system develops from the lung bud, which appears in the ventral wall of the foregut about four weeks into development. The lung bud forms the trachea and two lateral growths known as the bronchial buds, which enlarge at the beginning of the fifth week to form the left and right main bronchi. These bronchi in turn form secondary (lobar) bronchi; three on the right and two on the left (reflecting the number of lung lobes). Tertiary bronchi form from secondary bronchi.

While the internal lining of the larynx originates from the lung bud, its cartilages and muscles originate from the fourth and sixth pharyngeal arches.

Three different kidney systems form in the developing embryo: the pronephros, the mesonephros and the metanephros. Only the metanephros develops into the permanent kidney. All three are derived from the intermediate mesoderm.

Between the fourth and seventh weeks of development, the urorectal septum divides the cloaca into the urogenital sinus and the anal canal. The upper part of the urogenital

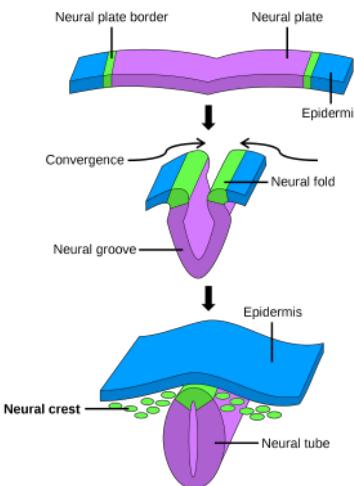


Figure 30.11: A diagram of the stages of neural tube formation.¹¹

sinus forms the bladder, while the lower part forms the urethra.

The superficial layer of the skin, the epidermis, is derived from the ectoderm. The deeper layer, the dermis, is derived from mesenchyme.

The formation of the epidermis begins in the second month of development and it acquires its definitive arrangement at the end of the fourth month. The ectoderm divides to form a flat layer of cells on the surface known as the periderm. Further division forms the individual layers of the epidermis.

The mesenchyme that will form the dermis is derived from three sources:

- The mesenchyme that forms the dermis in the limbs and body wall derives from the lateral plate mesoderm
- The mesenchyme that forms the dermis in the back derives from paraxial mesoderm
- The mesenchyme that forms the dermis in the face and neck derives from neural crest cells

Late in the fourth week, the superior part of the neural tube bends ventrally as the cephalic flexure at the level of the future midbrain—the mesencephalon. Above the mesencephalon is the prosencephalon (future forebrain) and beneath it is the rhombencephalon (future hindbrain).

Cranial neural crest cells migrate to the pharyngeal arches as neural stem cells, where they develop in the process of neurogenesis into neurons.

The optical vesicle (which eventually becomes the optic nerve, retina and iris) forms at the basal plate of the prosen-

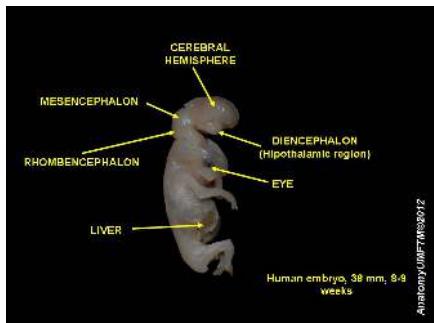


Figure 30.12: An 8 week old human embryo.¹²

cephalon. The alar plate of the prosencephalon expands to form the cerebral hemispheres (the telencephalon) whilst its basal plate becomes the diencephalon. Finally, the optic vesicle grows to form an optic outgrowth.

Fetal development is the third of the three stages of prenatal development, following from the initial germinal stage (preembryonic stage), and stage of embryonic development. These stages are also referred to in pregnancy as terms or trimesters.

From the 10th week of gestation (8th week of development), the developing organism is called a fetus.

All major structures are already formed in the fetus, but they continue to grow and develop. Since the precursors of all the major organs are created by this time, the fetal period is described both by organ and by a list of changes by weeks of gestational age.

Because the precursors of the organs are now formed, the fetus is not as sensitive to damage from environmental exposure as the embryo was. Instead, toxic exposure often causes physiological abnormalities or minor congenital malformation.

30.4 Aging

A number of characteristic aging symptoms are experienced by a majority or by a significant proportion of humans during their lifetimes.

- Teenagers lose the young child's ability to hear high-frequency sounds above 20 kHz.
- Wrinkles develop mainly due to photoageing, particularly affecting sun-exposed areas (face).
- After peaking in the mid-20s, female fertility declines.
- After age 30 the mass of human body is decreased until 70 years and then shows damping oscillations.
- Muscles have reduced capacity of responding to exercise or injury and loss of muscle mass and strength

(sarcopenia) is common. VO₂ max and maximum heart rate decline.

- People over 35 years of age are at increasing risk for losing strength in the ciliary muscle which leads to difficulty focusing on close objects, or presbyopia. Most people experience presbyopia by age 45–50. The cause is lens hardening by decreasing levels of α-crystallin, a process which may be sped up by higher temperatures.
- Around age 50, hair turns grey. Pattern hair loss by the age of 50 affects about 30–50% of males and a quarter of females.
- Menopause typically occurs between 44 and 58 years of age.
- In the 60–64 age cohort, the incidence of osteoarthritis rises to 53%. Only 20% however report disabling osteoarthritis at this age.
- Almost half of people older than 75 have hearing loss (presbycusis) inhibiting spoken communication. Many vertebrates such as fish, birds and amphibians do not suffer presbycusis in old age as they are able to regenerate their cochlear sensory cells, whereas mammals including humans have genetically lost this ability.
- By age 80, more than half of all Americans either have a cataract or have had cataract surgery.
- Frailty, a syndrome of decreased strength, physical activity, physical performance and energy, affects 25% of those over 85.
- Atherosclerosis is classified as an aging disease. It leads to cardiovascular disease (for example stroke and heart attack) which globally is the most common cause of death. Vessel aging causes vascular remodeling and loss of arterial elasticity and as a result causes the stiffness of the vasculature.
- Recent evidence suggests that age-related risk of death plateaus after age 105. The maximum human lifespan is suggested to be 115 years. The oldest reliably recorded human was Jeanne Calment who died in 1997 at 122.

