

BIOLOGY FOR ENGINEERS

LECTURE NOTES

CO1

The basic elements and their bonding ability, macromolecules and their structures and their role and general law of thermodynamics in biological systems.

SI No	Topics	Page
1	Introducing the topic	2
2	Elements of Life, Explanation about the important elements in any organism. Electronegativity of the element, Importance of carbon.	3
3	Different types of bonds, Some examples of different types of bonding in the biological systems, Water and phospholipids	7
4	Carbohydrates and ATP	17
5	Proteins and their structure	23
6	Enzymes and how enzyme works	28
7	Bioenergetics Pathways and pH	31

Introduction to the course

Purpose of Biology for Engineers

“As engineers, we are always looking into how to modify or improve an existing system. We must hold the key to combining the fields of engineering and biology, and this course is designed to provide a logical understanding about biology from an engineering perspective. So our aim is to understand the logical principles of biology, inspire the ideas from it to have a useful creativity in the engineering field”

Elements of Life, Explanation about the important elements in any organism.

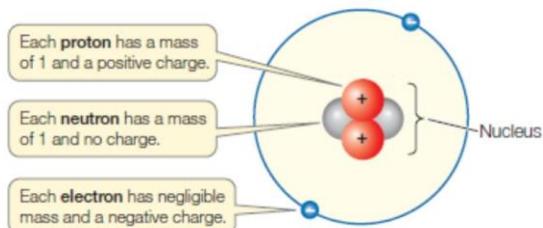
Electronegativity of the element, Importance of carbon.

Chemistry of life:

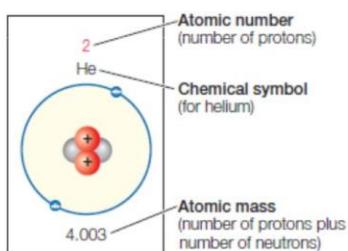
The atomic composition of the cell: H = 63%, O = 24% C = 10%, N = 1.4 %, P = 0.2 % & S = < 0.1%

Trace amount: Ca, Cl, K, Na, Mg, Mn, Fe, Se, I etc

Let us understand what an element is and where these elements related to life are present in the periodic table by simple pictures.



2.1 The Helium Atom This representation of a helium atom is called a Bohr model. It exaggerates the space occupied by the nucleus. In reality, although the nucleus accounts for virtually all of the atomic mass, it occupies only about 1/10,000 of the atom's volume. The Bohr model is also inaccurate in that it represents the electron as a discrete particle in a defined orbit around the nucleus.



2.2 The Periodic Table The periodic table groups the elements according to their physical and chemical properties. Elements 1–92 occur in nature; elements with atomic numbers above 92 were created in the laboratory.

1 H 1.0079	4 Be 9.012	Elements in the same vertical columns have similar properties because they have the same number of electrons in their outermost shell.												2 He 4.003													
3 Li 6.941	11 Na 22.990	12 Mg 24.305	21 Sc 44.956	22 Ti 47.88	23 V 50.942	24 Cr 51.996	25 Mn 54.938	26 Fe 55.847	27 Co 58.933	28 Ni 58.69	29 Cu 63.546	30 Zn 65.38	31 Ga 69.72	32 Ge 72.59	33 As 74.922	34 Se 78.96	35 Br 79.909	36 Kr 83.80									
19 K 39.098	20 Ca 40.08	38 Rb 85.4778	39 Sr 87.62	40 Y 88.906	41 Zr 91.22	42 Nb 92.906	43 Mo (99)	44 Ru 101.07	45 Rh 102.906	46 Pd 106.4	47 Ag 107.870	48 Cd 112.41	49 In 114.82	50 Sn 118.69	51 Te 121.75	52 I 127.60	53 Xe 131.30										
55 Cs 132.905	56 Ba 137.34	71 Lu 174.97	72 Hf 178.49	73 Ta 180.948	74 W 183.85	75 Re 186.207	76 Os 190.2	77 Ir 192.2	78 Pt 195.08	79 Au 196.967	80 Hg 200.59	81 Tl 204.37	82 Pb 207.19	83 Bi 208.980	84 Po (209)	85 At (210)	86 Rn (222)										
87 Fr (223)	88 Ra (260)	103 Lr (260)	104 Rf (261)	105 Db (262)	106 Sg (266)	107 Bh (269)	108 Hs (268)	109 Mt (269)	110 Nh (269)	111 Nh (272)	112 Nh (277)	113 Nh (285)	114 Nh (289)	115 Nh (289)	116 Nh (289)	117 Nh (293)	118 Nh (293)										
Masses in parentheses indicate unstable elements that decay rapidly to form other elements.															Elements without a chemical symbol are as yet unnamed.												
Lanthanide series															Actinide series												
57 La 138.906	58 Ce 140.12	59 Pr 140.9077	60 Nd 144.24	61 Pm (145)	62 Sm 150.36	63 Eu 151.96	64 Gd 157.25	65 Tb 158.924	66 Dy 162.50	67 Ho 164.930	68 Er 167.26	69 Tm 168.934	70 Yb 173.04	89 Ac 227.028	90 Th 232.038	91 Pa 231.0359	92 U 238.02	93 Np 237.0482	94 Pu (244)	95 Am (243)	96 Cm (247)	97 Bk (247)	98 Cf (251)	99 Es (252)	100 Fm (257)	101 Md (258)	102 No (259)

When we look at the composition of elements, C, N, O & H constitutes more than 95%. They are also lightest elements in the periodic table (As we go down in the periodic table, atomic number is going to increase so as atomic mass)

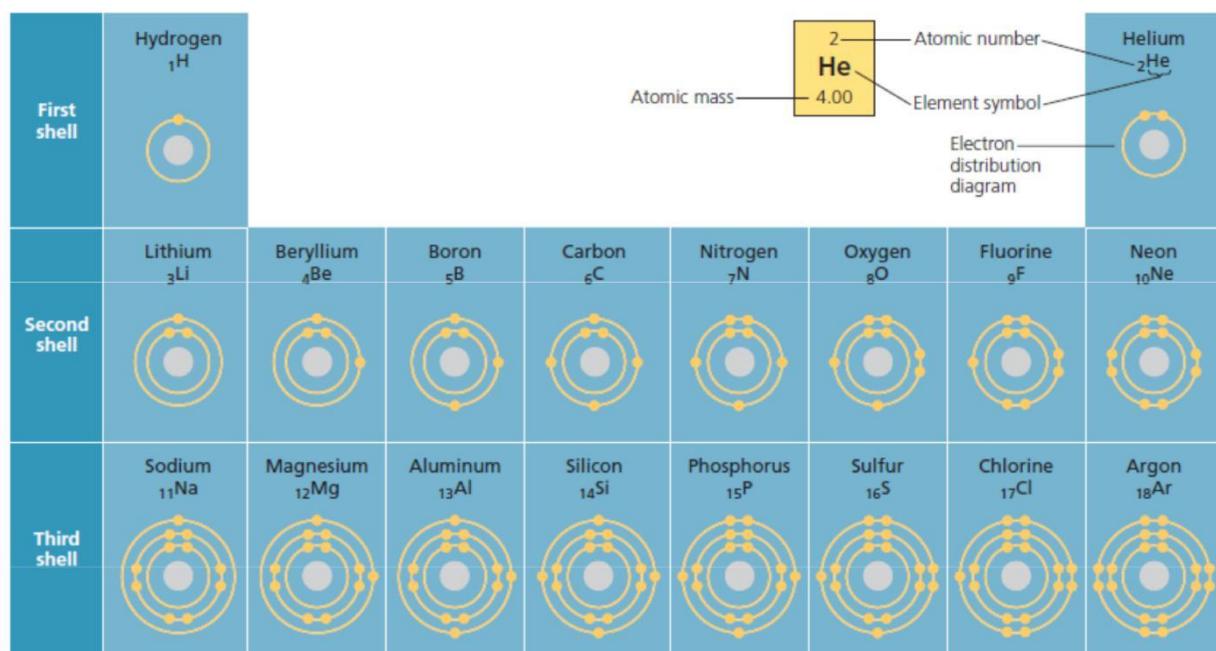
Suppose we want to design a moving machine like car what are the critical things we look for before selecting material to construct a body of the car? One of the criteria should be Materials used should not be heavy (fuel efficiency is going to decrease). The backbone element of life is carbon and carbon is the appropriate element to become a backbone element of life, since no element is present above carbon in periodic table which can have similar properties and lighter than carbon.

Important properties of these elements are their ability to form bond with other elements to form compounds/molecules. Let us see the molecular composition of the life. 80% is water & the dry weight of remaining 20% contains 50% protein, 15% carbohydrates, 10% lipids & fats & 15% nucleic acids.

Formation of these molecules and the interaction between these molecules depends on the chemical properties of the important six elements we mentioned it before. In this course, we just

learn Chemistry to logically understand the structure and functions of Biomolecules and their interaction.

Two important properties we are concentrating in this course to understand the elements used in life are valency and electronegativity. The Valency of an atom is unpaired electron in the outer orbital of the shell. This gives an opportunity for the element to combine with other element.



▲ Figure 2.9 Electron distribution diagrams for the first 18 elements in the periodic table. In a standard periodic table (see Appendix B), information for each element is presented as shown for helium in the inset. In the diagrams in this table, electrons are represented as yellow dots and electron

shells as concentric circles. These diagrams are a convenient way to picture the distribution of an atom's electrons among its electron shells, but these simplified models do not accurately represent the shape of the atom or the location of its electrons. The elements are arranged in rows, each representing the filling of an

electron shell. As electrons are added, they occupy the lowest available shell.

? *What is the atomic number of magnesium? How many protons and electrons does it have? How many electron shells? How many valence electrons?*

Valency of H = 1, C=4, O = 2, N = 3 or 4, P = 3 or 5 and S = 2 (As shown in the figure), 4, 6

Covalent Bond: sharing of a pair of valence electrons by two similar or dissimilar elements.

C–C single bond. Energy required to break them is equal to 80 Kcal/mol

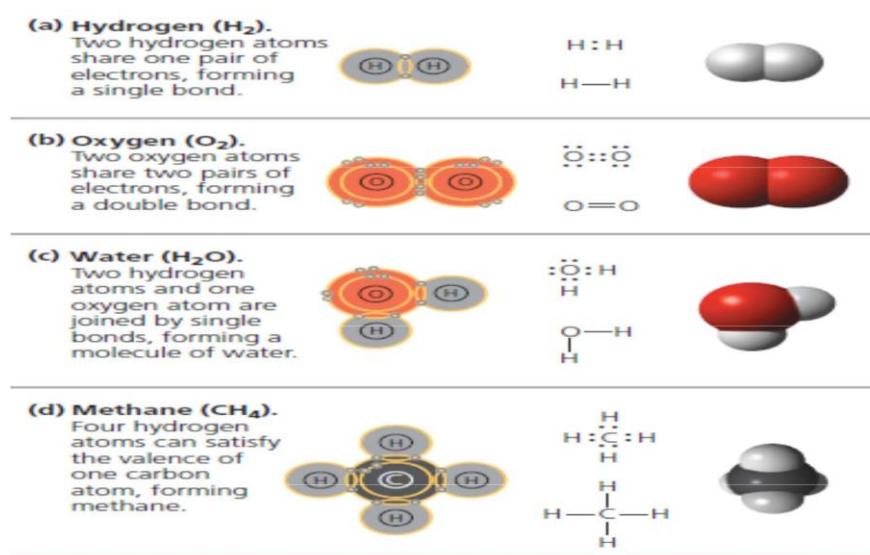
C=C double bond (more energy is required to break them compared to single bond between them)

C≡C triple bond (more energy is required to break them compared to double bond)

Covalent bonds are very strong. Suppose if we compare covalent bond strength to say random energy fluctuation in daily life- random thermal fluctuations at room temperature are on the order of 0.6 kilocalories per mole. Covalent bonds are extremely stable, usually, unless something is attacking them and breaking them.

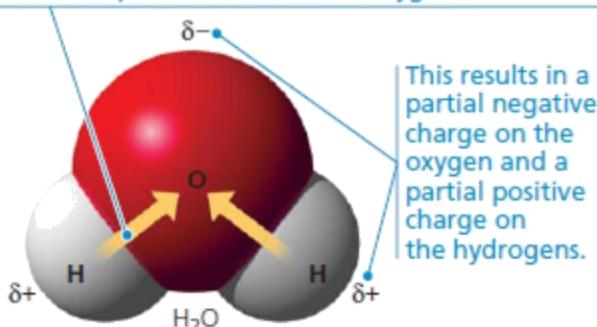
Different types of bonds, Some examples of different types of bonding in the biological systems, Water & phospholipid their importance in existence of life

Atoms in a molecule attract shared electrons in varying degrees, depending on the element. The attraction of a particular element for the electrons of a covalent bond is called its electronegativity. The more electronegative this element is, the more strongly it pulls shared electrons towards itself. If an element bonded to more electronegative element, the electrons of the bond are not shared equally therefore, There exists a polarity between them. This type of bond is called polar covalent bond. Such bonds vary in their polarity depending on the relative electronegativity of the two elements. For example the bond between the hydrogen atom and oxygen atom in water molecule is quite polar.



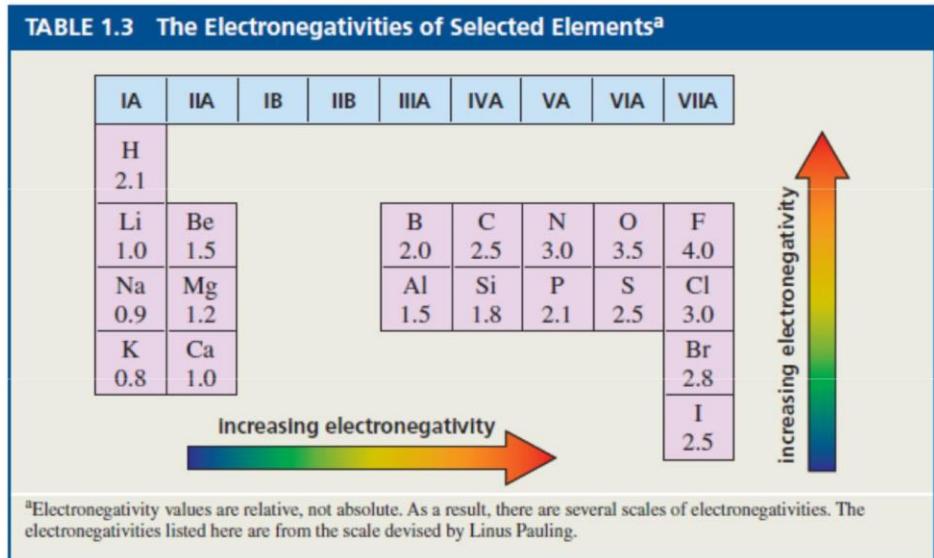
▲ **Figure 2.12 Covalent bonding in four molecules.** The number of electrons required to complete an atom's valence shell generally determines how many covalent bonds that atom will form. This figure shows several ways of indicating covalent bonds.

Because oxygen (O) is more electronegative than hydrogen (H), shared electrons are pulled more toward oxygen.



▲ **Figure 2.13 Polar covalent bonds in a water molecule.**

How do we know that polarity exist between two atoms of elements? What is the measuring way?



Please refer the figure above. Suppose if the difference in electronegativity between the two atoms is 0.5 and more, there exist polarity.

Example: carbon and hydrogen C-H the difference in electronegativity is 0.4 so it is non-polar. Carbon and oxygen C-O The difference is 1.0 therefore polarity exist.

Ionic Bond: transfer of electrons from one atom to another atom to form bond. The atom should form ions i.e. it should be in ionic state (positively charged or negatively charged) before it forms bond with another oppositely charged ion. The bonds are strong as long as it is not disturbed. If it is disturbed it becomes fragile. Ex: Once the water is added to NaCl, the ionic bond breaks.

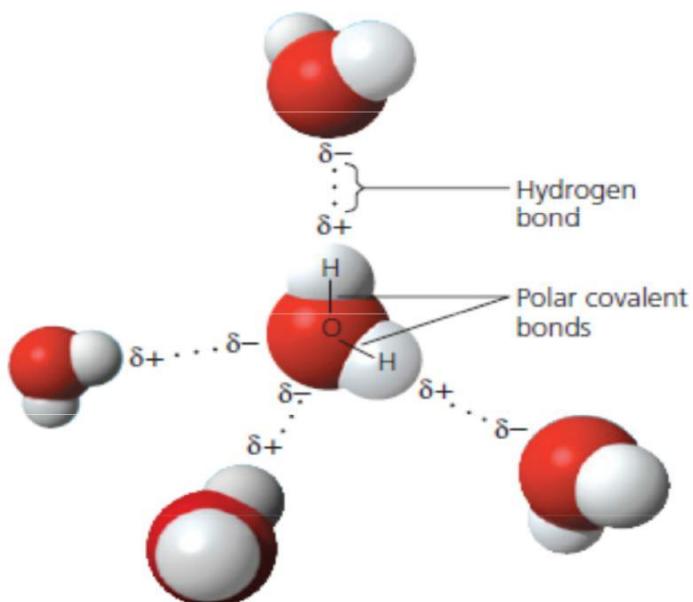
Water and Phospholipid

Life on earth began in water and evolved there for three billion years before spreading onto land. Although most of the water in liquid form, it is also in solid form and gaseous form.

Water is the only solvent, it is present in all the three phases and interchange of the phases will enormously affect the life on the earth. We look into some of the important properties of water that make earth suitable for life.

Polar covalent bonds in water molecules results in hydrogen bonding

The hydrogen bonds form, break and re-form with great frequency. Each lasts only a few trillionths of a second, but the molecules are constantly forming new hydrogen bonds with a succession of partners. Therefore at any instant, all water molecules are hydrogen bonded to their neighbours.

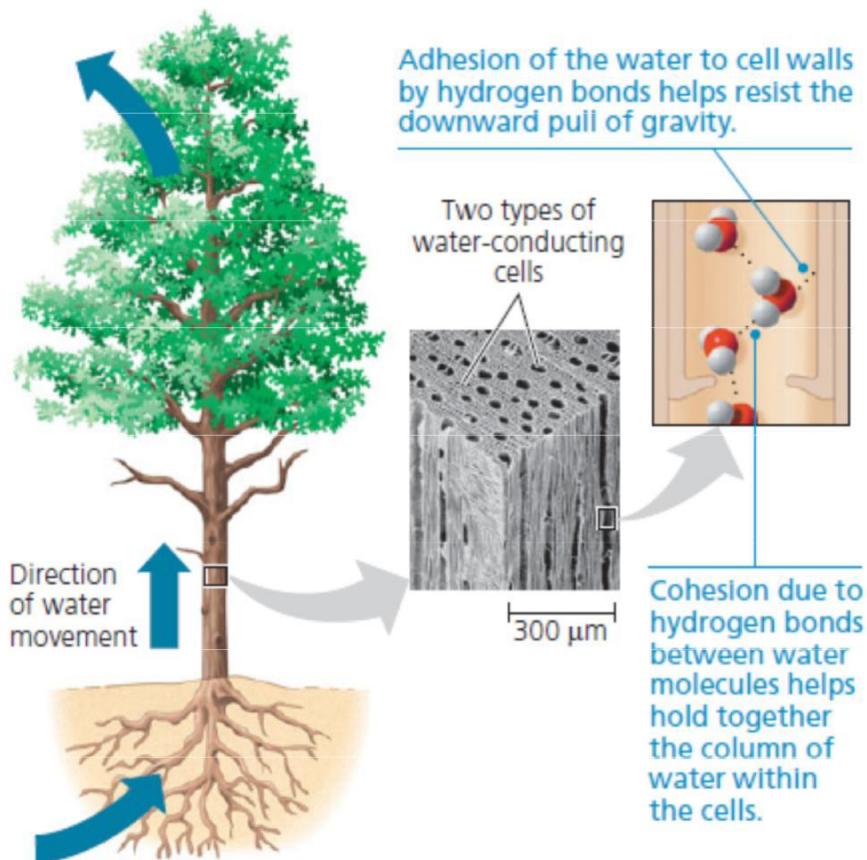


▲ **Figure 3.2 Hydrogen bonds between water molecules.**

The charged regions in a water molecule are due to its polar covalent bonds. Oppositely charged regions of neighboring water molecules are attracted to each other, forming hydrogen bonds. Each molecule can hydrogen-bond to multiple partners, and these associations are constantly changing.

Cohesive and adhesive properties of water

Water molecules stay close to each other as a result of hydrogen bonding. Although the arrangement of molecules in a sample of liquid water is constantly changing, at any given moment many of the molecules are linked by multiple hydrogen bonds. These linkages make water more structured than most other liquids. Collectively, the hydrogen bonds hold the substance together, a phenomenon called cohesion.



▲ Figure 3.3 Water transport in plants. Evaporation from leaves pulls water upward from the roots through water-conducting cells. Because of the properties of cohesion and adhesion, the tallest trees can transport water more than 100 m upward—approximately one-quarter the height of the Empire State Building in New York City.

Adhesion, the clinging of one substance to another, also plays a role. Adhesion of water to cell walls by hydrogen bonds helps counter the downward pull of gravity

Moderation of Temperature by Water

Water moderates air temperature by absorbing heat from air that is warmer and releasing the stored heat to air that is cooler. Water is effective as a heat bank because it can absorb or release a relatively large amount of heat with only a slight change in its own temperature. How it can do that?

The ability of water to stabilize temperature stems from its relatively high specific heat. The specific heat of water is 1 calorie per gram and per degree Celsius, abbreviated as 1 cal/g °C. Compared with most other substances, water has an unusually high specific heat. Because of the high specific heat of water relative to other materials, water will change its temperature less when

it absorbs or loses a given amount of heat. We can trace water's high specific heat, like many of its other properties, to hydrogen bonding. Heat must be absorbed in order to break hydrogen bonds; by the same token, heat is released when hydrogen bonds form. A calorie of heat causes a relatively small change in the temperature of water because much of the heat is used to disrupt hydrogen bonds before the water molecules can begin moving faster. And when the temperature of water drops slightly, many additional hydrogen bonds form, releasing a considerable amount of energy in the form of heat.

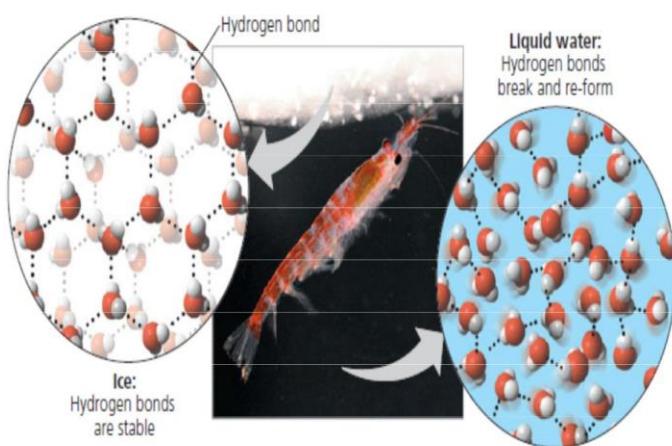
What is the relevance of water's high specific heat to life on Earth? A large body of water can absorb and store a huge amount of heat from the sun in the daytime and during summer while warming up only a few degrees. At night and during winter, the gradually cooling water can warm the air. This is the reason coastal areas generally have milder climates than inland regions. The high specific heat of water also tends to stabilize ocean temperatures, creating a favorable environment for marine life. Thus, because of its high specific heat, the water that covers most of Earth keeps temperature fluctuations on land and in water within limits that permit life.

Floating of Ice on Liquid Water

Water is one of the few substances that are less dense as a solid than as a liquid. In other words, ice floats on liquid water. While other materials contract and become denser when they solidify, water expands. How it helps for life?

► **Figure 3.6** Ice: crystalline structure and floating barrier. In ice, each molecule is hydrogen-bonded to four neighbors in a three-dimensional crystal. Because the crystal is spacious, ice has fewer molecules than an equal volume of liquid water. In other words, ice is less dense than liquid water. Floating ice becomes a barrier that protects the liquid water below from the colder air. The marine organism shown here is a type of shrimp called krill; it was photographed beneath floating ice in the Southern Ocean near Antarctica.

WHAT IF? If water did not form hydrogen bonds, what would happen to the shrimp's environment?



Water: The Solvent of Life

Water is a very good solvent. Many reactions take place in an organism. For almost all the reactions, water acts as a solvent.

Possible Evolution of Life on Other Planets with Water

Biologists who look for life elsewhere in the universe have concentrated their search on planets that might have water. To date, more than 200 planets have been found outside our solar system, and there is evidence for the presence of water vapor on one or two of them. In our own solar system, Mars has been most compelling to biologists as a focus of study. Like Earth, Mars has an ice cap at both poles.



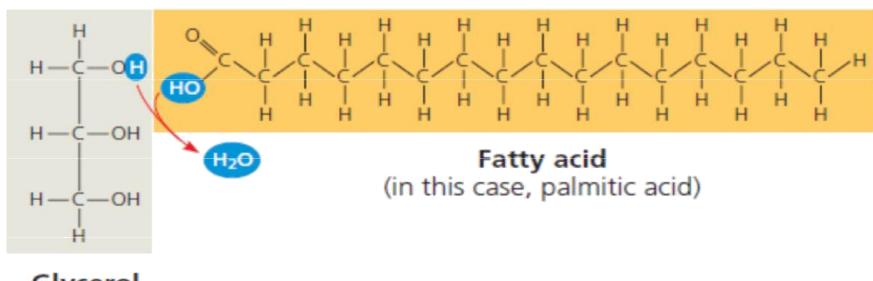
◀ **Figure 3.9 Subsurface ice and morning frost on Mars.** This photograph was taken by the Mars lander *Phoenix* in 2008. The trench was scraped by a robotic arm, uncovering ice (white in rectangle near bottom) below the surface material. Frost also appears as a white coating in several places in the upper half of the image. This photograph was colorized by NASA to highlight the ice.

Lipids and phospholipids:

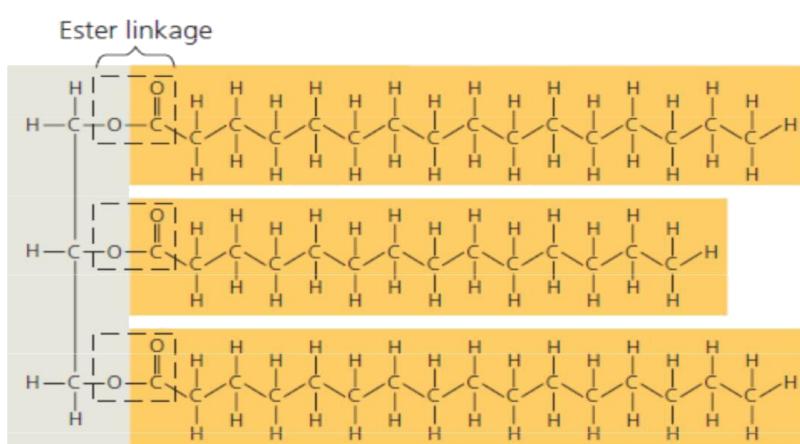
Lipids are hydrophobic molecules. The hydrophobic behavior of lipids is based on their molecular structure. Although they may have some polar bonds associated with oxygen, lipids consist mostly of hydrocarbon regions. Lipids are varied in form and function. They include waxes and certain pigments, but we will focus on the most biologically important types of lipids: fats & phospholipids.

A fatty acid has a long carbon skeleton, usually 16 or 18 carbon atoms in length. The carbon at one end of the skeleton is part of a carboxyl group, the functional group that gives these

molecules the name fatty acid. The rest of the skeleton consists of a hydrocarbon chain. The relatively nonpolar C–H bonds in the hydrocarbon chains of fatty acids are the reason fats are hydrophobic. Fats separate from water because the water molecules hydrogen bond to one another and exclude the fats. In making a fat, three fatty acid molecules are each joined to glycerol by an ester linkage, a bond between a hydroxyl group and a carboxyl group. The resulting fat, also called a triacylglycerol, thus consists of three fatty acids linked to one glycerol molecule.



(a) One of three dehydration reactions in the synthesis of a fat



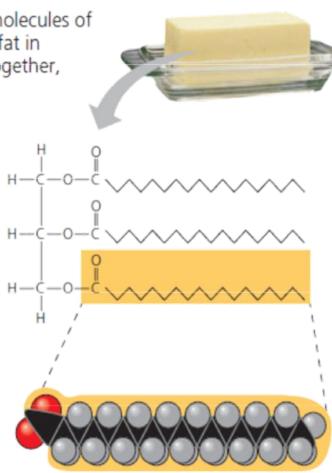
(b) Fat molecule (triacylglycerol)

The terms saturated fats and unsaturated fats are commonly used in the context of nutrition. These terms refer to the structure of the hydrocarbon chains of the fatty acids. If there are no double bonds between carbon atoms composing a chain, then as many hydrogen atoms as possible are bonded to the carbon skeleton. Such a structure is said to be saturated with hydrogen, and the resulting fatty acid therefore called a saturated fatty acid. An unsaturated fatty acid has one or more double bonds, with one fewer hydrogen atom on each double-bonded carbon. Nearly all double bonds in naturally occurring fatty acids are cis double bonds, which cause a kink in the hydrocarbon chain wherever they occur

(a) Saturated fat

At room temperature, the molecules of a saturated fat, such as the fat in butter, are packed closely together, forming a solid.

Structural formula of a saturated fat molecule (Each hydrocarbon chain is represented as a zigzag line, where each bend represents a carbon atom and hydrogens are not shown.)

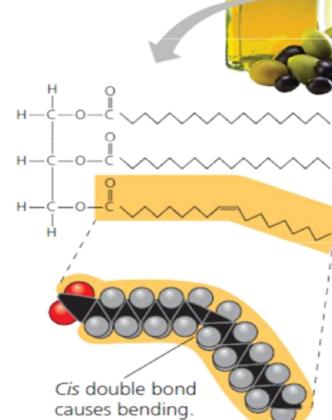


Space-filling model of stearic acid, a saturated fatty acid (red = oxygen, black = carbon, gray = hydrogen)

(b) Unsaturated fat

At room temperature, the molecules of an unsaturated fat such as olive oil cannot pack together closely enough to solidify because of the kinks in some of their fatty acid hydrocarbon chains.

Structural formula of an unsaturated fat molecule

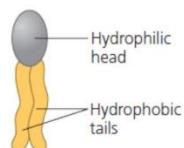
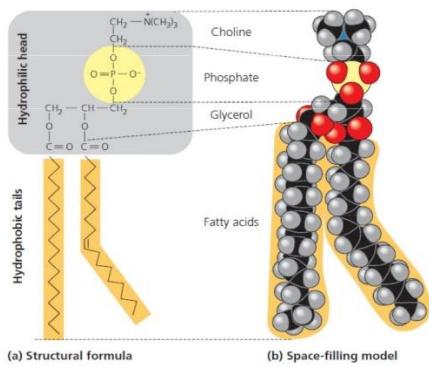


Space-filling model of oleic acid, an unsaturated fatty acid

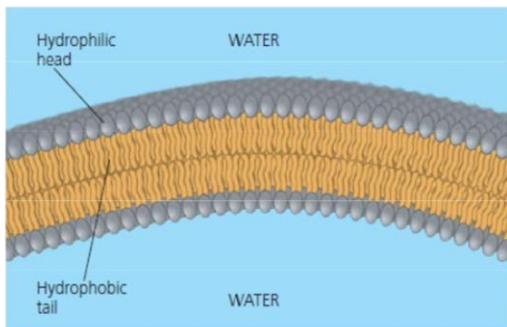
Cis double bond causes bending.

Phospholipids are essential for cells because they make up cell membranes. Phospholipid is similar to a fat molecule but has only two fatty acids attached to glycerol rather than three. The third hydroxyl group of glycerol is joined to a phosphate group, which has a negative electrical charge in the cell. Additional small molecules, which are usually charged or polar, can be linked to the phosphate group to form a variety of phospholipids. The two ends of phospholipids show different behaviour toward water. The hydrocarbon tails are hydrophobic and are excluded from water. However, the phosphate group and its attachments form a hydrophilic head that has an affinity for water. These strange behaviour molecules are called amphipathic molecules. When phospholipids are added to water, they self-assemble into double-layered structures called "bilayers," shielding their hydrophobic portions from water. At the surface of a cell, phospholipids are arranged in a similar bilayer. The hydrophilic heads of the molecules are on the outside of the

bilayer, in contact with the aqueous solutions inside and outside of the cell. The hydrophobic tails point toward the interior of the bilayer, away from the water. The phospholipid bilayer forms a boundary between the cell and its external environment; in fact, cells could not exist without phospholipids



(c) Phospholipid symbol

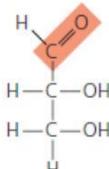
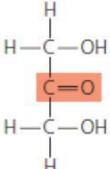
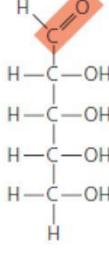
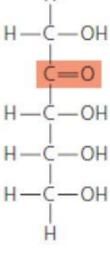
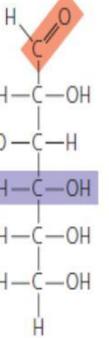
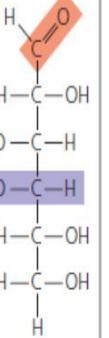
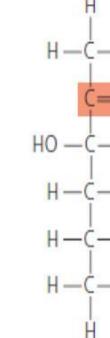


▲ Figure 5.13 Bilayer structure formed by self-assembly of phospholipids in an aqueous environment. The phospholipid bilayer shown here is the main fabric of biological membranes. Note that the hydrophilic heads of the phospholipids are in contact with water in this structure, whereas the hydrophobic tails are in contact with each other and remote from water.

Carbohydrates & ATP

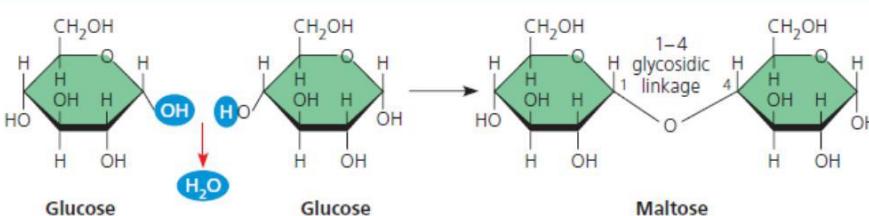
Carbohydrates include both sugars and polymers of sugars. The simplest carbohydrates are the monosaccharides, or simple sugars; these are the monomers from which more complex carbohydrates are constructed. Disaccharides are double sugars, consisting of two monosaccharides joined by a covalent bond. Carbohydrates also include macromolecules called polysaccharides, polymers composed of many sugar building blocks.

Monosaccharides (from the Greek monos, single, and sacchar, sugar) generally have molecular formulas that are some multiple of the unit CH_2O . Glucose ($\text{C}_6\text{H}_{12}\text{O}_6$) (multiple of six), the most common monosaccharide, is of central importance in the chemistry of life. In the structure of glucose, we can see the trademarks of a sugar: The molecule has a carbonyl group (C=O) and multiple hydroxyl groups ($—\text{OH}$). Depending on the location of the carbonyl group, a sugar is either an aldose (aldehyde sugar) or a ketose (ketone sugar). Glucose, for example, is an aldose; fructose, an isomer of glucose, is a ketose. (Most names for sugars end in -ose.) Another criterion for classifying sugars is the size of the carbon skeleton, which ranges from three to seven carbons long. Glucose, fructose, and other sugars that have six carbons are called hexoses. Trioses (three-carbon sugars) and pentoses (five-carbon sugars) are also common.

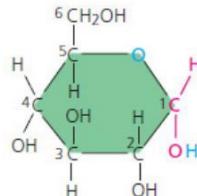
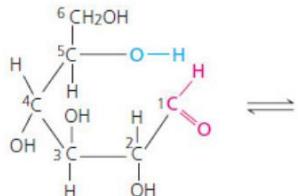
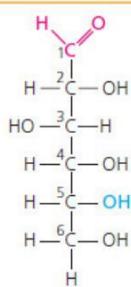
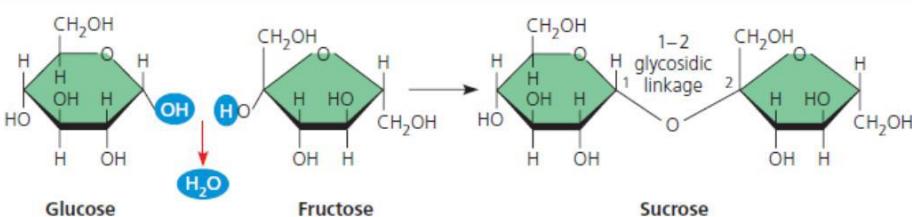
Aldoses (Aldehyde Sugars) Carbonyl group at end of carbon skeleton	Ketoses (Ketone Sugars) Carbonyl group within carbon skeleton
Trioses: 3-carbon sugars ($\text{C}_3\text{H}_6\text{O}_3$)	
 Glyceraldehyde An initial breakdown product of glucose	 Dihydroxyacetone An initial breakdown product of glucose
Pentoses: 5-carbon sugars ($\text{C}_5\text{H}_{10}\text{O}_5$)	
 Ribose A component of RNA	 Ribulose An intermediate in photosynthesis
Hexoses: 6-carbon sugars ($\text{C}_6\text{H}_{12}\text{O}_6$)	
 Glucose Energy sources for organisms	 Galactose Energy sources for organisms
	 Fructose An energy source for organisms

A disaccharide consists of two monosaccharides joined by a glycosidic linkage, a covalent bond formed between two monosaccharides by a dehydration reaction. For example, maltose is a disaccharide formed by the linking of two molecules of glucose. Also known as malt sugar, maltose is an ingredient used in brewing beer. The most prevalent disaccharide is sucrose, which is table sugar. Its two monomers are glucose and fructose. Plants generally transport carbohydrates from leaves to roots and other nonphotosynthetic organs in the form of sucrose. Lactose, the sugar present in milk, is another disaccharide, in this case a glucose molecule joined to a galactose molecule.

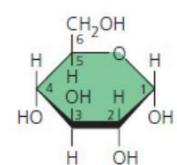
(a) Dehydration reaction in the synthesis of maltose. The bonding of two glucose units forms maltose. The glycosidic linkage joins the number 1 carbon of one glucose to the number 4 carbon of the second glucose. Joining the glucose monomers in a different way would result in a different disaccharide.



(b) Dehydration reaction in the synthesis of sucrose. Sucrose is a disaccharide formed from glucose and fructose. Notice that fructose, though a hexose like glucose, forms a five-sided ring.



(a) Linear and ring forms. Chemical equilibrium between the linear and ring structures greatly favors the formation of rings. The carbons of the sugar are numbered 1 to 6, as shown. To form the glucose ring, carbon 1 bonds to the oxygen attached to carbon 5.



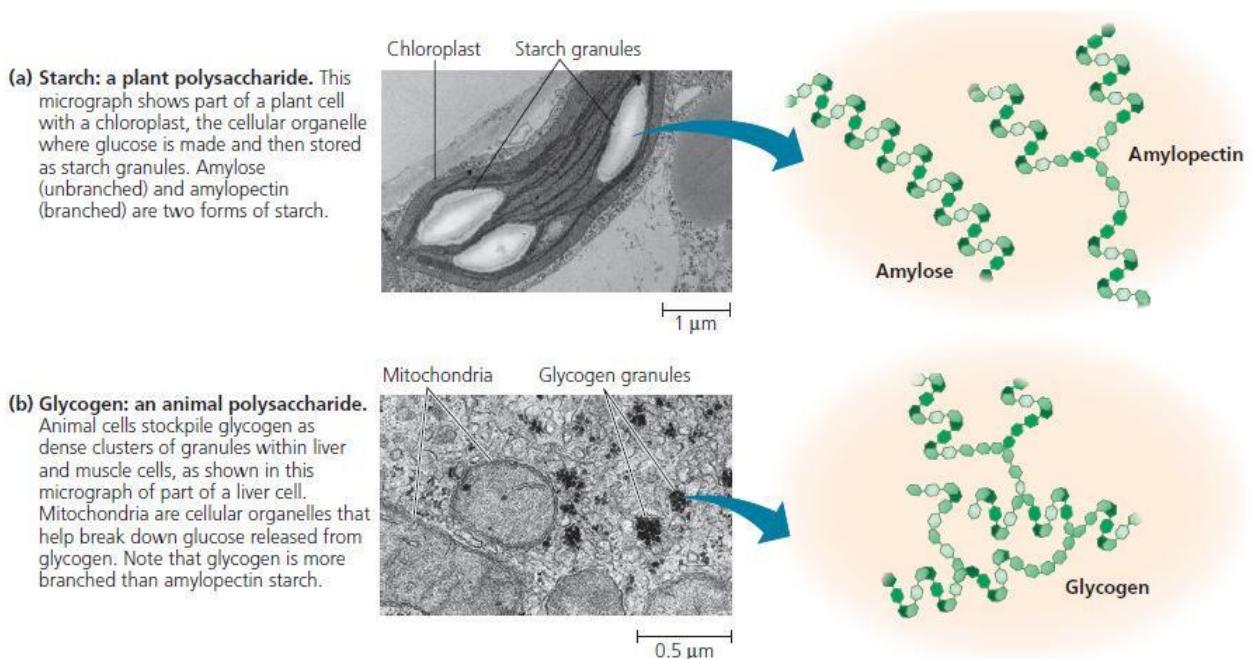
(b) Abbreviated ring structure. Each corner represents a carbon. The ring's thicker edge indicates that you are looking at the ring edge-on; the components attached to the ring lie above or below the plane of the ring.

Polysaccharides are macromolecules, polymers with a few hundred to a few thousand monosaccharides joined by glycosidic linkages. Some polysaccharides serve as storage material, hydrolyzed as needed to provide sugar for cells. Other polysaccharides serve as building material for structures that protect the cell or the whole organism. The architecture and function of a

polysaccharide are determined by its sugar monomers and by the positions of its glycosidic linkages.

Both plants and animals store sugars for later use in the form of storage polysaccharides. Plants store starch, a polymer of glucose monomers, as granules within cellular structures known as plastids, which include chloroplasts. Synthesizing starch enables the plant to stockpile surplus glucose. Because glucose is a major cellular fuel, starch represents stored energy. The sugar can later be withdrawn from this carbohydrate “bank” by hydrolysis, which breaks the bonds between the glucose monomers. Most animals, including humans, also have enzymes that can hydrolyze plant starch, making glucose available as a nutrient for cells. Potato tubers and grains—the fruits of wheat, maize (corn), rice, and other grasses—are the major sources of starch in the human diet.

