Self-study Notes on OpenStax AP Biology Books and Other Resources

Kedar Mhaswade

2020-21

1 Unit 1: The Chemistry of Life

1.1 The Study of Life

1.2 The Chemical Foundation of Life

1.3 Biological Macromolecules

1.3.1 Introduction

FOOD PROVIDES the body with the nutrients it needs to survive. Many of these nutrients are "macro" or large molecules necessary for and built by living things. They are *necessary* because humans have evolved to use them for survival:

- amino acids in protein for healthy bone and muscle,

- fats to build new cells, store energy, and digest properly,

- carbohydrates to drive all living processes, and

- nucleic acids to carry genetic information.

Even so, an *imbalance* in our consumption of these macromolecules may lead to health problems. My personal view here is that since all humans are different (however similar they may be) and there are many theories of a general human diet and human feelings are involved in defining what

"health" means, it is best to spend some time understanding oneself and one's reactions to dietary components. Such a knowledge about oneself is often illusive because humans are very complex multicellular organisms. Whereas macromolecules are absolutely essetial, there are nutrients like vitamins and minerals that have now become equally essential (in much smaller quantities as compared to macromolecules identified above), perhaps not for survival, but to *manage* health problems and age gracefully.

1.3.2   Synthesis of Biological Macromolecules

Biological Macromolecules (BMm's henceforth) are made of atoms of 6 elements: S, P, O, N, C, H. A majority of cell's *dry mass* is made of BMm's.

1.3.2.1   Dehydration Synthesis

BMm's are *polymers* which are chains of *monomers* joined by covalent bonds. A common polymer is the disaccharide *maltose*: $C_{12}H_{22}O_{11}$ which is formed when two glucose $(C_6H_{12}O_6)$ molecules combine releasing a water molecule:
$2\,C_6H_{12}O_{11} \longrightarrow C_{12}H_{22}O_{11} + H_2O$

2
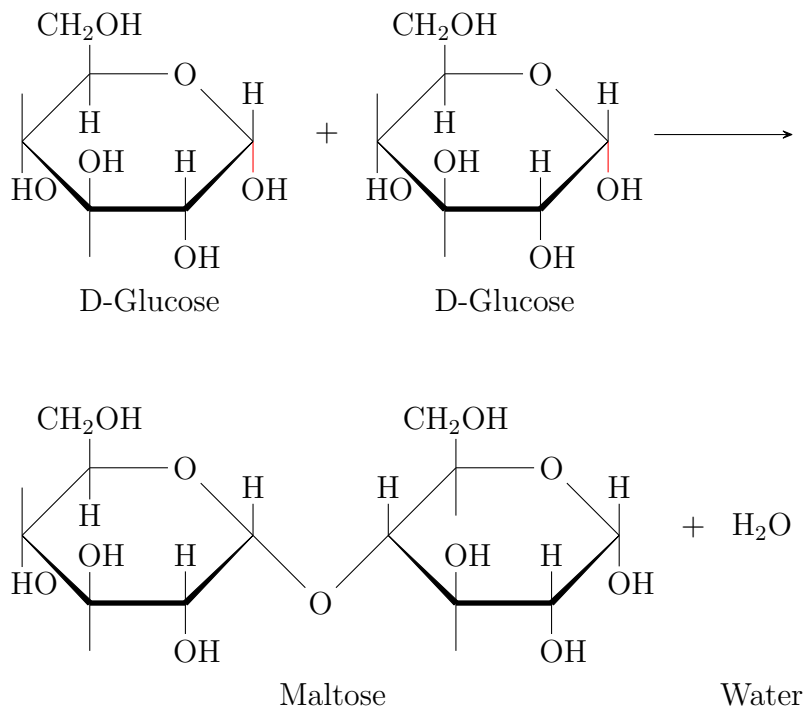
The monomers share electrons as they form covalent bonds. It is called "dehydration synthesis" because it literally means formation (of a polymer, in this case) while losing water.

1.3.2.2 Hydrolysis

An almost opposite of Dehydration Synthesis is Hydrolysis which literally means separation (-lysis) with water (hydro-). Thus, the energy contained in bonds of polymers is released when they react with water (water is split into hydrogen and hydroxyl ions) and monomers are created.

Dehydration Synthesis *requires* energy and releases water to form [polymer] bonds.

Hydrolysis requires water and *releases* energy stored in [polymer] bonds.

1.3.2.3 Enzymes

Both these kinds of reaction require enzymes that are *specific* to the BMm's:

Table 1: Enzymes for Hydrolysis of BMm's

BMm	Enzymes
Carbohydrates	amylase, sucrase, lactase, or maltase
Proteins	pepsin, peptidase, hydrochloric acid
Lipids	lipase

We wonder how life formed on Earth. In 1953, Stanley Miller and Harold Urey developed an apparatus to model early conditions on earth. They wanted to test if *organic molecules could form from inorganic precursors believed to exist very early in Earth's history.* They used boiling water to mimic early Earth's oceans. Steam from the "ocean" combined with methane, ammonia, and hydrogen gases from the early Earth's atmosphere and was exposed to electrical sparks to act as lightning. As the gas mixture cooled and condensed, *it was found to contain organic compounds, such as amino acids and nucleotides.* According to the **abiogenesis theory** (the original evolution of life or living organisms from inorganic or inanimate substances), these organic molecules came together to form the earliest form of life about 3.5 billion years ago.

1.3.3 Carbohydrates

The Wikipedia page on Carbohydrates is comprehensive. A few things to know about carbohydrates:

- A carbohydrate consists of carbon, hydrogen, and oxygen atoms, usually (*but not always*) with a 2:1 hydrogen-oxygen atom ratio and the empirical formula $C_m(H_2O)_n$.

- In biochemistry, the term *carbohydrates* is synonymous with the term *saccharides* which contain sugars, starch, and cellulose.

- Four main groups (of these, the first two are commonly referred to as sugars):

  1. monosaccharides (simple sugars: $(CH_2O)_n$): Simplest of carbohydrates. **They cannot be hydrolyzed into simpler compounds**. Common examples: glyceraldehyde, ribose, glucose, fructose, and galactose.
  2. disaccharides: Formed when two monosaccharide molecules are joined with the (covalent) glycosidic bonds. Common examples: sucrose, lactose, and maltose.
  3. oligosaccharides: Consist of a few to several monosaccharide molecules (as a rule of thumb: 3 to 10). Common examples: glycolipids.