Dementia becomes more common with age. About 3% of people between the ages of 65 and 74, 19% between 75 and 84, and nearly half of those over 85 years of age have dementia. The spectrum ranges from mild cognitive impairment to the neurodegenerative diseases of Alzheimer's disease, cerebrovascular disease, Parkinson's disease and Lou Gehrig's disease. Furthermore, many types of memory decline with aging, but not semantic memory or general knowledge such as vocabulary definitions, which typically increases or remains steady until late adulthood. Intelligence declines with age, though the rate varies depending on the type and may in fact remain steady throughout most of the lifespan, dropping suddenly only as people near the

end of their lives. Individual variations in rate of cognitive decline may therefore be explained in terms of people having different lengths of life. There are changes to the brain: after 20 years of age there is a 10% reduction each decade in the total length of the brain's myelinated axons.

Age can result in visual impairment, whereby non-verbal communication is reduced, which can lead to isolation and possible depression. Older adults, however, may not suffer depression as much as younger adults, and were paradoxically found to have improved mood despite declining physical health. Macular degeneration causes vision loss and increases with age, affecting nearly 12% of those above the age of 80. This degeneration is caused by systemic changes in the circulation of waste products and by growth of abnormal vessels around the retina.

A distinction can be made between "proximal aging" (age-based effects that come about because of factors in the recent past) and "distal aging" (age-based differences that can be traced to a cause in a person's early life, such as childhood poliomyelitis).

Aging is among the greatest known risk factors for most human diseases. Of the roughly 150,000 people who die each day across the globe, about two thirds—100,000 per day—die from age-related causes. In industrialized nations, the proportion is higher, reaching 90%. In humans, aging represents the accumulation of changes in a human being over time and can encompass physical, psychological, and social changes. Reaction time, for example, may slow with age, while knowledge of world events and wisdom may expand.

The causes of aging are uncertain; current theories are assigned to the damage concept, whereby the accumulation of damage (such as DNA oxidation) may cause biological systems to fail, or to the programmed aging concept, whereby internal processes (such as DNA methylation) may cause aging. Programmed aging should not be confused with programmed cell death (apoptosis).

At present, researchers are only just beginning to understand the biological basis of aging even in relatively simple and short-lived organisms such as yeast. Less still is known of mammalian aging, in part due to the much longer lives of even small mammals such as the mouse (around 3 years). A model organism for studying of aging is the nematode *C. elegans*. Thanks to its short lifespan of 2–3 weeks, our ability to easily perform genetic manipulations or to suppress gene activity with RNA interference, or other factors. Most known mutations and RNA interference targets that extend lifespan were first discovered in *C. elegans*.

The factors proposed to influence biological aging fall into two main categories, programmed and damage-

related. Programmed factors follow a biological timetable, perhaps one that might be a continuation of the one that regulates childhood growth and development. This regulation would depend on changes in gene expression that affect the systems responsible for maintenance, repair and defense responses. Damage-related factors include internal and environmental assaults to living organisms that induce cumulative damage at various levels. A third, novel, concept is that aging is mediated by vicious cycles.

In a detailed review, Lopez-Otin and colleagues (2013), who discuss aging through the lens of the damage theory, propose nine metabolic "hallmarks" of aging in various organisms but especially mammals:

- genomic instability (mutations accumulated in nuclear DNA, in mtDNA, and in the nuclear lamina)
- telomere attrition (the authors note that artificial telomerase confers non-cancerous immortality to otherwise mortal cells)
- epigenetic alterations (including DNA methylation patterns, post-translational modification of histones, and chromatin remodelling)
- loss of proteostasis (protein folding and proteolysis)
- deregulated nutrient sensing (relating to the Growth hormone/Insulin-like growth factor 1 signalling pathway, which is the most conserved aging-controlling pathway in evolution and among its targets are the FOXO3/Sirtuin transcription factors and the mTOR complexes, probably responsive to caloric restriction)
- mitochondrial dysfunction (the authors point out however that a causal link between aging and increased mitochondrial production of reactive oxygen species is no longer supported by recent research)
- cellular senescence (accumulation of no longer dividing cells in certain tissues, a process induced especially by p16INK4a/Rb and p19ARF/p53 to stop cancerous cells from proliferating)
- stem cell exhaustion (in the authors' view caused by damage factors such as those listed above)
- altered intercellular communication (encompassing especially inflammation but possibly also other intercellular interactions)

There are three main metabolic pathways which can influence the rate of aging, discussed below:

- the FOXO3/Sirtuin pathway, probably responsive to caloric restriction
- the Growth hormone/Insulin-like growth factor 1 signalling pathway
- the activity levels of the electron transport chain in mitochondria and (in plants) in chloroplasts.

It is likely that most of these pathways affect aging separately, because targeting them simultaneously leads to additive increases in lifespan.