Most of the glucose monomers in starch are joined by 1–4 linkages (number 1 carbon to number 4 carbon), like the glucose units in maltose. The simplest form of starch, amylose, is unbranched. Amylopectin, a more complex starch, is a branched polymer with 1–6 linkages at the branch points. Animals store a polysaccharide called glycogen, a polymer of glucose that is like amylopectin but more extensively branched. Humans and other vertebrates store glycogen mainly in liver and muscle cells. Hydrolysis of glycogen in these cells releases glucose when the demand for sugar increases.



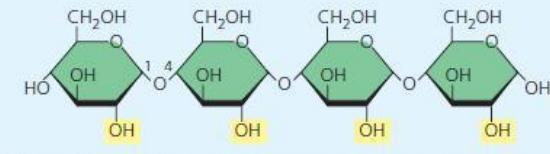
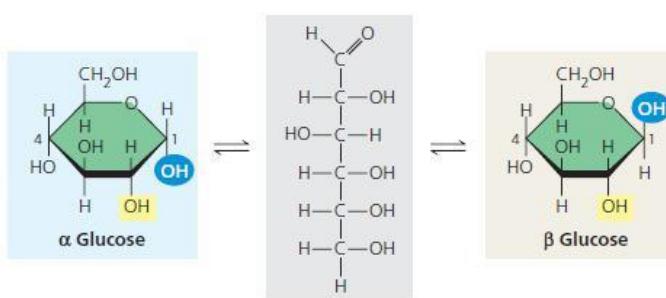
Structural Polysaccharides

Organisms build strong materials from structural polysaccharides. For example, the polysaccharide called cellulose is a major component of the tough walls that enclose plant cells. On a global scale, plants produce almost 10^{14} kg (100 billion tons) of cellulose per year; it is the most abundant organic compound on Earth. Like starch, cellulose is a polymer of glucose, but the glycosidic linkages in these two polymers differ. The difference is based on the fact that there are actually two slightly different ring structures for glucose. When glucose forms a ring, the hydroxyl group attached to the number 1 carbon is positioned either below or above the plane of the ring. These two ring forms for glucose are called alpha (α) and beta (β), respectively. In starch, all the glucose monomers are in the α configuration. The glucose monomers of cellulose are all in the β configuration, making every glucose monomer “upside down” with respect to its neighbors.

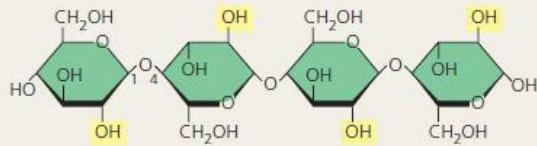
The differing glycosidic linkages in starch and cellulose give the two molecules distinct three-dimensional shapes. Whereas certain starch molecules are largely helical, a cellulose molecule is straight. Cellulose is never branched, and some hydroxyl groups on its glucose monomers are free to hydrogen-bond with the hydroxyls of other cellulose molecules lying parallel to it. In plant cell walls, parallel cellulose molecules held together in this way are grouped into units called microfibrils. These cable-like microfibrils are a strong building material for plants and an important substance for humans because cellulose is the major constituent of paper and the only component of cotton.

Enzymes that digest starch by hydrolyzing its α linkages are unable to hydrolyze the β linkages of cellulose because of the distinctly different shapes of these two molecules. In fact, few organisms possess enzymes that can digest cellulose. Animals, including humans, do not; the cellulose in our food passes through the digestive tract and is eliminated with the feces. Along the way, the cellulose abrades the wall of the digestive tract and stimulates the lining to secrete mucus, which aids in the smooth passage of food through the tract. Thus, although cellulose is not a nutrient for humans, it is an important part of a healthful diet. Most fresh fruits, vegetables, and whole grains are rich in cellulose. On food packages, “insoluble fiber” refers mainly to cellulose.

(a) α and β glucose ring structures. These two interconvertible forms of glucose differ in the placement of the hydroxyl group (highlighted in blue) attached to the number 1 carbon.



(b) Starch: 1-4 linkage of α glucose monomers. All monomers are in the same orientation. Compare the positions of the —OH groups highlighted in yellow with those in cellulose (c).

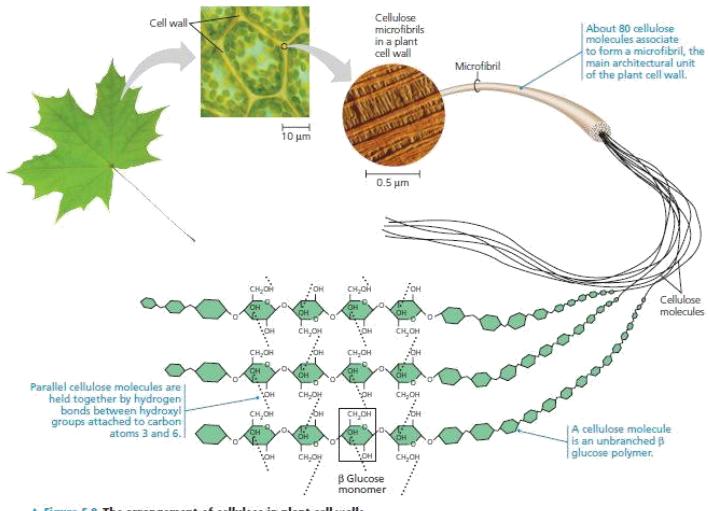


(c) Cellulose: 1-4 linkage of β glucose monomers. In cellulose, every β glucose monomer is upside down with respect to its neighbors.

Some microorganisms can digest cellulose, breaking it down into glucose monomers. A cow harbors cellulose digesting prokaryotes and protists in its stomach. These microbes hydrolyze the cellulose of hay and grass and convert the glucose to other compounds that nourish the cow. Similarly, a termite, which is unable to digest cellulose by itself, has prokaryotes or protists living in its gut that can make a meal of wood.

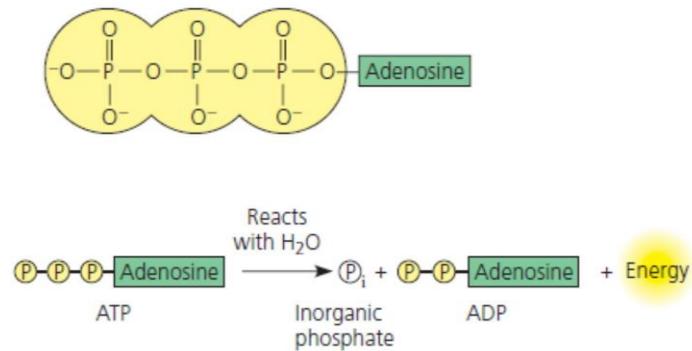
Adenosine triphosphate (ATP)

ATP is an important source of energy for cellular processes. ATP consists of an organic molecule called adenosine attached to a string of three phosphate groups: Where three phosphates are present in series, as in ATP, one phosphate may be split off as a result of a reaction with water. This inorganic phosphate ion, HOPO_3^{2-} , is often referred as a phosphate group.



▲ Figure 5.8 The arrangement of cellulose in plant cell walls.

While losing this group, it releases energy equivalent to 7.3 Kcal/mol. Energy required for most of the chemical reactions in the cell is used by releasing phosphate group. Energy is also stored in the form of ATP by the cell.

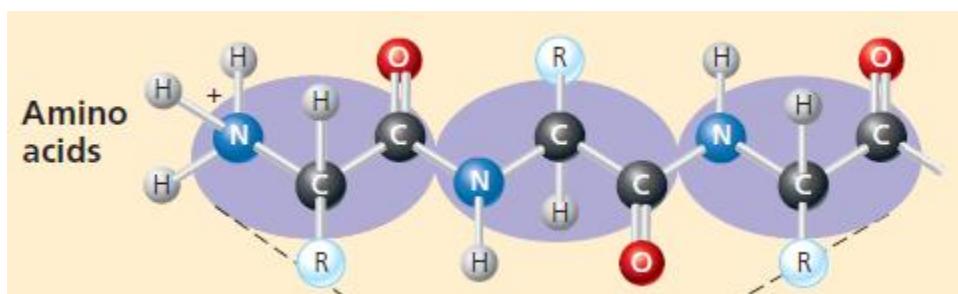


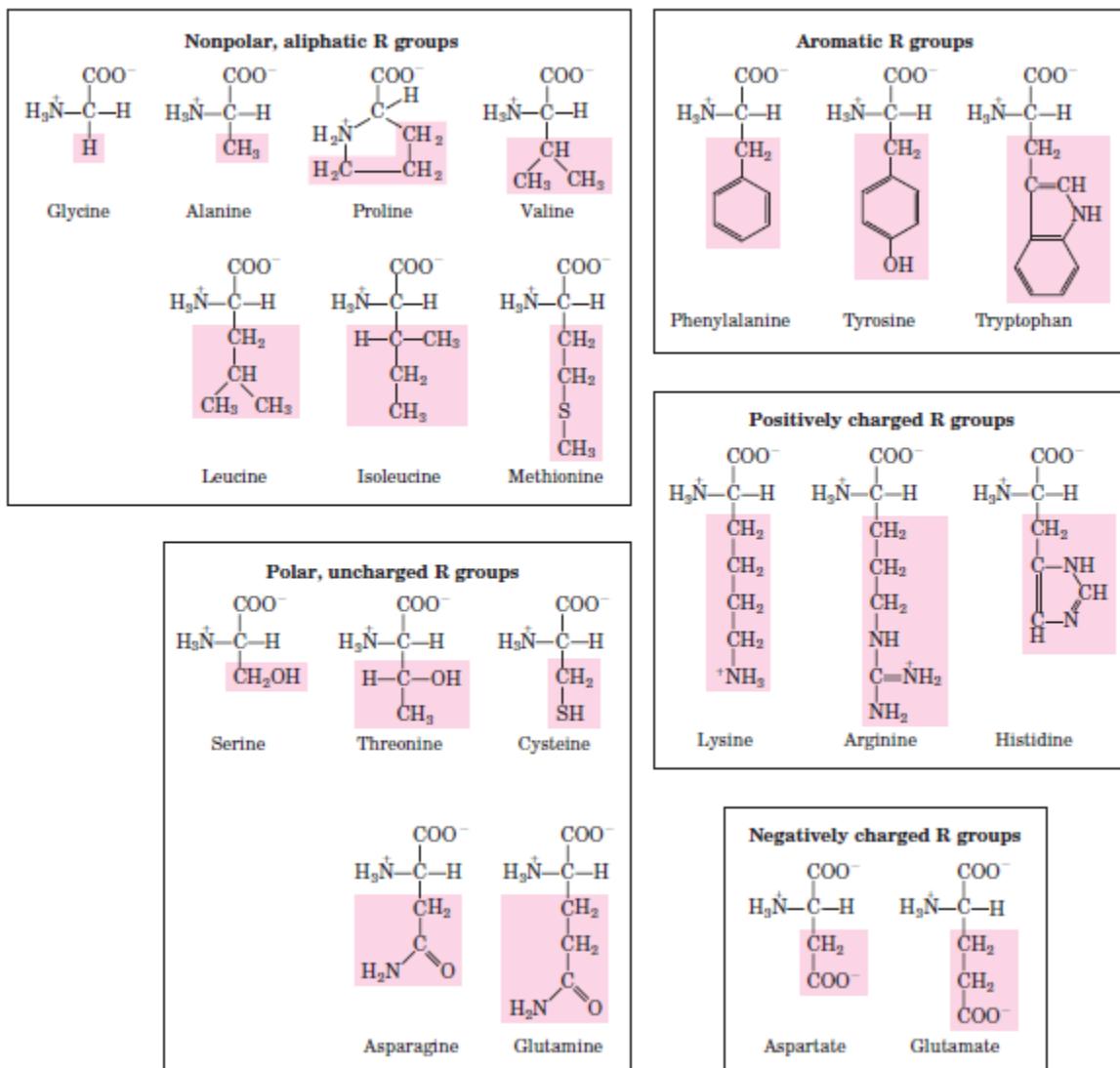
Proteins and their structure

Proteins are the most abundant biological macromolecules, occurring in all parts of the cell. Proteins also occur in great variety; thousands of different kinds, ranging in size from relatively small peptides to huge polymers with molecular weights in millions. Nearly every dynamic function of a living being depends on proteins. In fact, the importance of proteins is underscored by their name, which comes from the Greek word *proteios*, meaning "first," or "primary." Proteins account for more than 50% of the dry mass of most cells, and they are instrumental in almost everything the organism does. Some proteins speed up chemical reactions, while others play a role in defense, storage, transport, cellular communication, movement, or structural support.

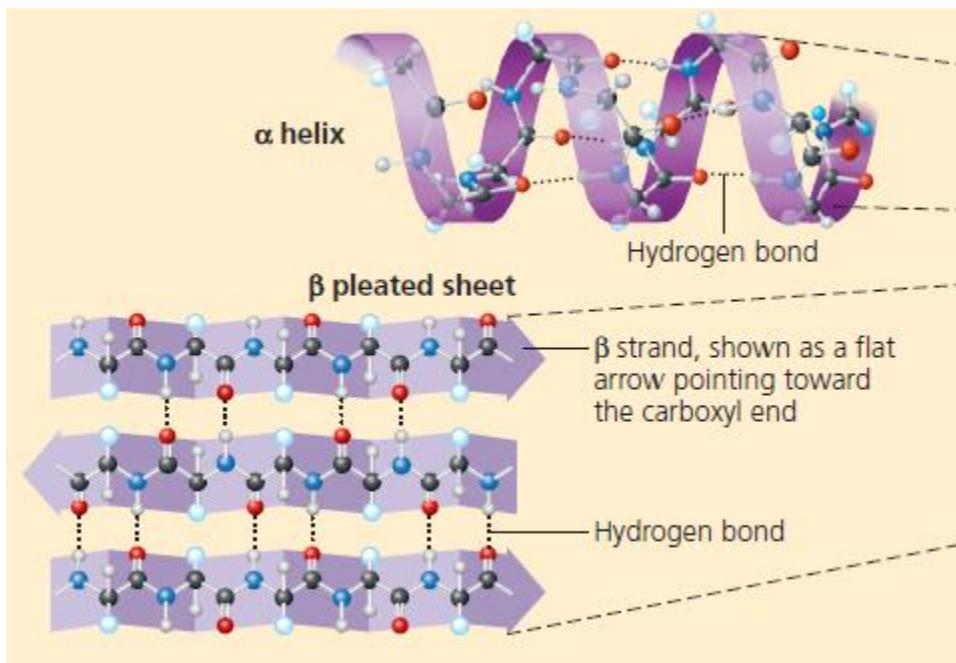
The basic building blocks of proteins are amino acids. Twenty different amino acids are commonly found in proteins. All 20 of the common amino acids are α -amino acids. They have a carboxyl group and an amino group bonded to the same carbon atom (α carbon). They differ from each other in their side chains, or **R groups**, which vary in structure, size, and electric charge, and which influence the solubility of the amino acids in water. The common amino acids of proteins have been assigned three-letter abbreviations and one-letter symbols, which are used as shorthand to indicate the composition and sequence of amino acids polymerized in proteins.

The **primary structure** of a protein is simply the linear arrangement, or *sequence*, of the amino acid residues that compose it. Many terms are used to denote the chains formed by the polymerization of amino acids. A short chain of amino acids linked by peptide bonds and having a defined sequence is called a **peptide**; longer chains are referred to as **polypeptides**. Peptides generally contain less than 20–30 amino acid residues, whereas polypeptides contain as many as 4000 residues. We generally reserve the term **protein** for a polypeptide (or for a complex of polypeptides) that has a well-defined three-dimensional structure.



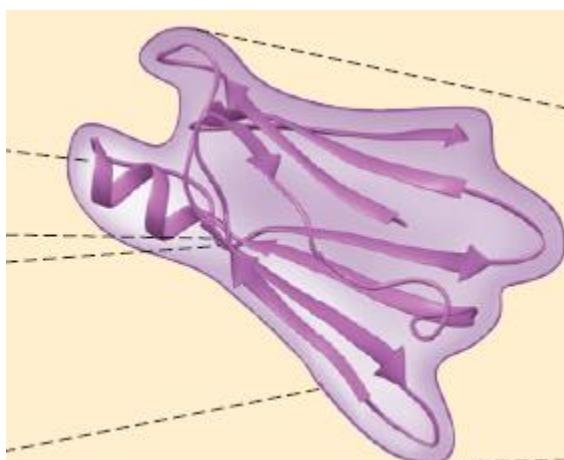
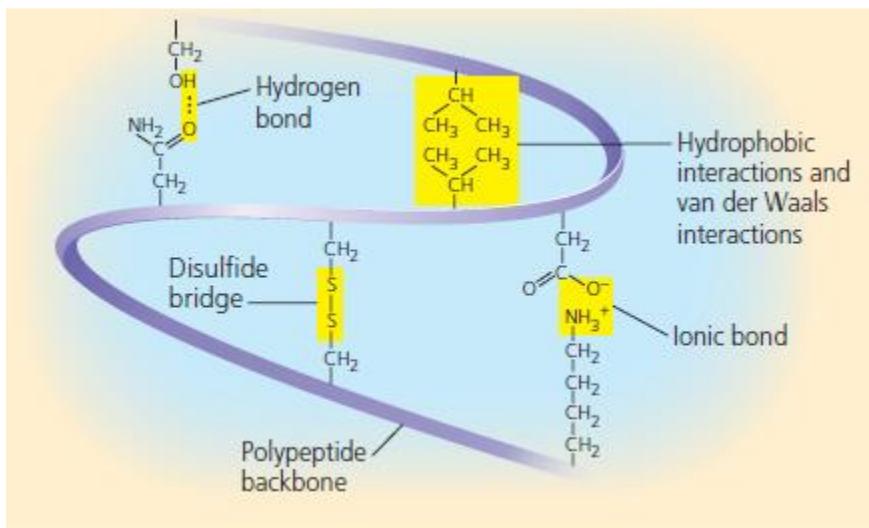


Secondary structure, are the result of hydrogen bonds between the repeating constituents of the polypeptide backbone (not the amino acid side chains). Within the backbone, the oxygen atoms have a partial negative charge, and the hydrogen atoms attached to the nitrogen have a partial positive charge; therefore, hydrogen bonds can form between these atoms. Individually, these hydrogen bonds are weak, but because they are repeated many times over a relatively long region of the polypeptide chain, they can support a particular shape for that part of the protein. One such secondary structure is the **α - helix**, a delicate coil held together by hydrogen bonding between every fourth amino acid. The other main type of secondary structure is the **β - pleated sheet**. In this structure two or more strands of the polypeptide chain lying side by side (called β strands) are connected by hydrogen bonds between parts of the two parallel polypeptide backbones.



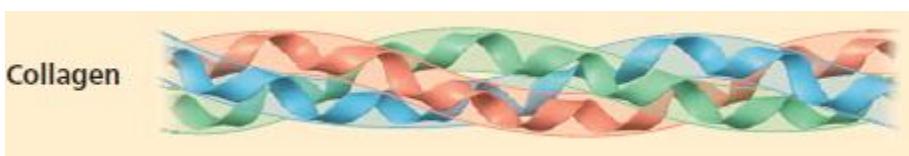
Turns Composed of three or four residues, turns are located on the surface of a protein, forming sharp bends that redirect the polypeptide backbone back toward the interior. These short, U-shaped secondary structures are stabilized by a hydrogen bond between their end residues. Glycine and proline are commonly present in turns. The lack of a large side chain in glycine and the presence of a built-in bend in proline allow the polypeptide backbone to fold into a tight U shape. Turns allow large proteins to fold into highly compact structures. A polypeptide backbone also may contain longer bends, or *loops*. In contrast with turns, which exhibit just a few well-defined structures, loops can be formed in many different ways.

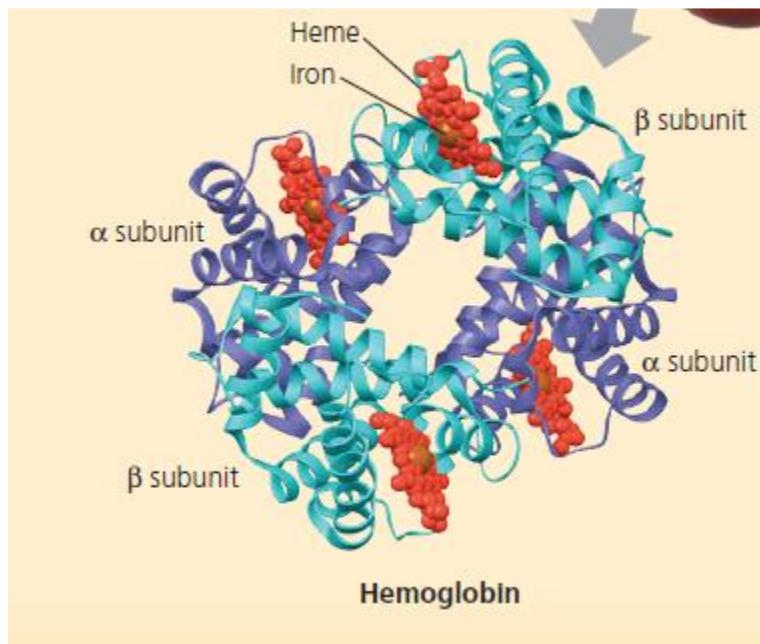
Tertiary structure is the overall shape of a polypeptide resulting from interactions between the side chains (R groups) of various amino acids. One type of interaction that contributes to tertiary structure is—somewhat misleadingly—called a **hydrophobic interaction**. As a polypeptide folds into its functional shape, amino acids with hydrophobic (nonpolar) side chains usually end up in clusters at the core of the protein, out of contact with water. Thus, a “hydrophobic interaction” is actually caused by the exclusion of nonpolar substances by water molecules. Once nonpolar amino acid side chains are close together, van der Waals interactions help hold them together. Meanwhile, hydrogen bonds between polar side chains and ionic bonds between positively and negatively charged side chains also help stabilize tertiary structure. These are all weak interactions in the aqueous cellular environment, but their cumulative effect helps give the protein a unique shape. Covalent bonds called **disulfide bridges** may further reinforce the shape of a protein. Disulfide bridges form where two cysteine monomers, which have sulfhydryl groups ($-SH$) on their side chains, are brought close together by the folding of the protein. The sulfur of one cysteine bonds to the sulfur of the second, and the disulfide bridge ($-S-S-$) rivets parts of the protein together. All of these different kinds of interactions can contribute to the tertiary structure of a protein.



Tertiary Structure of a Protein.

Quaternary structure is the overall protein structure that results from the aggregation of these polypeptide subunits. Example: collagen, which is a fibrous protein that has three identical helical polypeptides intertwined into a larger triple helix, giving the long fibers great strength. This suits collagen fibers to their function as the girders of connective tissue in skin, bone, tendons, ligaments, and other body parts. Collagen accounts for 40% of the protein in a human body. Hemoglobin, the oxygen-binding protein of red blood cells is another example of a globular protein with quaternary structure. It consists of four polypeptide subunits, two of one kind (α) and two of another kind (β). Both α and β subunits consist primarily of α -helical secondary structure. Each subunit has a nonpolypeptide component, called heme, with an iron atom that binds oxygen.





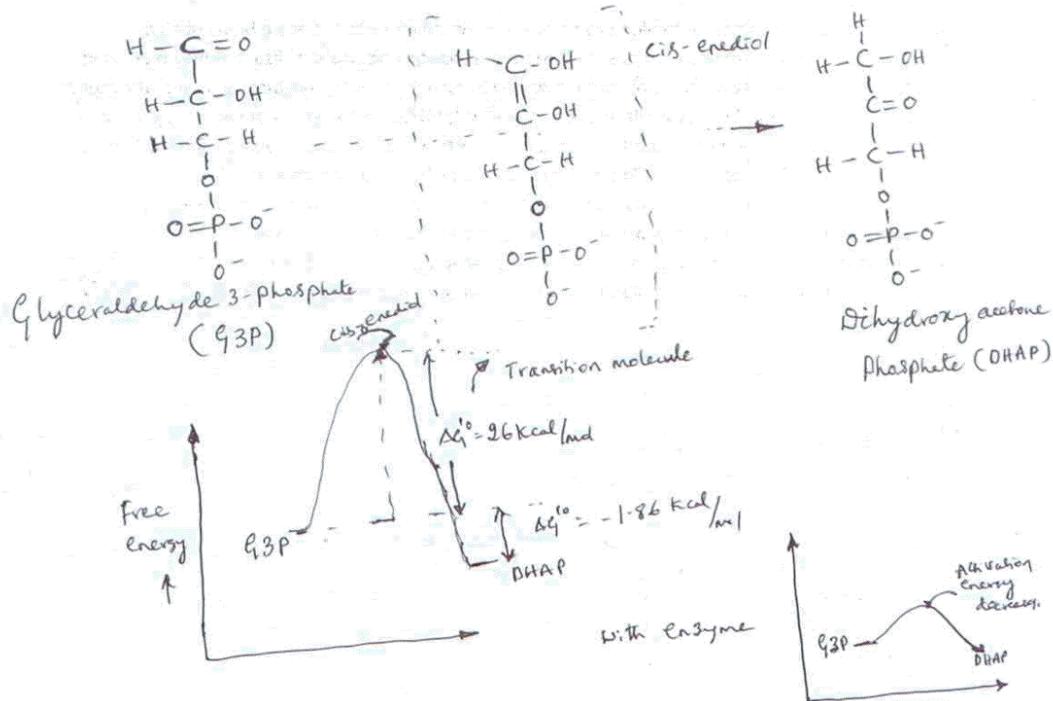
Enzymes and how enzymes functions

Enzymes

Enzymes are the most remarkable and highly specialized proteins. Enzymes have extraordinary catalytic power, often far greater than that of synthetic or inorganic catalysts. They have a high degree of specificity for their substrates, they accelerate chemical reactions tremendously, and they function in aqueous solutions under very mild conditions of temperature and pH. Like all catalysts, enzymes do not affect the extent of a reaction, which is determined by the change in free energy ΔG° between reactants and products. For reactions that are energetically favorable, enzymes increase the reaction rate by lowering the activation energy.

Edward Buchner (Nobel prize in chemistry-1907) initially showed that fermentation can be done with yeast juice rather than yeast itself. The word enzyme came from the latin word 'Zyma' means yeast and 'enzyme' means something in yeast.

To understand more about enzymes, we will take an example of an enzyme which is present in almost all organisms, called triose phosphate isomerase. The enzyme converts one triose sugar with phosphate group (glyceraldehyde 3-phosphate (G3P)) to another form of triose sugar with phosphate group (Dihydroxy acetone phosphate (DHAP)) which is energetically favourable reaction ($\Delta G^\circ = -1.86$ Kcal/mol). While converting from G3P to DHAP, it has to go through a molecule called cis-enediol (transition molecule). It is energetically unfavourable molecule ($\Delta G^\circ = + 26$ Kcal/mol). It is so unstable, once cis-enediol forms, the phosphate group comes out of molecule especially in presence of water. Though formation of DHAP from G3P is energetically favourable, since it has to go through a transition molecule, which is energetically unfavourable, the formation of product in normal condition is impossible. Enzyme helps to stabilize this transition state and helps in formation of product.

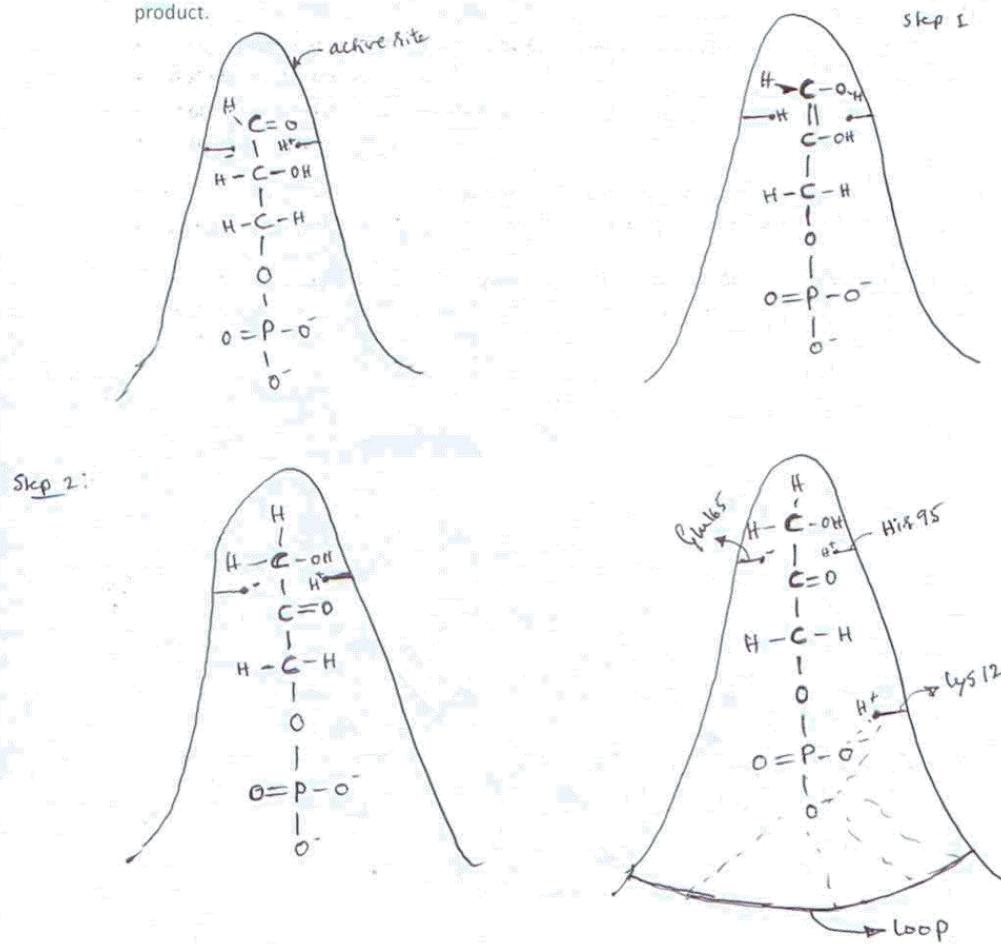


Important points about enzymes:

- It does not change energetic favourability
- It does change activation barrier (reduces activation barrier)
- It stabilizes the transition state
- It prevents side reaction to occur
- It speeds up the reaction

Triose phosphate isomerase is a perfect enzyme! The rate limiting step is the rate at which molecules diffuse into it. Once the substrate is in its active site, the product will be formed.

How does triose phosphate isomerase works? It is 250 amino acid polypeptide chain and it is present in the form of dimer. Few amino acids in the active site helps to convert substrate to product.



His95, i.e. the 95th aminoacid is histidine. It is positively charged aminoacid. It is exactly present at the position to give H⁺ ion to first carbon of G3P. Glutamic acid is present at 165th position. The negatively charged ion is appropriately present to take H⁺ from the second carbon of G3P. The resulting molecule is cis-enediol. It is highly unstable molecule.

In the second step, H⁺ ion from the second carbon is taken by His95 to become again positive charge and H⁺ ion from the glutamic acid to the first carbon atom of the molecule. The resulting molecule is DHAP.

To stabilize the phosphate group (which is negatively charged), lysine is present at the 12th aminoacid. The positively charged lysine electrostatically interacts with the phosphate group and helps in stabilizing it.

When this is in the cis-enediol state, the phosphate will normally come off pretty spontaneously & quickly. It will lose the phosphate and turn into what's called methylglyoxal. Actually it turns out that it's even going to float away in water if it's not sufficiently well bound.

A loop part of the protein closes down on the active site. And four polar amino acids make hydrogen bonds with the molecule & it prevents this intermediate from floating away. Because it's closed down, it actually prevents water getting into active site and so it protects it.

Let us analyze active site with different amino acids and whether the enzyme can able to convert G3P to DHAP?

If we make triosephosphate isomerase and substituted, instead of at position 165 a glutamic acid, an aspartic acid, would it work? Turns out it'll work 1,000 times worse. It will just work, it's about 1,000-fold worse in speed. Check the difference in structure of Glu & Asp, little one extra carbon bond positioning in the right place makes a difference of 1,000-fold to the speed of the enzyme.

Suppose we change lysine at 12th position to a non-charged amino acid. Let's change it to Leucine, a hydrophobic amino acid. Does the enzyme work? No. Turns out the enzyme doesn't even work. We can't get it to catalyze at all! But what if we change it to another positive amino acid? How about arginine? If we substitute an arginine instead of lysine, well, it still works, but it's about 200 times worse, because the positive charge is not in the ideal place.

Suppose if we remove the loop that closes the active site, and keep everything else in its position, does it work? It is about 100,000 times worse in speed. That loop really matters.

Bioenergetics

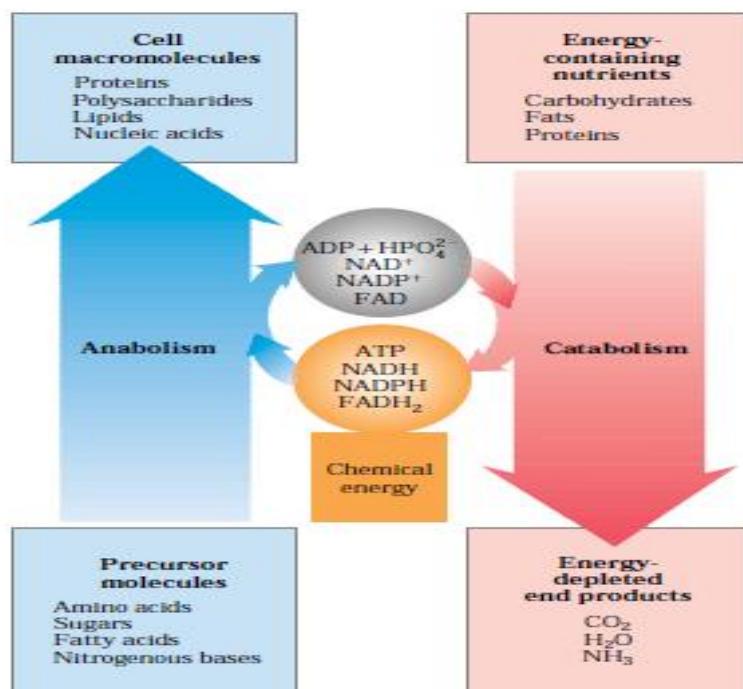
Biochemical Pathways

Living organisms can be divided into two large groups according to the chemical form in which they obtain carbon from the environment. **Autotrophs** can use carbon dioxide from the atmosphere as their sole source of carbon, from which they construct all their carbon containing biomolecules. **Heterotrophs** cannot use atmospheric carbon dioxide and must obtain carbon from their environment in the form of relatively complex organic molecules such as glucose. Multicellular animals and most microorganisms are heterotrophic.

Metabolism, the sum of all the chemical transformations taking place in a cell or organism, occurs through a series of enzyme-catalyzed reactions that constitute **metabolic pathways**. Each of the consecutive steps in a metabolic pathway brings about a specific, small chemical change, usually the removal, transfer, or addition of a particular atom or functional group. The precursor is converted into a product through a series of metabolic intermediates called metabolites. The term intermediary metabolism is often applied to the combined activities of all the metabolic pathways that interconvert precursors, metabolites, and products of low molecular weight.

Catabolism is the degradative phase of metabolism in which organic nutrient molecules (carbohydrates, fats, and proteins) are converted into smaller, simpler end products. Catabolic pathways release energy, some of which is conserved in the formation of ATP and reduced electron carriers (NADH, NADPH, and FADH₂); the rest is lost as heat.

In **anabolism**, also called biosynthesis, small, simple precursors are built up into larger and more complex molecules, including lipids, polysaccharides, proteins, and nucleic acids. Anabolic reactions require an input of energy, generally in the form of the phosphoryl group transfer potential of ATP and the reducing power of NADH, NADPH, and FADH₂.



For metabolism, catabolism and anabolism to occur in the cell, the substrate has to go through **biochemical pathways**. The goal of the lecture is to understand the energetics of typical pathway reactions and **the basic principles of energetics of any biochemical reaction**.

In all the heterotrophs glycolysis (breaking up of sugar) takes place. In this biochemical pathway glucose is converted in to pyruvate. To understand the energetics of typical pathways of reaction, glycolysis is studied as an example. Before that, we will understand the basics of chemical reaction energetics.

Energetics of the reaction: Because biological systems are generally held at constant temperature and pressure, it is possible to predict the direction of a chemical reaction from the change in the free energy G , named after J. W. Gibbs, who showed that “all systems change in such a way that free energy [G] is minimized.” In the case of a chemical reaction, reactants \leftrightarrow products, the change in free energy ΔG is given by

$$\Delta G_{\text{products}} - \Delta G_{\text{reactants}}$$

The relation of ΔG to the direction of any chemical reaction can be summarized in three statements:

- If ΔG is negative, the forward reaction (from left to right as written) will tend to occur spontaneously.
- If ΔG is positive, the reverse reaction (from right to left as written) will tend to occur.
- If ΔG is zero, both forward and reverse reactions occur at equal rates; the reaction is at equilibrium.

The standard free-energy change of a reaction ΔG° is the value of the change in free energy under the conditions of 298 K (25°C), 1 atm pressure, pH 7.0 (as in pure water), and initial concentrations of 1 M for all reactants and products except protons, which are kept at 10^{-7} M (pH 7.0). Most biological reactions differ from standard conditions, particularly in the concentrations of reactants, which are normally less than 1 M.

The actual change in free energy ΔG during a reaction is influenced by temperature, pressure, and the initial concentrations of reactants and products and usually differs from ΔG° . Most biological reactions—like others that take place in aqueous solutions—also are affected by the pH of the solution. We can estimate free-energy changes for different temperatures and initial concentrations, using the equation

$$\Delta G = \Delta G^\circ + RT \ln [Q] = \Delta G^\circ + RT \ln [\text{product}/\text{reactant}]$$

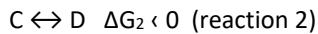
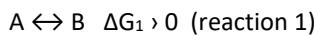
Where R is the gas constant of 1.987 cal/(degree·mol), T is the temperature (in degrees Kelvin), and Q is the initial concentration ratio of products to reactants.

Regardless of the ΔG° for a particular biochemical reaction, it will proceed spontaneously within cells only if ΔG is negative, given the usual intracellular concentrations of reactants and products. For example, the conversion of glyceraldehyde 3-phosphate (G3P) to dihydroxyacetone phosphate (DHAP), $\text{G3P} \leftrightarrow \text{DHAP}$ has a ΔG° of -1.840 kcal/mol. If the initial concentrations of G3P and DHAP are equal, then $\Delta G = G^\circ$, because $RT \ln 1 = 0$; in this situation, the reversible reaction $\text{G3P} \leftrightarrow \text{DHAP}$ will proceed in the direction of DHAP formation until equilibrium is reached. However, if the initial [DHAP] is 0.1 M and the initial [G3P] is 0.001 M, with other conditions being standard, then Q in Equation given equals $0.1/0.001 = 100$, giving a ΔG of + 0.887 kcal/mol. Under these conditions, the reaction will proceed in the direction of formation of G3P.

The ΔG for a reaction is independent of the reaction rate. Indeed, under usual physiological conditions, few, if any, of the biochemical reactions needed to sustain life would occur without some mechanism for increasing reaction rates. As we described in the previous lecture, the rates of reactions in biological systems are usually determined by the activity of **enzymes**, the protein catalysts that accelerate the formation of products from reactants without altering the value of ΔG .

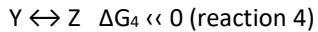
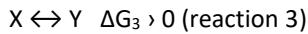
Cellular processes use two tricks in energetics to make energetically unfavorable reactions into favorable ones.

Trick 1) An Unfavorable Chemical Reaction Can Proceed If It Is directly coupled with an Energetically Favorable Reaction: Many processes in cells are energetically unfavorable ($\Delta G > 0$) and will not proceed spontaneously. Cells can carry out an energy-requiring reaction ($\Delta G_1 > 0$) by coupling it to an energy-releasing reaction ($\Delta G_2 < 0$) if the sum of the two reactions has a net negative ΔG . Suppose



In most of the directly coupled reactions, cellular processes use $ATP \leftrightarrow ADP + P_i \quad \Delta G = -7.3 \text{ kcal/mol}$ as reaction 2.

Trick 2) An Unfavorable Chemical Reaction Can Proceed If It Is indirectly coupled with an Energetically Favorable Reaction: Suppose



If the concentration of Y is zero and concentration of X increases, due to very high concentration difference (analyze with the equation $\Delta G = \Delta G^\circ + RT \ln [Q] = \Delta G^\circ + RT \ln [\text{product}/\text{reactant}]$) some amount of Y is going to form. Since the reaction $Y \leftrightarrow Z \quad \Delta G_2 <> 0$, Y will be converted to Z spontaneously.

Now let us analyze the bioenergetics of glycolysis pathway. In glycolysis 6 carbon sugar glucose is converted into 3 carbon pyruvate. In anaerobic condition, in some organisms pyruvate will be converted to ethanol and carbon dioxide. In some organisms, it is converted into lactate. First one billion years after the existence of life, this was the only pathway used by cellular processes to produce energy. The step-by-step conversion of glucose to pyruvate is given in the figure (not necessary to remember). Analyze the pathway- wherever ΔG° becomes positive, direct or indirect coupling of reaction takes place in the pathway and cellular processes keep ΔG always negative or zero.