4. polysaccharides: These long chains of monosaccharide molecules (as a rule of thumb: more than 10) are the most abundant carbohydrates in human food. Two main types:

    (a) Storage polysaccharides: starch, glycogen, and galactogen.
    (b) Structural polysaccharides: cellulose and chitin.

- Common names – fructose (a monosaccharide): *fruit sugar*, glucose (a monosaccharide): *starch sugar*, sucrose (a disaccharide): *cane* or *beet sugar*, lactose (a disaccharide): *milk sugar*.

-

1.3.3.1   Monosaccharides

Monosaccharides are the simplest form of sugar. They cannot be further *hydrolyzed* to simpler chemical compounds. They are classified according to number of carbon atoms. They are also classified according to the functional group they contain. Based on the number of carbon atoms they are classified as triose, tetrose, pentose, hexose and so on. Based on the functional group, they are classified as either aldose (aldehyde group: $R-CHO$) or ketose (ketone group: $R-C=O-R'$).

1.3.4   Lipids

1.3.5   Proteins

Proteins are long chains (sequences) of amino acids. There are 20 common amino acids. An amino acid is a monomer; itcontains an amino group ($-NH_2$), a carboxyl group ($-COOH$), and a variable group. Suppose that a certain protein is a sequence of **500** amino acids, then since we have 20 choices to pick from for each of the 500 amino acid molecules, there are 20^{500} (a huge number!) choices to make the protein. The following diagram shows a simple protein that is a sequence of just 8 amino acids: This hypotheti-

Table 2: An Example Protein: a Sequence of 8 Amino Acids

AA-15	AA-02	AA-20	AA-11	AA-03	AA-10	AA-07	AA-18

cal case of all proteins being a sequence of just 8 amino acids would yield

$20^8 = 25,600,000,000$ different proteins. Table 2 just shows one possible arrangement. Although amino acids have names, we have simply referred to them as AA-1 to AA-20 in that example. In nature, proteins are arbitrarily long sequences of amino acids.

In proteins, each amino acid molecule is linked to the next by a *peptide bond* formed by dehydration synthesis (not unlike the formation of polysaccharides from monosaccharides). This makes protein a *polypeptide*. Proteins are really the workhorses in organisms. They carry out the following functions:

- Act as enzymes that catalyze chemical reactions

- Provide structural support (along with cellulose which is available only in plants)

- Regulate the passage of substances across the cell membrane (even viruses and other pathogens may *attach* to proteins to enter cells)

- Help protect against (infectious) diseases (by participating in antibody creation)

- Communicate with other cells; receptors which are proteins receive and transduce *signals*

Characteristic property of proteins is their *folding* or three-dimensional structure. Predicting Protein Folding from its linear structure is very useful for research and it is also a major computational problem. Recently, AI techniques have been used for this purpose [9]. Protein Folding occurs at four levels:

1. Primary: Decides the unique (linear) sequence of amino acids; a single amino acid change in a protein can alter its structure and function

2. Secondary: Consists of the local folding of the polypeptide by hydrogen bond formation leading to α helix and β pleated sheet *conformations*

3. Tertiary: Various interactions give rise to the three-dimensional structure

4. Quaternary: May be formed from several polypeptide interactions

Proteins may stop functioning when they are *denatured* which depends upon environmental conditions like temperature and pH.

1.3.5.1 Types and Functions of Proteins

Each cell may contain thousands of proteins. Their structures, like their function, may vary greatly.

Table 3: Types and Functions of Proteins

Type of Protein	Examples	Functions
Digestive Enzymes	Amylase, lipase, pepsin, and trypsin	Help in digestion of food by catabolizing nutrients into monomeric units
Transport	Hemoglobin, albumin	Carry substances in blood or lymph throughout the body
Structural	Actin, tubulin, keratin	Construct different structues like e.g. cytoskeleton
Hormone	Insulin, thyroxine	Coordinate activities in different body systems by participating in cell signaling
Defense	Immunoglobulins	Protect the body from pathogens
Contractile	Actin, myosin	Effect muscle contraction
Storage	Legume storing proteins, albumin	Provide nourishment in early development of the embryo and the seedling

1.3.5.2 Amino Acids

Amino acids are the monomers that make up proteins. Each amino acid has the same fundamental structure, which consists of a central carbon atom, also known as the alpha (α) carbon, bonded to an amino group (NH_2), a carboxyl group (COOH), and to a hydrogen atom. Every amino acid also has another atom or group of atoms bonded to the central atom known as the R group.

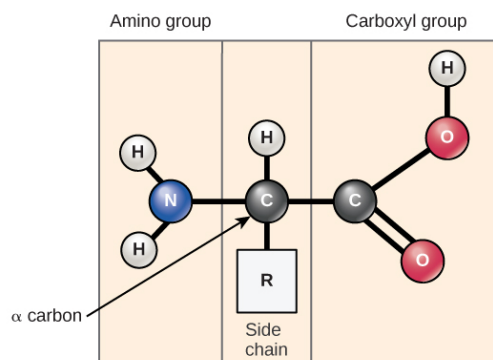Amino acids are acids since they can be looked upon as $R-COOH$.

Figure 1: The General Structure of a Protein (From OpenStax)

1.3.6 Nucleic Acids

Nucleic acids (DNA and RNA) contain phosphorus in addition to C, H, O, and N. **Conserved through evolution in all organisms, nucleic acids store and transmit hereditary information**. A cell (nucleus) contains DNA spread across various chromosomes. In humans, each chromosome comes as a *pair*, that is, if we numbered chromosomes 1, 2, ..., then humans have a pair of each number. A chromosome is, in turn, a pair of distinct DNA molecules connected by hydrogen bonds.

1 chromosome = 2 DNA molecules

All matter is made of atoms and molecules and bonds between them. Nucleic acids are polymers. One molecule of a nucleic acid is made of several other molecules. Think of this or any other polymer as a hierarchical structure that becomes progressively bigger and more complex.

One molecule of a nucleic acid could be made of tens of thousands of the monomeric *Nucleotide* molecules. Each nucleotide molecule in turn is made of

- one molecule of *Nucleoside* which is made of

  - one molecule of a nitrogenous base (Nucleobase) and
  - one molecule of a pentose sugar (ribose in RNA and deoxy-ribose in DNA)

- one, two, or three molecules of a phosphate

8

The nucleotide molecules are held together by phosphodiester bonds. A *nucleoside* is either

- Adenine (both DNA and RNA),

- Cytosine (both DNA and RNA),

- Guanine, (both DNA and RNA),

- Thymine (only DNA), or

- Uracil (only RNA)

There are differences between RNA and DNA. There is only one form of DNA, whereas RNA has various forms: messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), micro RNA (miRNA). In DNA, there is its signature *double helical structure.*

1.3.6.1 The Central Dogma of Molecular Biology

> ...Once the 'information' has passed into protein *it cannot get out again.* In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. **Information means here the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein**.