↓
Ethanol + CO₂
(yeast)

↓
Lactate
(human)

Glycolysis

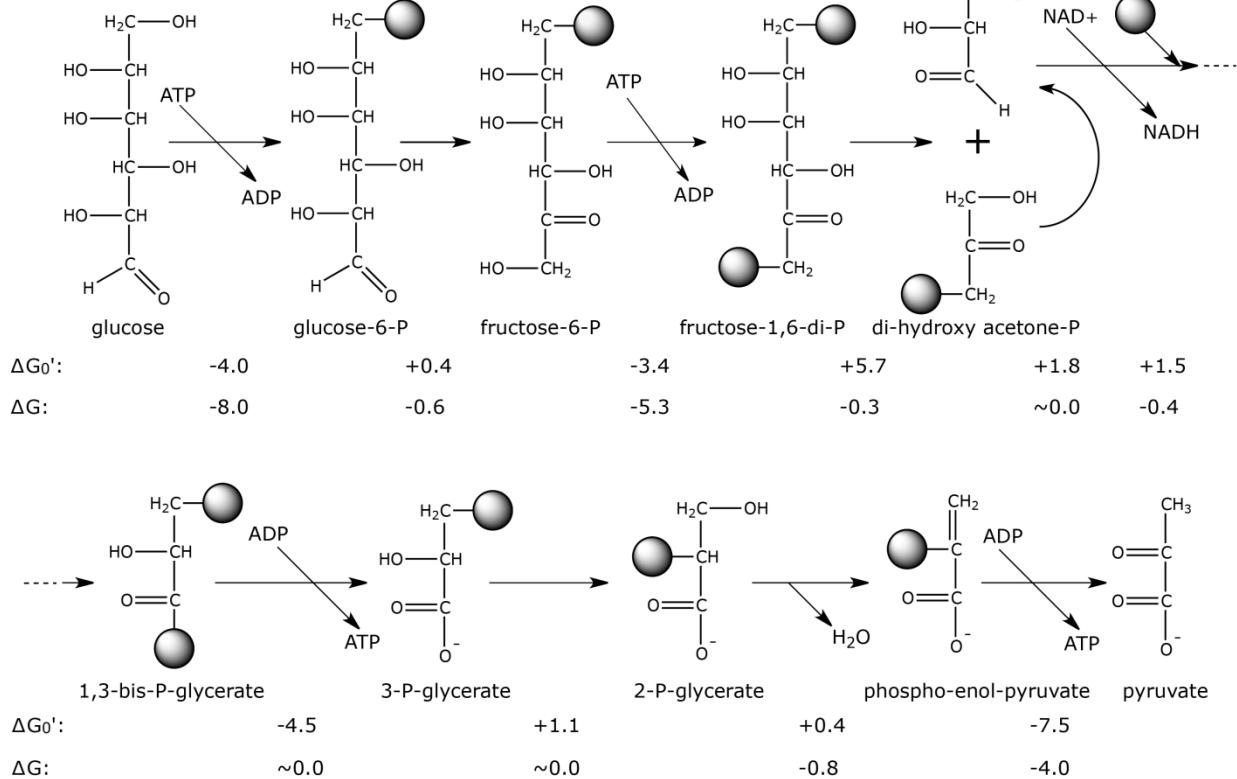
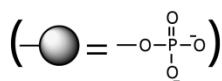
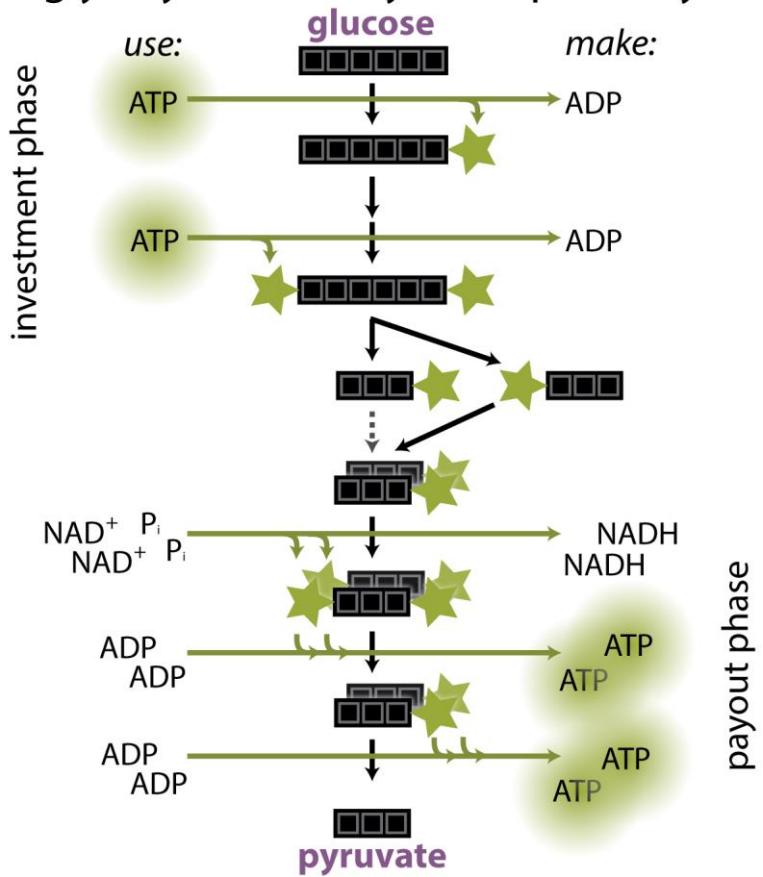
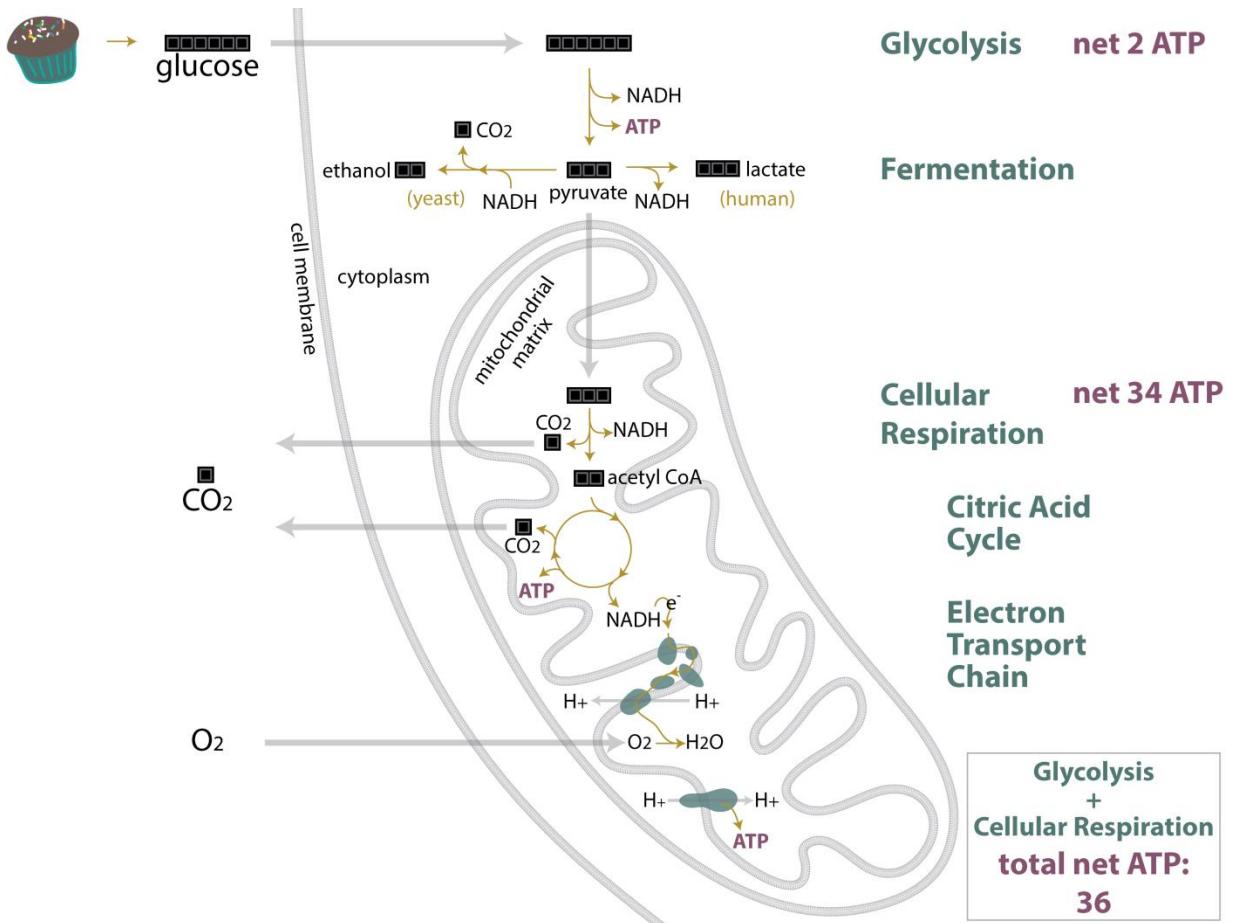


figure by Brian White

Calculate the amount of ATP produced by the glycolysis pathway. The net ATP produced is 4-2= 2. Two ATP's are invested to produce 4 ATPs.

glycolysis: an enzymatic pathway





After glycolysis pyruvate moves into mitochondria in aerobic organisms and produces 34 ATPs by utilizing oxygen as shown in the figure. The biochemical pathways of glycolysis, respiration, anabolic reactions are interrelated as shown in the picture below.

pH

When some substances dissolve in water, they release *hydrogen ions* (H^+), which are actually single, positively charged protons. Hydrogen ions can attach to other molecules and change their properties. For example, the protons in “acid rain” can damage plants, and you probably have experienced the excess of hydrogen ions that we call “acid indigestion.” Here we will examine the properties of acids (defined as substances that release H^+) and bases (defined as substances which accept H^+). We will distinguish between strong and weak acids and bases and provide a quantitative means for stating the concentration of H^+ in solutions: the **pH** scale.

ACIDS RELEASE H⁺ When hydrochloric acid (HCl) is added to water, it dissolves, releasing the ions H⁺ and Cl⁻:

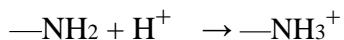


Because its H⁺ concentration has increased, such a solution is *acidic*. Acids are substances that release H⁺ ions in solution. HCl is an acid, as is H₂SO₄ (sulfuric acid). One molecule of sulfuric acid will ionize to yield two H⁺ and one SO₄²⁻. Biological compounds that contain —COOH (the carboxyl group) are also acids because —COOH → —COO⁻ + H⁺. Acids that fully ionize in solution, such as HCl and H₂SO₄ are called *strong acids*. However, not all acids ionize fully in water. For example, if acetic acid (CH₃COOH) is added to water, some will dissociate into two ions (CH₃COO⁻ and H⁺), but some of the original acetic acid remains as well. Because the reaction is *not complete*, acetic acid is a *weak acid*.

BASES ACCEPT H⁺ Bases are substances that *accept* H⁺ in solution. Just as with acids, there are strong and weak bases. If NaOH (sodium hydroxide) is added to water, it dissolves and ionizes, releasing OH⁻ and Na⁺ ions:



Because the concentration of OH⁻ increases and OH⁻ absorbs H⁺ to form water (OH⁻ + H⁺ → H₂O), such a solution is *basic*. Because this reaction is complete, NaOH is a *strong base*. Weak bases include the bicarbonate ion (HCO₃⁻), which can accept a H⁺ ion and become carbonic acid (H₂CO₃), and ammonia (NH₃), which can accept a H⁺ and become an ammonium ion (NH₄⁺). Biological compounds that contain —NH₂ (the amino group) are also bases because



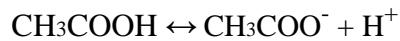
ACID–BASE REACTIONS MAY BE REVERSIBLE, when acetic acid is dissolved in water, two reactions happen. First, the acetic acid forms its ions:



Then, once the ions are formed, some of them re-form acetic acid:



This pair of reactions is reversible. A **reversible reaction** can proceed in either direction—left to right or right to left—depending on the relative starting concentrations of the reactants and products. The formula for a reversible reaction can be as



In terms of acids and bases, there are two types of reactions, depending on the extent of the reversibility:

- The ionization of strong acids and bases in water is virtually irreversible.
- The ionization of weak acids and bases in water is somewhat reversible.

WATER IS A WEAK ACID AND A WEAK BASE The water molecule has a slight but significant tendency to ionize into a hydroxide ion (OH^-) and a hydrogen ion (H^+). Actually, two water molecules participate in this reaction. One of the two molecules “captures” a hydrogen ion from the other, forming a hydroxide ion and a hydronium ion:



The hydronium ion is, in effect, a hydrogen ion bound to a water molecule. For simplicity, biochemists tend to use a modified representation of the ionization of water:



The ionization of water is important to all living creatures. This fact may seem surprising, since only about one water molecule in 500 million is ionized at any given time. But this is less surprising if we focus on the abundance of water in living systems, and the reactive nature of the H^+ ions produced by ionization.

- Pure water has a H^+ concentration of $10^{-7} M$.
- A $1 M \text{ HCl}$ solution has a H^+ concentration of $1 M$.
- A $1 M \text{ NaOH}$ solution has a H^+ concentration of $10^{-14} M$.

This is a very wide range of numbers to work with—think about the decimals! It is easier to work with the *logarithm* of the H^+ concentration, because logarithms compress this range as shown in figure.

Since the H^+ concentration of pure water is $10^{-7} M$, its pH is $-\log(10^{-7}) = -(-7)$, or 7. A smaller negative logarithm means a larger number. In practical terms, a lower pH means a higher H^+ concentration, or greater acidity. In $1 M \text{ HCl}$, the H^+ concentration is $1 M$, so the pH is the negative logarithm of 1 ($-\log 10^0$), or 0. The pH of $1 M \text{ NaOH}$ is the negative logarithm of 10^{-14} , or 14. A solution with a pH of less than 7 is acidic—it contains more H^+ ions than OH^- ions. A solution with a pH of 7 is *neutral* (without net charge), and a solution with a pH value greater than 7 is basic.

Why is this discussion of pH so important in biology? Many biologically important molecules contain charged groups (e.g., $-\text{COO}^-$) that can interact with the polar regions of water to form their structures. But these groups can combine with H^+ or other ions in their environment to form uncharged groups (e.g., $-\text{COOH}$). These uncharged groups have much less tendency to

interact with water. If such a group is part of a larger molecule, it might now induce the molecule to fold in such a way that it stays away from water because it is hydrophobic. In a more acidic environment, a negatively charged group such as —COO^- is more likely to combine with H^+ . So the pH of a biological tissue is a key to the three-dimensional structures of many of its constituent molecules. Organisms do all they can to minimize changes in the pH of their watery medium. An important way to do this is with buffers.

BUFFERS The maintenance of internal constancy—*homeostasis*—is a hallmark of all living things and extends to pH. As we mentioned earlier, if biological molecules lose or gain H^+ ions their properties can change, thus upsetting homeostasis. Internal constancy is achieved with buffers: solutions that maintain a relatively constant pH even when substantial amounts of acid or base are added. How does this work?

A **buffer** is a solution of a weak acid and its corresponding base—for example, carbonic acid (H_2CO_3) and bicarbonate ions (HCO_3^-). If an acid is added to a solution containing this buffer, not all the H^+ ions from the acid stay in solution. Instead, many of them combine with the bicarbonate ions to produce more carbonic acid:



This reaction uses up some of the H^+ ions in the solution and decreases the acidifying effect of the added acid. If a base is added, the reaction essentially reverses. Some of the carbonic acid ionizes to produce bicarbonate ions and more H^+ , which counteracts some of the added base. In this way, the buffer minimizes the effect that an added acid or base has on pH. This buffering system is present in the blood, where it is important for preventing significant changes in pH that could disrupt the ability of the blood to carry vital oxygen to tissues. A given amount of acid or base causes a smaller pH change in a buffered solution than in a non-buffered one. Buffers illustrate an important chemical principle of reversible reactions, called the *law of mass action*. Addition of a reactant on one side of a reversible system drives the reaction in the direction that uses up that compound. In the case of buffers, addition of an acid drives the reaction in one direction; addition of a base drives the reaction in the other direction. We use a buffer to relieve the common problem of indigestion. The lining of the stomach constantly secretes hydrochloric acid, making the stomach contents acidic. Excessive stomach acid

inhibits digestion and causes discomfort. We can relieve this discomfort by ingesting a salt such as NaHCO_3 (“bicarbonate of soda”), which acts as a buffer.



CO2

Mendel's rules for inheritance, chromosomal theory of inheritance, relationship of Mendelian inheritance to meiosis and pedigree analysis of genetic disorder diseases.

SI No.	Class	Page No.
8	Mendelian concept of inheritance and Logic of Mendel	2
9	Monohybrid cross and segregation	7
10	Terminologies, Back cross and test cross, Dihybrid cross, Law of independent assortment.	12
11	Chromosomes and cell division, chromosomal theory of inheritance	19
12	Morgan's experiment, X-linked inheritance	23
13	Pedigree analysis	30

THE LOGIC OF MENDEL

All of us have some characters similar to our parents or grand parents. Many of us like to look like or behave like our parents. Our grandparents, most of them are farmers, selected the best seeds for their next season. All of them knew that the best seeds have the best characters. Characters are passed to next generation through seeds. Seeds comes from flowers. Bees visit flowers for honey. Flowers contain honey fine powders.

Here we are trying to understand logically, how a normal man like Mendel who had lowest grades in Biology could do such meticulously planned experiments? Above all he had failed for a teaching certificate in natural sciences! On the other hand Mendel studied physics, mathematics and chemistry along with important aspects of biology. This mathematical back ground enabled Mendel to plan his experiments, draw out a theory and experimentally evaluate it.

We are logically going to understand what brought Mendel to understand the concept. Remember to consider the context that are looking at this from Mendel's perspective. He knew nothing about meiosis, never considered chromosomes or whether they had anything to do with what he was investigating. The word "gene" was invented after his death. All he did was breed peas, and do some high quality thinking. All this done between 1856 and 1863!

It is also interesting to see what is Mendel's previous knowledge? Sometimes early in his life he worked as a gardener and done the beekeeping. He also studied practical and theoretical philosophy and physics at the University of Olomouc, Czech republic. He happened to see the research of hereditary traits of plants and animals, in the department of natural history and agriculture. He also worked as a substitute high school teacher. He failed in getting a certificate for teacher and send to University of Vienna for further study on a sponsorship. He returned and again worked as physics teacher, but failed on the oral part of examination. Finally he was taken the superior priest of the monastery.

Mendel's work can be divided into the following steps: (A) Preparation for experiments (B) Choice of experimental material (C) Planning and execution of experiment (D) Interpretation of experimental results and (E) Further testing of his observations.

A. Preparation for Experiments

How do to an experiment? The laboratory should be accessible! That is Mendel approached the head of the department of Natural history and agriculture where he was working as a priest for permission to use the 2 hectare experimental garden intended to study differences in plants. His colleges conducted studies on the heredity of sheep. Why he wanted the entire full 2 hectares rather a few cents or pots? If we need to have a reliable results, the sample size should be very high! Therefore how much land you need to cultivate around 20,000 pea plants? Ask a farmer!

B. Choice of the experimental material

First he decided to work on plants? Do you know why? Plants attain reproductive age very soon compared to many animals, the number of plantlets from a plant will be very high, Many plants

can be grown together and above all plants reproduce through seeds which we will get from flowers!. Now the question is which plant? Shall we take a mango tree? Here comes the advantage of critical observation. The feasibility. It is Pea plants, because very easy to cultivate, flowers are big (think why?), number of plantlets produced from one flower will be more and above all its generation time is less. But the most important thing? What is the aim of the experiment? To study variation in plants? So we need to study variation. Examples of variation? Flower color, seed color, seed size and shape, flower position etc. So the experimental plant should have easily recognizable variations. From the Mendel's observations it is the Pea.

C. Planning and execution of the experiment

Selection of variations: He selected the following characters: 1. Seed shape [Spherical Vs Wrinkled]; 2. Seed color [Yellow Vs Green], 3. Flower color [Purple Vs white], 4. Pod shape [inflated vs constricted], 5. Pod color [green pod vs yellow], 6. Flower position [Axial vs Terminal] and 7. Plant size [Tall Vs Dwarf] (**The first logic of Mendel**: Easy recognizable separate characters for observation -that had well-defined, contrasting alternative traits)

Why he selected seven characters? How did he do? He selected seven variations simultaneously in each experiment or separate experiments for separate variations? What you feel? He conducted different experiments for different characters first of all to see that whether he is getting the same kind of results in each case, secondly it is easy to follow the inheritance of one character at a time rather all together. That was the **second logic** of Mendel.

This third logic of Mendel from the knowledge he may got from his practical knowledge on agriculture and theoretical knowledge he got from mathematics. Does Mendel randomly selected seeds? We should know which characters seeds are inherited? We are studying variation, should we select seeds which show consistent inheritance or variable inheritance? The choice should be of consistent inheritance. He knew that it is possible to raise plants with consistent inheritance by crossing sibling plants. That is Mendel used well defined seeds as the starting material in all his experiments involving thousands of plants. So if we want to test the fuel efficiency of 10 different bikes, we should use the same experimental conditions in all the cases against one variability. This is the spectacular logic he applied and the basis of Mendel's success.

The **fourth logic is experimental**. How to do controlled breeding? We want to cross a tall plant with a dwarf plant. So we need to ensure that only the pollen grain of the tall plant is falling on the stigma of the dwarf plant and vice versa. How to do this? Simply bag the flower bud, cut the stamens of the experimental flower, take the pollen grain from the other flower using a brush, "paint" on the stigma of the initial flower, bag it again to avoid unwanted entry of other pollen grain. (See the animation). Get the seeds from the flower and observe for the variation intended to study!

Here is the **fifth logic**. He just crossed only once. All other experiments involved self-pollination. i.e., he allowed mixing of two different characters only once. Why not twice or thrice? As simple as this, first let us start from simple things! So do the cross only once, avoid complexity and evaluate the results.

Now comes the **sixth and final logic**. Looking behind a character how many generations we should follow after crossing? Again comes the simplicity and reproducibility. Just followed the first generation after crossing.

So in short he started with the same experimental plan is as follows:

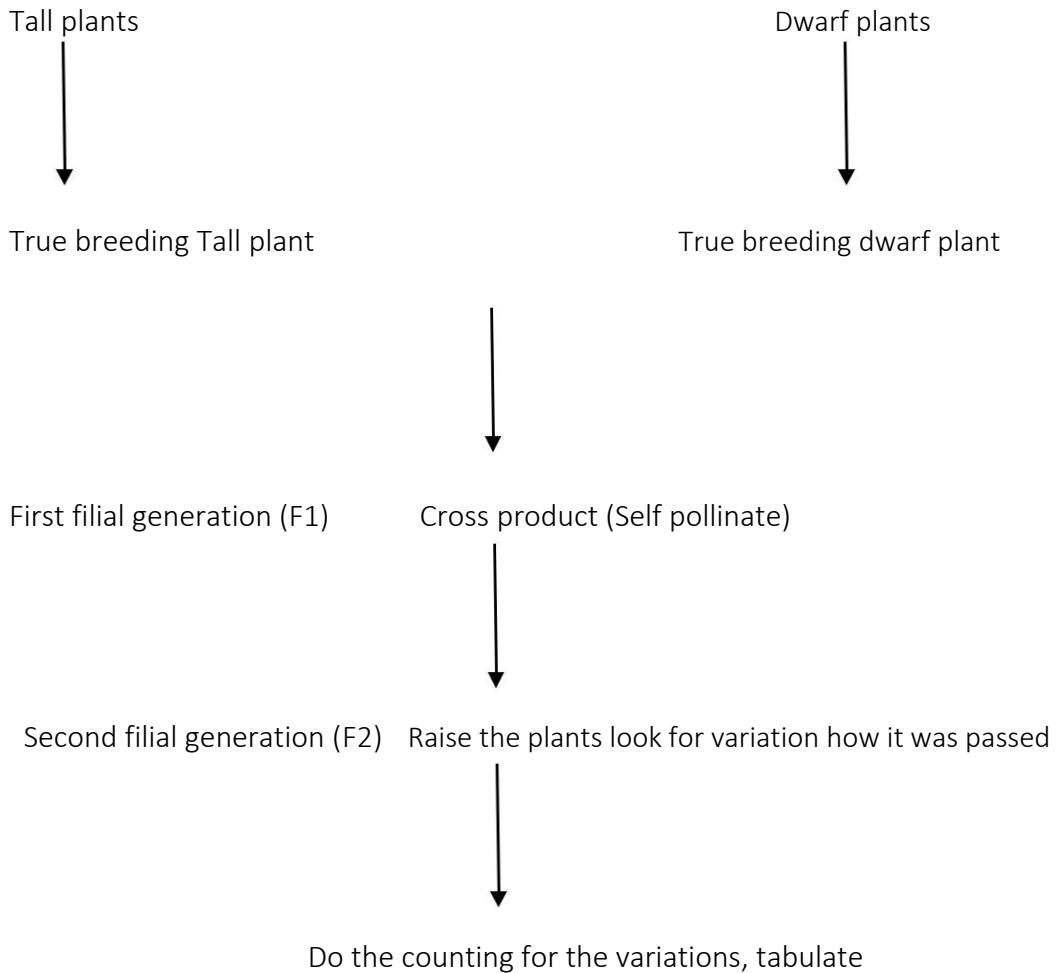


Figure 1

The seven variations observed in Pea plant that Mendel followed: Are they easily recognizable?

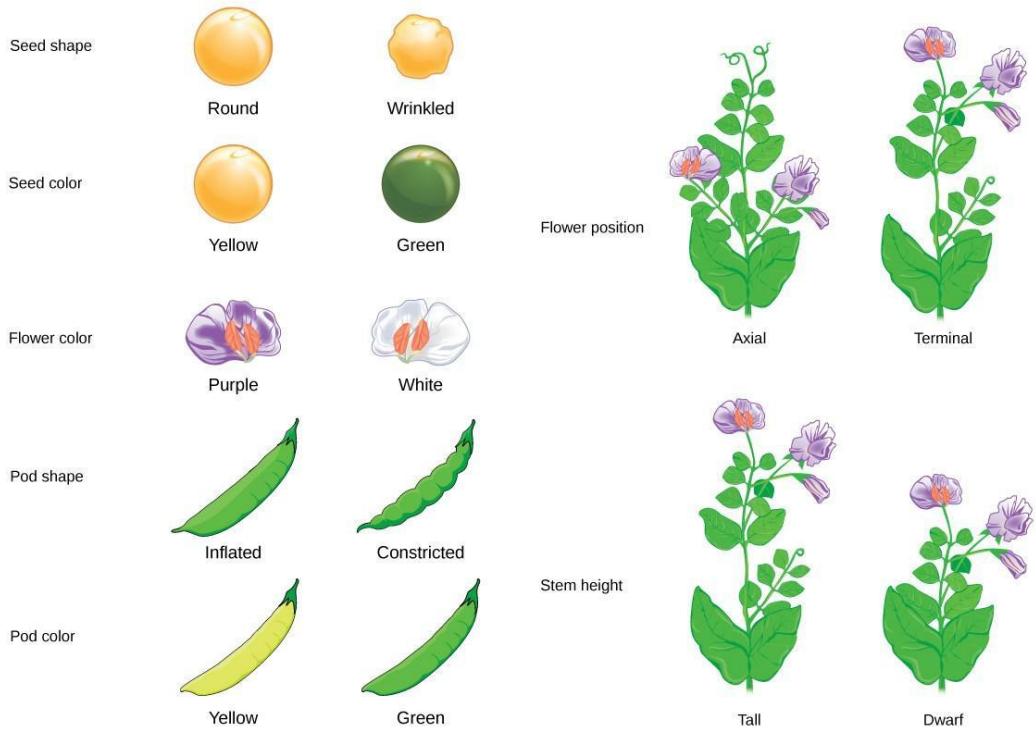


Figure 2 Flower structure of Pea: Observe the reproductive parts of the flower.

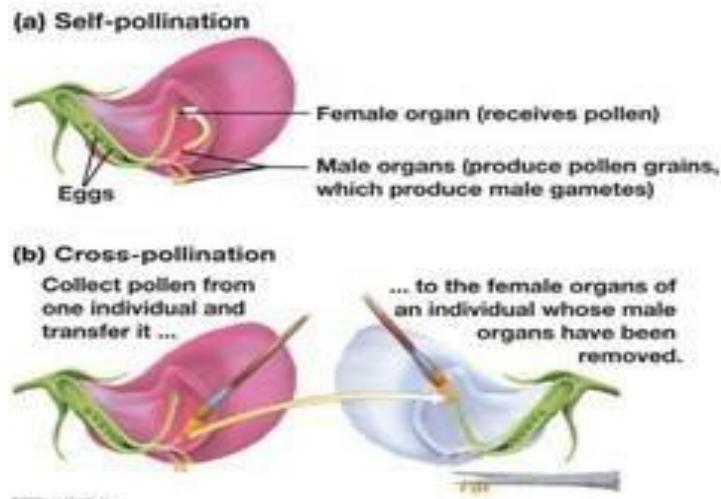


Photo of Mendel's Experimental Model i.e. Green Peas:

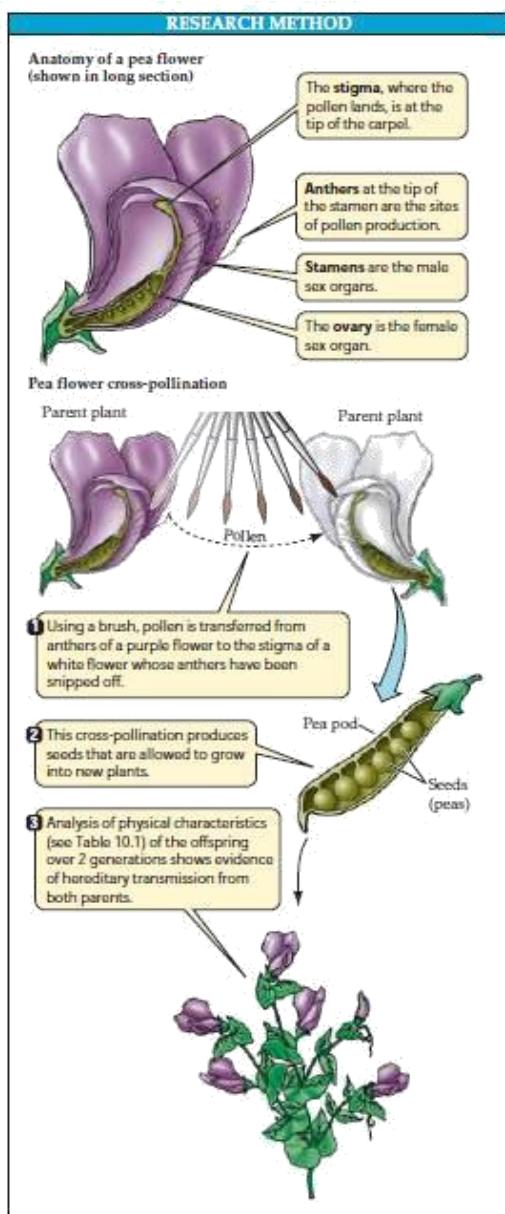


Mendel's controlled experiments and elucidation of the results

MONOHYBRID CROSS AND SEGREGATION

THE RESEARCH METHOD OF MENDEL

All of us know that pollen grains of one flower falls on the stigma of another flower or same flower. From the stigma the male nucleus of the pollen grain reaches the ovary and fertilizes an egg to produce a zygote. In plants fertilized zygote produces the seed and the ovary produces the fruit. The problem faced by Mendel is controlled pollination. Suppose he wanted to cross a tall plant with a dwarf plant. In this case the pollen grains of dwarf plant only should fall on the stigma of the dwarf plant or vice versa. How to achieve this? (Solve this problem with the help of the following figure and the animation provided)



10.1 A Controlled Cross between Two Plants Plants were widely used in early genetic studies because it is easy to control which individuals mate with which. Mendel used the garden pea (*Pisum sativum*) in many of his experiments.

EXPERIMENTS RESULTS OF MENDEL

We have seen what Mendel's logics were and how he executed the experiments. After his experiments the results were carefully tabulated by him. Analyze the following table showing the tabulation of Mendel

10.1 Mendel's Results from Monohybrid Crosses

PARENTAL GENERATION PHENOTYPES DOMINANT	RECESSIVE	F1 Phenotypes	F2 GENERATION PHENOTYPES			RATIO
			DOMINANT	RECESSIVE	TOTAL	
Spherical seeds × Wrinkled seeds		Spherical	5,474	1,850	7,324	2.96:1
Yellow seeds × Green seeds		Yellow	6,022	2,001	8,023	3.01:1
Purple flowers × White flowers		Purple	705	224	929	3.15:1
Inflated pods × Constricted pods		Inflated	882	299	1,181	2.95:1
Green pods × Yellow pods		Green	428	152	580	2.82:1
Axial flowers × Terminal flowers		Axial	651	207	858	3.14:1
Tall stems × Dwarf stems (1 m)	(0.3 m)	Tall	787	277	1,064	2.84:1

What you find here?

- (A) Only one character appeared in the F1 generation
- (B) Both characters were appeared in the F2 generation, but not in equal percentage
- (C) A character which disappeared in the F1, reappeared in F2
- (D) The results are consistent in all the seven characters
- (E) There is no blending of characters

The analysis of the results clearly indicate the following

The character which appeared in the F1 is having the higher percentage in F2 i.e. almost three times to that of the disappeared character in F1. What you will conclude from these results? The character which is appeared in the F1 is also **dominating** in F2 also. This is the **Dominant character**. The other or alternative character which is disappearing in the F1 generation is also less in number compared to dominant character. This is **Recessive** character. (Law of Dominance).

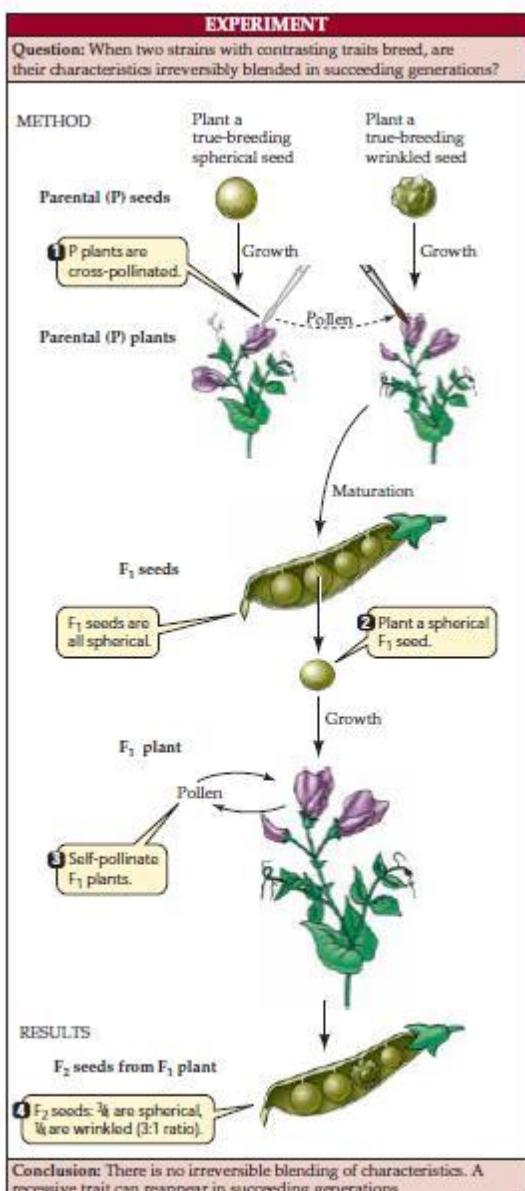
Many of us, including Mendel, expected a blended phenotype in F1. But it didn't happen. Why? Can't the characters blend? The experimental results in all the seven characters studied by Mendel didn't show any blending (i.e. when we cross a Tall and Dwarf plant we can expect a plant intermediate between tall and dwarf; but it didn't happen). This further means that some "units" which functions as discrete particles are responsible for characters. Since the character which disappeared in the F1 reappeared

in F₂ it is logic to conclude that these units occur as pairs. It means that two discrete units are responsible for the “Tall” character. Otherwise any given individual can be homozygous (Dominant/Dominant or Recessive/Recessive) or heterozygous (Dominant/Recessive). We can express the alternate forms of characters (dominant or recessive) as alleles. Also we know at present that character means a gene. So alleles are alternate forms (in fact variables) of a gene.

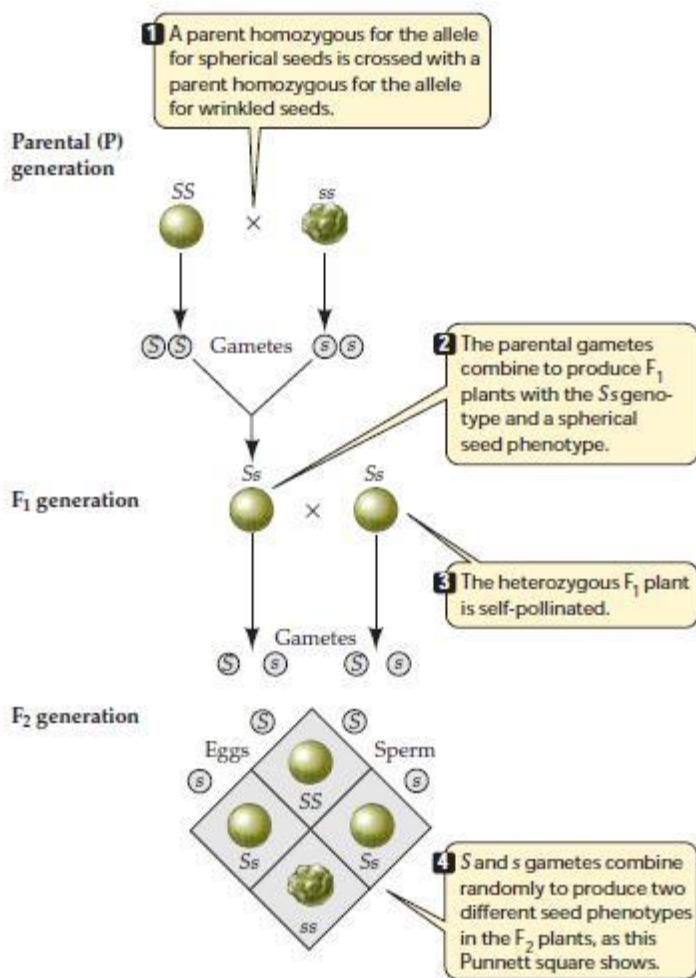
The factors can be represented as any letter forms (similar to polynomials). For example a tall plant can have TT or Tt. A dwarf plant can be represented as tt. This binomial expression is the genotype. In conclusion the “units” (or factors) are discrete, they never blend and are responsible for passing character from one generation to next ie inheritance.

Why two units of inheritance? Not three or four? We have father and mother. i.e. we have characters both from our father and mother. A character is represented by two units. One unit inherited from father and one from mother. So the gamete contains one unit. This is the core idea of Mendel’s inheritance. At present the “units” proposed by Mendel is known as gene.

The following figures illustrates a typical Mendelian cross.



Now study the following cross which can be represented in the form of a Punnett square (A simple grid representing all possible gametes and combinations. Given the credit to Reginald Crundall Punnett, a British geneticist)



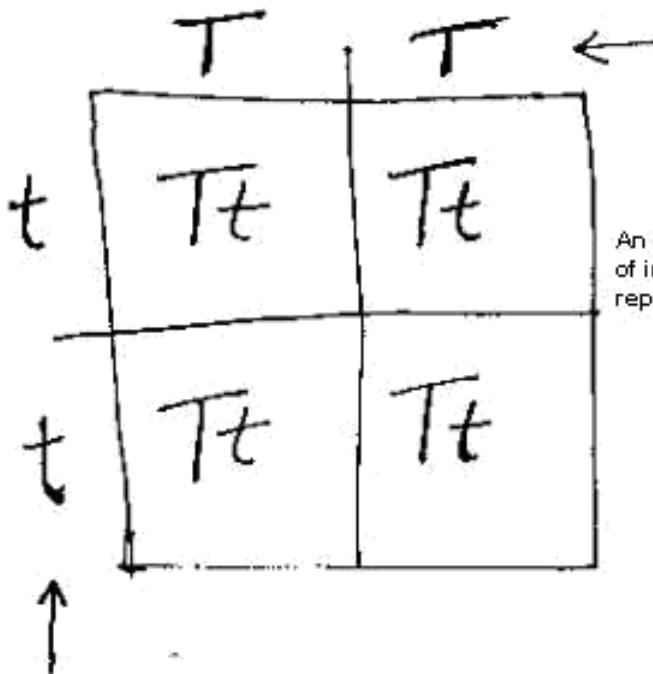
10.4 Mendel's Explanation of Experiment 1 Mendel concluded that inheritance depends on factors from each parent, and that these factors are discrete units that do not blend in the offspring.

In conclusion the units of inheritance are never blended, but segregated independently during reproduction (Law of segregation). This is the second core idea of Mendel's theory.

Another illustration is here for you regarding the cross between a Tall plant and a Dwarf plant. This illustration represents Mendel's core ideas.

tall parent (TT)

short parent (tt)



Units of inheritance from
Tall parent.
Each unit is segregated into
a gamete.

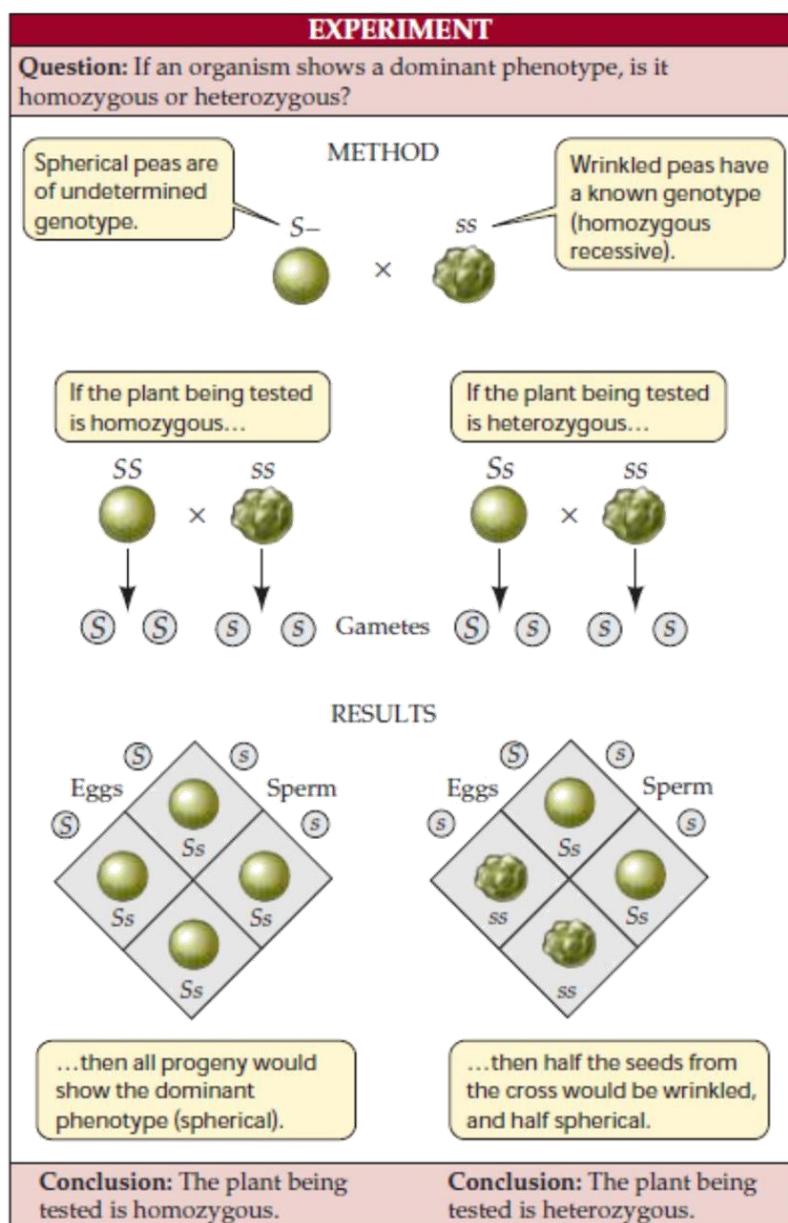
An individual contains two units
of inheritance for a character, can be
represented as two letter codes.

Units of inheritance from dwarf parent. Each unit is segregated into a gamete

Terminologies, Back cross and Test Cross, Dihybrid cross

Verification of Mendel's hypothesis:

He did verification experiments for his hypothesis, as illustrated below.



10.6 Homozygous or Heterozygous? An individual with a dominant phenotype may be homozygous or heterozygous. Its genotype can be determined by crossing it with a homozygous recessive plant and observing the phenotypes of the progeny produced. This procedure is known as a test cross.

Similarly the following figure illustrates a dihybrid test cross

Mendel has modelled one experiment, he executed and finally he tested it. This type of cross is known as a test cross (Crossing the unknown genotype with the recessive parent). It can predict the genotype to be tested based on the phenotypic ratio of the cross output.

Probability laws govern Mendelian inheritance: The Study of dihybrid cross

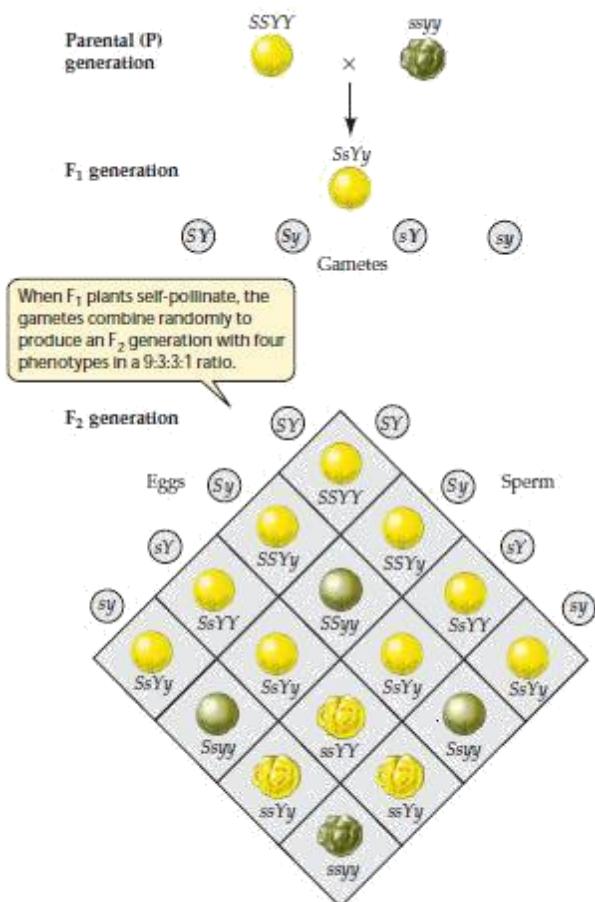
Initially Mendel did all his experiments by analyzing only one character at a time – monohybrid cross. Based on his results, he has tested his hypothesis of dominance and segregation. After this he wanted to study the inheritance of two characters at a time – the dihybrid cross

The experiment is planned in such a way to analyze the following:

Whether the alleles maintain the association they had in the parental generation: For this he crossed pure breeding spherical seed and yellow seed color pea plant with a wrinkled seed green seed color pea plant. If the alleles maintain the association, he is expecting only the parental types in the F₂ generations (Why not F₁ generation?)

If the alleles maintain the association, the F₁ gametes will be SY and sy. As a result the probability of Spherical and yellow seed peas: wrinkled and green seed peas will be 3:1 (i.e. only two phenotypes). If they segregate independently he was expecting four different phenotypes.

The experiment and the results are illustrated as below



10.7 Independent Assortment The 16 possible combinations of gametes in this dihybrid cross result in 9 different genotypes. Because *S* and *Y* are dominant over *s* and *y*, respectively, the 9 genotypes result in 4 phenotypes in a ratio of 9:3:3:1. These results show that the two genes segregate independently.

He didn't get a 3:1 ratio in F₂ generation. New types were obtained in F₂. It means that the alleles didn't maintain the same association as seen in the parental types, rather they assorted independently (Law of Independent Assortment).

Mendel and his Mathematics predictions.

You have a 1 rupee coin and 5 rupee coin. probability of getting a tail in both cases? independent?)

Probability of getting a 1 rupee tail = $\frac{1}{2}$. Probability of getting a 5 rupee tail = $\frac{1}{2}$

You are going to toss it together. What is the (Are the two events linked to each other or

Hence the probability of getting both tail = $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ i.e. 25%.

In a homozygote (SS), the probability of producing a S gamete is 1

In a heterozygote (Ss), the probability of producing a S gamete is $\frac{1}{2}$ and s gamete is also $\frac{1}{2}$

Now consider the F₂ generation. The probable gametes here are S and s.

Hence the probability of getting SS is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ i.e. 25% are homozygous dominant

The probability of getting ss is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ i.e. 15% are homozygous recessive

Adding probabilities: What is the probability of getting Ss and sS?

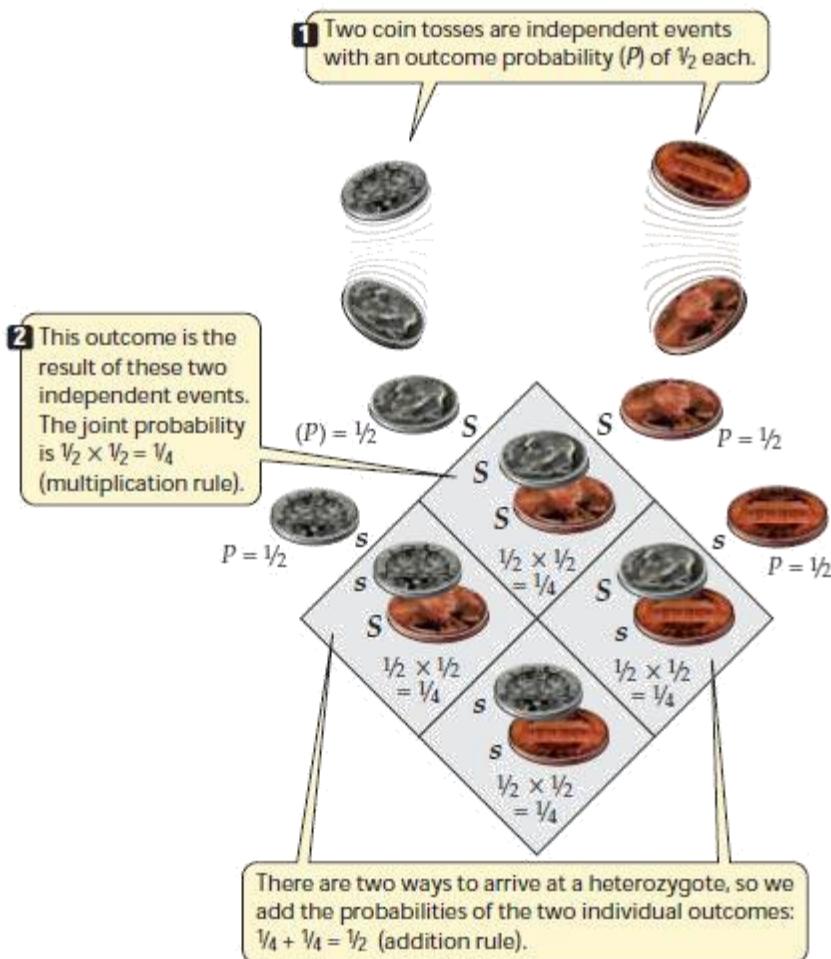
Probability of Ss (S from sperm and s from egg) = $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

Probability of sS (s from sperm and S from egg) = $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

Both Ss and sS are heterozygotes and will have the same phenotype. Hence added probability is $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ ie 50% will be heterozygotes.

Now can we calculate the probabilities in dihybrid cross?

In F₂ generation, the probabilities are illustrated below



10.9 Using Probability Calculations in Genetics The probability of any given combination of alleles from a sperm and an egg appearing in the offspring of a cross can be obtained by multiplying the probabilities of each event. Since a heterozygote can be formed in two ways, these two probabilities are added together.

Now what is the probability of getting an SS homozygote ? i.e. $\frac{1}{4}$

The probability of getting heterozygote (i.e. Ss or sS) is $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

The added probability (i.e. spherical seed) = $\frac{3}{4}$

Now calculate the probability of yellow seed using the above reasoning? It will be $\frac{3}{4}$

Hence what is the added probability of getting a spherical seed and yellow seed = $\frac{3}{4} \times \frac{3}{4} = \frac{9}{16}$
 (Since both events are independent i.e. independent assortment)

What will be the probability of getting a yellow and wrinkled seed?

Probability of yellow seed = $\frac{3}{4}$

Probability of wrinkled seed = $\frac{1}{4}$

Hence the added probability = $\frac{3}{4} \times \frac{1}{4} = \frac{3}{16}$

Using the same logic it is easy to calculate the probability of wrinkled yellow seed is $\frac{3}{16}$ and wrinkled green seed is $\frac{1}{16}$.

Mendel did all these statistical problems. Because of his mathematical knowledge, he could easily predict, the ratio obtained in F₂ generation of monohybrid and dihybrid crosses are simply a statistical event and the factors are independent of each other.

You should understand both; i.e. doing a genetic problem by using probability and by using a Punnett square.

One of Mendel's major contributions to the science of genetics was his use of the rules of statistics and probability to analyze his masses of data from hundreds of crosses producing thousands of plants. His mathematical analyses led to clear patterns in the data, and then to his hypotheses. Ever since Mendel, geneticists have used simple mathematics in the same ways that Mendel did.

It is also possible to test the dihybrid genotypes as illustrated below. Similar to monohybrid test cross, it is possible to predict the genotype of a phenotype by crossing with a true recessive parent. The prediction is based on the characteristic phenotypic ratio we will get in this cross.

	[Double heterozygous F ₁ Dihybrid] <i>Yellow Round</i>	[Double homozygous recessive parent] <i>Green Wrinkled</i>		
Test cross Genotypes	$YyRr$	$yyrr$		
Types of Gametes	$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow$ <i>YR Yr yR yr</i>	\downarrow <i>yr</i>		
	<i>YR Yr yR yr</i>			
Test cross Progeny	$YyRr$ <i>Yellow Round</i>	$Yyrr$ <i>Yellow Wrinkled</i>	$yyRr$ <i>Green Round</i>	$yyrr$ <i>Green Wrinkled</i>
	1 25%	1 25%	1 25%	1 25%

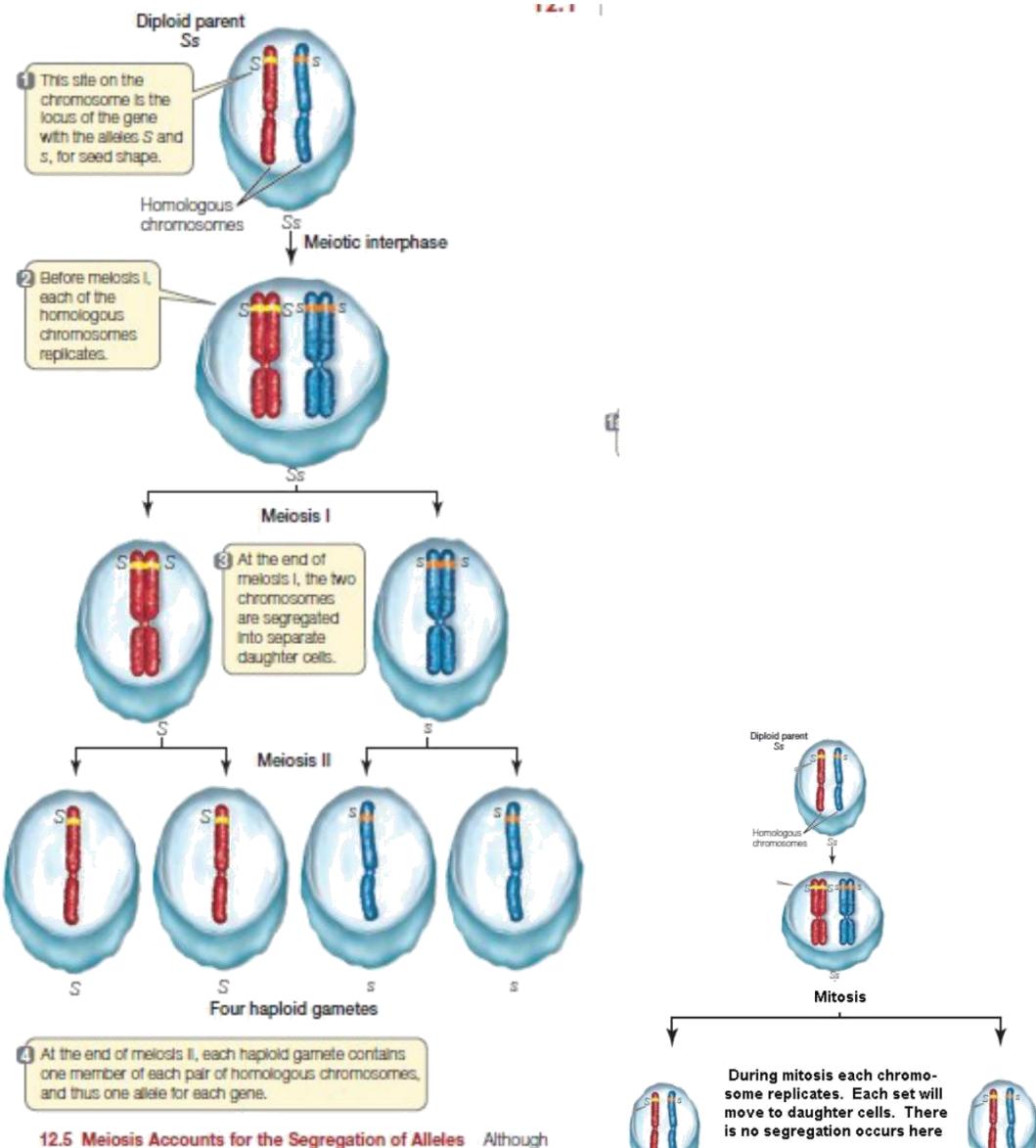
Chromosomes and cell Division, Chromosomal Theory

Mendel experiments: Did he predict chromosomes?

(Meiosis accounts for segregation)

The segregation of Mendelian factors is because of meiosis

Mendel proposed mechanisms of heredity. Mendel had no knowledge of chromosomes or meiosis. But he speculated that cells contained some type of **factor** that carried traits from one generation to the next. The scientific importance of Mendel's work remained unrecognized for several years. Most probably Mendel believed that for each character there is a factor. Currently we know that this factor is a gene (or an allele) that is located on a chromosome. They show characteristic segregation and independent assortment are due to meiosis (Illustrated below).

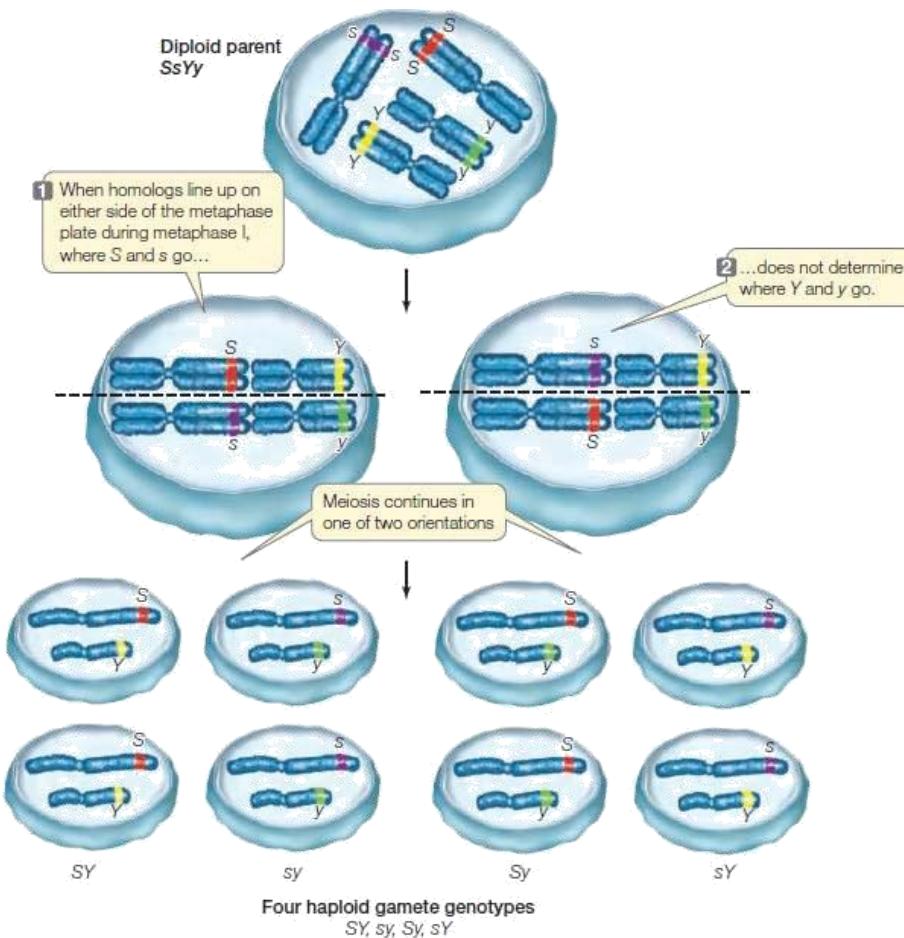


12.5 Meiosis Accounts for the Segregation of Alleles Although Mendel had no knowledge of chromosomes or meiosis, we now know that a pair of alleles resides on homologous chromosomes, and that those alleles segregate during meiosis.

So if we have a cell with a genotype Ss it should produce two types gametes, i.e. one type with S and another type with s . Mendel said the two alleles will segregate. Our current knowledge is that meiosis accounts for segregation. This is exactly matches what Mendel speculated from his results. He also proved that each factor segregates when traits passed from one generation to another generation.

What happens in mitosis? A cell with genotype Ss just produces two daughter cells with the same genotype as illustrated above and there is no segregation.

Now you will see how alleles assort independently during meiosis when we consider a dihybrid cross



Mendelian principles doesn't applies to all cases of inheritance.

Whether all the inheritance follows Mendelian pattern? Human have several traits like hair pattern, skin color, tasting ability, shape of ear and so on. How many chromosomes we have? 23 pairs. So if Mendel's rules we apply, we should have only 23 chromosomes. Hence it becomes clear that a chromosome can contain more than one factor. Now we have to think that who is the luckiest Man? It is Mendel. He selected seven characters. Each character was regulated by a gene and they were located in seven different chromosomes. Now we know that Pea plant has seven chromosomes. Suppose if the traits selected by Mendel resides on the same chromosome, he will not get a 9:3:3:1 ratio as expected.

In short if we get a ratio of 9:3:3:1, we can assume that the genes we selected are located on different chromosome. If we are not getting this ratio, then genes may be on same chromosome. The genes on the same chromosome means that they are linked.

The seven chromosomes of *Pisum sativum*. Luckily the seven genes for the selected traits by Mendel was located on seven different chromosomes leading to his success in modelling his hypothesis and successful testing. If it was not, Mendel might have failed in his efforts. Hence many people believes that Mendel was the luckiest person.

Chromosome Theory of Inheritance

Sutton and Boveri in 1902 correlated Mendel's conclusions about genes (or inherited traits) to the behavior of chromosomes during mitosis and meiosis.

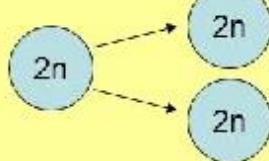
Sutton is credited with first proposing the chromosome theory of inheritance:

- Chromosomes are in pairs
- Homologous chromosomes separate during meiosis so that alleles are segregated
- Meiotic products have one of each homologous chromosome but not both
- Fertilization restores the pairs of chromosomes
- Genes are located on chromosomes

Two types of cell Division

MITOSIS

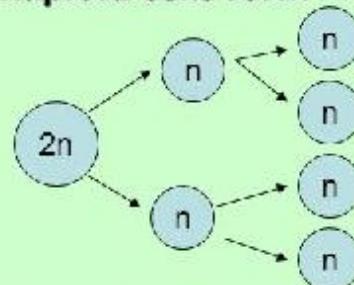
- For growth, maintenance & repair
- $2n$ cell \rightarrow $2n$ cells
(46 chromosomes \rightarrow 46 chromosomes)
D. Ploidy
- 2 diploid cells form



- Occurs in somatic cells in the human body!

MEIOSIS

- For gamete formation – sperm & egg
- $2n$ cell \rightarrow n cells
(46 chromosomes \rightarrow 23 chromosomes)
haploid.
- 4 haploid cells form



- Occurs only in gonads (ovaries and testes)
- Cause of most existing genetic variation

Morgan's experiments

Morgan and his *Drosophila*: Mendel's hypothesis is rejected in Morgan's experiments

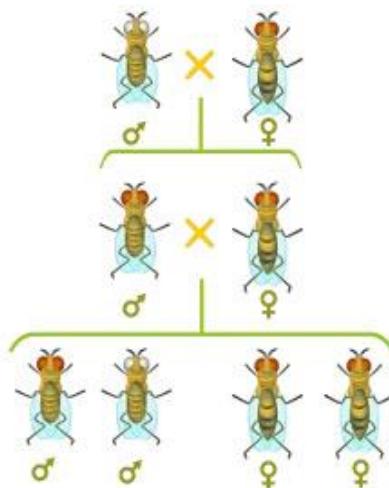
[Chromosomal theory, Connecting Mendel to Morgan, Linkage and crossing over]

Thomas Hunt Morgan and his students of Columbia University did pioneering works to explain heredity from the beginning of 1909. He explored the Mendel theories in *Drosophila melanogaster*, the fruit fly as the experimental organism. He selected fruit fly because of its small size, easy to grow and breed and its short generation time.

Thomas H. Morgan correctly perceived that the success of genetic investigators depended critically upon the choice of the organism to be investigated. Much of the work in the early years had centred upon agricultural plants and animals: we knew how to grow successive generations of them, and the information had direct practical bearing. Morgan abandoned agricultural utility in favor of experimental utility-plants just took too long between generations, and they took up too much space. Morgan wanted an organism with which one could carry out many crosses, with many progeny, easily and quickly. With this in mind, he began to investigate the genetics of Drosophila. No genetic varieties were available in Drosophila, so Morgan set out to find them. He obtained his first mutant in 1910, from normal red eyes to white. At last he could set out to examine Mendelian segregation.

MORGAN'S FRUIT FLY CROSSES

First, Morgan crossed the white-eyed male he had found to a normal female, and he looked to see which trait was dominant in the F1 generation: all the progeny had red eyes. Now, would the white-eye trait reappear, segregating in the F2 progeny as Mendel had predicted? In the F2, there were 3470 red-eyed flies and 782 white-eyed flies, roughly a 3:1 ratio. Allowing for some deficiency in recessives, this was not unlike what Mendel's theory predicted. But in this first experiment, there was a result that was not predicted by Mendel's theory: all the white-eyed flies were male!



At this point, Morgan had never seen a white-eyed fly that was female. Morgan preferred a straightforward test: if any of the F2 females carried the white-eye trait but did not show it, then

it should be revealed by a test cross to the recessive parent. It was. Crossing red-eyed F₂ females back to the original white-eyed male, he obtained 129 red-eyed females, 132 red-eyed males, and 88 white-eyed females, 86 white-eyed males.

Again, this was a rather poor fit to the expected 1:1:1:1 ratio due to a deficiency in recessives. The important thing, however, was that there were fully 88 white-eyed female flies. Clearly, it was not impossible to be female and white-eyed. Why, then, were there no white-eyed females in the original cross?

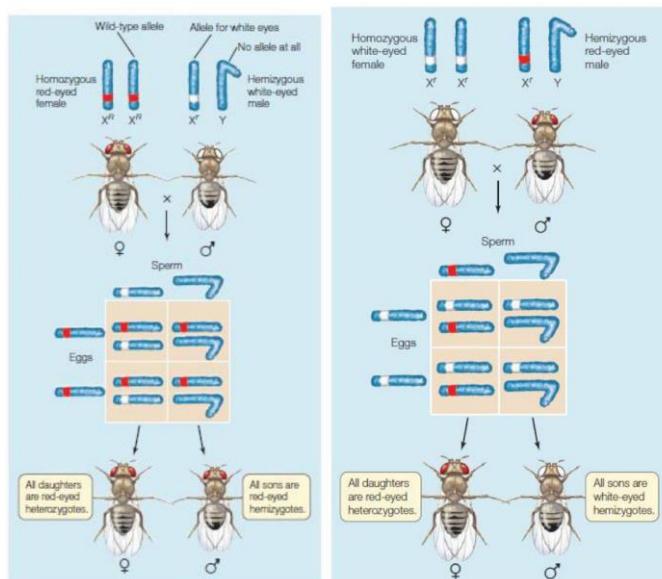
X AND Y CHROMOSOMES

We know that in mammals and many other animals sex is determined by chromosomes i.e. XX will be female and XY will be male. Thus, sperm may contain either an X or a Y chromosome, while all the female gametes will contain a copy of the X chromosome. In forming a zygote, sperm that carry an X chromosome will produce an XX zygote (female), while sperm that carry a Y chromosome will produce an XY zygote (male). This simple model explained the 1:1 proportions of males to females usually observed, as well as the correspondence of sex with chromosome cytology.

SEX LINKAGE

This theory provided a really simple explanation of Morgan's result, and he was quick to see it: what if the white-eye trait resided on the X chromosome? Morgan had only to assume that the Y chromosome did not have this gene (it was later shown to carry almost no functional genes). Knowing from his previous crosses that white-eye is a recessive trait, the results he obtained could be seen to be a natural consequence of Mendelian segregation!

Thus, a typically Mendelian trait, white-eye, is associated with an unambiguously chromosomal trait, "sex." This result provided the first firm experimental confirmation of the chromosomal theory of inheritance. This association of a visible trait that exhibited Mendelian segregation with the sex chromosome (sex linkage) was the first case in which a specific Mendelian gene could be said to reside on a specific chromosome. It firmly established the fusion of the Mendelian and chromosomal theories, marking the beginning of modern genetics.



In the above cross, the normal allele is red, the recessive allele is white. Red is dominant over white. Whenever the white male is crossed with a true breeding red female the result is both male and female flies are red eyed. Whenever a red male is crossed with a true breeding white female all male offspring's are whited eyed and female flies are red eyed. So the gene for the trait eye color in *Drosophila* resides on X linked chromosome. This inheritance is X linked recessive. We can show X linked inheritance in the pedigree chart illustrated below.

Did you note three things (1) in females, both X chromosomes should carry the recessive allele for the expression of white eye color. Hence this is an X linked recessive trait. (2) Males have only one X chromosome. Hence the trait will express even if the X chromosome contains the recessive allele (3) X linked recessive traits are more frequently occurs in males compared to females. The reason is that males have only one X chromosome. Hence the recessive allele will express.

Recombination of linked genes: crossing over

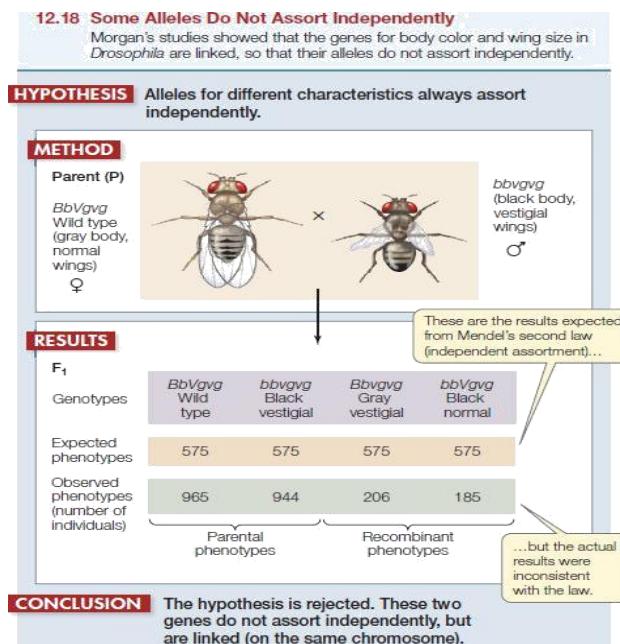
From the independent assortment of chromosomes revealed that the traits that do not match those of either parent for e.g.: the cross between a pea plant with yellow-round seeds that is heterozygous for both seed color and seed shape (a dihybrid $YyRr$) and a plant with green-wrinkled seeds homozygous for both recessive alleles (test cross) 1:1:1:1 half of the offspring are called parental types and another with new combinations of seed shape and color are called recombinant type with 50% frequency of recombination. The proof that the genes were located on chromosomes was provided by single small fly.

Thomas Hunt Morgan's drosophila dihybrid experiments for the body color and wing size. Wild type flies have gray bodies and normal-sized wings. In addition to these flies, Morgan had managed to obtain, through breeding, doubly mutant flies black body and wings much smaller than normal, called vestigial wings. Mutant alleles are recessive to the wild – type alleles. Morgan wanted to know whether the genes for body color and wing size were genetically linked, and if so, how this affected their inheritance. The alleles for body color are b_{-} (gray) and b (black), and those for wing size are vg_{-} (normal) and vg (vestigial). Morgan mated true-breeding P (parental) generation flies—wild-type flies with black, vestigial-winged flies—to produce heterozygous F1 dihybrids ($b_{-} b\ vg_{-} vg$), all of which are wild-type in appearance. He then mated wild-type F1 dihybrid females with black, vestigial-winged males. This testcross will reveal the genotype of the eggs made by the dihybrid female.

The resulting flies had a much higher proportion of the combinations of traits seen in the P generation flies (called parental phenotypes) than would be expected if the two genes assorted independently. Morgan thus concluded that body color and wing size are usually inherited together in specific (parental) combinations because the genes for these characters are near each other on the same chromosome. **The Predicted ratios if genes are located on different chromosomes were 1:1:1:1. If the genes are located on the same chromosomes and parental alleles are always inherited together then the ratio is 1:1:0:0.**

However, both of the combinations of traits not seen in the P generation (nonparental phenotypes) were also produced in Morgan's experiments, suggesting that the body-color and wing-size alleles are not always linked genetically. To understand this conclusion, we need to further explore **genetic recombination**, the production of offspring with combinations of traits that differ from those found in either parent.

Since most offspring had a parental (P generation) phenotype, Morgan concluded that the genes for body color and wing size are genetically linked on the same chromosome. However, the production of a relatively small number of offspring with non parental phenotypes indicated that some mechanism occasionally breaks the linkage between specific alleles of genes on the same chromosome.



What Morgan expected is a 1:1:1:1 ratio (Recollect Mendel's dihybrid test cross ratio)

$$\text{Total individuals} = 965 + 944 + 206 + 185 = 2300$$

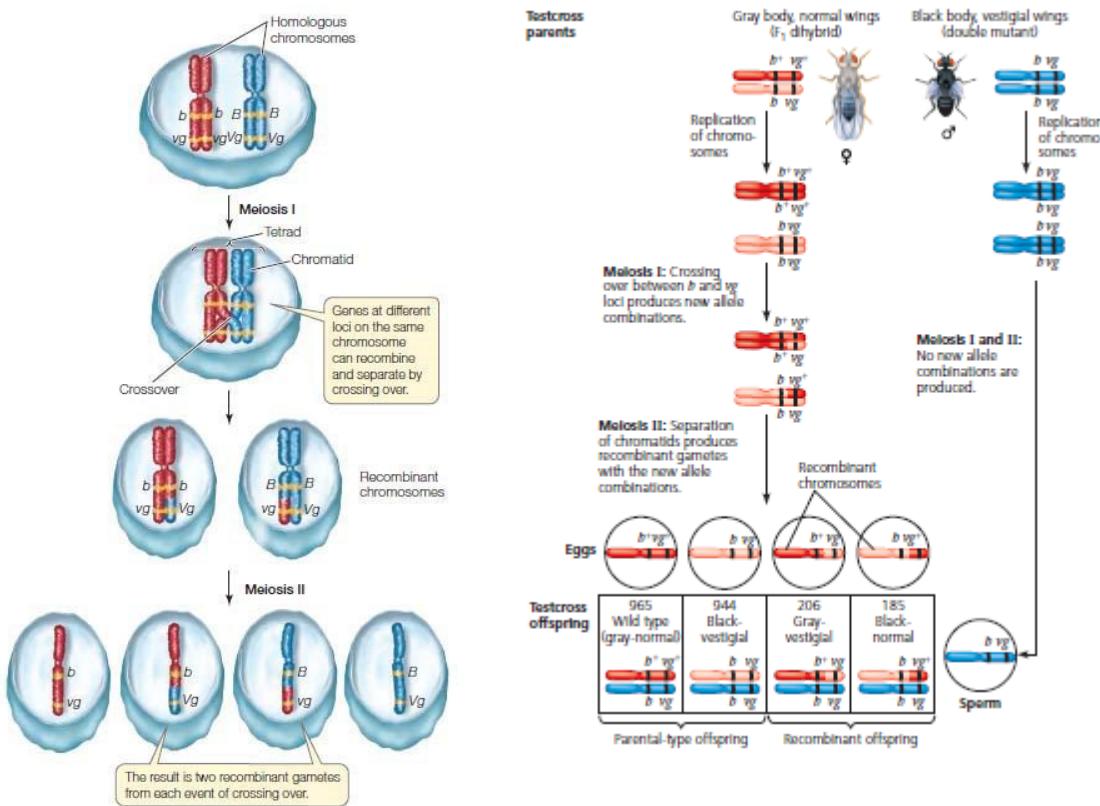
$$\text{Parental types} = 965 + 944 = 1909$$

$$\text{Non parental types or recombinant types} = 391$$

Recombination frequency = (Recombinant types / Total individuals) X 100
 $= (391/2300) \times 100 = 17\%$ (We can also write recombination frequency as 0.17 assuming that maximum recombination is 1)

Now we have to see why the new phenotypes (non-parental phenotypes) occurs?

The new phenotypes appear because of exchange of genes between homologous chromosomes that occurs during meiosis (swapping). This event is known as crossing over. Look at the following illustration to understand the process.



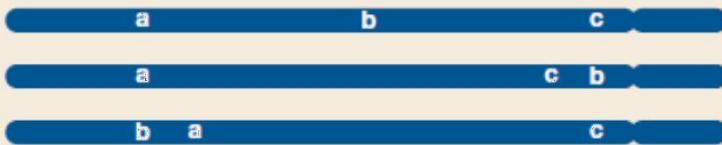
If we consider Morgan's *Drosophila* testcross result offspring from the testcross for body color and wing size most of the offspring (>50%) had parental phenotypes and about 17% of offspring were recombinants. This suggested that the two genes were on the same chromosome.

With these results, Morgan proposed that some process must occasionally break the physical connection between specific alleles of genes on the same chromosome. And this process is called **crossing over** which accounts for the recombination of linked genes. When replicating the homologous chromosomes are paired during prophase of meiosis I, an exchange of end portions of two non-sister chromatids takes places leading to crossover.

Towards the genetic map

The probability of recombination between two loci increases with distance. Morgan's found recombination frequencies of many genes through experiments and used these frequencies to construct a genetic map or mapping the genes. A genetic map tells the distance between two genes. The following illustration helps us to find how to do a genetic map. It is measured in terms of centimorgan or cM.

4.1 At the outset, we have no idea of the individual distances between the genes, and there are several possible sequences (*a-b-c*, *a-c-b*, *b-a-c*).



We make a cross *AABB* × *aabb*, and obtain an *F₁* generation with a genotype *AaBb*. We test cross these *AaBb* individuals with *aabb*. Here are the genotypes of the first 1,000 progeny:

450 *AaBb*, 450 *aabb*, 50 *Aabb*, and 50 *aaBb*.
(parental types) (recombinant types)

2 How far apart are the *a* and *b* genes?

What is the recombinant frequency? Which are the recombinant types, and which are the parental types?

Recombinant frequency (*a* to *b*) = $(50 + 50)/1,000 = 0.1$
So the map distance is

$$\text{Map distance} = 100 \times \text{recombinant frequency} = \\ 100 \times 0.1 = 10 \text{ cM}$$



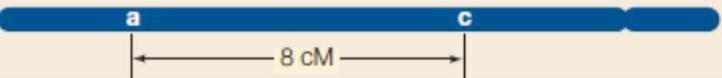
3 How far apart are the *a* and *c* genes?

Now we make a cross *AACC* × *aacc*, obtain an *F₁* generation, and test cross it, obtaining

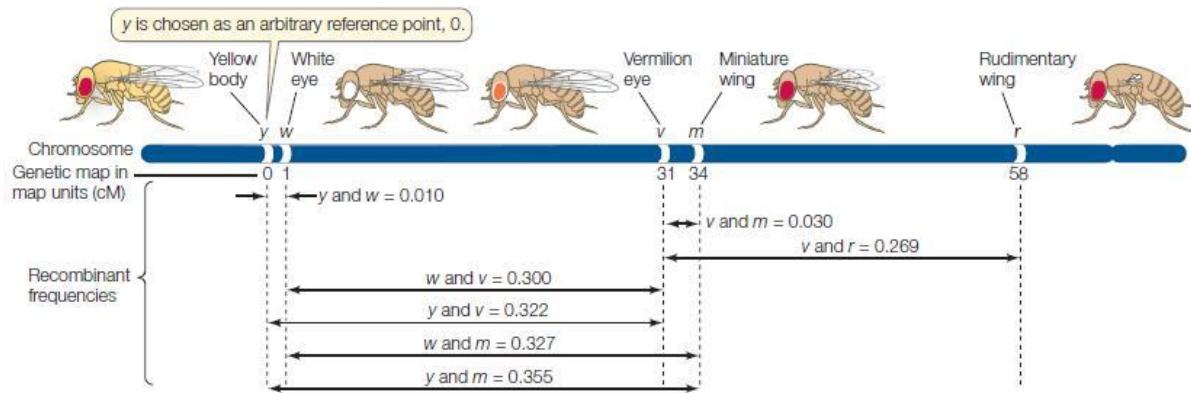
460 *AaCc*, 460 *aacc*, 40 *Aacc*, and 40 *aaCc*

Recombinant frequency (*a* to *c*) = $(40 + 40)/1,000 = 0.08$

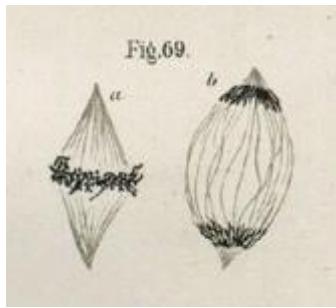
$$\text{Map distance} = 100 \times \text{recombinant frequency} = \\ 100 \times 0.08 = 8 \text{ cM}$$



The recombination frequencies can be used for making genetic maps. Morgan's group conducted several crosses in *Drosophila*. After finding out the frequency he was able to apply for construction of genetic map because the more the distance between two loci, the more will be the recombination. It means the distance between genes can be calculated based on this. The unit is cM (Centimorgan) or map units (1cM = 1 map unit). The following illustration shows an illustration of genetic mapping by Morgan.



Flemming was a German military physician. He found cells contains the coloured genetic material, the chromosomes (Chrome = color; some = body). This is in fact the factor represented by Mendel. Even he discovered that chromosomes splits longitudinally during cell division (His illustration is given below). This is what happens during mitosis. We know, in meiosis the longitudinal splitting happens after crossing over.



The chromosomal theory was not the work of a single scientist by Mendel or Morgan. Many people experimented over decades on it. Indeed, the first logic steps were initiated during 1860 by the mathematician Mendel and evolutionary biologist Charles Darwin. The probable mechanism of transmission from one generation to next was speculated by the discovery of chromosomes by Walther Flemming, a German biologist. Now to connect between chromosomes and heredity. This was done by Boveri, Sutton and Morgan during the dawn of 20th century. Thus the chromosomal theory came out which experimentally proved that chromosomes are responsible for transmission of trait from one generation to next.

In fact Mendel was a Physicist (and philosopher), Darwin was a naturalist, Morgan was a zoologist. Above all Flemming was a military physician!!

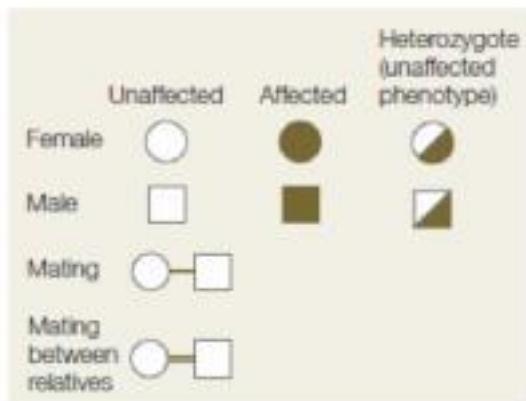
INHERITANCE AND PEDIGREE

Life has evolved on earth gradually. Most of the life forms have two different sex, what we say is a male and a female or a + strain and a – strain. Why life preferred two genders? Male and Female? It would have been simple for the life if only one gender is existing and all of them will reproduce.

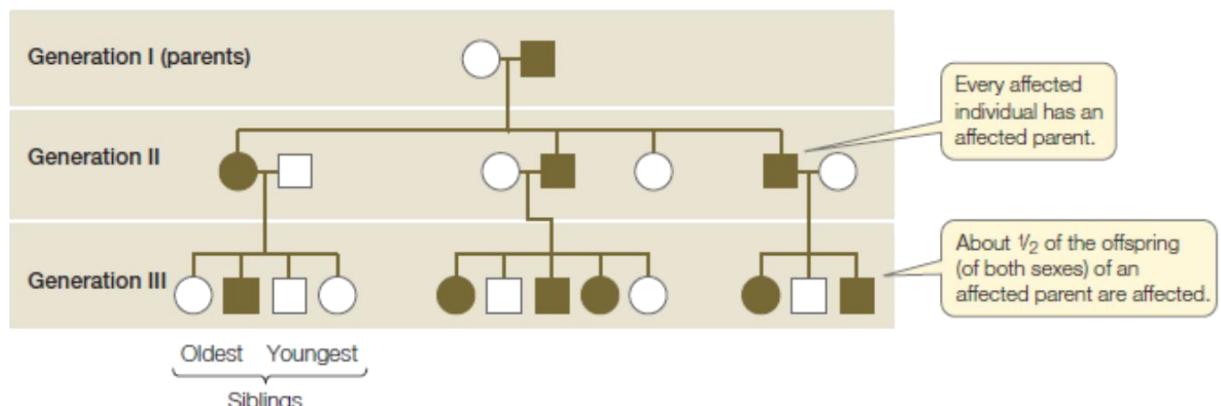
The advantage lies on the recombination event. During sexual reproduction, the chromosome number of the gametes are reduced into half through meiosis. We have seen that meiosis accounts for segregation and assortment. We have also seen the non-parental genotypes appeared in the F₂ generation of dihybrid cross. It means sexual reproduction gives an opportunity for variation through meiotic recombination. Hence life systems are not static, they are dynamic. They are evolving. The most perfect life machine will always be preferred by the nature. Others will disappear, the survival of the fittest. It means the best character is inherited over the generations.