– As stated by Francis Crick [5]

1.3.6.2 DNA and RNA

Both DNA and RNA are, of course, examples of a *nucleic acid.* This *genetic material* has different functions. DNA, which stands for the *Deoxyribose Nucleic Acid* has, one oxygen atom **less than** (per repetitive nucleotide unit) RNA, which stands for *RiboNucleic Acid.* Figures 2 and 3 show their chemical structures.

The *strand* of several nucleotides in DNA looks like Figure 4.

The structures of nucleic acids reveal that they have nitrogenous base and phosphate. You may wonder why it is called an acid if it has a nitrogenous
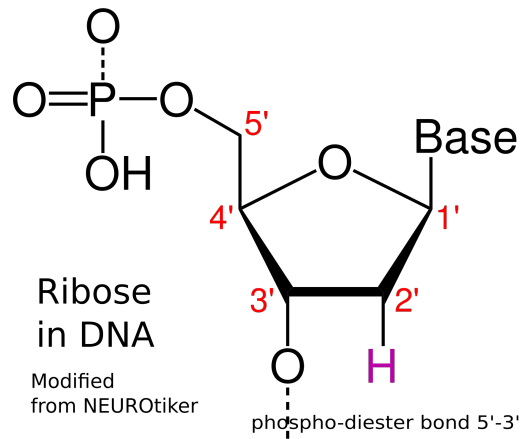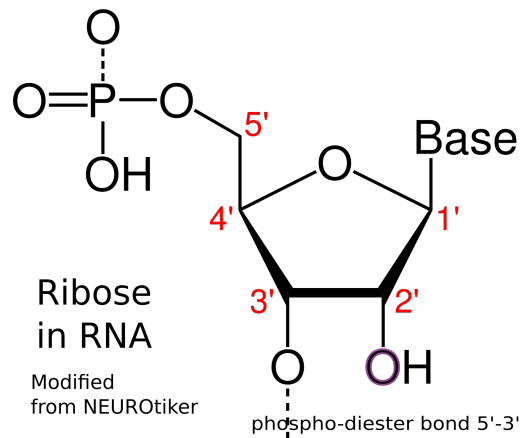
Figure 2: DNA

Figure 3: RNA

base. Looking at the *phosphodiester bonds* between the nucleotide molecules in a DNA or RNA molecule should help decide. See Figure 4. The phosphate group (blurred yellow oval in Figure 4) is similar to the phosphoric acid (H_3PO_4) molecule which has 3 hydrogen atoms. In DNA or RNA nucleotide chain, the two hydrogen atoms (from the phosphoric acid) are replaced by (the 5' and the 3') carbon atoms on the ribose sugar molecule. This makes the remaining hydrogen extremely reactive and it is lost rather easily under neutral conditions. As the Figure 4 shows, the oxygen on the phosphate
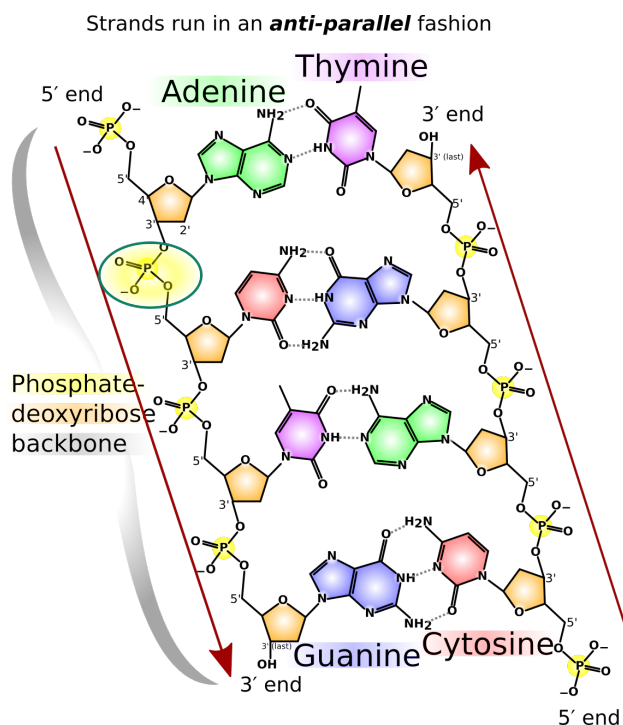
Figure 4: A DNA Strand

group assumes a negative charge because the molecule literally gets *deprotonated*. This makes the DNA or RNA molecule acidic in nature as a whole in spite of the fact that there are a lot of nitrogenous base molecules in it[1].

The nitrogenous bases, important components of nucleotides, are organic molecules and are so named because they contain carbon and nitrogen. They are bases because they contain an amino group that has the **potential of binding an extra hydrogen, thus decreasing the hydrogen ion concentration in its environment, making it more basic**. Each nucleotide in DNA contains one of four possible nitrogenous bases: adenine (A), guanine (G) cytosine (C), and thymine (T).

Adenine and guanine are classified as purines. The primary structure of a purine is two carbon-nitrogen rings. Cytosine, thymine, and uracil are classified as pyrimidines which have a single carbon-nitrogen ring as their primary structure. Each of these basic carbon-nitrogen rings has different functional

---

[1]DNA is *double stranded* and there are separate hydrogen bonds between the nitrogenous bases (purines and pyrimidines) as well.

groups attached to it. In molecular biology shorthand, the nitrogenous bases are simply known by their symbols A, T, G, C, and U. DNA contains A, T, G, and C whereas RNA contains A, U, G, and C.

The pentose sugar in DNA is deoxyribose, and in RNA, the sugar is ribose. The difference between the sugars is the presence of the hydroxyl group on the second carbon of the ribose (Figure 3) and hydrogen on the second carbon of the deoxyribose (Figure 2). The carbon atoms of the sugar molecule are numbered as 1', 2', 3', 4', and 5' (1' is read "one prime"). The phosphate residue is attached to the hydroxyl group of the 5' carbon of one sugar and the hydroxyl group of the 3' carbon of the sugar of the next nucleotide, which forms a 5'–3' phosphodiester linkage. The phosphodiester linkage is not formed by simple dehydration reaction like the other linkages connecting monomers in macromolecules: its formation involves the removal of two phosphate groups. A polynucleotide may have thousands of such phosphodiester linkages.