As we have seen with the Mendel and Morgan inheritance can be dominant or recessive

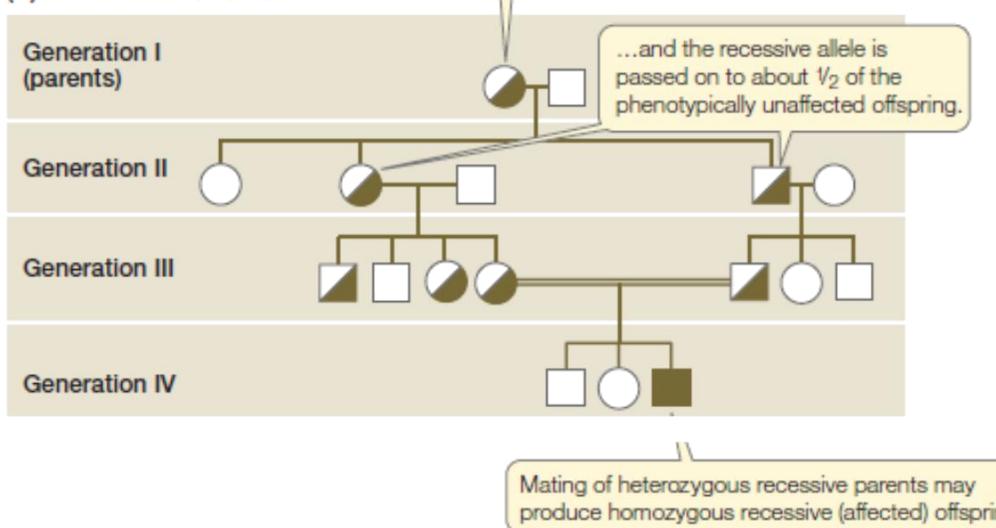
Now let us go through the different types of inheritance. Before discussing that we should see how to represent the inheritance in the form of a diagram. This is known as pedigree. The basic rules of pedigree chart is give below.



(A) Dominant inheritance



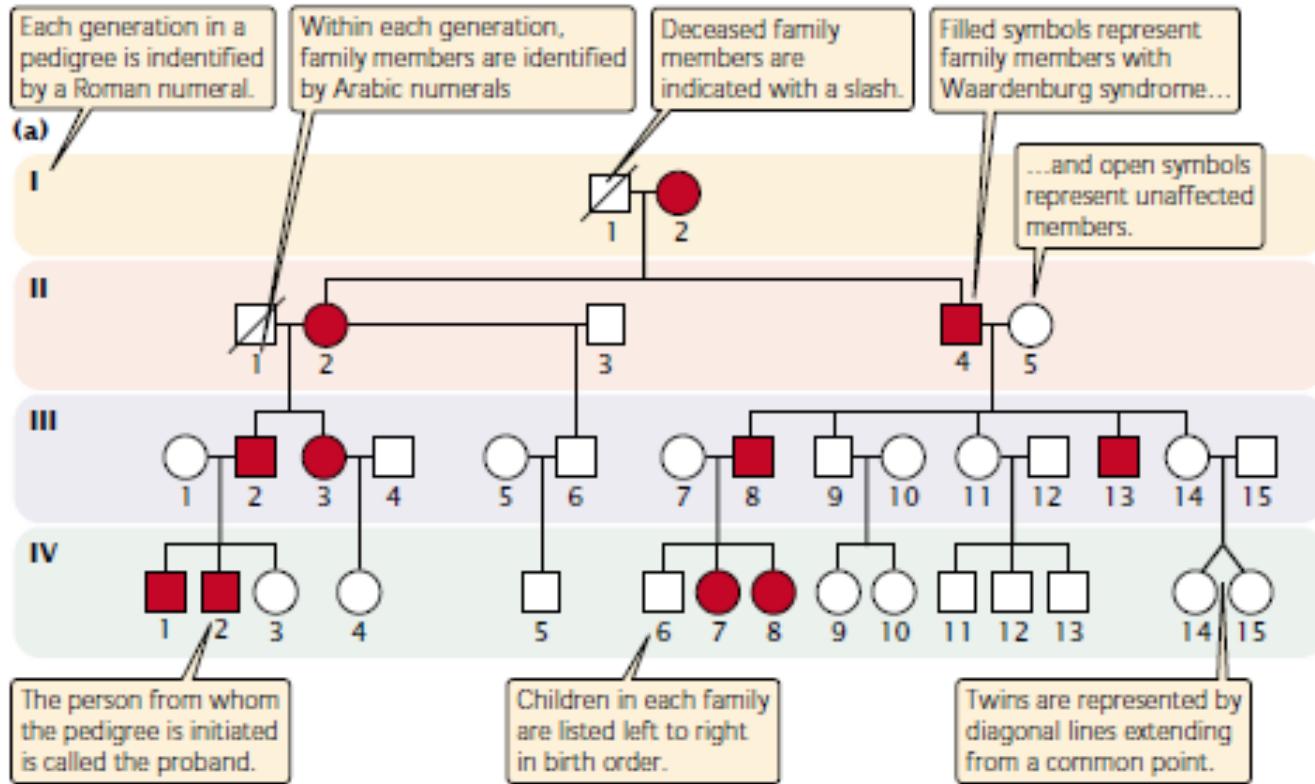
(B) Recessive inheritance

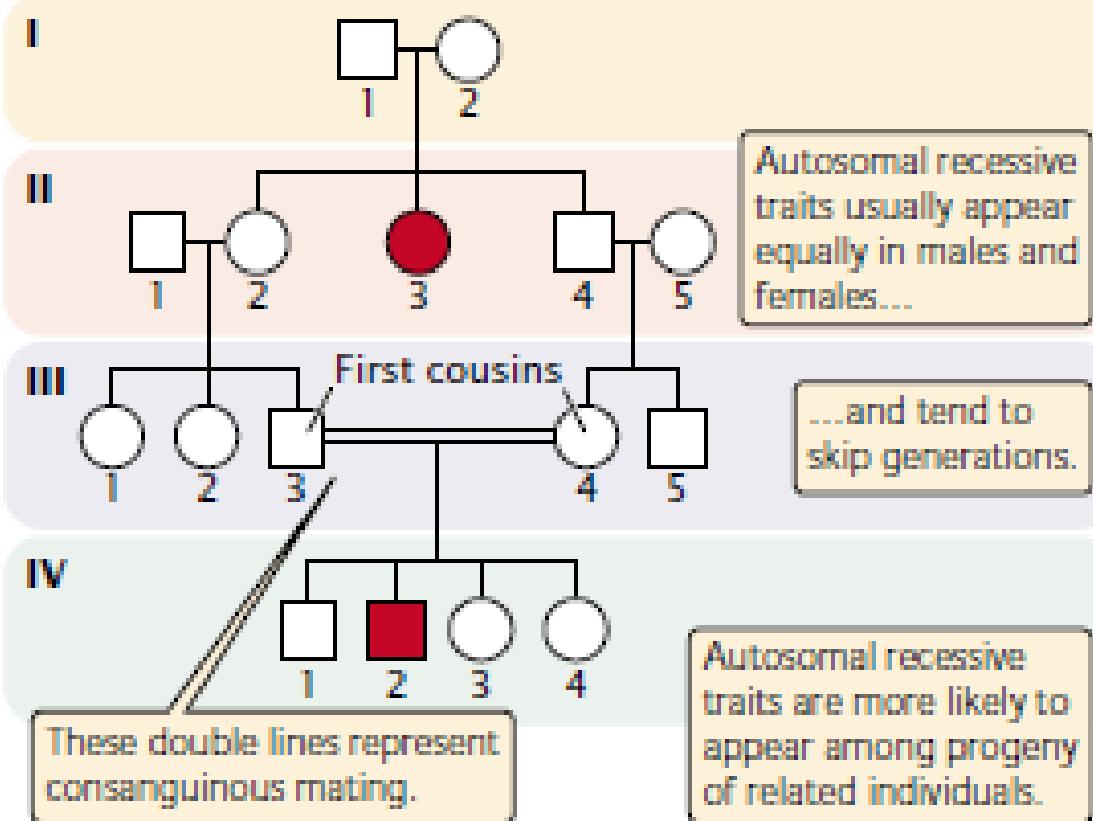


Now calculate the probability of the children getting affected in a cross . The male can be homozygous dominant or heterozygous. If he is homozygous dominant, the probability of an unaffected child is zero. This probability will be 50% if he is heterozygous in each child birth. Now

see the probability in the case of a recessive inheritance . The probability of getting the child having the recessive character is zero, because we know that the expression of a recessive character occurs only in case of homozygous condition of the alleles. But the unaffected child may be a carrier of the allele, even though he or she is not expressing the trait. The above types of inheritance illustrated here appears on both sexes. Hence the factor or the gene is resided on the autosomes. So an autosomal inheritance can be autosomal dominant or autosomal recessive.

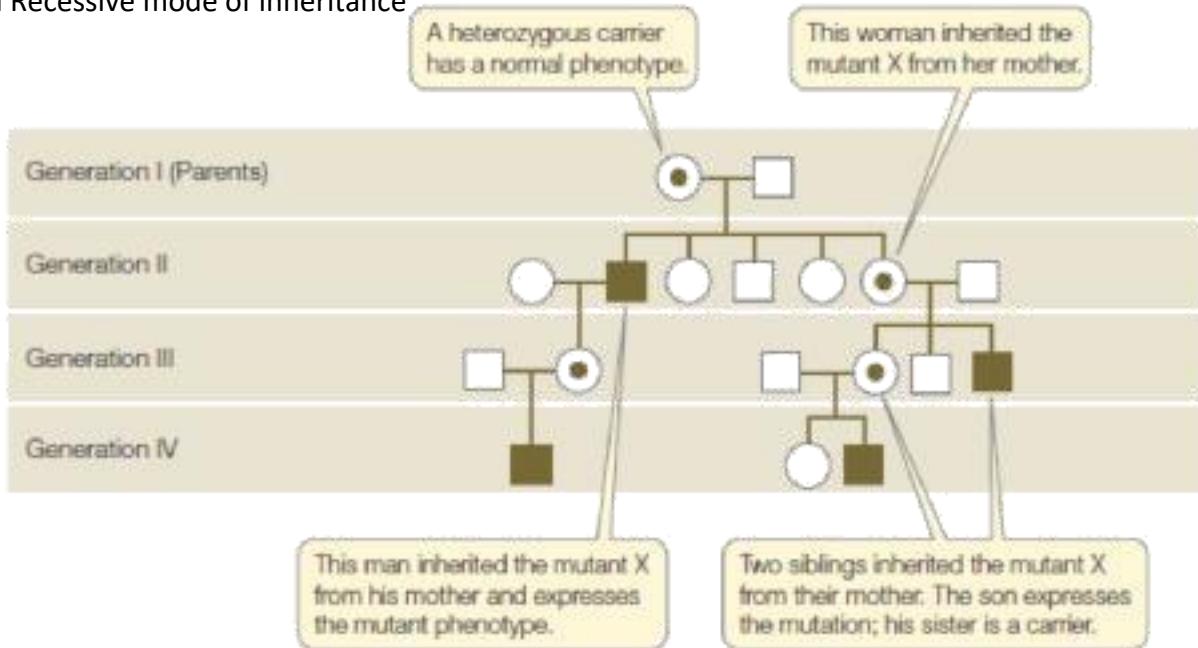
Autosomal dominant trait





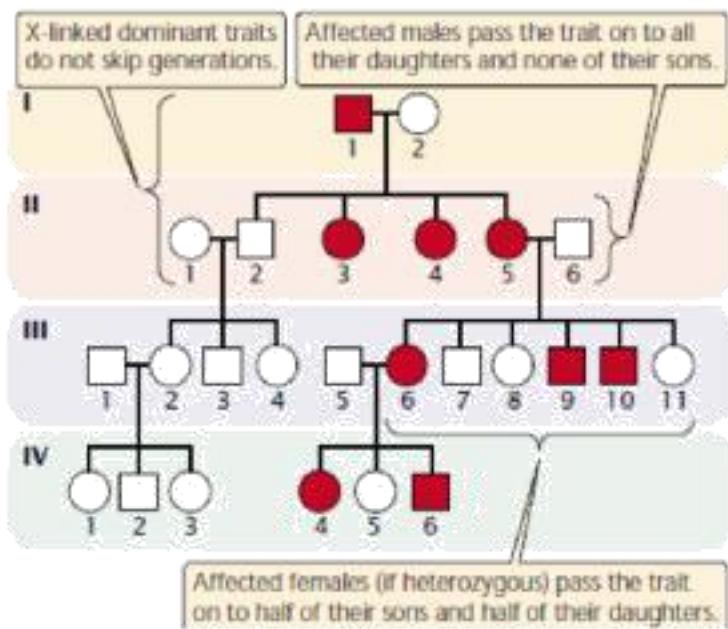
6.4 Autosomal recessive traits normally appear with equal frequency in both sexes and seem to skip generations.

X Linked Recessive mode of inheritance



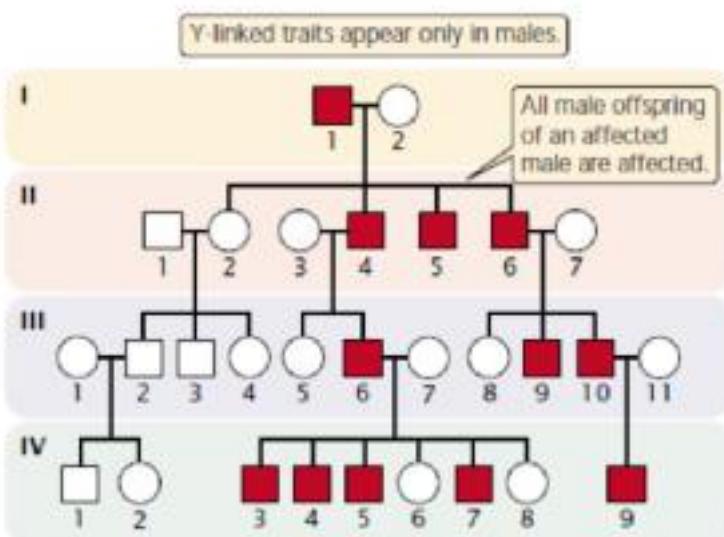
Did you noted three things (1) In females, both X chromosomes should carry the recessive allele for the expression of white eye color. Hence this is an X linked recessive trait. (2) Males have only one X chromosome. Hence the trait will express even if the X chromosome contains the recessive allele (3) X linked recessive traits are more frequently occurs in males compared to females. The reason is that males have only one X chromosome. Hence the recessive allele will express.

In some other cases, X linked characters may appear in dominant pattern also. Illustration is given below



6.9 X-linked dominant traits affect both males and females. An affected male must have an affected mother.

Some of the traits only appear in males because the gene for this trait are located on Y chromosome. They are passed by father to all of his sons, but not daughter. The illustration is given below



6.10 Y-linked trait:
passed from a father to all his sons.

The following table explains the features of different patterns of inheritance

Table 6.1 Pedigree characteristics of autosomal recessive, autosomal dominant, X-linked recessive, X-linked dominant, and Y-linked traits		
Autosomal recessive trait		X-linked dominant trait
1. Appears in both sexes with equal frequency. 2. Trait tends to skip generations. 3. Affected offspring are usually born to unaffected parents. 4. When both parents are heterozygous, approximately 1/4 of the offspring will be affected. 5. Appears more frequently among the children of consanguineous marriages.	5. When one parent is affected (heterozygous) and the other parent is unaffected, approximately 1/2 of the offspring will be affected. 6. Unaffected parents do not transmit the trait.	1. Both males and females are affected; often more females than males are affected. 2. Does not skip generations. Affected sons must have an affected mother; affected daughters must have either an affected mother or an affected father. 3. Affected fathers will pass the trait on to all their daughters. 4. Affected mothers (if heterozygous) will pass the trait on to 1/2 of their sons and 1/2 of their daughters.
Autosomal dominant trait	X-linked recessive trait	Y-linked trait
1. Appears in both sexes with equal frequency. 2. Both sexes transmit the trait to their offspring. 3. Does not skip generations. 4. Affected offspring must have an affected parent, unless they possess a new mutation.	1. More males than females are affected. 2. Affected sons are usually born to unaffected mothers; thus, the trait skips generations. 3. A carrier (heterozygous) mother produces approximately 1/2 affected sons. 4. Is never passed from father to son. 5. All daughters of affected fathers are carriers.	1. Only males are affected. 2. Is passed from father to all sons. 3. Does not skip generations.

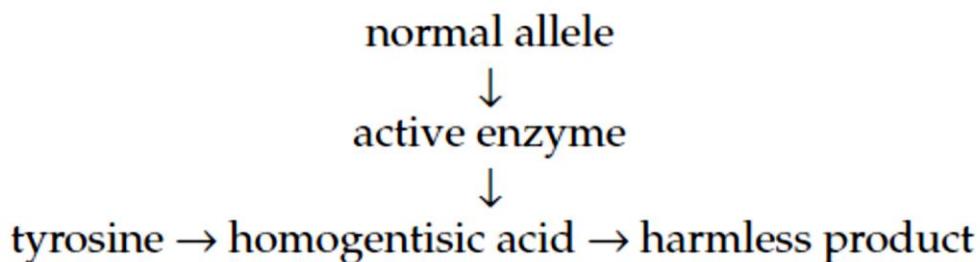
Diseases or traits can pass from one generation to another generation. Many of them follow typical Mendelian inheritance. The following table gives few examples

Trait or disease	Type of inheritance
Color blindness	X linked recessive
Hemophilia	X linked recessive
Huntington disease	Autosomal dominant
Sickle cell anemia	Autosomal recessive
Tongue rolling	Autosomal dominant
Hand clasping	Clasp your hands together. Notice whether your left or your right thumb is on top. If the left thumb is on top you have the dominant trait (C), the right thumb is recessive
Alport syndrome	X linked dominant
ADP/ATP translocase	The gene responsible for this enzyme that moves ADP into and ATP out of mitochondria has been linked to Y chromosome

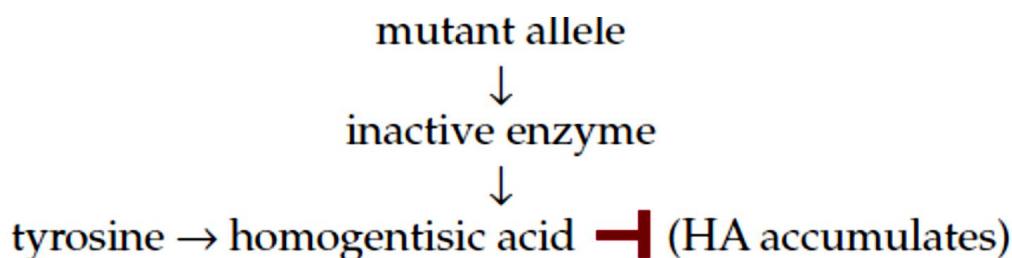
First observation for inheritance being related to a metabolic pathway

Archibald Garrod was an English physician. He observed that many children exhibited a symptom - the urine turns dark brown immediately after urination, i.e. when exposed to air. Further analysis indicated that the frequency of the disease was more common in children of consanguineous (within the family marriage e.g. Cousin-Cousin, uncle-niece) marriages. Simultaneously, Garrod found the concepts involved in rediscovery of Mendelian inheritance inspiring. A pedigree chart allowed Garrod to determine that the couples had a recessive allele causing the child to be homozygous recessive.

Biochemical composition of the urine was investigated by Garrod and observed the following pattern for normal individuals or heterozygous individuals.



So in homozygous recessive cases the reaction is blocked.



This was an indication that inheritance is responsible for turning the urine black. Garrod speculated this fact, but was unable to identify the enzyme or the gene. However an exact confirmation required several years of study. In 1958, the enzyme was identified –homogentisic acid oxidase and in 1996 the gene was identified.

Chromosomes contain both proteins and DNA: What is the evidence that which chemical component carries the genetic information?

Chromosome is a combination of two chemicals: DNA and Proteins:

The chromosome is a dynamic structure in the sense that it condenses and expands during various stages of the cell cycle. Chromosome is a mixture of two different components (i) DNA and (ii) proteins in higher quantity compared to DNA. In fact, DNA is bound to proteins. This unique combination accounts for dynamicity of the structure. What we see or represent for a

chromosome is the most condensed state of chromatin fibers (DNA fiber). This can be visually seen during the metaphase of the cell division (Fig 1).

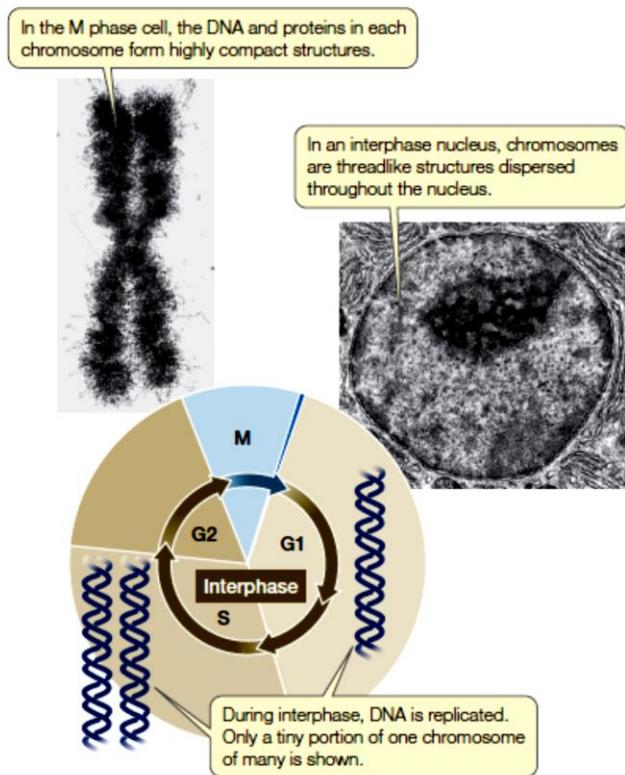


Figure 1: Dynamicity of the chromatin fibers – they expand and contract according to the stage of the cell cycle: M=Mitotic phase (Division phase), G1 and G2 are gap phases and synthesis of the raw materials occurs during the Synthesis (S phase). [Figure adapted from Sadava et al, Life: The science of Biology, 9th edition, Page 211, Fig 11.8]

The circumstantial evidences and logic to assume that DNA is the genetic material:

Living creatures exhibit a great diversity. Similar traits are observed in living forms of the same kind (species), while differences are observed between different species. It means that the genetic composition of living forms of one kind differs from that of the other kind. This exactly means their amount will differ. So can we prove it? This theory was proven by Robert Feulgen, who developed a red colored dye which binds to DNA. It stains DNA material red inside the nucleus. So the intensity of the red color is an approximate estimate of the DNA it contains. This dye is known as Feulgen stain. The Feulgen staining techniques presented the following information: It was in the right position (inside the nucleus) and the color intensity varied between two species.

The need of more cause and effect evidence!

The Feulgen staining only provided a circumstantial evidence. We should prove with a cause and effect situation that DNA carries the genetic information and not the proteins. How to do it?

Frederick Griffith was a physician from England. He was working with bacteria, which are visible only under a microscope. Pneumonia was taking many lives during his time. So he wanted to develop a vaccine for pneumonia. He found that there exists two forms (strains) of the bacteria which causes pneumonia, *Streptococcus pneumonia*. These strains are: Smooth (S) and Rough (R) forms.

Smooth forms are capable of causing the disease, while rough forms do not cause the disease. The reason is that the smooth forms are hidden inside a proteinaceous cover, so it can cheat the firewall (The immune system), while rough forms are not able to utilize that trick, as they lack the protein coat. Hence the firewall will definitely catch and eliminate them. Now he planned and executed the experiments as illustrated in figure 2.

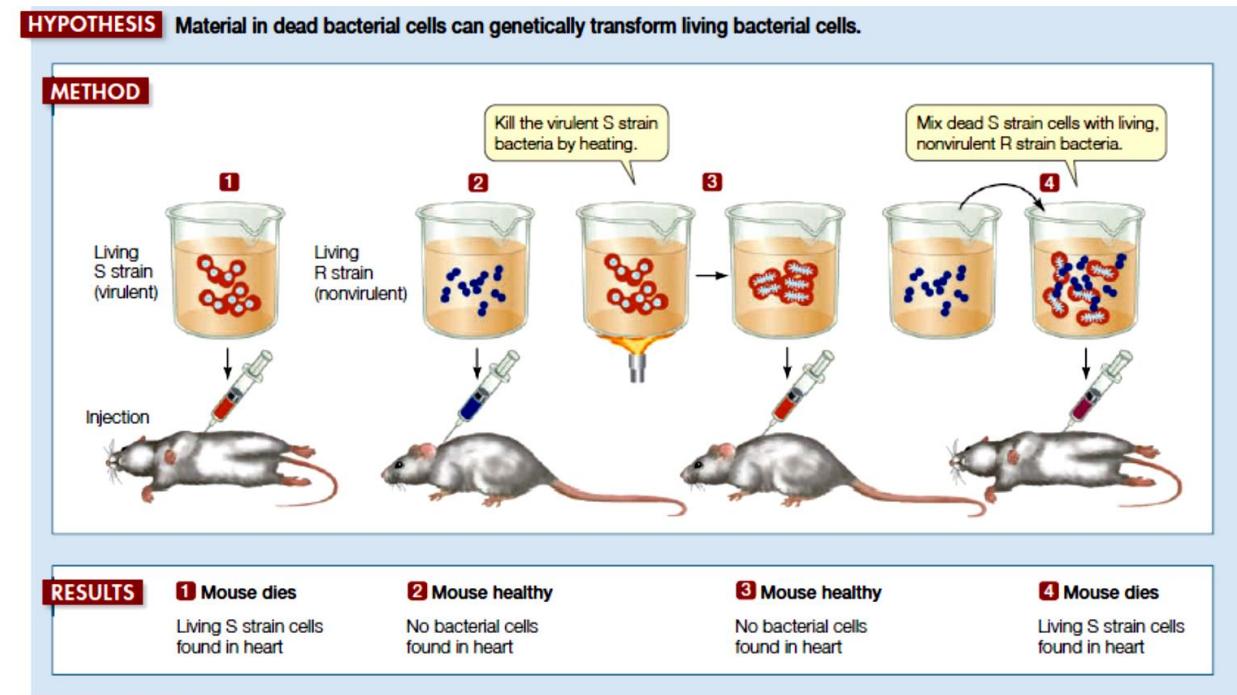


Figure 2: The experiments conducted by Griffith with the bacteria and mouse. [Figure adapted from Sadava et al, Life: The science of Biology, 9th edition, Page 268, Fig 13.2].

The above figure illustrates that in the experiment 4 the material present in the S form transforms the material present in the R form. So we can say that some transforming principle is responsible for this change from an R form to S form of bacteria. How to identify this transforming principle? The scientific group led by Oswald Avery of Rockefeller University cracked this problem. Their experiment is illustrated in Figure 3.

Experiments by Avery, MacLeod, and McCarty showed that DNA from the virulent S strain of pneumococcus was responsible for the transformation in Griffith's experiments (see Figure 13.1).

HYPOTHESIS The chemical nature of the transforming substance from pneumococcus is DNA.

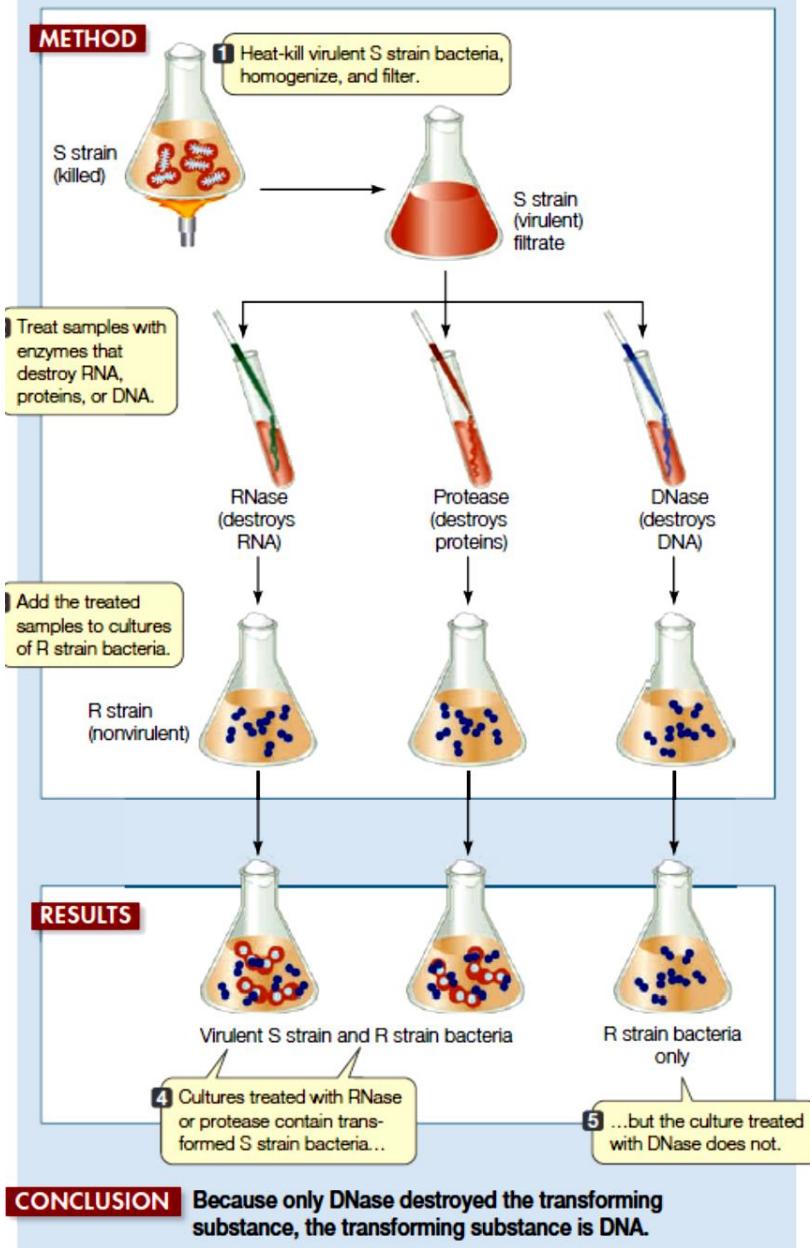


Figure 3: Identifying the “transforming principle”. [Figure adapted from Sadava et al, *Life: The science of Biology*, 9th edition, Page 269, Fig 13.1]

The above work was published without much impact in 1944 because, many were not aware of the fact that DNA is complex enough to give the diverse output. Moreover many were still wondering whether microscopic small creatures, like bacteria, has genes in it.

The impact of this work was intensified after another experimental work published in 1952 by Alfred Hershey and Martha Chase at Carnegie Laboratory of Genetics. They were trying to determine whether DNA or protein contains the genetic material by using a bacteriophage (a virus that attacks and kills the bacteria). Why they have selected a virus? Because virus is composed of just two components that we are trying to sort, the protein cover and the DNA inside it (Figure 4).

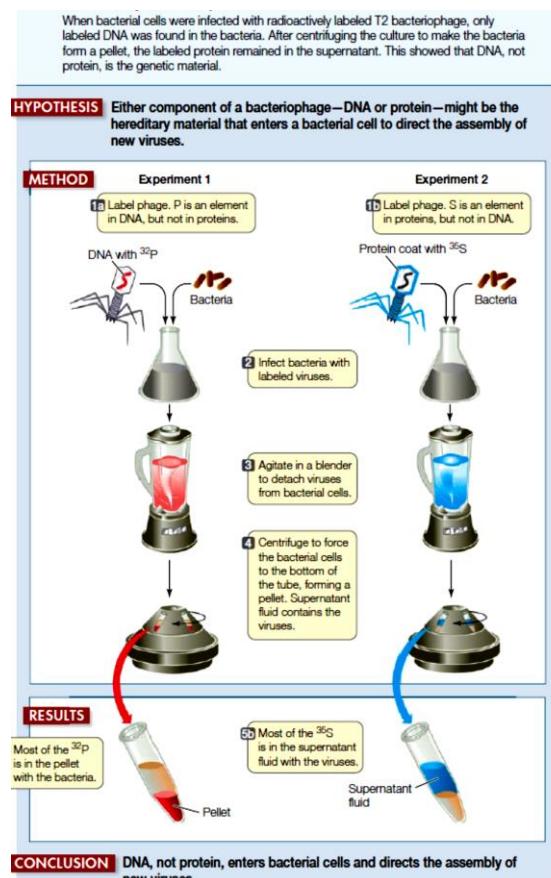


Figure 4: Hershey and Chase experiment. [Figure adapted from Sadava et al, *Life: The science of Biology*, 9th edition, Page 271, Fig 13.4]

DNA structure:

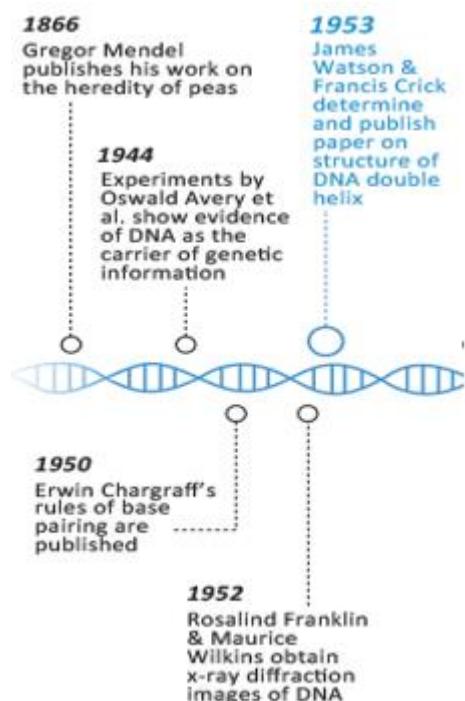


Figure 1- Timeline of events in the discovery of DNA structure. Image courtesy www.genscript.com

The race for DNA structure:

The structure of DNA remained elusive until experimental evidence of many types were considered together in a theoretical framework. Figure 1 indicates the various experiments which acted as steps towards the discovery of the correct DNA structure.

The most crucial evidence was obtained using X-ray crystallography. Some chemical substances, when they are isolated and purified, can be made to form crystals. The positions of atoms in a crystallized substance can be inferred from the diffraction pattern of X-rays passing through the substance.

The events that provided information about this vital molecule are described in the following text.

Chemical composition of DNA:

Biochemists knew that DNA was a polymer of nucleotides. Each nucleotide consists of a molecule of the sugar deoxyribose, a phosphate group, and a nitrogen containing base.

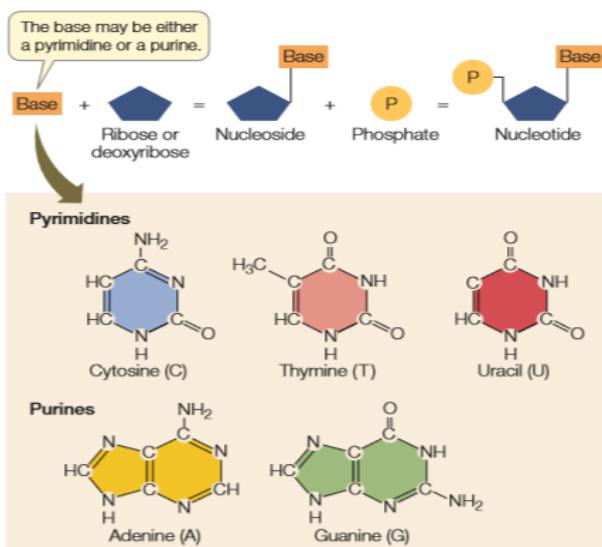


Figure 2- Chemical composition of DNA monomers. The only differences among the four nucleotides of DNA are their nitrogenous bases: the purines adenine (A) and guanine (G), and the pyrimidines cytosine (C) and thymine (T). Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

In 1950, biochemist Erwin Chargaff reported that DNA from many different species—and from different sources within a single organism—exhibits certain regularities. In almost all DNA, the following rule holds: The amount of adenine equals the amount of thymine ($A = T$), and the amount of guanine equals the amount of cytosine ($G = C$).

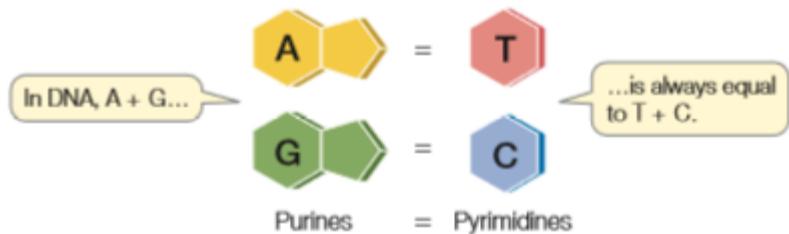


Figure 3- Chargaff's rule- In DNA, total abundance of purines is equal to total abundance of pyrimidines. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

In 1952, Rosalind Franklin was able to obtain an X-ray diffraction pattern for certain DNA fibers. This experiment provided the greatest help required by the scientists to deduce the DNA structure.

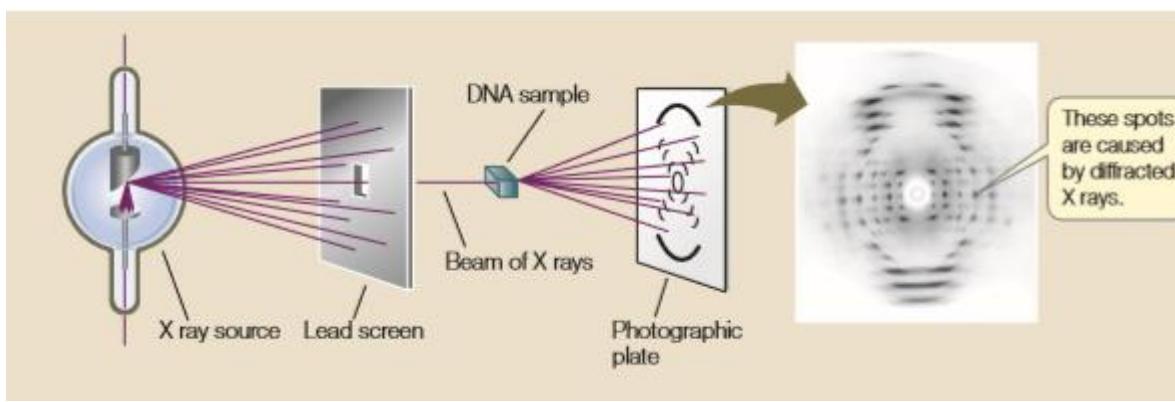


Figure 4- The positions of atoms in a crystallized chemical substance can be inferred by the pattern of diffraction of X rays passed through it. The pattern of DNA is both highly regular and repetitive. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

Around the same time, Linus Pauling proposed a triple stranded helical structure for DNA. Linus Pauling had discovered the helical nature of protein folding and deduced that the same folding pattern may be followed by DNA. The structure proposed by Linus Pauling had the phosphate groups of the nucleotides facing inside the helical core. However, such a structure would result in the phosphate group repulsion (due to the negatively charged oxygen groups). Almost unbelievable that the man who had such a command over chemical bonds would get this wrong.

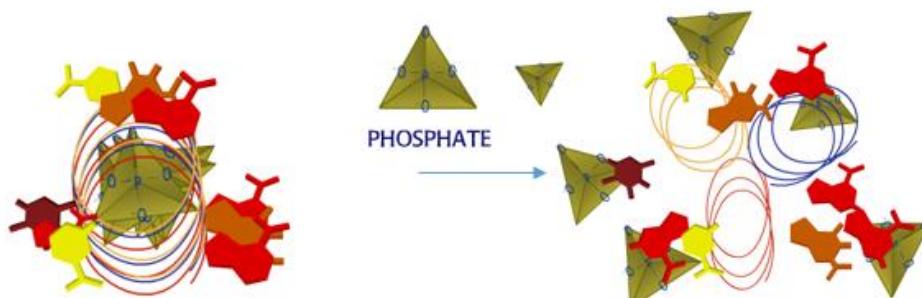


Figure 5- DNA structure as proposed by Linus Pauling (Top view). The negatively charged oxygen atoms repel each other and would cause the strands to disassociate. Image courtesy- <http://www.dnai.org/>

Double stranded structure discovery by Watson and Crick:

The English physicist Francis Crick and the American geneticist James D. Watson, who were both then at the Cavendish Laboratory of Cambridge University, used model building to solve the structure of DNA.

Watson and Crick attempted to combine all that had been learned so far about DNA structure into a single coherent model. Rosalind Franklin's crystallography results (see Figure 4) convinced Watson and Crick that the DNA molecule must be helical (cylindrically spiral). Density measurements and previous model building results suggested that there are two polynucleotide chains in the molecule. Modeling studies also showed that the strands run in opposite directions, that is, they are antiparallel; that two strands would not fit together in the model if they were parallel.