1.3.6.3 DNA Double-Helix Structure

DNA has a double-helix structure. The sugar and phosphate lie on the outside of the helix, forming the backbone of the DNA. The nitrogenous bases are stacked in the interior, like the steps of a staircase, in pairs; the pairs are bound to each other by hydrogen bonds (See Figure 4). Every base pair in the double helix is separated from the next base pair by[2] 0.34 nm. The two strands of the helix run in opposite directions, meaning that the 5' carbon end of one strand will face the 3' carbon end of its matching strand. (This is referred to as *antiparallel* orientation and is important to DNA replication and in many nucleic acid interactions.)

Erwin Chargaff studied many species for the *amount* of purines and pyrimidines their cells have.

---

[2]Charles Mallery [6] reports that one full turn of the helix is about 3.4 nanometers, or 10 base pairs long.

Chargaff's rules state that DNA from any species of any organism should have a 1:1 stoichiometric ratio (base pair rule) of pyrimidine and purine bases and, more specifically, that the amount of guanine should be equal to cytosine and the amount of adenine should be equal to thymine. This pattern is found in both strands of the DNA. They were discovered by Austrian-born chemist Erwin Chargaff in the late 1940s.

Wikipedia entry on Chargaff's rules

Mallery [6] states the rules thus:

1. A with T: the purine adenine (A) always pairs with the pyrimidine thymine (T)

2. C with G: the pyrimidine cytosine (C) always pairs with the purine guanine (G)

This is consistent with there not being enough space (20 Å) for two purines to fit within the helix and too much space for two pyrimidines to get close enough to each other to form hydrogen bonds between them.

But why not A with C and G with T?

The answer: only with A and T and with C and G are there opportunities to establish hydrogen bonds between them (*two* between A and T; *three* between C and G). The ability to form hydrogen bonds makes the base pairs structurally more stable.

Like in the case of any scientific discovery that becomes an accepted fact, one should see Chargaff's findings in the light of the prevalent scientific understanding at the time. Although Chargaff was *in the vicinity* of making the discovery that DNA has a *double helical* structure, he was not able to do it. This is an interesting example of how creativity[3] is illusive in a scientific endeavor:

---

[3]Of course, several other factors contribute to a discovery and its acceptance by the scientific community.

Our conversation covered many topics. Erwin Chargaff's name came up and Francis found it strange that Chargaff did not discover base-pairing in the light of his observations on the base ratios in DNA. However, *this may be more surprising in hindsight.* Just looking at the data, there is much fluctuation, about 10 percent, about the 1 to 1 ratios. So *even recognizing and suggesting the 1 to 1 ratio was a sharp observation.* Francis then added that Chargaff's mind might have not wandered toward pairing because he, that is, Chargaff *was thinking in terms of one chain rather than two.* In a single chain, nothing would prompt one's thinking toward pairing. Once you know that two chains need be considered, pairing enters one's thinking more naturally. Francis seemed careful not to use the word helix at this point, as if placing himself into Chargaff's position. Even if Chargaff was thinking about the significance of the 1 to 1 ratios, about the possible meaning of such a ratio in a nucleotide chain, it's not surprising that nothing suggested itself for a reasonable solution.

– Francis Crick Interview by Hargittai and Hargittai [7]
(emphasis mine)

1.3.6.4   RNA

Ribonucleic acid, or RNA, is mainly involved in the *process of protein synthesis under the direction of DNA*. RNA is usually single-stranded and is made of ribonucleotides that are linked by phosphodiester bonds. A ribonucleotide in the RNA chain contains ribose (the pentose sugar), one of the four nitrogenous bases (A, **U**, G, and C), and the phosphate group.

There are four major *functional* types of RNA:

1. messenger RNA (mRNA) : literally *carries* the message or the information from DNA; its sequence is complementary to the DNA from which it is copied.

2. ribosomal RNA (rRNA): a constituent of the ribosomes that binds the mRNA.

3. transfer RNA (tRNA)a: a small RNA (70-90 nucleotides) that *transfers* correct amino acid to be inserted into the polypeptide chain that is the protein that an expressed gene encodes.

4. mircroRNA (miRNA): the smallest RNA that regulates the *gene expression.*

We stated the *central dogma* above. It appears to be true of eukaryotes although there are exceptions. Not everyone agrees with the assertion that information cannot come out of proteins. In fact, Eugene Koonin states the following in a Springer Nature article ([8]):

> Central Dogma of molecular biology is invalid as an 'absolute' principle: transfer of information from proteins (and specifically from protein sequences) to the genome does exist. This is not to deny that the Central Dogma does capture the principal route of information transfer in biology: the main flow information does follow the path in Figure 1 [DNA → RNA → protein], and elaborate mechanisms ensuring acceptable fidelity operate on each step. And, there is a major discontinuity between the levels of RNA and protein because during translation because the coupling of amino acids with the cognate tRNAs does not involve direct recognition but rather requires dedicated enzymes, aminoacyl-tRNA synthetases, that recognize and connect the partners. The opposite direction of information flow, from proteins to the genome, is asymmetrical (not a simple reversion), much more modest quantitatively and intrinsically stochastic but nevertheless appears to be important in evolution.
>
> *– Eugene Koonin* [8]

2 Unit 2: The Cell

2.1 Cell Structure

2.1.0.1 Introduction

Cells are the fundamental building blocks of all organisms. There are, of course, many kinds of cell and each kind of cell has a *specific* function to perform. There's tremendous variety of functions that cells perform and the organisms that they create are also diverse (e.g. sponges, bacteria, fungi, plants, animals) in nature. Even so, many cells share a few fundamental characteristics.

A cell's development and growth to perform its eventual tasks is called its "specialization". In humans, before a cell develops into its specialized type, it is called a *stem cell*. Thus, stem cells do not undergo changes involved in specialization. Stem cells are called *pluripotent* (that means they are capable of giving rise to several different cell types). There is a great potential in stem cell research to address human illnesses.

2.1.1 Studying Cells

Cells of a kind — form $\rightarrow$ Tissue — combine to form $\rightarrow$ Organ (e.g. stomach, heart, or brain) — make up $\rightarrow$ Organ System(s) (e.g. digestive system, endocrine system) — work together to form $\rightarrow$ an Organism.

2.1.2 Prokaryotic Cells

2.1.3 Eukaryotic Cells

2.1.4 The Endomembrane System and Proteins

2.1.5 Cytoskeleton

2.1.6 Connections between Cells and Cellular Activities

2.2 Structure and Function of Plasma Membranes

2.2.0.1 Introduction

The plasma membrane (the PM, henceforth), which is also called the cell membrane, has many functions; but, the most basic one is to *define the borders* and *act as gatekeeper* for the cell.

The PM is a selectively permeable membrane that allows only certain molecules to freely enter cells. Other molecules may require assistance. A striking example of this is the (membrane) protein called NPC1 (Niemann-Pick Type C1) which mediates the intracellular cholesterol trafficking in mammals. NPC1 is therefore called a *cholesterol transporter*. A mutation in the NPC1 gene (that encodes the NPC1 protein) may prevent the entry of cholesterol into cells that results in cholesterol accumulation in intercellular space. But this problem may be a boon in disguise because if such an organism is infected with the Ebola virus, since the NPC1 protein's cholesterol

transportation is impared, its cells are impervious to the virus even when it binds with NPC1.

2.3 Metabolism

2.3.1 ATP: Adenosine Tri-Phospate

- Even exergonic reactions (those that *release* energy) require a small amount of *activation energy* to proceed.

- Products of endergonic reactions (those that *require* energy input) have more energy than that of their reactants. Where does the cell provide this energy from?

- ATP is the molecule that provides the energy required for the endergonic reactions as well as the small activation energy required for the exergonic reactions.

- ATP is a *relatively* simple molecule: $C_{10}H_{16}N_5O_{13}P_3$

  – It has adenosine bound to three phosphate groups that are named alpha- (closest to ribose), beta-, and gamma-phosphate (farthest from ribose).

  – Adenosine is a *nucleoside*—a compound commonly found in DNA or RNA, consisting of a purine or pyrimidine base linked to a sugar.

- ATP *hydrolyzes* to ADP (Adenosine Di-Phosphate) and inorganic phosphate. Regeneration of ATP requires energy to be supplied. Thus, hydrolysis of ATP is reversible:
  $\text{ATP} + \text{H}_2\text{O} \rightleftharpoons \text{ADP} + \text{P}_i + \text{free energy}$

- Cellular conditions differ from standard conditions. In cellular conditions, the ΔG (change in Gibbs free energy) of ATP's hydrolysis is -57 kJ/mol = -14 kcal/mol.

2.4 Cellular Respiration

2.4.0.1 Introduction

Energy enters an organism's body in one form and is converted into another form that can fuel the organism's life functions. In the process of photosynthesis, plants and other photosynthetic producers take in energy in the form of light (solar energy) and convert it into chemical energy, glucose, which stores this energy in its chemical bonds. Then, a series of metabolic pathways, collectively called cellular respiration, extract the energy from the carbon–carbon bonds of glucose and convert it into a form that all living things can use—both producers, such as plants, and consumers, such as animals.

2.4.1 Energy in Living Systems

2.4.2 Glycolysis

2.4.3 Oxidation of Pyruvate and the Citric Acid Cycle

2.4.4 Oxidative Phosphorylation

2.4.5 Metabolism without Oxygen

2.4.6 Connections of Carbohydrate, Protein, and Lipid Metabolic Pathways

2.4.7 Regulation of Cellular Respiration

2.5 Photosynthesis

Many organisms access stored energy by *eating* or *ingesting (consuming)* other organisms: जीवो जीवस्य जीवनम्l. All of this energy can be traced back to photosynthesis.

2.5.1 Overview

Photosynthesis is the *only* biological process that can capture the energy in sunlight and convert it into the energy stored in the covalent bonds of sugar molecules. Additionally, it acts as a source of oxygen necessary for *many* living organisms.

Photoautotrophs (organisms that use light to synthesize their own food) such as plants, algae(अल्जी), and cyanobacteria are the only ones that can perform photosynthesis.

Chemoautotrophs (organisms that use energy stored in inorganic molecules, and not sunlight, to synthesize their own food) found in deep sea represent another class of *autotrophs*.

Animals, fungi, and most other bacteria are termed *Heterotrophs* because they have to rely on the sugars produced by Photoautotrophs.

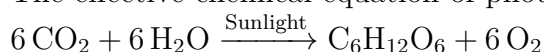### 2.5.1.1   Main Structures and Summary

Photosynthesis is a multi-step process that requires

1. Specific wavelengths of visible sunlight and

2. Carbon dioxide (low in energy) and water as substrates.

It

1. Releases oxygen and

2. Produces glyceraldehyde 3-phosphate (G3P) which can be synthesized into different sugar molecules (G3P has three carbon atoms; two G3P molecules form one glucose molecule). G3P is formed when one CH2OH from glycerol is replaced in an aldehyde and then a hydrogen is replaced by phosphate (H2PO4-).

The effective chemical equation of photosynthesis is:
$$6\,CO_2 + 6\,H_2O \xrightarrow{\text{Sunlight}} C_6H_{12}O_6 + 6\,O_2$$
But note that there are several steps involved in this transformation.

2.5.1.2   Basic Photosynthetic Structures

1. In plants, it generally takes place in *leaves* which have several layers of cells. The layer for photosynthesis is called *mesophyll* (Greek: *meso*: in the middle, *-phyll*: leaf or pigment)

2. The gas exchange takes place through *stomata* (Greek: *stoma*: mouth or opening)
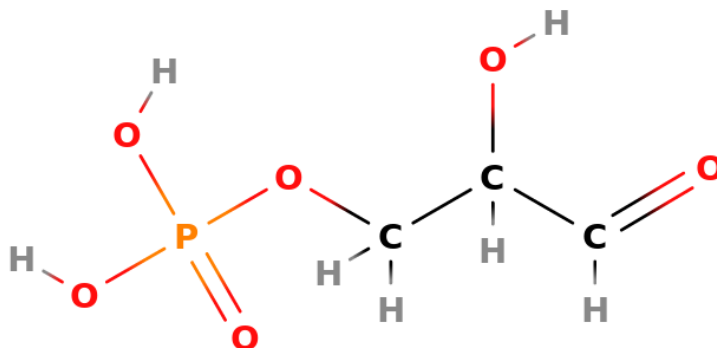
Figure 5: Glyceraldehyde-3 phosphate

2.6 Cell Communication

2.6.1 Signaling Molecules and Cellular Receptors

2.6.1.1 Introduction

In order to respond properly to external stimuli, cells have developed *sophisticated* mechanisms of communication that can

1. receive a *message*,

2. transfer the information across the plasma membrane, and

3. produce changes in the cell as a response

Both single-celled (yes, single-celled) and multicellular organisms communicate with each other at a cellular level and across the organisms.