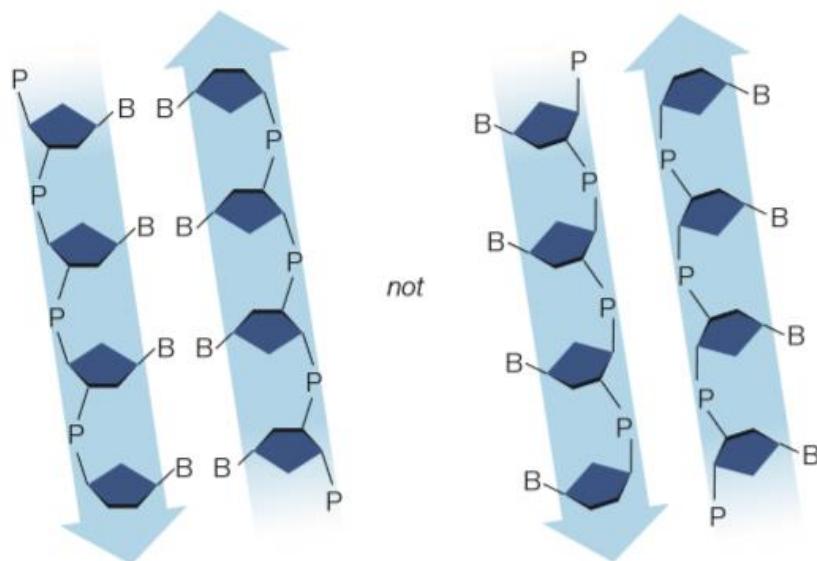


Figure 6- DNA model proposed by Watson and Crick made several assumptions. The nucleotide bases are on the interior of the two strands, with a sugar-phosphate backbone on the outside. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

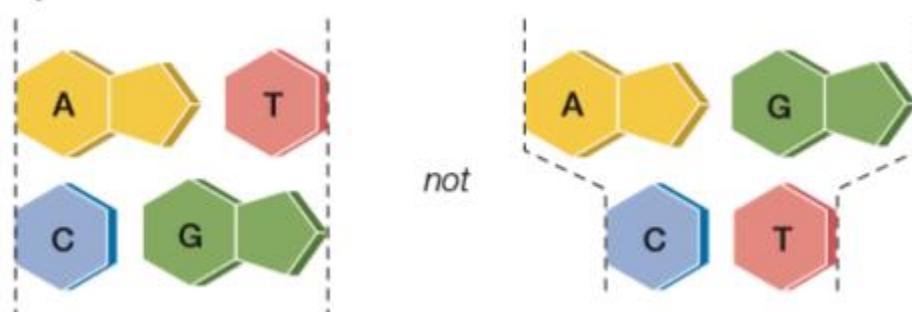


Figure 7-To satisfy Chargaff's rule (purines = pyrimidines), a purine on one strand is always paired with a pyrimidine on the opposite strand. These base pairs (A-T and G-C) have the same width down the double helix, a uniformity shown by x-ray diffraction. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

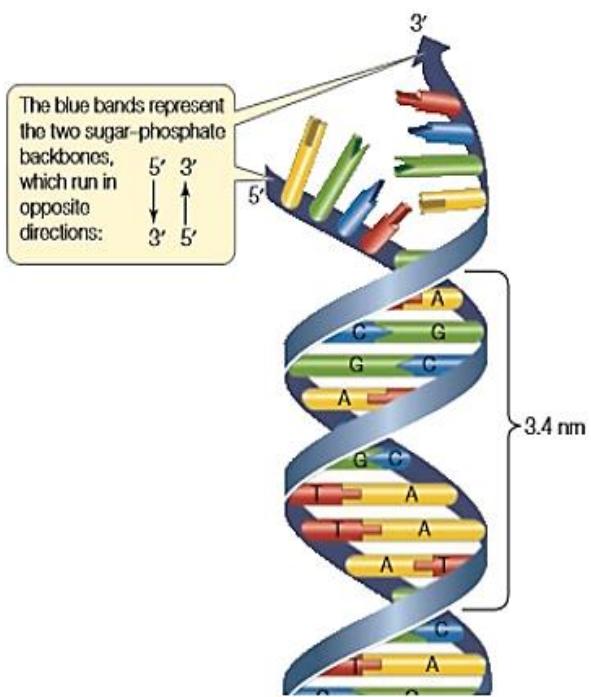


Figure 8- In late February of 1953, Crick and Watson built a model out of tin that established the general structure of DNA. This structure explained all the known chemical properties of DNA, and it opened the door to understanding its biological functions. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

Important properties of DNA structure:

1. The double stranded helix has a uniform diameter
2. The two strands run in opposite direction (antiparallel).
3. The backbone of each strand is made up of sugar phosphate groups linked by phosphodiester bonds
4. The two strands are held together by hydrogen bonding between the nitrogenous bases.

Meselson and Stahl experiment:

This experiment was instrumental in proving that DNA follows a semiconservative model of replication. A cell while undergoing division requires that old cell to produce copies of DNA that can be transferred to the new cells. The new cells receive a copy of the parent cell's DNA.

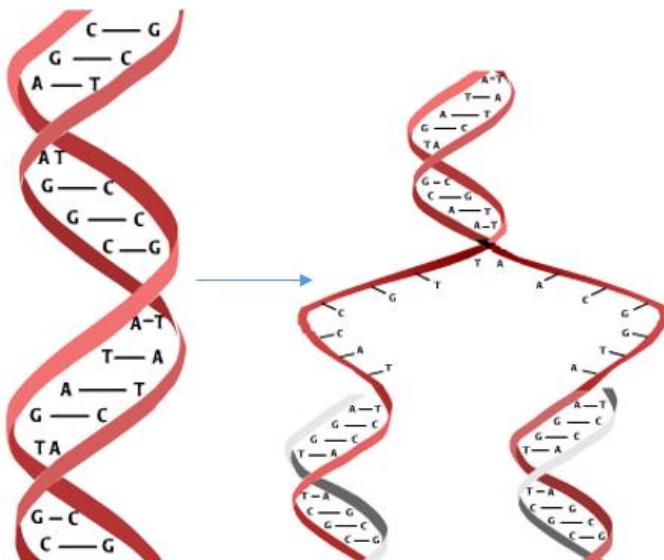


Figure 9- Watson and Crick suggested a semiconservative model of replication, wherein each parental strand acts as a template for synthesizing a new complementary strand. In this model, each daughter molecule consists of one old strand (from parent molecule) and one newly synthesized strand. Image courtesy- www.dnai.org

Meselson and Stahl made clever use of radiolabelled (¹⁵N) heavy isotope of nucleotides and a density gradient of Cesium chloride (CsCl) to provide evidence for the semiconservative model of replication.

They collected some of the bacteria after each division and extracted DNA from the samples. To separate the DNA from the cells at different generation on basis of density, they developed a density gradient solution in a test tube using CsCl. They found that the density gradient was different in each bacterial generation:

- At the time of the transfer to the ^{14}N medium, the DNA was uniformly labeled with ^{15}N , and hence formed a single band corresponding with dense DNA.
- After one generation in the ^{14}N medium, when the DNA had been duplicated once, all the DNA was of intermediate density.
- After two generations, there were two equally large DNA bands: one of low density and one of intermediate density.
- In samples from subsequent generations, the proportion of low-density DNA increased steadily.

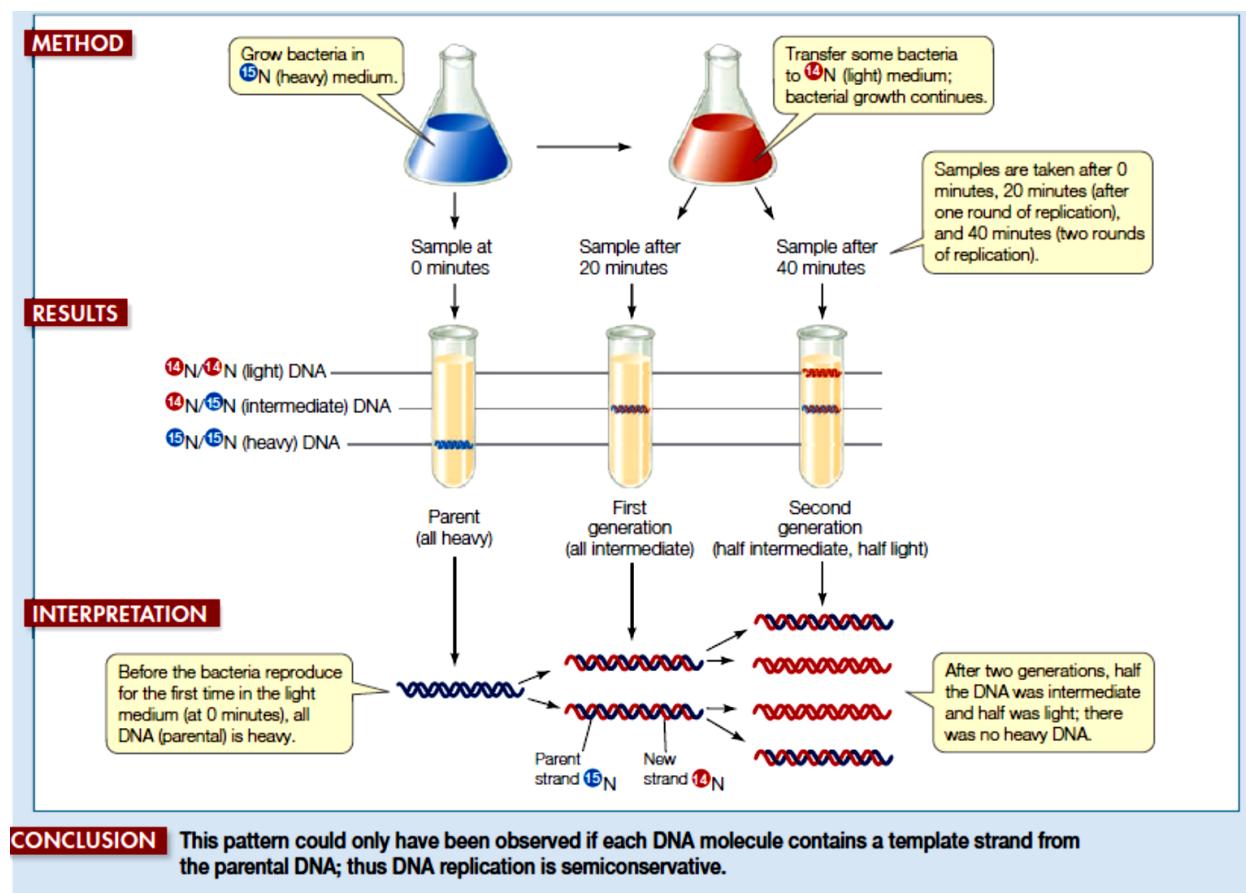


Figure 10- Meselson and Stahl experiment to prove that DNA replicates in a semiconservative manner. The researchers grew another *E. coli* culture on ^{15}N medium, then transferred it to normal ^{14}N medium and allowed the bacteria to continue growth. Image courtesy- Sadava et al, *Life: The science of Biology*, 9th edition.

The results of this experiment can be explained only by the semiconservative model of DNA replication. In the first round of DNA replication in the ^{14}N medium, the strands of the double helix—both heavy with ^{15}N —separated. Each strand then acted as the template for a second strand, which contained only ^{14}N and hence was less dense. Each double helix then consisted of one ^{15}N strand and one ^{14}N strand, and was of intermediate density. In the second replication, the ^{14}N -containing strands directed the synthesis of partners with ^{14}N , creating low-density DNA, and the ^{15}N strands formed new ^{14}N partners. The crucial observation demonstrating the semiconservative model was that intermediate-density DNA ($^{15}\text{N}-^{14}\text{N}$)

appeared in the first generation and continued to appear in subsequent generations. With the other models, the results would have been quite different:

- If conservative replication had occurred, the first generation would have had both high-density DNA ($^{15}\text{N}-^{15}\text{N}$) and low-density DNA ($^{14}\text{N}-^{14}\text{N}$), but no intermediate density DNA.
- If dispersive replication had occurred, the density of the new DNA would have been intermediate, but DNA of this density would not continue to appear in subsequent generations.

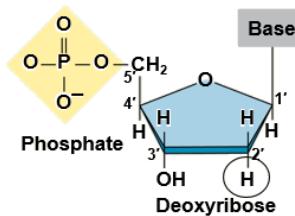


Figure 11- Monomer of DNA (nucleotide monophosphate). The monomers exist within the strand in monophosphate form.

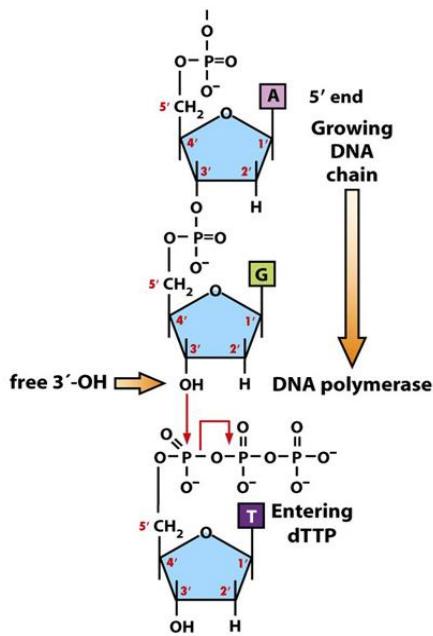


Figure 12- During the formation of a phosphodiester bond, the incoming monomers are in nucleotide triphosphate form. The cleavage of the beta and gamma phosphate provides the energy needed for phosphodiester bond formation. This reaction is catalyzed by DNA polymerase.

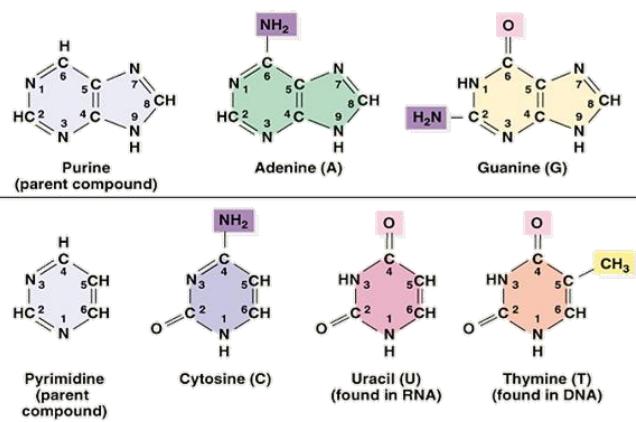


Figure 13- Nitrogenous bases of DNA and RNA. Uracil is present in RNA instead of Thymine.

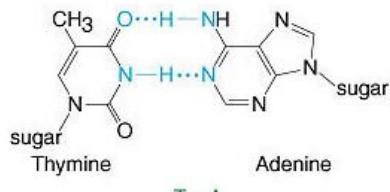
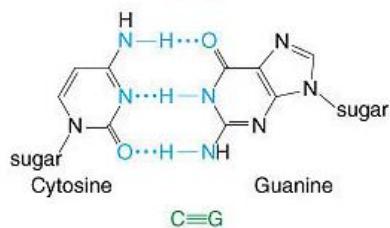


Figure 14- Hydrogen bonding between the nitrogenous bases of DNA.



Pairs of complementary bases form hydrogen bonds that hold the two strands of the DNA double helix together.

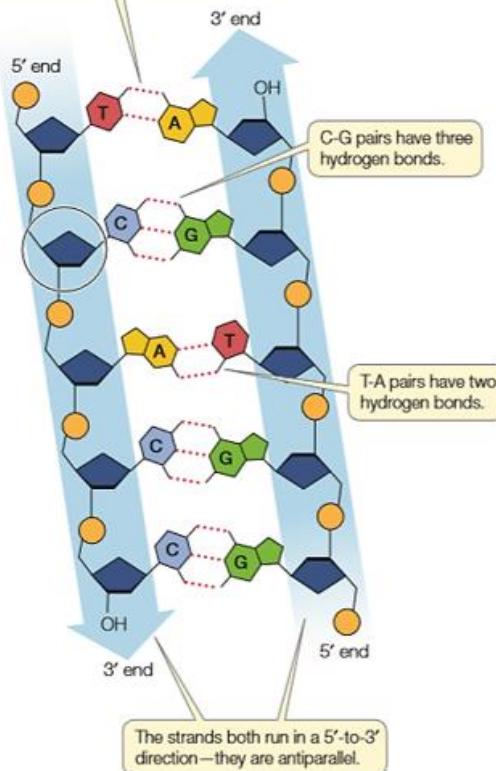


Figure 15- The purines (A and G) pair with the pyrimidines (T and C, respectively) to form base pairs that are equal in size and resemble the rungs on a ladder whose sides are formed by the sugar-phosphate backbones. The deoxyribose sugar (left) is where the 3' and 5' carbons are located. The two strands are antiparallel. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

This concludes the chase for discovery of DNA structure. The next section describes the process of DNA replication.

Kornberg, Arthur, an American physician and biochemist, was the first to discover how monomers of deoxyribonucleic acid (DNA) duplicate within bacterial cells and also the first to devise a cell free method for replicating DNA. For these achievements he shared the 1959 Nobel Prize in physiology.

Arthur Kornberg isolated a new enzyme from *E. coli*, "DNA polymerase", which had the property to assemble nucleotides and manufacture DNA. The cell free setup for replicating DNA was developed *in vitro* (in a test tube) by providing a pool of free nucleotides, a DNA primer (he used calf thymus DNA), a source of magnesium ions, and ATP.

Kornberg used DNA polymerase to verify one of the essential elements of the Watson - Crick Model of DNA structure: DNA is always polymerized in the 5' to 3' direction (H-CH₂ sugar phosphate bonds to H-O sugar phosphate bond; new nucleotides are added at the 3' end).

The findings of Kornberg experiment provided conclusive evidence that the double stranded helical model provided by Watson- Crick was accurate.

Kornberg designed the following experiment to determine the growing end for DNA replication process

The reaction used certain key ingredients as follows-

1. Labelled nucleotides (contained the radioactive phosphorus isotope ³²P)
2. DNA primer (he used calf thymus DNA),
3. a source of magnesium ions, and ATP

The procedure for the assignment is as follows:

Initiation of the reaction using unlabelled nucleotide as the nucleotide precursor.

Add radioactive (³²P) nucleotide for a brief period, and then quickly stop the reaction.

At the completion of the experiment, Kornberg observed that ³²P nucleotides attached only to the 3'-OH end of the growing strand.

The results of the experiment clearly indicated that the newly synthesized strand grows at the 3'-OH end.

DNA replication:

Semiconservative DNA replication in the cell involves a number of different enzymes and other proteins. It takes place in two general steps:

- The DNA double helix is unwound to separate the two template strands and make them available for new base pairing.
- As new nucleotides form complementary base pairs with template DNA, they are covalently linked together by phosphodiester bonds, forming a polymer whose base sequence is complementary to the bases in the template strand.

The process of replication occurs within a region called the replication bubble, which contains all the enzymes required for replication and the DNA strands to be replicated. The replication bubble consists of two replication forks, in which the synthesis of DNA proceeds in the two opposite directions.

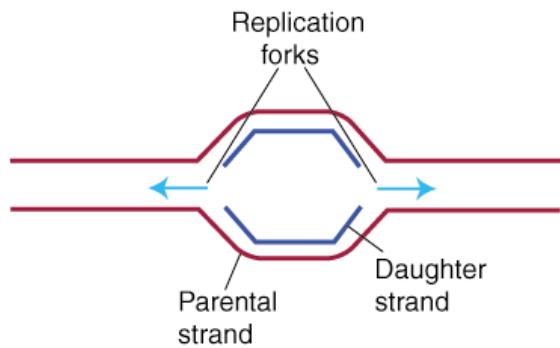


Figure 19- Replication bubble with two replication forks proceeding in opposite directions

Replication is carried out due to the efficient working of various enzymes involved in DNA replication.

DNA is replicated through the interaction of the template strand with a huge protein complex called the replication complex, which contains at least four proteins, including DNA polymerase. All chromosomes

have at least one region called the origin of replication (ori), to which the replication complex binds with high specificity.

The first event at the origin of replication is the localized unwinding and separation (denaturation) of the DNA strands. There are several forces that hold the two strands together, including hydrogen bonding and the hydrophobic interactions of the bases. An enzyme called **DNA helicase** uses energy from ATP hydrolysis to unwind and separate the strands. Proteins called **single-strand binding proteins** bind to the unwound strands to keep them from reassociating into a double helix. This process makes each of the two template strands available for complementary base pairing.

Three types of DNA polymerase exist which assist in the process of DNA replication.

- DNA polymerase I
- DNA polymerase II
- DNA polymerase III

The function of these DNA polymerases is revealed as the process of DNA replication is studied further.

The first DNA polymerase to act in the replication complex is **DNA polymerase III**, whose function is to extend the new strand.

A DNA polymerase elongates a polynucleotide strand by covalently linking new nucleotides to a previously existing strand. However, it cannot start this process without a short “starter” strand, called a primer. In DNA replication, the primer is usually a short single strand of RNA (Figure 14). This RNA primer strand is complementary to the DNA template, and is synthesized one nucleotide at a time by an enzyme called a primase. The DNA polymerase III then adds nucleotides to the 3' end of the primer and continues until the replication of that section of DNA has been completed. Thus the basic function of DNA polymerase III is to extend the new strand starting from one end of the RNA primer.

DNA polymerase I (discovered by Arthur Kornberg) degrades the RNA primer, and adds DNA in its place.

This action of DNA polymerase I causes the resulting new strand to have a small gap. This gap is filled in by DNA ligases. When DNA replication is complete, each new strand consists only of DNA.

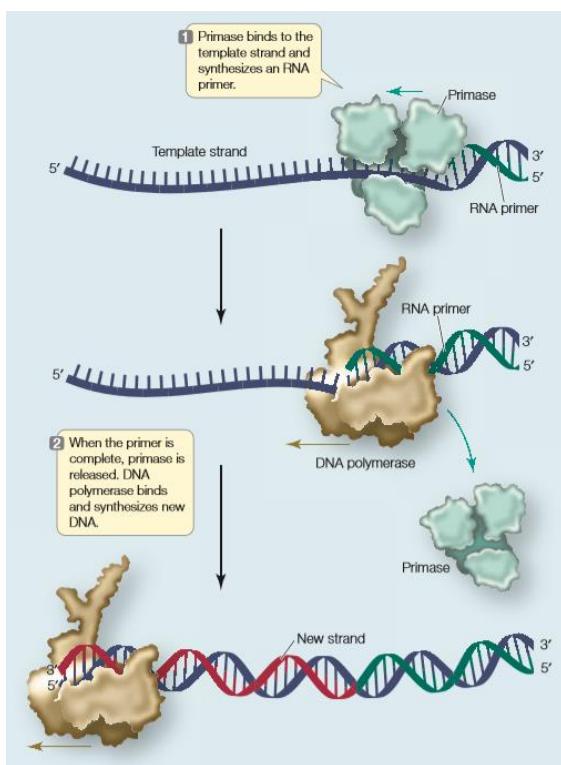


Figure 20- DNA polymerase requires a primer to initiate replication. Primase is the enzyme which provides an RNA primer. DNA polymerase attaches nucleotides to the end of this RNA primer and extends the new strand. Note: the new strand is DNA in nature and not RNA. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

The two daughter strands resulting from a double stranded parent grow differently. The DNA double helix is antiparallel in nature, therefore:

- One newly replicating strand (the leading strand) is oriented so that it can grow continuously at its 3'end as the fork opens up.
- The other new strand (the lagging strand) is oriented so that as the fork opens up, its exposed 3'end gets farther and farther away from the fork, and an unreplicated gap is formed. This gap would get bigger and bigger if there were not a special mechanism to overcome this problem.

Synthesis of the lagging strand requires the synthesis of relatively small, discontinuous stretches of sequence. These discontinuous stretches are synthesized just as the leading strand is, by the addition of new nucleotides one at a time to the 3' end of the new strand, but the synthesis of this new strand moves in the direction opposite to that in which the replication fork is moving. These stretches of new DNA are called Okazaki fragments (after their discoverer, the Japanese biochemist Reiji Okazaki). While the leading strand grows continuously “forward,” the lagging strand grows in shorter, “backward” stretches with gaps between them. A single primer is needed for synthesis of the leading strand, but each Okazaki fragment requires its own primer to be synthesized by the primase. DNA polymerase III then synthesizes an Okazaki fragment by adding nucleotides to one primer until it reaches the primer of the previous fragment. At this point, DNA polymerase I removes the old primer and replaces it with DNA. Left behind is a tiny nick—the final phosphodiester linkage between the adjacent Okazaki fragments is missing. The enzyme DNA ligase catalyzes the formation of that bond, linking the fragments and making the lagging strand whole.

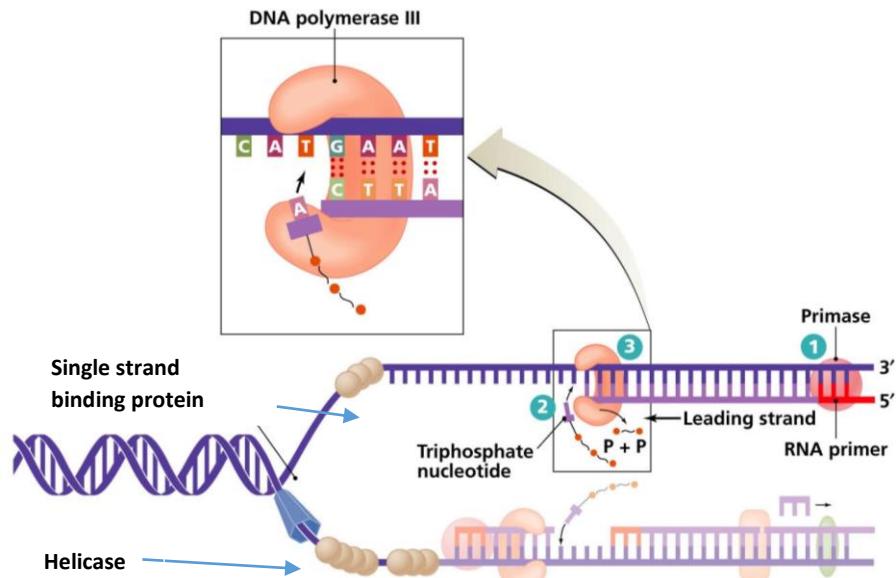


Figure 21- Synthesis of the leading strand. Image courtesy- Biology, by Campbell et al.

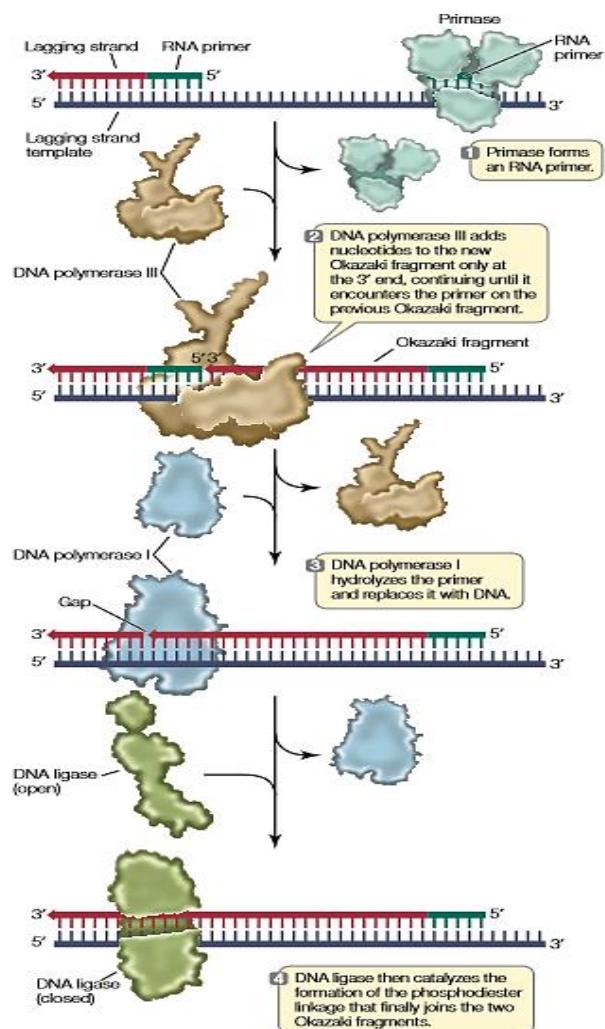


Figure 22- Synthesis of the lagging strand. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

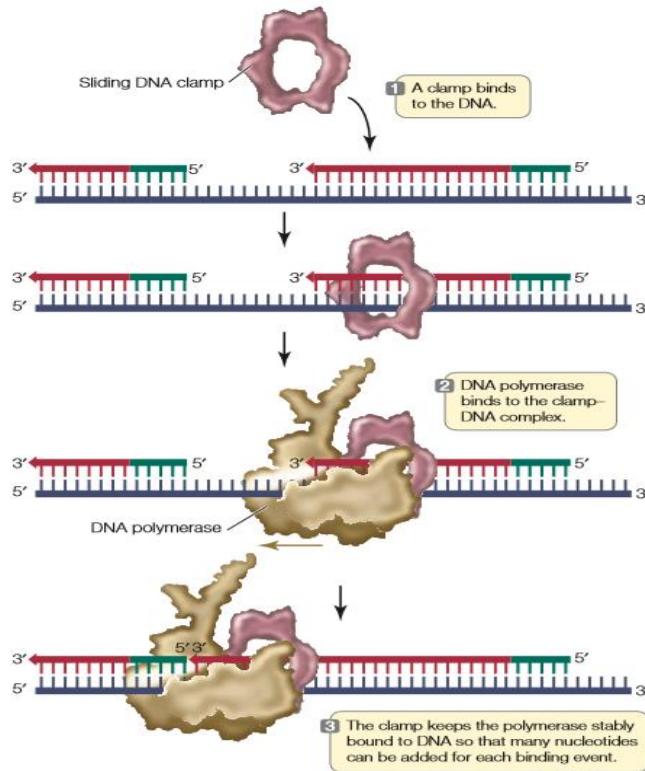


Figure 23- Sliding DNA clamp increases the efficiency of polymerization by keeping the enzyme bound to the substrate, so the enzyme does not have to repeatedly bind to template and substrate. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

During DNA replication, the progress of the replication fork generates positive supercoils ahead of the replication machinery and negative supercoils behind it. The DNA can be supercoiled to such an extent that if left unchecked it could impede the progress of the protein machinery involved. This is prevented by **DNA topoisomerase**, which makes single-stranded nicks to relax the helix.

Small circular chromosomes, such as those of bacteria (consisting of 1–4 million base pairs), have a single origin of replication. Two replication forks form at this ori, and as the DNA moves through the replication complex, the replication forks extend around the circle. Two interlocking circular DNA molecules are formed, and they are separated by an enzyme called DNA topoisomerase. DNA polymerases are very fast. In *E. coli*, replication can be as fast as 1,000 bases per second, and it takes 20–40 minutes to replicate the bacterium's 4.7 million base pairs.

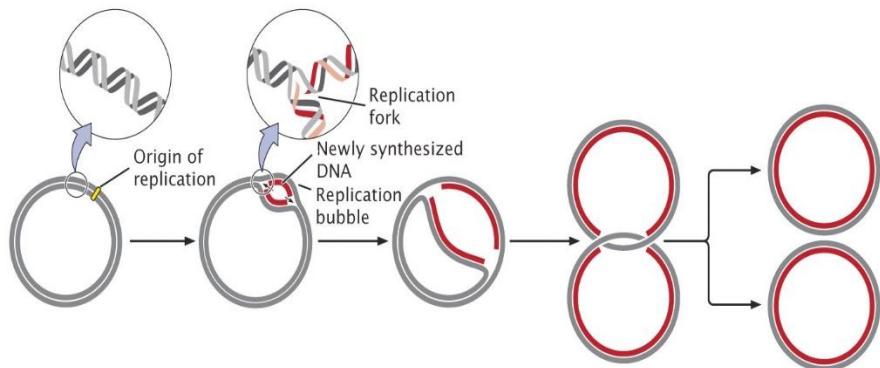
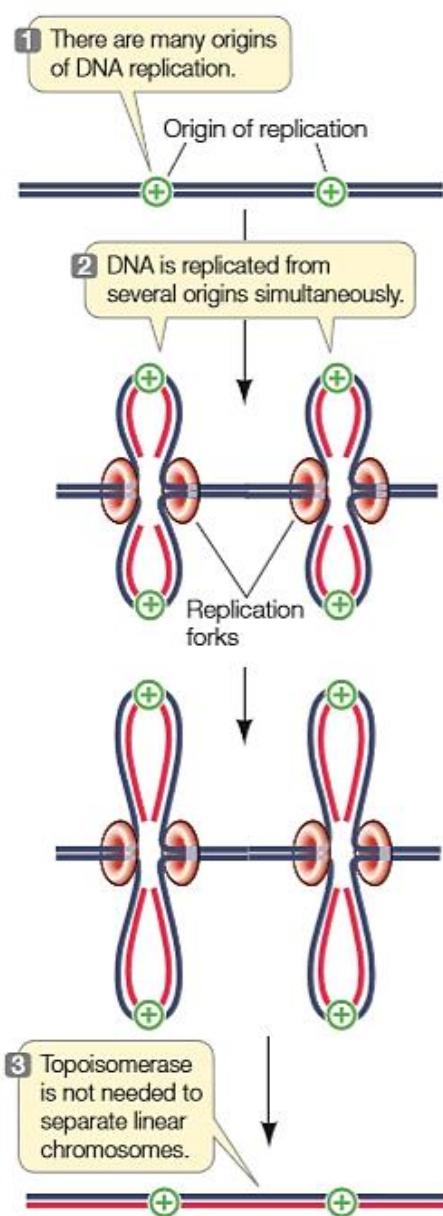


Figure 24- Replication in a bacteria having circular DNA.



Human DNA polymerases are slower than those of *E. coli*, and can replicate DNA at a rate of about 50 bases per second. Human chromosomes are much larger than those of bacteria (about 80 million base pairs) and linear. Large linear chromosomes such as those of humans contain hundreds of origins of replication. Numerous replication complexes bind to these sites at the same time and catalyze simultaneous replication. Thus there are many replication forks in eukaryotic DNA.

Figure 25- Larger linear chromosomes, typical of nuclear DNA in eukaryotes, have many origins of replication. Image courtesy- Sadava et al, *Life: The science of Biology*, 9th edition.

Activity of telomerase:

Replication of the lagging strand occurs by the addition of Okazaki fragments to RNA primers. When the terminal RNA primer is removed, no DNA can be synthesized to replace it because there is no 3' end to extend. So the new chromosome has a bit of single-stranded DNA at each end. This situation activates a mechanism for cutting off the single stranded region, along with some of the intact double-stranded DNA. This is a part of DNA repair mechanism. Thus the chromosome becomes slightly shorter with each cell division.

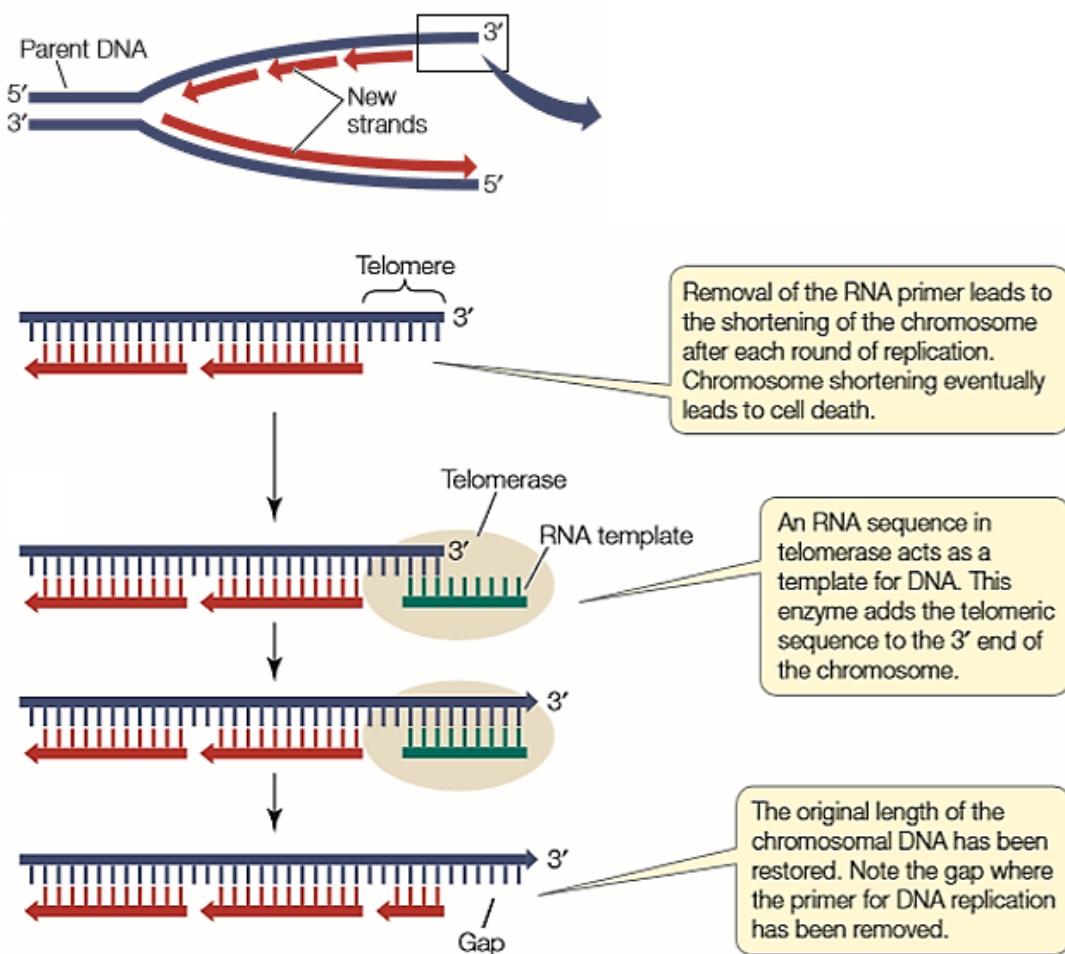


Figure 26- Removal of the RNA primer at the 3' end of the template for the lagging strand leaves a region of DNA—the telomere—unreplicated. In continuously dividing cells, the enzyme telomerase binds to the 3' end and extends the lagging strand of DNA, so the chromosome does not get shorter. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

An enzyme, appropriately called telomerase, catalyzes the addition of any lost telomeric sequences. Telomerase contains an RNA sequence that acts as a template for the telomeric DNA repeat sequence.

DNA Replication proofreading:

DNA must be accurately replicated and faithfully maintained. The price of failure can be great; the accurate transmission of genetic information is essential for the functioning and even the life of a single cell or multicellular organism. Yet the replication of DNA is not perfectly accurate. So the problem arises in preserving life.

DNA repair mechanisms help to preserve life. DNA polymerases initially make significant numbers of mistakes in assembling polynucleotide strands. Without DNA repair, the observed error rate of one for every 10^5 bases replicated would result in about 60,000 mutations every time a human cell divided. Fortunately, our cells can repair damaged nucleotides and DNA replication errors, so that very few errors end up in the replicated DNA.

Cells have at least three DNA repair mechanisms at their disposal:

- A **proofreading** mechanism corrects errors in replication as DNA polymerase makes them.
- A **mismatch repair** mechanism scans DNA immediately after it has been replicated and corrects any base-pairing mismatches.
- An excision repair mechanism removes abnormal bases that have formed because of chemical damage and replaces them with functional bases.

Most DNA polymerases perform a proofreading function each time they introduce a new nucleotide into a growing DNA strand. When a DNA polymerase recognizes a mispairing of bases, it removes the improperly introduced nucleotide and tries again. The error rate for this process is only about 1 in 10,000 repaired base pairs, and it lowers the overall error rate for replication to about one error in every 10^{10} bases replicated.

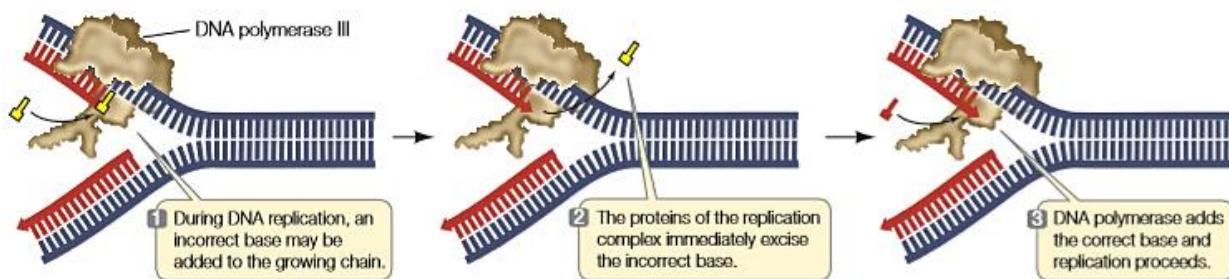


Figure 27- Proofreading activity during DNA replication. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

After the DNA has been replicated, a second set of proteins surveys the newly replicated molecule and looks for mismatched base pairs that were missed in proofreading. **Mismatch repair** mechanism might detect an A-C base pair instead of an A-T pair. The repair mechanism “knows” whether the A-C pair should be repaired by removing the C and replacing it with T or by removing the A and replacing it with G (it can detect the “wrong” base) because a DNA strand is chemically modified some time after replication. In prokaryotes, methyl groups ($-CH_3$) are added to some adenines. In eukaryotes, cytosine bases are methylated. Immediately after replication, methylation has not yet occurred on the newly replicated strand, so the new strand is “marked” (distinguished by being unmethylated) as the one in which errors

should be corrected. When mismatch repair fails, DNA sequences are altered. One form of colon cancer arises in part from a failure of mismatch repair.

Excision repair mechanisms deal with damage due to high-energy radiation, chemicals from the environment, and random spontaneous chemical reactions. Individuals who suffer from a condition known as xeroderma pigmentosum lack an excision repair mechanism that normally corrects the damage caused by ultraviolet radiation. They can develop skin cancers after even a brief exposure to sunlight.

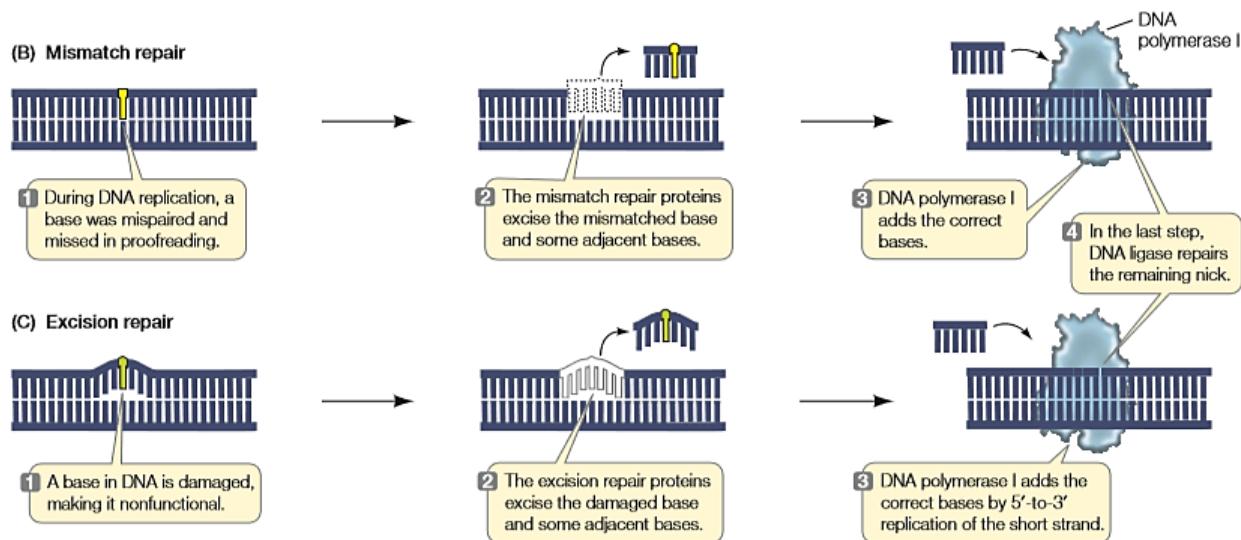


Figure 28- DNA repair mechanisms. Image courtesy- Sadava et al, *Life: The science of Biology*, 9th edition.