Communication within a cell is called the *intracellular* communication, whereas the communication between two or more cells is called the *intercellular* communication.

The small, usually volatile, and soluble signaling molecules that *bind* to another specific molecule (and deliver a signal in the process) are called *ligands.*

Ligands bind or interact with proteins called the *receptor proteins* in the *target cells* (the cells that would be affected by the "signals"). Specific ligands bind with specific receptors.

2.7 Cell Reproduction

2.7.1 Introduction

A human, like every sexually reproducing organism, begins life as a fertilized egg (embryo) or zygote. All multicellular organisms use cell division (even after they are fully grown) for growth and maintenance and repair of cells and tissues. Cell division is *regulated closely.* Unicellular organisms also use cell division for their reproduction.

Not all cells in the human body can reproduce *to repair tissues.* Most nerve cells, for example, are not capable of regeneration. This means that people who have damaged their nerve cells or nervous system are often left paralyzed[4].

2.7.2 Cell Division

The cell cycle is an orderly sequence of events. It describes the stages of a cell's life from the division of a single parent cell to the production of two new genetically identical daughter cells.

2.7.2.1 Genomic DNA

A cell's DNA, packaged as a double-stranded DNA molecule, is called its *genome.*

IN PROKARYOTES, the genome is composed of a single DNA molecule. It's found in an area called *nucleoid.* Sometimes, smaller loops of DNA called *plasmids* are also present. Prokaryotes exchange plasmids with other prokaryotes. *Antibiotic resistance* may be developed in a bacteria colony as a result of plasmid exchange between a resistant donor and recipient cells.

IN EUKARYOTES, the genome is composed of several *double stranded* linear DNA molecules. Each species has a *characteristic number* of chromosomes. Human somatic (body) cells have 46 chromosomes, whereas human gametes (sperm or egg cells) have 23 chromosomes each.

Matched pairs of chromosomes in a diploid organism are called homologous chromosomes. Each copy of a homologous chromosome originates in a different parent. The individual homologous chromosome therefore is not identical to its copy. **Regions of chromosome are identified as**

---

[4]This may change because of research

genes. Genes are the *functional units* of chromosome and they determine the phenotypical (externally visible) characteristics of an individual by *encoding specific proteins.*

A location on a chromosome is called **locus** and a gene is said to be located at some locus. Since there are population variations, two different humans are more likely to have variations in the genes. Although we do not have the DNA sequence of each human, there are statistical models that determine the form or version of each gene. **Each form or version of a gene at a given locus is called an allele**.

2.7.3 The Cell Cycle

2.7.4 Control of the Cell Cycle

2.7.5 Cancer and the Cell Cycle

2.7.6 Prokaryotic Cell Division

3   Unit 3: Genetics

3.1   Meiosis and Sexual Reproduction

3.1.0.1 Introduction

THE ABILITY to reproduce "in kind" (the offspring closely resembles the parent(s)) in a basic characteristics of all living things. The Dr. Seuss story "Horton hatches the egg", however lyrical and sentimental, is ascientific (that, of course, does not reduce its literary value).

Many unicellular organisms, such as yeast, and a few multicellular organisms, such as sponges, **can produce genetically identical clones of themselves through cell division**. However, many single-celled organisms and most multicellular organisms reproduce regularly using "Sexual Reproduction", a method requiring two parents. Sexual reproduction occurs through

1. the *production by each parent of a **haploid** cell each containing only half of the required genetic information* (meiosis), and

2. the *fusion of these two haploid cells to form a single, unique diploid cell with a complete set of genetic information* (fertilization of an egg).

Variation or *Variability* is an important component of the *evolutionary success* of species. A vast majority of eukaryotes use **meiosis** and **fertilization** to "reproduce". But like Yashiro et.al. have argued in [2] the prevalence of sexual reproduction remains an enigma. *Queens* of Asian termites can reproduce both sexually and asexually (asexual reproduction may be better termed "cloning") by a method known as *parthenogenesis* (The prefix parthen- means virgin in greek; development of embryo from an unfertilized egg).

3.1.1 The Process of Meiosis

3.1.1.1 Introduction

In *mitosis* cells divide to grow, replace other cells, and *reproduce asexually*. **Without mutation**, or **changes in the DNA**, the daughter cells produced by mitosis receive a set of genetic instructions that is identical to that of the parent cell. Because changes in genes drive both the unity and diversity of life, organisms without genetic variation cannot evolve through natural selection. Evolution occurs only because organisms have developed ways to vary their genetic material. This occurs through

- mutations in DNA,

- recombination of genes during meiosis, and

- meiosis followed by fertilization in sexually reproducing organisms.

Sexual reproduction requires that diploid (2n) organisms produce haploid (1n) cells through meiosis and that these haploid cells fuse to form new, diploid offspring. The *union of these two haploid cells*, one from each parent, is *fertilization*.

The eukaryotic DNA is contained in chromosomes, and chromosomes occur in *homologous pairs* (homologues). At fertilization, the male parent contributes one member of each homologous pair to the offspring, and the female parent contributes the other. With the exception of the sex chromosomes, homologous chromosomes contain the same genes, but *these genes can have different variations, called alleles*. (For example, you might have inherited an allele for brown eyes from your father and an allele for blue eyes from your mother.) As in mitosis, homologous chromosomes are duplicated during the S-stage (synthesis) of interphase. However, unlike mitosis, in which

there is just one nuclear division, meiosis has two complete rounds of nuclear division—meiosis I and meiosis II. These result in four nuclei and (usually) four daughter cells, each with half the number of chromosomes as the parent cell (1n). The first division, meiosis I, separates homologous chromosomes, and the second division, meiosis II, separates chromatids.

> During meiosis, DNA replicates ONCE but divides TWICE, whereas in mitosis, DNA replicates ONCE but divides only ONCE.

It is important to note that [3] meiosis is the process by which sexually reproducing organisms generate gametes (sex cells). However, the main function of meiosis is **the reduction of the ploidy (number of chromosomes) of the gametes from diploid (2n, or two sets of 23 chromosomes) to haploid (1n or one set of 23 chromosomes)**.

Both males and females use meiosis to produce their gametes, although there are some key differences between the sexes at certain stages. In females, the process of meiosis is called **oogenesis**, since it produces oocytes and ultimately yields mature ova(eggs). The male counterpart is **spermatogenesis**, the production of sperm. While they occur at different times and different locations depending on the sex, both processes begin meiosis in essentially the same way. **Meiosis occurs in the primordial germ cells, cells specified for sexual reproduction and separate from the body's normal somatic cells** [3].
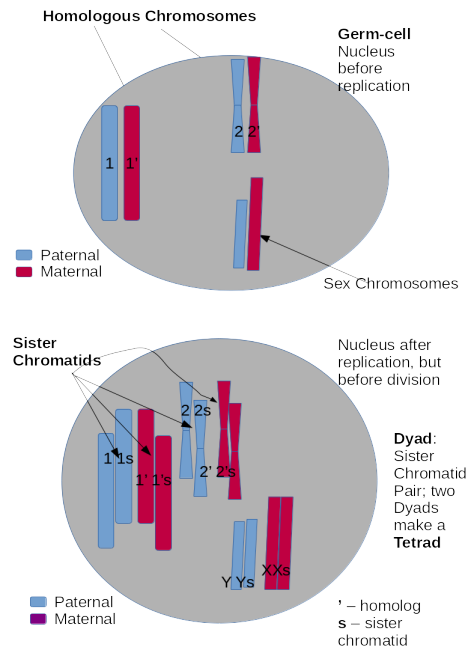
See Figure 6 for some terminology.

Figure 6: Meiosis terminology for a germ cell

The Wikipedia page has many good figures and explanation of meiosis. It is important to realize that the **germ cells that undergo meiosis are different from the sperm and egg cells**. Germ cells are diploid (2n) cells.

Meiosis proceeds in two distinct stages called Meiosis I and Meiosis II. **Steps in Meiosis I**:

1. **Prophase I**. Mitosis preceeds meiosis. A germ cell divides by mitosis. Figure shows the **diploid** oogonium (egg cell) or spermatogonium (sperm). This is the Prophase I. At this stage, as you can see, there are sister chromatids for each chromosome (including the X- and the Y-chromosome). The synapse formation is shown below. Note the difference between homologous chromosomes and sister chromatids. Figure 8 shows the replication of the genetic material in the diploid cell. Located at intervals along the synaptonemal complex are large protein assemblies called **recombination nodules**. These assemblies mark the points of later chiasmata and mediate the multistep pro-
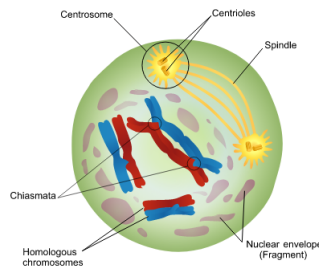
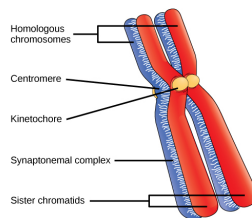Figure 7: Prophase-I: chromosomes condense and crossover begings to happen

Figure 8: Prophase-I: Synaptonemal complex, centromere, and synapse formation

cess of *crossover*—or *genetic recombination*—between the *non-sister chromatids*. Near the recombination nodule on each chromatid, the double-stranded DNA is cleaved, the cut ends are modified, and **a new connection is made between the non-sister chromatids**. A chiasma (pl. chiasmata) is the point of contact, the physical link, between two (non-sister) chromatids belonging to homologous chromosomes. At a given chiasma, an exchange of genetic material (a *chromosomal crossover*) can occur between both chromatids, but this is much more frequent during meiosis than mitosis. In meiosis, absence of a chiasma generally results in improper chromosomal segregation and **aneuploidy** (presence of an abnormal number of chromosomes in a cell, for example a human cell having 45 or 47 chromosomes instead of the usual 46.)

At the end of prophase I, the pairs are held together only at the chiasmata (Figure 9) and are called tetrads because the four sister chromatids of each pair of homologous chromosomes are now visible.
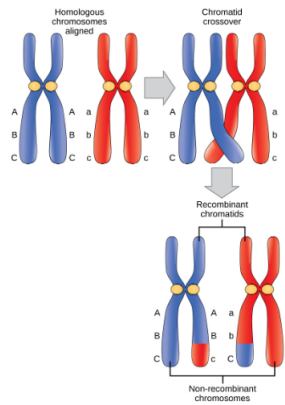
26

Figure 9: Prophase I: Crossover in action

The crossover events are the first source of genetic variation in the nuclei produced by meiosis. A single crossover event between homologous non-sister chromatids leads to a reciprocal exchange of equivalent DNA between a maternal chromosome and a paternal chromosome. Now, when that sister chromatid is moved into a gamete cell it will carry some DNA from one parent of the individual and some DNA from the other parent. **The sister recombinant chromatid has a combination of maternal and paternal genes that did not exist before the crossover**. Multiple crossovers in an arm of the chromosome have the same effect, exchanging segments of DNA to create recombinant chromosomes.

2. **Metaphase I**. During metaphase I, the homologous chromosomes are arranged in the center of the cell with the kinetochores facing opposite poles. The homologous pairs orient themselves *randomly* at the equator. **The crossover events are the first source of genetic variation in the nuclei produced by meiosis**. A single crossover event between homologous non-sister chromatids leads to a reciprocal exchange of equivalent DNA between a maternal chromosome and a paternal chromosome. Now, when that sister chromatid is moved into a gamete cell it will carry some DNA from one parent of the individual and some DNA from the other parent. **The sister recombinant chromatid has a combination of maternal and paternal genes that did not exist before the crossover**. Multiple crossovers in an
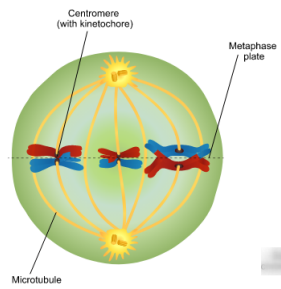
27

Figure 10: Metaphase I: Chromosome pairs move to the equator of the cell

arm of the chromosome have the same effect, exchanging segments of DNA to create recombinant chromosomes.

3.2 Mendel's Experiments and Heredity

3.2.1 Mendel's Experiments and the Laws of Probability

Johann Gregor Mendel conducted experiments to demonstrate the existence of genes long before the term was coined. His experiments are an exemplary lesson in the "design and execution of scientific experiments". Mendel is regarded as the father of modern genetics.