Transcription

RNA (ribonucleic acid) is a key intermediary between a DNA sequence and a polypeptide. RNA is an informational polynucleotide similar to DNA, but it differs from DNA in three ways:

- RNA generally consists of only one polynucleotide strand.
- The sugar molecule found in RNA is ribose, rather than the deoxyribose found in DNA.
- Although three of the nitrogenous bases (adenine, guanine, and cytosine) in RNA are identical to those in DNA, the fourth base in RNA is uracil (U), which is similar to thymine but lacks the methyl ($-CH_3$) group

In the process of RNA synthesis, the information contained in DNA is transcribed into RNA. During transcription, the information in a DNA sequence (a gene) is copied into a complementary RNA sequence. The process occurs in the nucleus and the resulting RNA is carried to the cytoplasm, where the protein synthesis occurs.

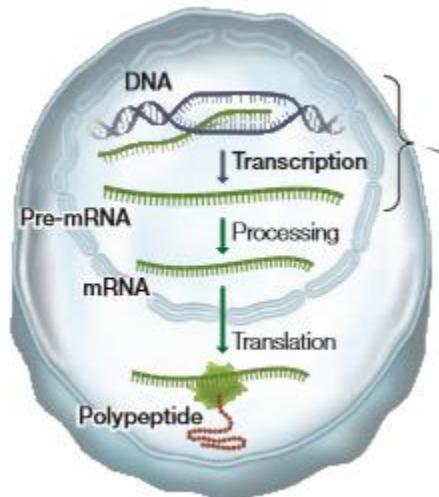


Figure 1- Sites of transcription and translation (protein synthesis) in a eukaryotic cell. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

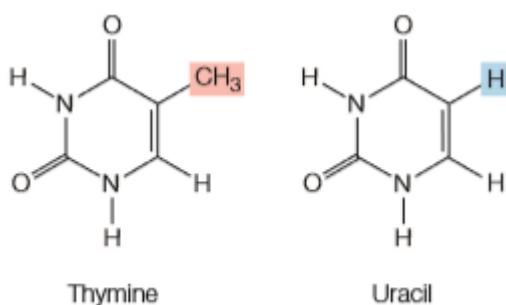


Figure 2- The nucleotide Thymine (present only in DNA) is replaced by Uracil (present only in RNA). The pairing of the ribonucleotides obeys the same complementary base-pairing rules as in DNA, except that adenine pairs with uracil instead of thymine. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

Single-stranded RNA can fold into complex shapes by internal base pairing. Three types of RNA participate in protein synthesis:

- Messenger RNA (mRNA) carries a copy of a gene sequence in DNA to the site of protein synthesis at the ribosome.
- Transfer RNA (tRNA) carries amino acids to the ribosome for assembly into polypeptides.
- Ribosomal RNA (rRNA) catalyzes peptide bond formation and provides a structural framework for the ribosome.

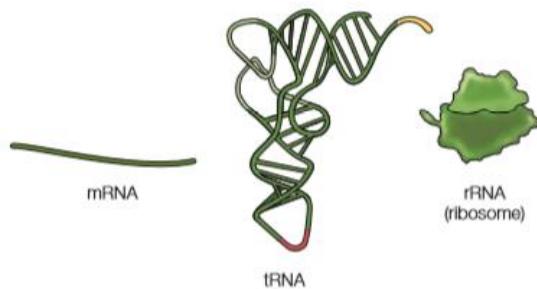


Figure 3- Types of RNA. Transcription is responsible for the synthesis of mRNA, tRNA and ribosomal RNA (rRNA), who play important roles in protein synthesis. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

Transcription requires several components:

- A DNA template for complementary base pairing; one of the two strands of DNA
- The appropriate nucleoside triphosphates (ATP, GTP, CTP, and UTP) to act as substrates
- An RNA polymerase enzyme- Like DNA polymerases, RNA polymerases are *processive*; that is, a single enzyme–template binding event results in the polymerization of hundreds of RNA bases. But unlike DNA polymerases, RNA polymerases do not require a primer and do not have a proofreading function.

Transcription occurs in three stages:

- a) initiation,
 - b) elongation, and
 - c) termination
-
- a) **INITIATION-** Transcription begins with initiation, which requires a promoter, a special sequence of DNA to which the RNA polymerase recognizes as a start site and binds very tightly. Eukaryotic genes generally have one promoter each, while in prokaryotes and viruses, several genes often share one promoter. Promoters are important control sequences that “tell” the RNA polymerase two things:

- Where to start transcription
- Which strand of DNA to transcribe (which of the two strands of DNA will act as template for producing a single RNA strand)

Part of each promoter is the initiation site, where transcription begins. Groups of nucleotides lying “upstream” from the initiation site (5' on the non-template strand, and 3' on the template strand) help the RNA polymerase bind.

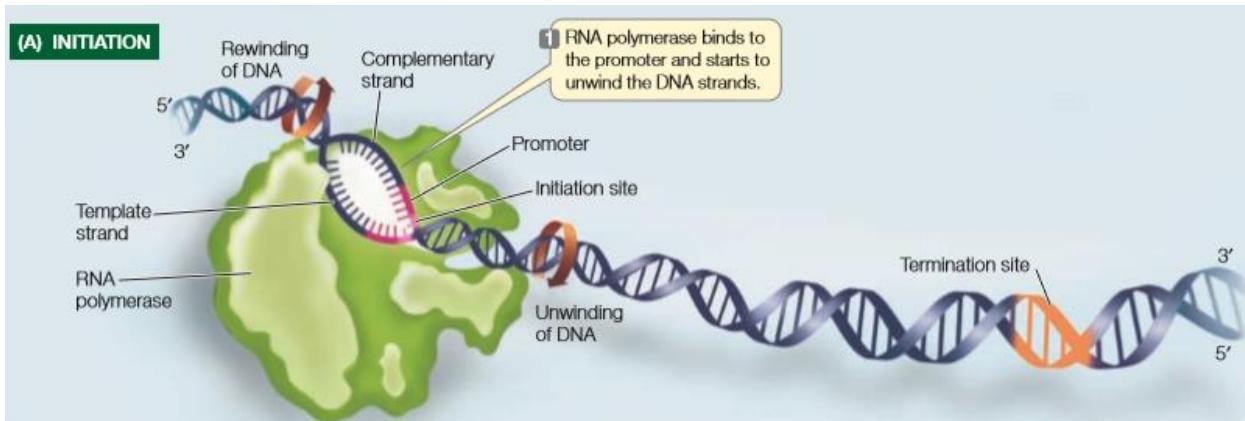


Figure 4- Initiation of RNA synthesis by RNA polymerase. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

- b) **ELONGATION-** Once RNA polymerase has bound to the promoter, it begins the process of elongation. RNA polymerase unwinds the DNA about 10 base pairs at a time and reads the template strand in the 3'-to-5' direction. Like DNA polymerase, RNA polymerase adds new nucleotides to the 3' end of the growing strand, but does not require a primer to get this process started. The RNA transcript produced is antiparallel to the DNA template strand.

Because RNA polymerases do not proofread, transcription errors occur at a rate of one for every 10^4 to 10^5 bases. Because many copies of RNA are made, however, and because they often have only a relatively short life span, these errors are not as potentially harmful as mutations in DNA.

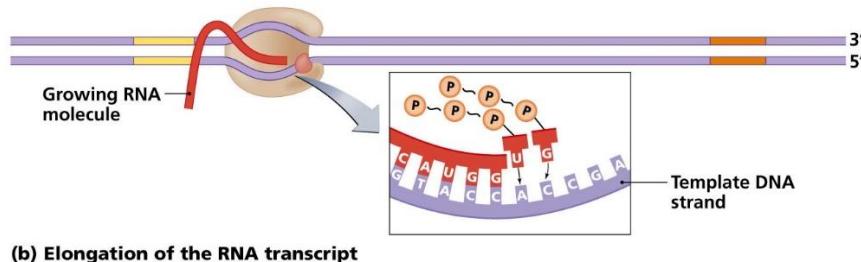


Figure 5- Successive addition of ribonucleotides to the 3' growing end of the newly synthesized RNA transcript.

- c) **TERMINATION**- Just as initiation sites in the DNA template strand specify the starting point for transcription, particular base sequences specify its termination. For some genes, the newly formed transcript falls away from the DNA template and the RNA polymerase. For others, a helper protein pulls the transcript away.

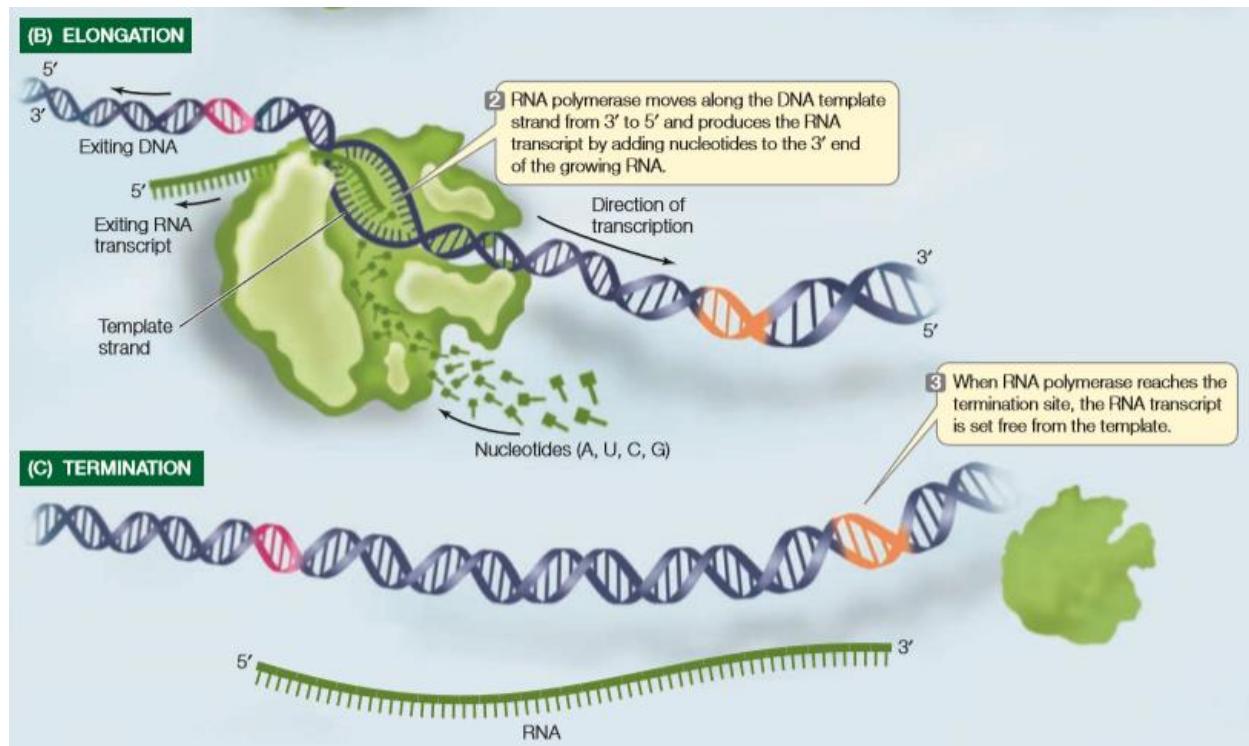


Figure 6- Process of elongation and termination in transcription of RNA. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

Difference in transcription between Prokaryotes and Eukaryotes

- **Initiation:**

In bacterial cells, the holoenzyme (RNA polymerase plus sigma) recognizes and binds directly to sequences in the promoter. In eukaryotic cells, promoter recognition is carried out by accessory proteins (transcription factors) that bind to the promoter and then recruit a specific RNA polymerase (I, II or III) to the promoter.

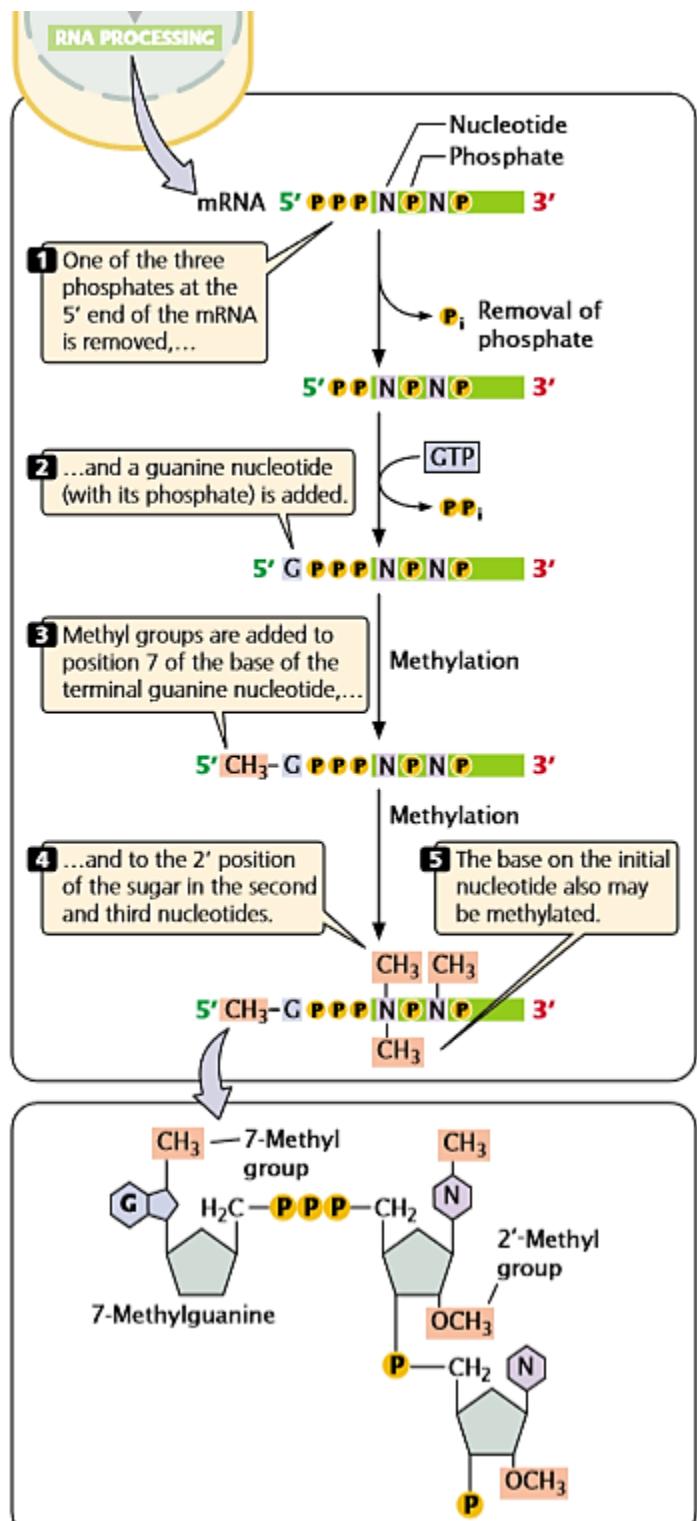
RNA processing:

In prokaryotes, several adjacent genes sometimes share one promoter; however, in eukaryotes, each gene has its own promoter, which usually precedes the coding region.

Eukaryotic genes undergo a systematic process called RNA processing to produce a mature mRNA from pre mRNA.

Eukaryotic genes may contain noncoding base sequences, called introns (intervening regions). One or more introns may be interspersed with the coding sequences, which are called exons (expressed regions). Both introns and exons appear in the primary mRNA transcript, called pre-mRNA, but the introns are removed by the time the mature mRNA—the mRNA that will be translated—leaves the nucleus (figure 1). Pre-mRNA processing involves cutting introns out of the pre-mRNA transcript and splicing together the remaining exon transcripts.

Eukaryotic gene transcripts are processed before translation: The primary transcript of a eukaryotic gene is modified in several ways before it leaves the nucleus: both ends of the pre mRNA are modified, and the introns are removed.



MODIFICATION AT BOTH ENDS Two steps in the processing of pre mRNA take place in the nucleus, one at each end of the molecule.

- A G cap is added to the 5' end of the pre-mRNA as it is transcribed. The G cap is chemically modified (methylated) guanosine triphosphate (GTP). It facilitates the binding of mRNA to the ribosome for translation, and it protects the mRNA from being digested by ribonucleases that break down RNAs.

Figure 7- Addition of 5' cap. Image courtesy- Genetics- : A Conceptual Approach; by Benjamin A. Pierce

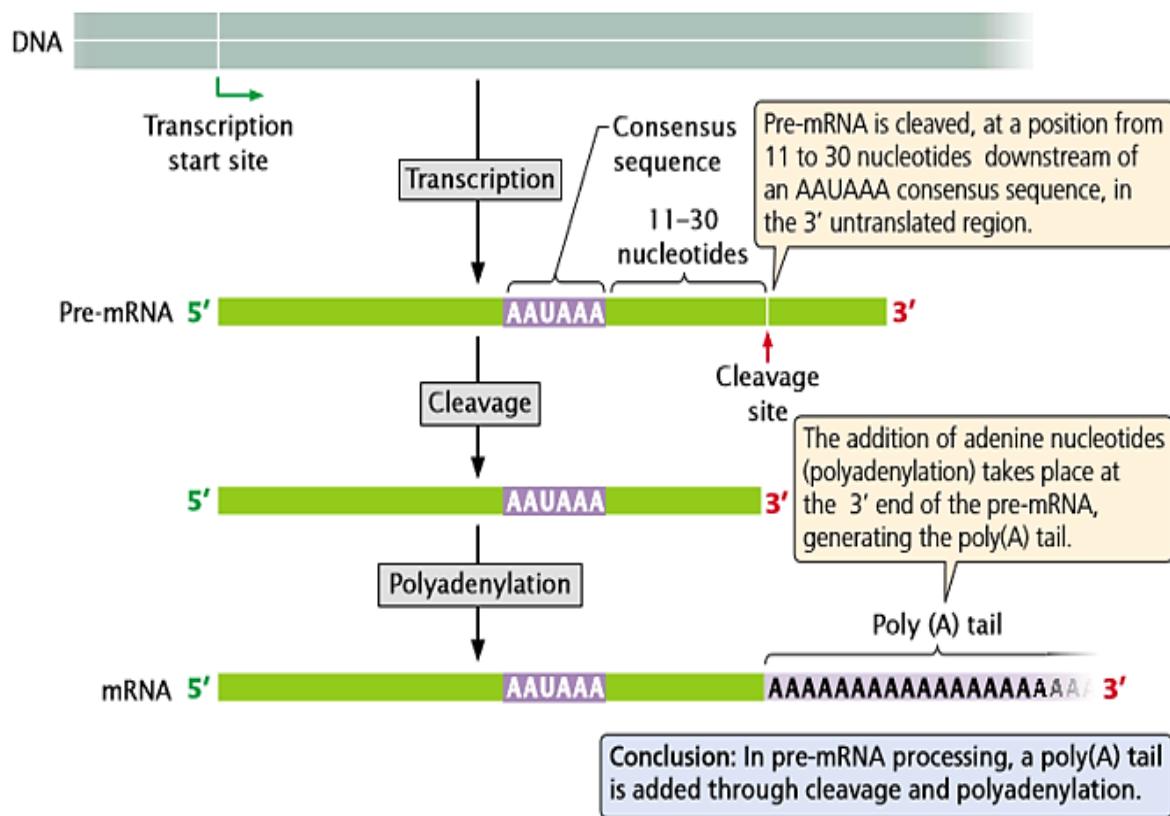
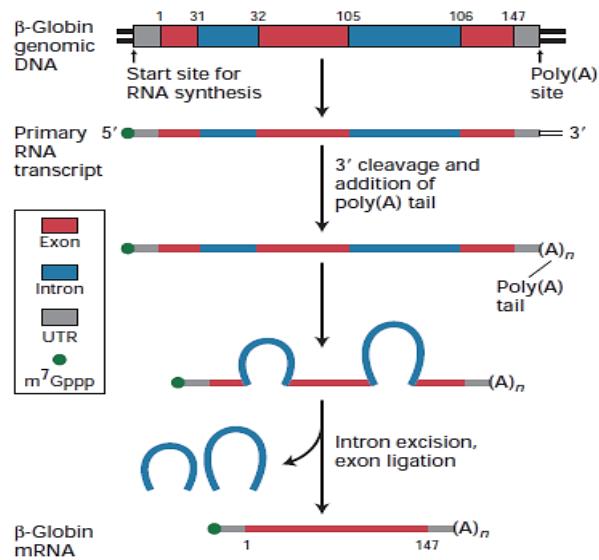


Figure 8- Addition of 50 to 250 adenine nucleotides at the 3' end is called Polyadenylation, which occurs after transcription is completed. Image courtesy- Genetics- : A Conceptual Approach; by Benjamin A. Pierce

- A poly A tail is added to the 3' end of the pre-mRNA at the end of transcription. In both prokaryotic and eukaryotic genes, transcription begins at a DNA sequence that is upstream (to the “left” on the DNA) of the first codon (i.e., at the promoter), and ends downstream (to the “right” on the DNA) of the termination codon. In eukaryotes, there is usually a “polyadenylation” sequence (AAUAAA) near the 3' end of the pre-mRNA, after the last codon. This sequence acts as a signal for an enzyme to cut the pre mRNA. Immediately after this cleavage, another enzyme adds 100 to 300 adenine nucleotides (a “poly A” sequence) to the 3' end of the pre-mRNA. This “tail” may assist in the export of the mRNA from the nucleus and is important for mRNA stability.

Splicing-



The next step in the processing of eukaryotic pre-mRNA within the nucleus is removal of the introns. If these RNA sequences were not removed, a very different amino acid sequence, and possibly a nonfunctional protein, would result. A process called RNA splicing removes the introns and splices the exons together.

Figure 9- The process of splicing beta globin gene. (UTR-untranslated region).

Alternative splicing

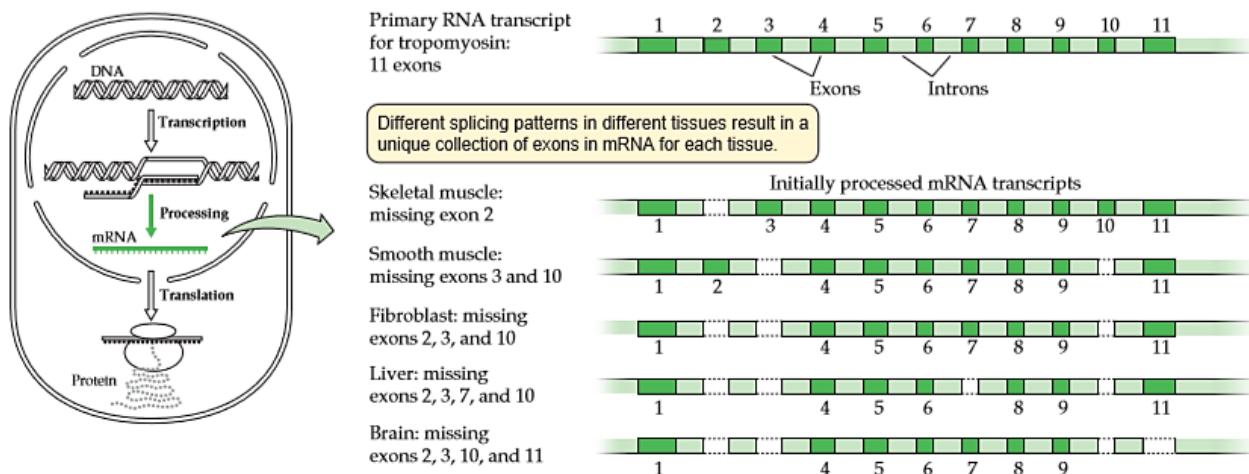


Figure 10- Alternative Splicing Results in Different mRNAs and Proteins In mammals, the protein tropomyosin is encoded by a gene that has 11 exons. Tropomyosin pre-mRNA is spliced differently in different tissues, resulting in five different forms of the protein.

Alternate splicing is a mechanism in which the differential splicing of the same pre-mRNA gives rise to different proteins, which may have different functions. It can be a deliberate mechanism for generating a family of different proteins from a single gene. For example, a single pre-mRNA for the structural protein tropomyosin is spliced differently in five different tissues to give five different mature mRNAs. These mRNAs are translated into the five different forms of tropomyosin found in these tissues: skeletal muscle, smooth muscle, fibroblast, liver, and brain.

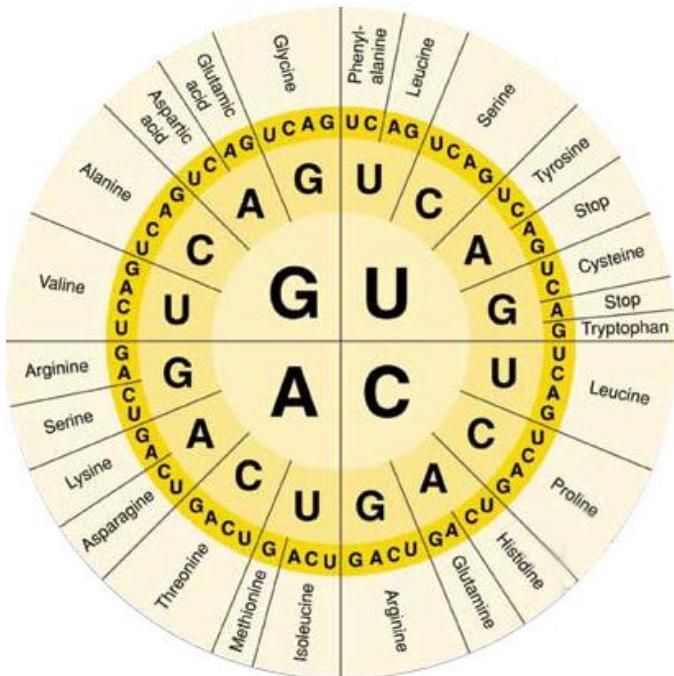
Translation

Genetic code

If genes are segments of DNA and if DNA is just a string of nucleotide pairs, then how does the sequence of nucleotide pairs dictate the sequence of amino acids in proteins?

Simple logic tells us that, if nucleotide pairs are the “letters” in a code, then a combination of letters can form “words” representing different amino acids. We must ask how the code is read. How many letters in the RNA make up a word, or codon, and which specific codon or codons represent each specific amino acid.

The logic is that the nucleotide code must be able to specify the placement of 20 amino acids. Since there are only four nucleotides, a code of single nucleotides would only represent four amino acids, such that A, C, G and U could be translated to encode amino acids. A doublet code could code for 16 amino acids (4×4). A triplet code could make a genetic code for 64 different combinations ($4 \times 4 \times 4$) genetic code and provide plenty of information in the DNA molecule to specify the placement of all 20 amino acids.



Also, this genetic code is redundant in nature. After the start and stop codons, the remaining 60 codons are far more than enough to code for the other 19 amino acids— and indeed there are repeats (Figure 1). Thus we say that the genetic code is redundant; that is, an amino acid may be represented by more than one codon.

Figure 1- Table of genetic code. Three letter code for amino acids.

Deciphering the genetic code:

In 1961, Marshall Nirenberg and Heinrich Matthaei mixed poly(U) with the protein synthesizing machinery of *E. coli* in vitro and observed the formation of a protein! The main excitement centered on the question of the amino acid sequence of this protein. It proved to be

polyphenylalanine—a string of phenylalanine molecules attached to form a polypeptide. This clearly meant that “words” consisting purely of U somehow caused the incorporation of phenylalanine.

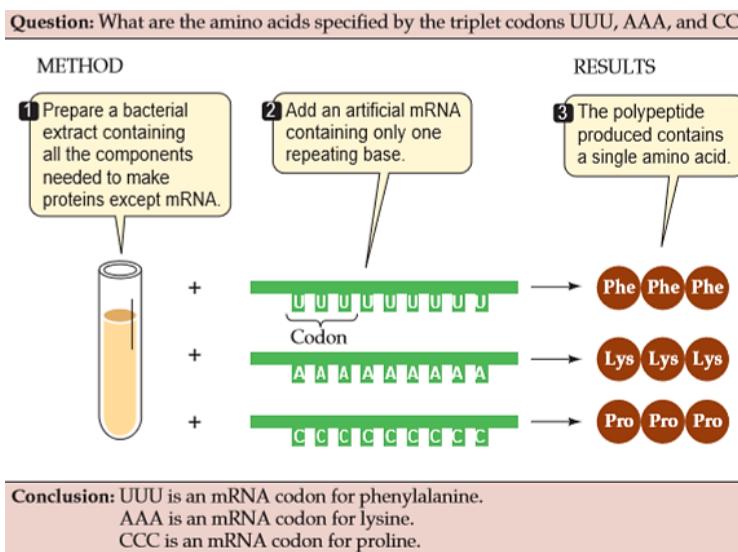


Figure 2- Nirenberg and Matthaei used a test-tube protein synthesis system to determine the amino acids specified by synthetic mRNAs of known codon composition. Image courtesy-Sadava et al, *Life: The science of Biology*, 7th edition.

decisively revealed the nature of the genetic code. He synthesized artificial messages more complex than Nierenberg's and analyzed the resulting polypeptides. His data are shown below. $(XY)_n$ means “XYXYXY ...”, and the resulting amino-acid couplet also repeats indefinitely (e.g., Ser-Leu-Ser-Leu-Ser-Leu ...).

Preparation for Translation: Linking RNAs, Amino Acids and Ribosomes

The translation of mRNA into proteins requires a molecule that links the information contained in mRNA codons with specific amino acids in proteins. That function is performed by tRNA. Two key events must take place to ensure that the protein made is the one specified by mRNA:

- tRNA must read mRNA correctly.
- tRNA must carry the amino acid that is correct for its reading of the mRNA.

Transfer RNAs carry specific amino acids and bind to specific codons

The codon in mRNA and the amino acid in a protein are related by way of an adapter—a specific tRNA with an attached amino acid. For each of the 20 amino acids, there is at least one specific type (species) of tRNA molecule.

The tRNA molecule has three functions:

It carries (“charged”) an amino acid, it associates with mRNA molecules, and it interacts with ribosomes.

At the 3' end of every tRNA molecule is a site to which its specific amino acid binds covalently. The charging of each tRNA with its correct amino acid is achieved by a family of activating enzymes, known more formally as aminoacyl-tRNA synthetases. At about the midpoint of tRNA is a group of three bases, called the anticodon that constitutes the site of complementary base pairing (hydrogen bonding) with mRNA. Each tRNA species has a unique anticodon, which is complementary to the mRNA codon for that tRNA's amino acid. At contact, the codon and the anticodon are antiparallel to each other. As an example of this process, consider the amino acid arginine:

- The DNA coding region for arginine is 3'-GCC-5', which is transcribed, by complementary base pairing, to the mRNA codon 5'-CGG-3'.
- That mRNA codon binds by complementary base pairing to a tRNA with the anticodon 3'-GCC-5', which is charged with arginine.

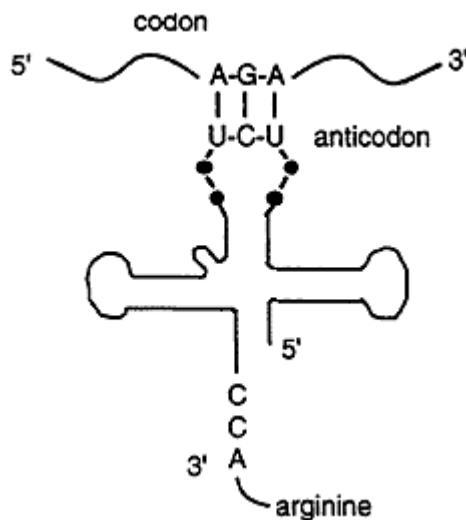


Figure 3- tRNA with a specific anticodon for arginine interacts with one arginine amino acid. This is called an activated tRNA. Activated tRNA specifically recognizes the codon on mRNA and produces hydrogen bonds at the site of anticodon-codon interaction. (Note the anticodon and codon are complementary to each other).

Ribosomes act as the workbench for translation:

Ribosomes are required for the translation of the genetic information in mRNA into a polypeptide chain. Each ribosome consists of two subunits, a large one and a small one. In eukaryotes, the large subunit consists of three different molecules of rRNA and about 45 different protein molecules, arranged in a precise pattern. The ribosomes of prokaryotes are somewhat smaller than those of eukaryotes, and their ribosomal proteins and RNAs are different.

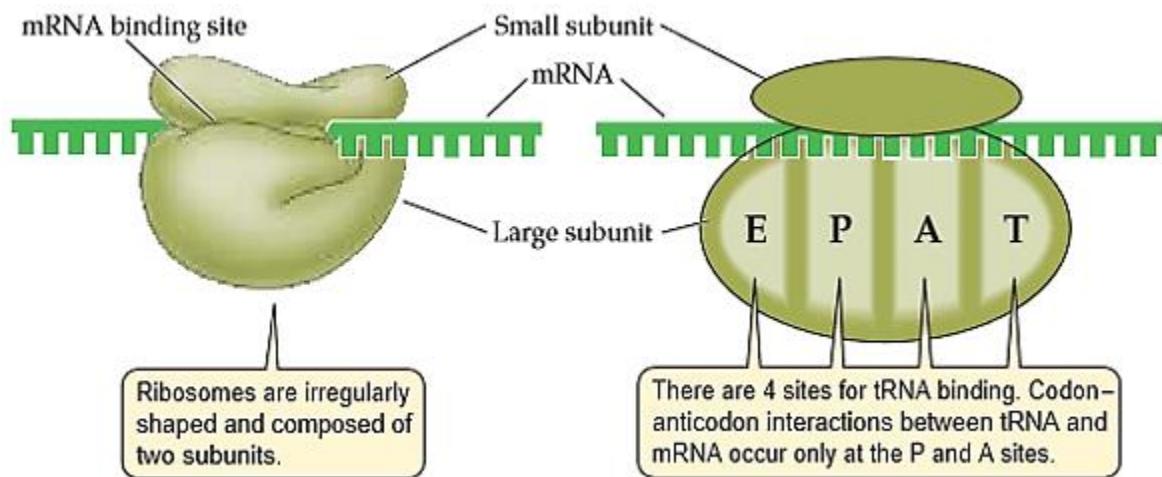


Figure 4- Ribosome Structure Each ribosome consists of a large and a small subunit. The subunits remain separate when they are not in use for protein synthesis. Its structure enables it to hold the mRNA and charged tRNAs in the right positions, thus allowing the growing polypeptide to be assembled efficiently. Image courtesy- Sadava et al, Life: The science of Biology, 7th edition.

A given ribosome does not specifically produce just one kind of protein. A ribosome can use any mRNA and all species of charged tRNAs, and thus can be used to make many different polypeptide products.

On the large subunit of the ribosome are four sites to which tRNA binds (Figure 4). A charged tRNA traverses these four sites:

- The T (transfer) site
- The A (amino acid) site
- The P (polypeptide) site
- The E (exit) site

An important role of the ribosome is to make sure that the mRNA-tRNA interactions are precise: that is, that a charged tRNA with the correct anticodon (e.g., 3'-UAC-5') binds to the appropriate codon in mRNA (e.g., 5'-AUG-3'). When this occurs, hydrogen bonds form between the base pairs. But these hydrogen bonds are not enough to hold the tRNA in place. The rRNA of the small ribosomal subunit plays a role in validating the three-base-pair match. If hydrogen bonds have not formed between all three base pairs, the tRNA must be the wrong one for that mRNA codon, and that tRNA is ejected from the ribosome.

Translation Process: RNA-Directed Polypeptide Synthesis

Like transcription, translation occurs in three steps: initiation, elongation, and termination.

Initiation:

The translation of mRNA begins with the formation of an initiation complex, which consists of a charged tRNA bearing what will be the first amino acid of the polypeptide chain and a small ribosomal subunit, both bound to the mRNA. The rRNA of the small ribosomal subunit binds to a complementary ribosome recognition sequence on the mRNA. This sequence is “upstream” (toward the 5’ end) of the actual start codon that begins translation.

The mRNA start codon in the genetic code is AUG. The anticodon of a methionine charged tRNA binds to this start codon by complementary base pairing to form the initiation complex. Thus the first amino acid in the chain is always methionine.

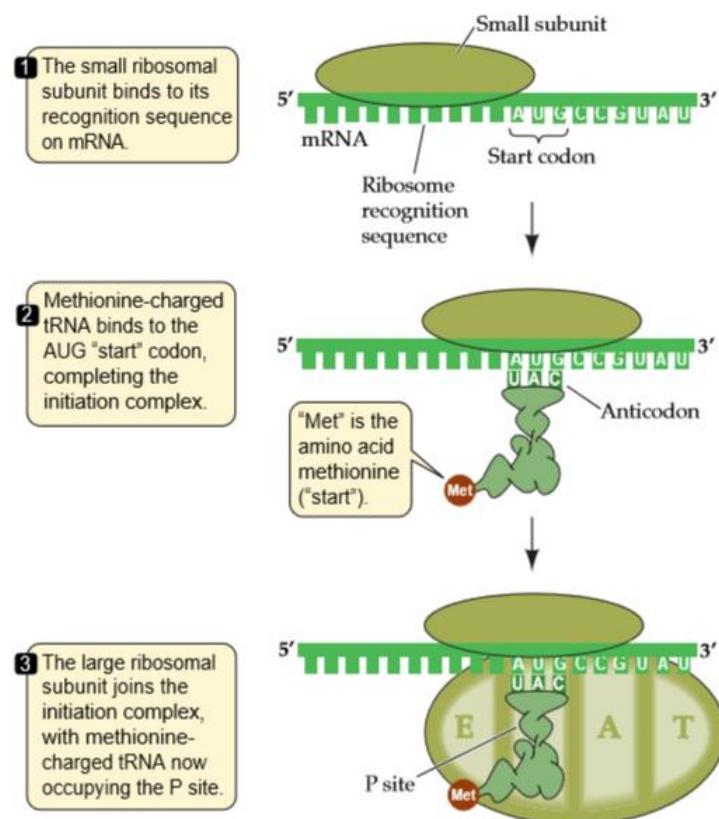


Figure 5- Initiation of translation begins with formation of an initiation complex (step 2). Image courtesy- Sadava et al, Life: The science of Biology, 7th edition.

Elongation:

The polypeptide elongates from the N terminus. A charged tRNA whose anticodon is complementary to the second codon on the mRNA now enters the open A site of the large ribosomal subunit. The large subunit then catalyzes two reactions:

- It breaks the bond between the tRNA in the P site and its amino acid.
- It catalyzes the formation of a peptide bond between that amino acid and the one attached to the tRNA in the A site.

Because the large subunit performs these two actions, it is said to have peptidyl transferase activity. In this way, methionine (the amino acid in the P site) becomes the N terminus of the new protein. The second amino acid is now bound to methionine, but remains attached to its tRNA by its carboxyl group ($-COOH$) in the A site.

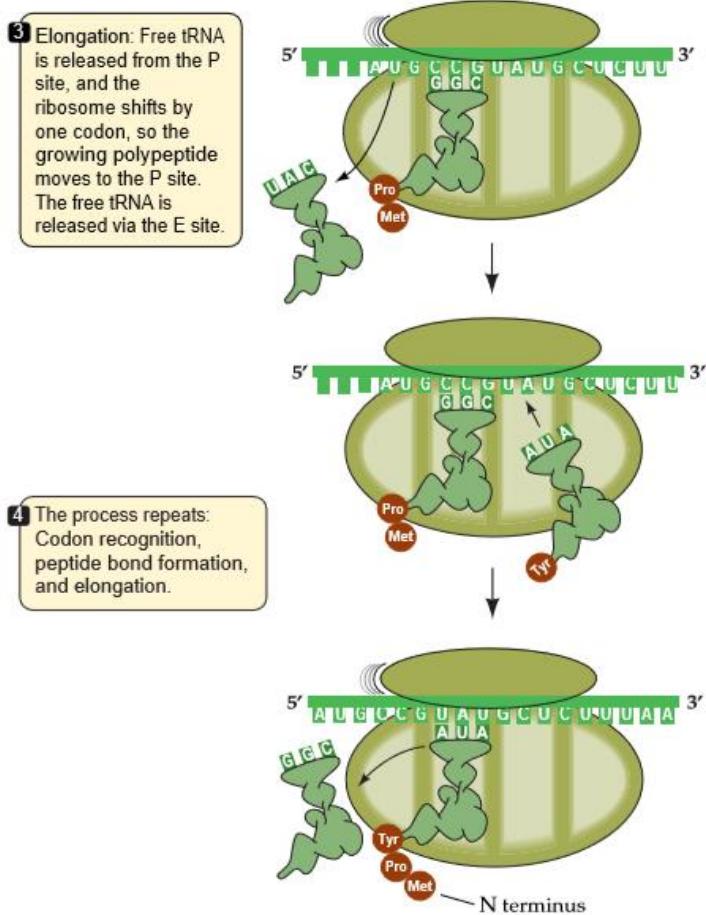


Figure 6- The incoming charged tRNA recognizes the specific codon on the mRNA and enters into the A site. Peptidyl transferase activity of the large subunit causes the transfer of amino acid on the first tRNA (at P site) on the second tRNA (at A site) and a peptide bond is formed between these amino acids. Image courtesy- Sadava et al, Life: The science of Biology, 7th edition.

After the first tRNA releases its methionine, it dissociates from the ribosome, returning to the cytosol to become charged with another methionine. The second tRNA, now bearing a dipeptide, is shifted to the P site as the ribosome moves one codon along the mRNA in the 5'-to-3' direction. The elongation process continues, and the polypeptide chain grows, as the steps are repeated. Elongation factors assist the elongation of a polypeptide chain.