3.2.2 Extra: Mendel's Experiments (Refers to Various Resources)

Mendel wanted to investigate if there was a *general law* for the formation and development of *hybrids*, something he noted had not been previously formulated [1]:

> Those who survey the work done in this department will arrive at the conviction that among all the numerous experiments made, *not one has been carried out to such an extent and in such a way as to make it possible to determine the number of different forms under which the offspring of hybrids appear*, or to arrange these forms with certainty according to their separate generations, or definitely to ascertain their statistical relations.
>
> —*Johann Gregor Mendel [1]*

Mendel and his colleagues carried out their experiments for nearly a decade (1856-1865). Mendel required [1] that the experimental plant must

1. possess constant differentiating factors

2. be capable, during flowering, of protecting from the influence of *foreign pollen*

After careful analysis of various options, Mendel chose "garden pea" (*Pisum Sativum L.*) to experiment with. He decided to experiment with the garden pea and examine following traits:

1. Seed shape: Round or Wrinkled (a *seed*[5] characteristic)

2. Cotyledon (the first leaves that sprout out of the seed upon germination) color: Yellow or Green (a seed characteristic)

---

[5]A characteristic that can be ascertained by simply analyzing the seed

3. Seed-coat color: Colored or White (a *plant*[6] *characteristic*)

4. Pod shape: Inflated or Constricted (a plant characteristic)

5. Pod color: Green or Yellow (a plant characteristic)

6. Flower position: Axial (along the stem) or Terminal (at the end of the stem) (a plant characteristic)

7. Stem length: Long (6–7 feet) or Short (less than a foot) (a plant characteristic)

An interesting and somewhat philosophical question arises: "Chicken first or egg first?" Mendel tried to answer it from what he had available: Generations of "true-bred" plants which were available in the monastery garden where he conducted the experiments.

There is a lot of terminology here and we need to understand just what the *key terms* mean before proceeding. We should also put ourselves into Mendel's shoes and try to reason the way he did. Explaining what Mendel did using the modern Genetics terms is likely to be confusing. For example, we shouldn't be using a term like *homozygous* because Mendel didn't even know what that term meant. Of course, at some point we should reconcile all of this in terms of contemporary knowledge.

Note that a seed belongs to the generation of the plant that comes *after* the [generation of the] plant whose fruit bears the seeds.

Here are the key terms in the order of their dependence (i.e. a term appearing later in the list will depend on the terms appearing before):

1. **Character**: This is a heritable feature, for example, stem length. Typically, an organism simultaneously exhibits several different Characters.

2. **Trait**: Each *variant* of a Character is called a Trait, for example, a plant could have "Tall" (stem length) Trait.

3. **True-bred**: This is an adjective (past participle of "True-breed"), typically applied to an organism. An organism is said to be True-bred from seed alone with respect to certain Traits when it passes down

---

[6]A characteristic that can only be ascertained by waiting for a plant to grow from the seed

those Traits to its offspring (more generally, descendants) *of many generations.* Since flowering plants have both female and male reproductive organs, true-bred Traits of many plants can be created by their *self-pollination.* (Note, however, that flowering plants like apple are notorious for not breeding true from seed; apple needs to be *grafted*). Mendel started with true-bred pea plants mainly because they were relatively much easier to true-breed from seed by self-pollination.

4. **Cross**: Fertilization of the egg by the pollen from another plant. Since a single flowering plant has, unlike humans, both male and female reproductive organs, Mendel had to prevent their self-pollination in order to carry out the desired crosses. In a way, a cross-bred offspring is opposite of a true-bred offspring.

5. **Monohybrid Cross**: Cross between two *true-bred* organisms of a species that differ in a single Trait. Mendel painstakingly carried out such crosses with peas (e.g. by crossing true-bred green-seed plants with true-bred yellow-seed plants).

6. **Dominance**: The ability of a certain Trait (of a given Characteristic) to express in

7. **Submissiveness**:

3.2.3   Characteristics and Traits

3.2.4   Laws of Inheritance

3.3   Modern Understandings of Inheritance

3.4   DNA Structure and Function

Each person's DNA is unique, and it is possible to detect differences among individuals within a species on the basis of these unique features. DNA analysis has many practical applications, including identifying criminals (forensics), determining paternity, tracing genealogy, identifying pathogens, researching archeological finds, tracing disease outbreaks, and studying human migration patterns. In the medical field, DNA is used in diagnostics, new vaccine development, and cancer therapy. It is often possible to determine predisposition to diseases by sequencing genes.

Sometimes an innocent person is erroneously convicted of a crime and sent to jail. Between 2000 and 2015, evidence from DNA was used to exonerate over 250 innocent people. Twenty of those people were on death row after being convicted of a murder they didn't commit. To learn more about the intense scientific and legal processes used to exonerate those wrongfully convicted, go to The Innocence Project website here.

3.4.1 Historical Basis of Modern Understanding

3.5 Genes and Proteins

3.6 Gene Regulation

3.7 Biotechnology and Genomics

4 Unit 4: Evolution

4.1 Evolution and Origin of Species

4.2 The Evolution of Populations

4.3 Phylogenies and the History of Life

5 Unit 5: Viruses

References

[1] Allan Franklin, A. W. F. Edwards, Daniel J. Fairbanks, Daniel L. Hartl. Ending the Fisher-Mendel Controversy. University of Pittsburgh Press, 2008. Page 2.

[2] Yashiro, Toshihisa and Matsuura, Kenji. (2014). Termite queens close the sperm gates of eggs to switch from sexual to asexual reproduction. Proceedings of the National Academy of Sciences. 111. 17212-17217. 10.1073/pnas.1412481111.

[3] Maayan, Inbar. "Meiosis in Humans". Embryo Project Encyclopedia (2011-03-24). ISSN: 1940-5030.

[4] Animal Cell Meiosis.

[5] Crick Francis. "On Protein Synthesis". CSHL Archives Repository.

[6] Mallery Charles. Biology 150 and Biology 255.

[7] Istaván Hargittai, Magdolna Hargittai. Candid Science VI: More Conversations With Famous Scientists. Imperial College Press. 2006. ISBN 1-86094-694-1.

[8] Eugene V. Koonin. "Does the central dogma still stand?" Biology Direct. 23 August 2012.

[9] AlphaFold: Using AI for Scientific Discovery. DeepMind.