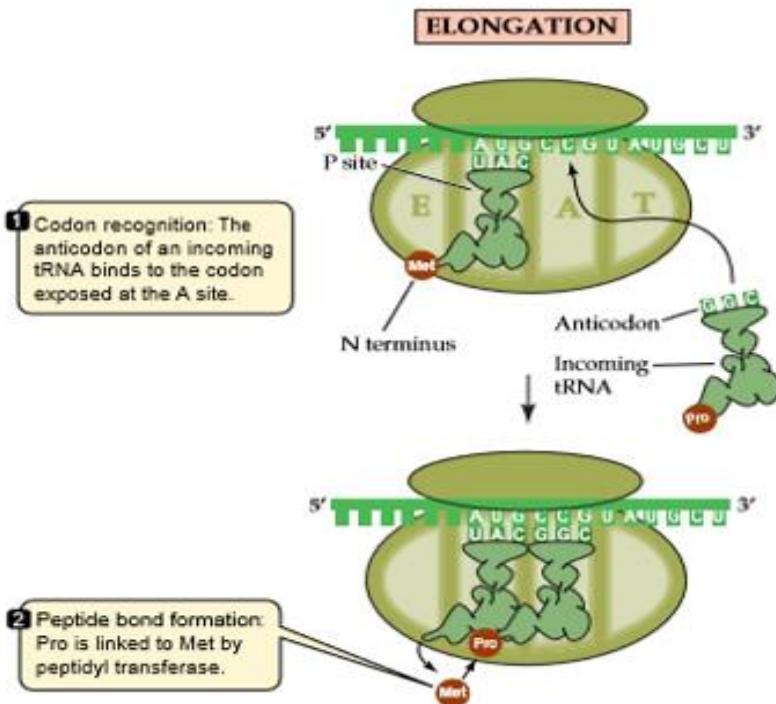


Figure 7- Elongation of polypeptide: (a) The next charged tRNA enters the open A site. (b) Its amino acid forms a peptide bond with the amino acid chain in the P site, so that it picks up the growing polypeptide chain from the tRNA in the P site. (c) The tRNA in the P site. Image courtesy- Sadava et al, Life: The science of Biology, 7th edition.

Termination:

The elongation cycle ends, and translation is terminated, when a stop codon—UAA, UAG, or UGA—enters the A site. These codons encode no amino acids, nor they bind tRNAs. Rather, they bind a protein release factor, which hydrolyzes the bond between the polypeptide and the tRNA in the P site. The newly completed protein thereupon separates from the ribosome. Its C terminus is the last amino acid to join the chain. Its N terminus, at least initially, is methionine, as a consequence of the AUG start codon. In its amino acid sequence, it contains information specifying its conformation, as well as its ultimate cellular destination.

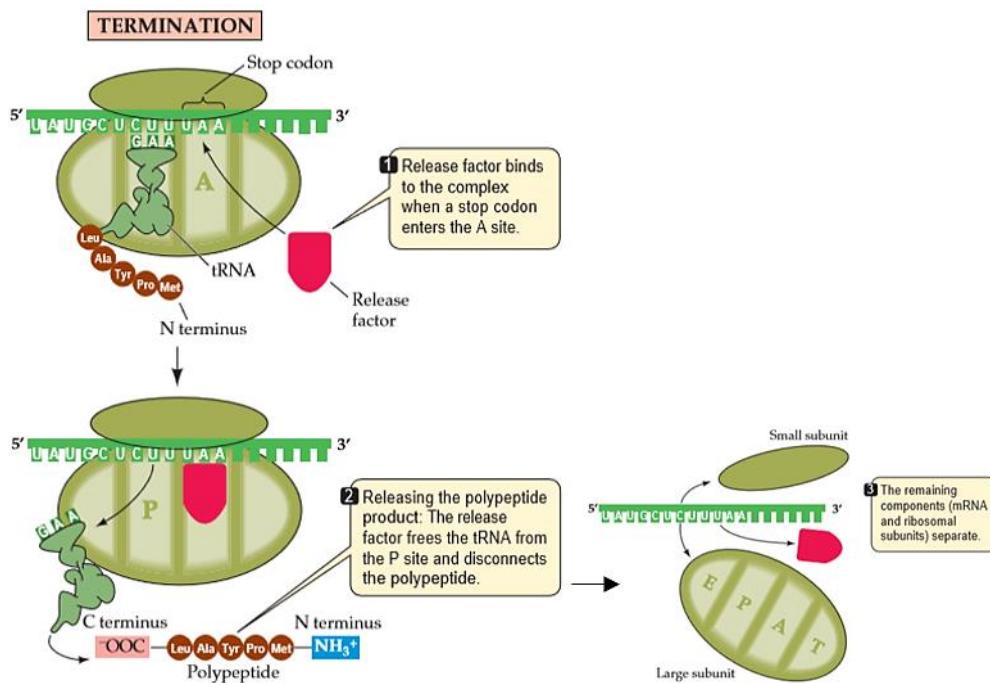


Figure 8- Events of termination of polypeptide chain. Image courtesy- Sadava et al, Life: The science of Biology, 7th edition.

	TRANSCRIPTION	TRANSLATION
Initiation	Promoter sequence in DNA	AUG start codon in mRNA
Termination	Terminator sequence in DNA	UAA, UAG, or UGA stop codon in mRNA

Figure 9- Signals that start and stop transcription and Translation. These act as a boundary for the mRNA and protein synthesis. Image courtesy- Sadava et al, Life: The science of Biology, 7th edition.

Evolution

Introduction

Evolution means change over a period of time. The idea that living beings may have evolved from simple compounds is a fascinating concept. Evolution is the process by which modern organisms have descended from ancient organisms. In other words, species are not constant, they change over time. The change observed so that the population is better adapted to their environment.

The study of evolution provides an insight into the investigation for nature of life, origins of life, diversity of various living beings and the similarities and differences in their structure and function.

The origin of life has different theories and these theories are uncertain. The major theories for the origin of life are:

1) Special creation

This theory is supported by most religions, civilizations. The basis of this theory is that life was created by a supernatural power at a particular time. While this theological approach concentrates on the reason behind creation of beings scientific theories concentrate on the how these beings came into existence. There is no intellectual conflict between scientific and theological theories as they are mutually exclusive realms of thought.

2) Spontaneous creation

The theory suggested that life arose from non-living matter. This theory was highly popular and coexisted with the special creation theory. Aristotle (384-322 BC) believed in this theory and said that life did not arise just from pre-existing parents but also from by spontaneous generation due to natural forces.

This theory fell from favor as more advancements in science. In 1688, Redi observed that the little worms which arise in decaying flesh were fly larvae. He supported the idea of biogenesis, which states that life can arise only from pre-existing life.

In 1765, Lazzaro Spallanzani observed that vegetables would not support growth of other life forms after intense heat treatment and sealing. Based on Spallanzani's work Louis Pasteur designed several experiments and finally disapproved the theory of spontaneous creation.

3) Cosmozoan

This theory does not suggest a mechanism for origin of life but favors the idea that life on earth has an extraterrestrial origin. In 1908, Arrhenius proposed the cosmozoan or panspermia theory. This theory assumed the existence of advanced civilization on other planets in our galaxy and life on earth and many other planets were infected from these advanced civilized planets.

4) Biochemical evolution

As per this theory, life arose as per the chemical and physical laws. In 1923, Oparin suggested that from the simple compound like nitrides, oxides, ammonia, methane many complex organic compounds were formed

gradually under the influence of electric charges, ultra-violet rays. The accumulation of the simple compounds in the oceans resulted in the production of the primeval soup from which life could have arisen.

In 1953, Stanley Miller a graduate student of Harold Urey designed an apparatus for stimulating condition prevalent on earth at the time of abiogenic evolution of organic substances. The apparatus has a spark chamber with two electrodes, a flask for boiling and a condenser. Miller used a mixture of methane ammonia, hydrogen and water. The mixture was exposed to electric discharges, following by condensation and then boiling. It was continued for 18 days. Miller was able to identify 15 amino acids, organic acid, ribose sugar and purine, adenine. Several theories were propounded to explain the evolution.

Lamarck's theory of evolution

Jean Baptiste Lamarck was a French naturalist who proposed a theory based on inheritance of acquired characteristics the offspring then adapt further, advancing evolution of the species.

He explained that the use and disuse of certain abilities led to the organism to gain or lose the ability. In support of Lamark's theory, some of the characters indeed passed from parents to offspring like development of strong biceps muscles in blacksmith, elongated body and loss of limbs in snakes due to continuous creeping through the holes and crevices, migration of both the eyes towards the upper side in flat fishes living on the bottom of sea, lengthening of neck in the giraffe due to its continuous use in reaching to the leaves and fruits of high rise tree. This theory put more importance on need of the animal and considers it strong enough to device ways to form organs needed for adaptation.

This theory was discredited as the use or disuse of all the abilities does not cause transfer abilities to next generation (eg. healthy parents need not always have healthy children). The experiments carried out by August Weismann proved to be a major criticism against Lamarkism. He had cut off the tails of rats for about 80 generations, but tailless offsprings were never born,

Theory of Natural selection:

In June 1831, the H.M.S. Beagle set sail from England and 22 year-old Charles Darwin took up an unpaid position on this voyage. Darwin had begun his studies as a medical student, then became a divinity student at Cambridge. But neither field excited him, much to his father's disappointment. Darwin became interested in geology and spent some time studying geology informally. After three years of surveying the South American coast, the Beagle reached San Cristobal (Chatham) in September 1835. The Beagle spent 5 weeks in the Galapagos.

The voyage helped Darwin to observe the variety of differences which occurred in the same species on these different islands. Also, he observed completely different species exclusive to some islands. Thus Darwin's greatest scientific contribution was that he could provide a logical insight into how and why evolution occurred. Charles Darwin propounded the theory of Natural selection. When there are differences in organism's abilities to survive and reproduce based upon inheritable traits natural selection takes place. As per Thomas Malthus, every generation in a species, more offspring are produced than actually survive due to limited resources. Survival of

any individual is not random, and it depends on hereditary factors. Those individuals with favorable inheritable traits will survive and reproduce. Those with less favorable inheritable traits will be eliminated. This will lead to a gradual change in the entire population, favorable hereditary variations accumulating over time and the species will change.

Darwin devoted 20 years in generating evidence to support his ideas. Wallace wrote to Darwin telling him his ideas on natural selection. This geared Darwin into publishing his ideas. Thomas Wallace and Darwin agreed to publish simultaneous papers. Darwin's book, *The Origin of Species*, was an immediate sensation.

Darwin has described his views on evolution as follows:

“Thus, from the war of nature, from famine and death, the most exalted object which we are capable of conceiving, namely, the production of higher animals, directly follows. There is grandeur in this view of life, with its several powers, having been originally breathed by the Creator into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being evolved.”

There are several observed cases of Natural selection.

1. Insects resistance to insecticides

Insecticides are sprayed on crops to protect them from attack of various insects. Some insects which are resistant to the insecticides, survive the use of insecticides on crops. Thus their progeny flourishes, while the others which are susceptible will be killed by the insecticides. This is a result of genetic variability in the population of insects. Those with the beneficial genetic makeup will survive and flourish.

2. Bacterial resistance to antibiotics

Antibiotics are targeted against bacteria. Some bacteria possess the antibiotic resistance gene in their plasmid. They can survive even in the presence of antibiotics. Bacteria have the capacity to transfer the plasmid to other bacteria by transformation. Therefore all the bacteria with the plasmid having antibiotic resistance gene survive in the presence of antibiotic.

3. Increased frequency of sickle cell anemia in Africans.

The protein in red blood cells (RBCs) that transports oxygen from the lung to metabolically active tissues, like muscle, where it is needed. The discovery of haemoglobin S (HbS) by Linus Pauling and colleagues in 1949 was the first demonstration that the production of an abnormal protein could be the cause of a genetic disorder. In 1956, Vernon Ingram identified the abnormality in the amino acid sequence of the β -globin chain (β 6Glu \rightarrow Val). This abnormality resulted in the normal concave cells gaining a sickled appearance.

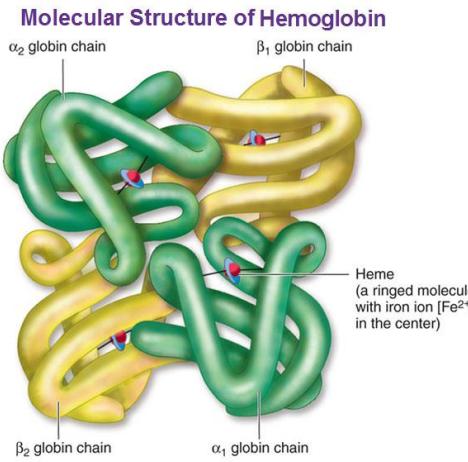


Figure 1- The structure of hemoglobin protein in the RBCs. Sickle cell anemia is produced due to mutation event in the beta chain of hemoglobin.

ANEMIA

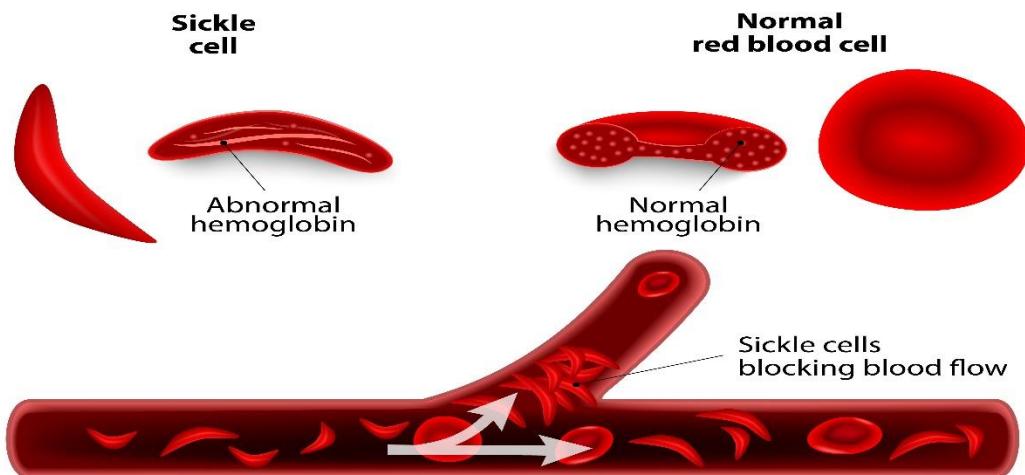


Figure 2- The abnormal cells of RBC as seen in Sickle cell Disease (SCD) compared with the normal RBCs. The sickle shape of the abnormal cells obstructs the blood flow and causes blood blockage in the thin capillaries causing extreme pain.

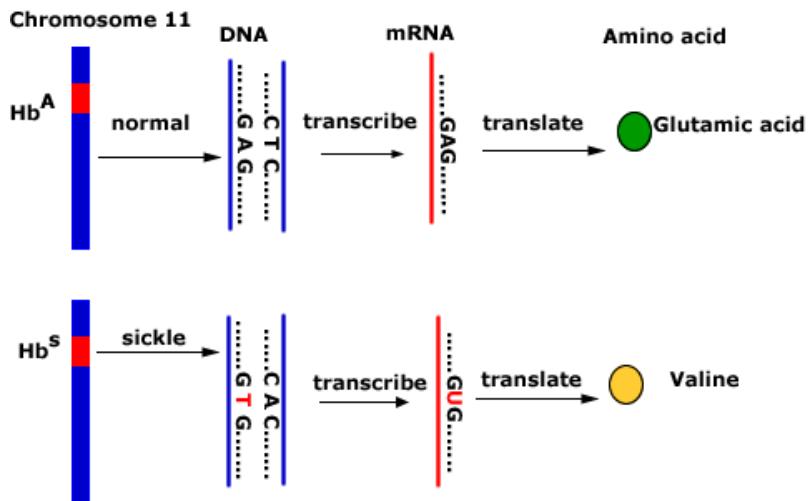


Figure 3- The gene for the beta chain of hemoglobin in normal RBCs codes for glutamic acid, which is hydrophilic in nature. The mutation in the gene coding for beta chain changes the codon on the mRNA to code for valine instead of glutamic acid. Valine is hydrophobic in nature.

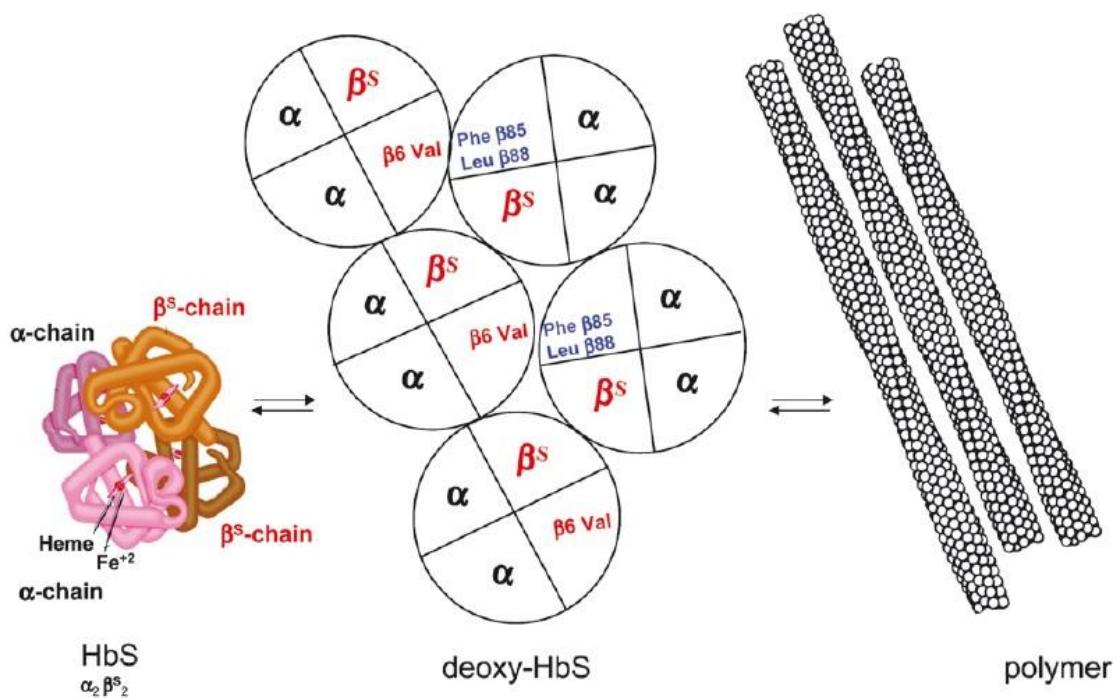


Figure 4-Basic pathophysiological mechanism of sickle cell disease: the polymerization of deoxy-HbS. The replacement of a glutamic acid by a valine residue at position 6 in the β -globin polypeptide chain characterizes the abnormal haemoglobin of SCD: HbS. The presence of hydrophobic valine in the beta chain acts as a hydrophobic pocket to which the other hydrophobic residues (phenylalanine and leucine) in the beta chain bind. At low oxygen pressure, deoxy-HbS polymerises and gets organised in long polymer fibres that deform, stiffen, and weaken the red blood cell

Inheritance of sickle cell anemia

The character of having a sickle cell is recessively inherited and follows the Mendelian pattern of inheritance for a recessive trait. The normal cell phenotype is dominant.

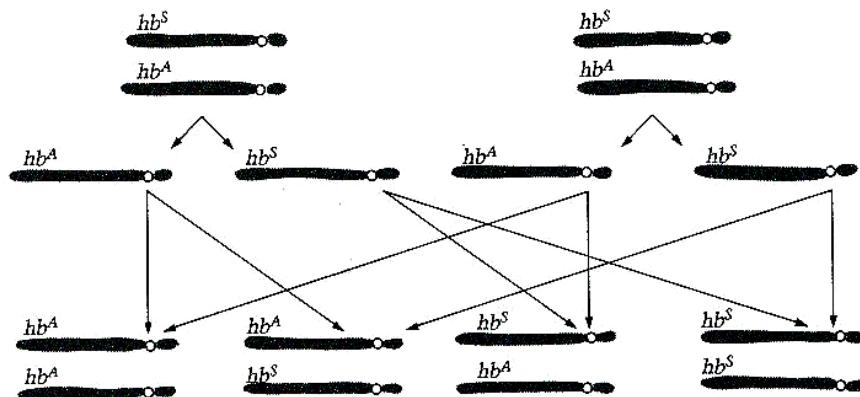


Figure 5- Inheritance of sickle cell disease. hb^A stands for normal RBC gene and hb^S stands for sickle shaped RBC gene. The recessive trait of a RBC being sickle shaped is not expressed phenotypically in a heterozygous individual. The phenotype of a sickle cell is expressed in a homozygous recessive individual. Thus the trait follows Mendelian inheritance pattern for a recessive character.

The sickle cell eventually bursts and dies. Under low oxygen conditions (high altitude or after rigorous exercise), a heterozygous individual for the sickle cell trait, RBCs start showing the sickled phenotype. These heterozygous individuals realize oxygen scarcity in their cells under such conditions as the sickling of cells reduces the oxygen carrying capacity of RBCs.

Sickle cell and resistance to malaria:

The reduced oxygen carrying capacity gives the heterozygous individual protection against malaria. Malaria pathogen completes a part of its early life cycle in the RBCs. The early stages of the malaria pathogen's development requires oxygen to complete its life cycle. Since in the heterozygous individuals the RBCs have a low capacity to carry oxygen, the malaria pathogen is not able to survive. Thus, the heterozygous and homozygous recessive individuals are protected from malaria.

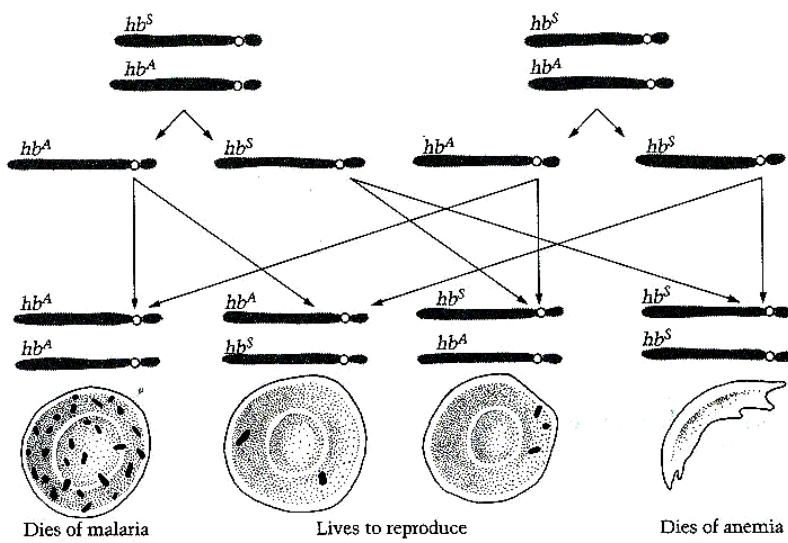


Figure 6- SCD and malaria resistance.

Evolution to protect against death by malaria:

The effect of natural selection is evident in the distribution of the sickle cell trait in areas where malaria is found to be indigenous. Thus, hinting that the process of evolution may have used the sickle cell trait as a protection against malaria, in these regions.

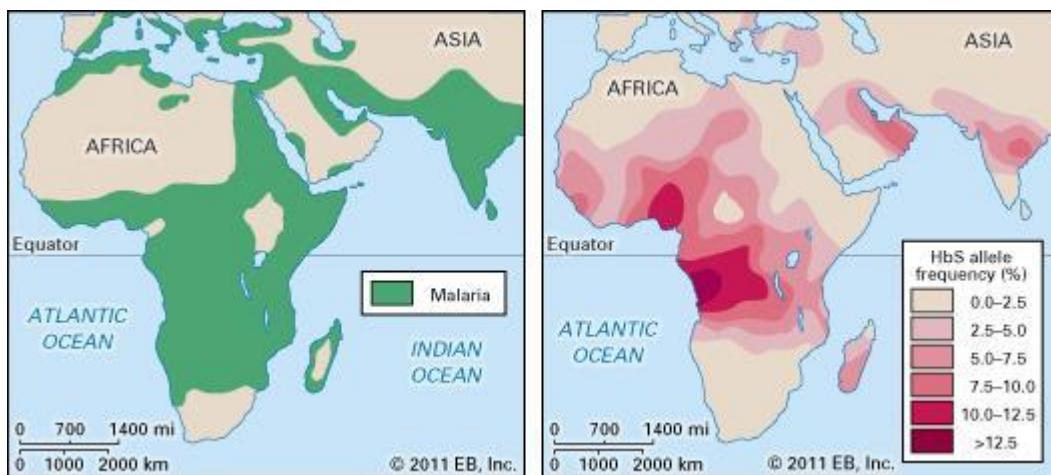


Figure 7- Geographic distribution of malaria and the appearance of sickle cell trait.

Darwin's theory of natural selection was had one major loophole. The theory could not explain the inheritance of traits from one generation to next. Existence of vestigial organs could not be explained using natural selection. Overspecialisation of some structures like antlers and tusks of elephants becomes a hindrance to these organisms. This fact that these structures that were hindrance to the organism being inherited could not be explained.

Importance of Population genetics:

Natural selection is understood on observing the changes which occur in a population rather than in an individual. Technically, evolution results from the change in gene frequencies within a population overtime. Therefore, in order to understand evolution, it would be important to describe those events that change gene frequencies in a population. Darwin could not explain how adaptive inheritable traits are passed on.

Biologists did not have a good understanding of the genetic details of how natural selection works until the field of transmission genetics was established in the early 1900s. At that time, the rediscovery of Gregor Mendel's publications paved the way for the development in the 1930s and 1940s of the field of population genetics. As the principles of evolution were integrated with the principles of modern genetics during this period, a new understanding of evolutionary biology—known as the Modern Synthesis—emerged. This was when biologists began to study mechanistic aspects of evolution as well as the broad evolutionary patterns that were so evident in nature.

-----1 hour

There are several evidence for evolution:

1. Fossil evidence: Fossils are the preserved remains of ancient organisms. The remains of the organisms can be found in preserved form in sap, mineral replacement, in ice, or traces e.g. footprints, molds. Fossils demonstrate the existence of intermediate forms of species, thus demonstrating evolution. The given figure shows how the ancient whales spent more time immersed, hence their nostrils where at the tip of the nose. Over the years, as they migrated to seas, the nostrils occupied a space higher on the skull. At present, whales can break the surface of sea because their nostrils are on the top beginning of the skull.

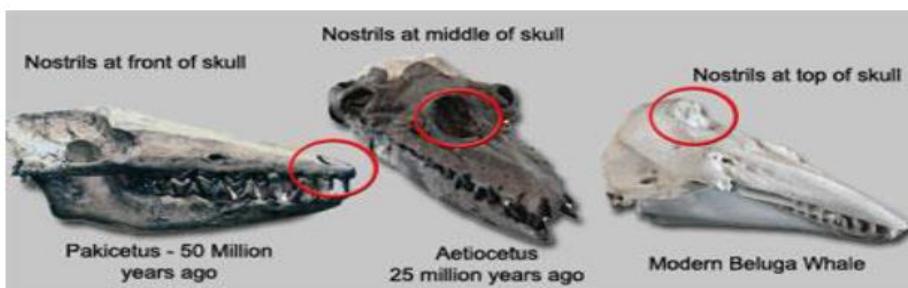


Figure 8- Ancient whales spent more time immersed, hence their nostrils where at the tip of the nose. Over the years, as they migrated to seas, the nostrils occupied a space higher on the skull. At present, whales can break the surface of sea because their nostrils are on the top

2. Embryonic evidence: The embryonic stages of various organisms share similar features for eg. the duck and chick embryo both have presence of webbed feet (Fig. 2). The later stages of development in the chick embryo causes the death of the layer in the interdigital zone since the chick has no use of webbed feet in its habitat.

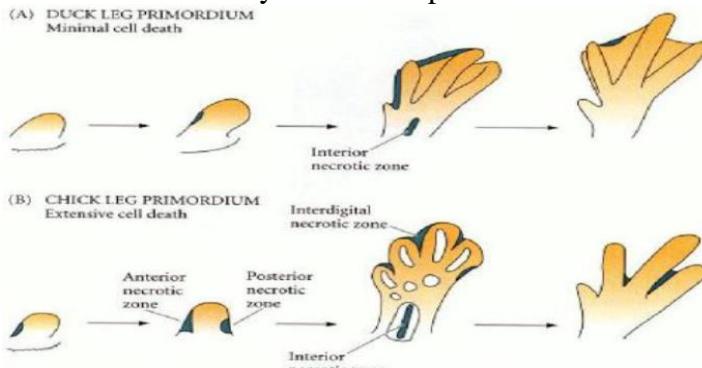


Figure 9- The embryo of duck and chick share a common structure initially. The molding of limbs in duck and chick embryo

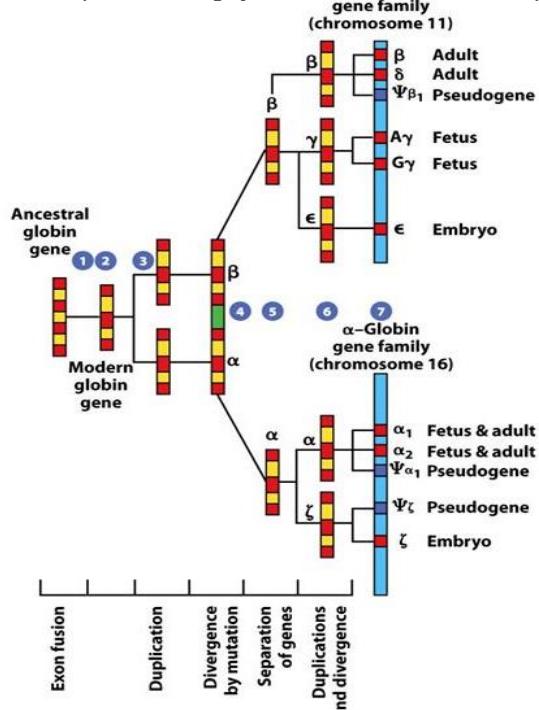


Figure 10-Evolution of globin gene

3. Genetic evidence: The DNA sequences in a gene family are usually different from one another. As long as at least one member encodes a functional protein, the other members may mutate in ways that change the functions of the proteins they encode. For evolution, the availability of multiple copies of a gene allows for selection of mutations that provide advantages under certain circumstances. If a mutated gene is useful, it may be selected for in succeeding generations. If the mutated gene is a total loss (pseudogene), the functional copy is still there to carry out its role. The presence of pseudogenes is an evidence of evolution in the gene. The gene family encoding the globins is a good example of the gene families found in vertebrates (Fig. 3). These proteins are found in hemoglobin and myoglobin (an oxygen-binding protein present in muscle). The globin genes all arose long ago from a single common ancestral gene. In adults, each hemoglobin molecule is a tetramer containing two identical α -globin subunits, two identical β -globin subunits, and four heme pigments. During human development, different members of the globin gene cluster are expressed at

different times and in different tissues. This differential gene expression has great physiological significance. Hemoglobin is a tetramer of four polypeptide units. A characteristic organization is realized while studying the globin genes from various organisms like mammal or fish, each of them contains three exons and two introns. Globin like polypeptides such as plant leghemoglobin and the muscle protein myoglobin, reveals the presence of four exons and three introns.

Billions of years ago, when the earth was devoid of life, there was also no oxygen in the air. The first life developed about 3.8 billion years ago, the water vapor, nitrogen, methane and ammonia were used for food and energy. It is expected that the metabolic reactions were catalyzed by metals such as iron and

magnesium. With the emergence of photosynthetic organisms around 3.5 billion years ago, the atmosphere of earth was remarkably changed. Oxygen produced as a result of photosynthesis was

released into the atmosphere and gradually came to be the most important constituent of atmosphere. Life forms appeared on earth utilizing this abundant oxygen on the atmosphere. The oxygen has to bind to a carrier and not supposed to react with that carrier. After binding it has to be transported to cells.

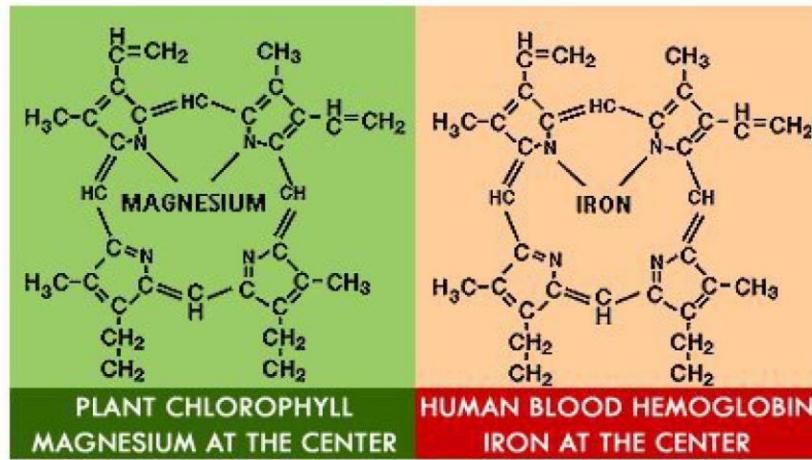


Figure 4- Comparison of Plant chlorophyll and hemoglobin structure (Image credit: <http://cascade.patch.com/groups/nicole-moseleys-blog/p/bp--hemoglobin-vs-chlorophyll>)

binds to oxygen in the lungs and gives blood its red color.

Chlorophyll, which is evolved earlier for photosynthesis, is a porphyrin ring containing magnesium. Another porphyrin ring containing iron is called heme. Heme bound to globin molecule is called hemoglobin. Hemoglobin

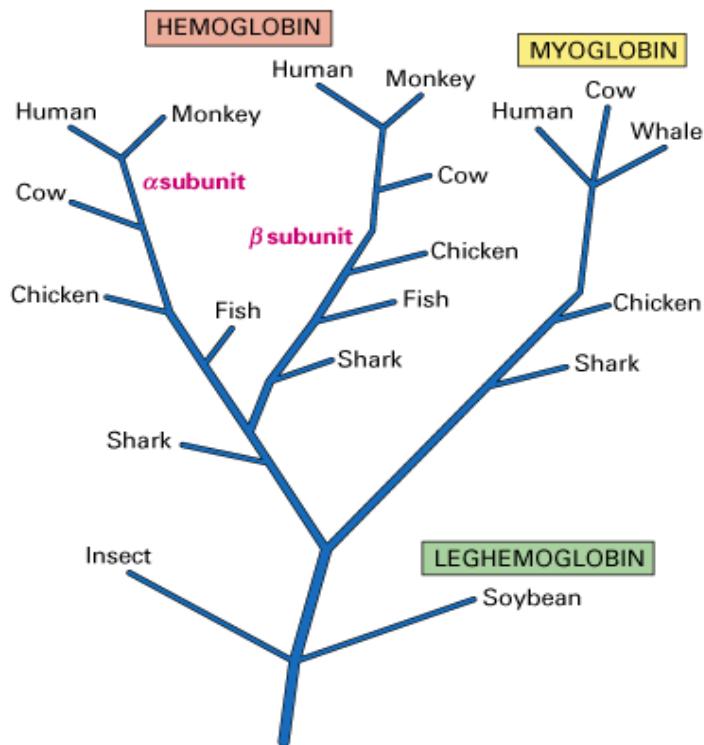


Figure 12-Phylogenetic tree showing the evolutionary history of globin protein

Myoglobin is a single polypeptide chain while vertebrate hemoglobins are tetrameric. Both are oxygen binding proteins. Leghemoglobin is also bind to oxygen, but found exclusively in root nodules of legume plants like soyabean.

We need to understand how a small change in a molecule has led to the diverse functions. For this we need to examine molecules such as haemoglobin, myoglobin, haemocyanin, leghaemoglobin and chlorophyll. All of them are modified porphyrins.

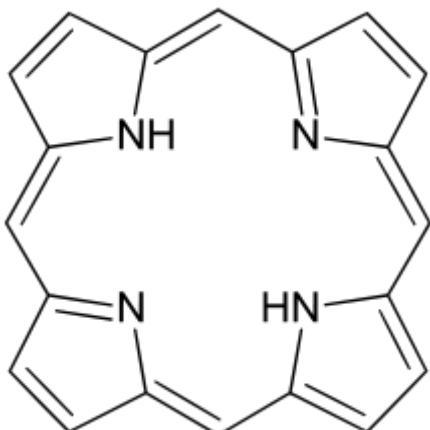


Figure 13-Porphyrin structure

They can accommodate different metal atoms at their centers. They are very stable due to their *conjugation* (alternating single and double bonds). Nature has utilized the properties of porphyrin ring in different ways for multiple functions.

Let us examine haemoglobin. As all of us know it is the oxygen carrying molecule of blood. O_2 is only marginally soluble ($< 0.0001\text{ M}$) in blood plasma at physiological pH. Hence by relying into the dissolved oxygen in the blood may not be feasible for the life systems to move on.

Therefore the evolution has occurred for the better transport of oxygen molecule. Animals requires the development of a mechanism to actively transport oxygen through the system. We have about 150 g of the protein which is known as haemoglobin (Hb) per liter of the blood. This is an effective oxygen carrier that the concentration of O_2 in the blood stream reaches 0.01 M — the same concentration as air. Once the Hb- O_2 complex reaches the tissue that consumes oxygen, the O_2 molecules are transferred to another protein — **myoglobin** (Mb) — which transports oxygen through the muscle tissue.

Now let us logically examine the haemoglobin molecule. Each haemoglobin is a combination of four polypeptide chains: two alpha and two beta subunits. Each chain has an associated heme group which can bind to oxygen. Thus one Hb molecule binds with four oxygen molecules. The structure of heme group is given below.

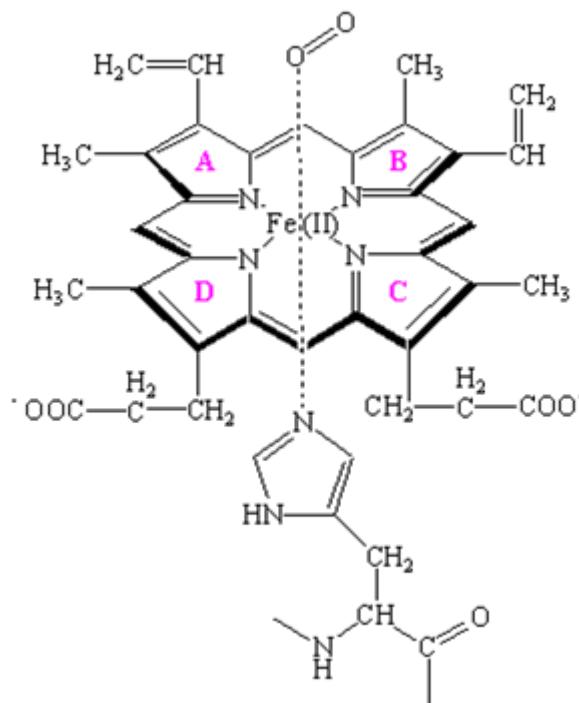


Figure 14-Myoglobin molecule

An Fe(II) atom is present at the centre of each heme. Four of the six coordination sites around this atom are occupied by nitrogen atoms from a planar **porphyrin** ring. The fifth coordination site is occupied by a nitrogen atom from a histidine side chain on one of the amino acids in the protein. The last coordination site is available to bind an O_2 molecule. Now we need to analyze why four protein parts? The distance between the iron atoms of adjacent hemes in hemoglobin is very large — between 250 and 370 nm — the act of binding an O_2 molecule at one of the four hemes in hemoglobin leads to a significant increase in the affinity for O_2 binding at the other hemes. Thus a cooperative interaction occurs that makes Hb an excellent carrier for oxygen.

The structure of myoglobin suggests that the oxygen-carrying heme group is buried inside the protein portion of this molecule (similar to Hb). This enables the pairs of hemes group from coming too close together. This is important, because these

proteins need to bind O₂ reversibly and the Fe(II) heme, by itself, cannot do this. When there is no globin to protect the heme, it reacts with oxygen to form an oxidized Fe(III) atom instead of an Fe(II)-O₂ complex.

What made the necessity of two oxygen binding molecules? Hb has four haeme group while myoglobin carries only one heme group. The affinity of Mb with oxygen is higher than the affinity of Hb with oxygen. This difference is related to its different role: whereas hemoglobin transports oxygen, myoglobin's function is to store oxygen.

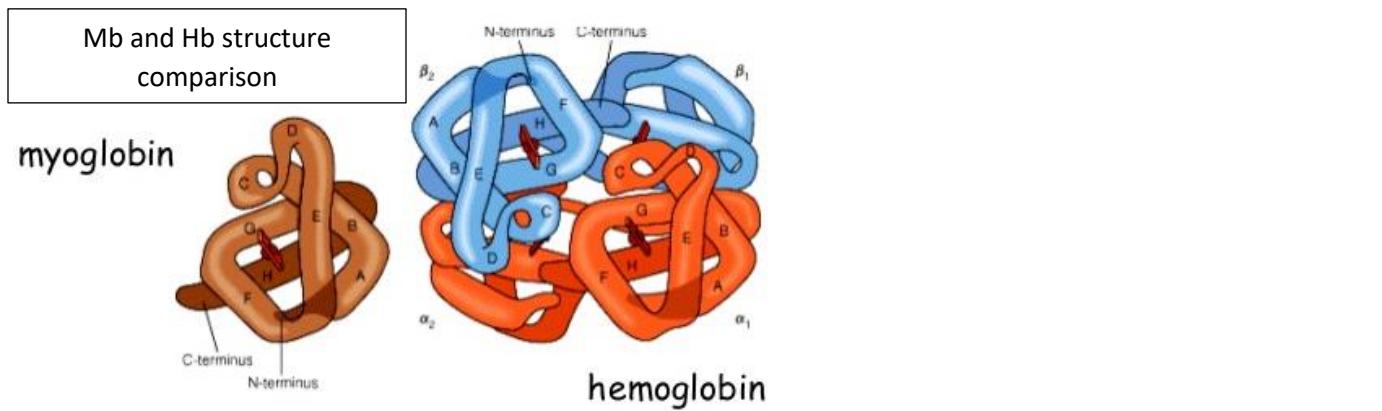


Figure 15-Mb and Hb subunits are structurally similar with 8 alpha helices and both containing a heme group. But Mb is monomeric while Hb is heterotetramer (2 alpha and 2 beta subunits)

Being large Hb can't enter muscle tissue. Further to prevent from clotting, blood has to move. Hence Hb is not an ideal structure for delivering oxygen in muscle tissues and its storing. Since myoglobin is a storage protein, its affinity towards oxygen should not differ with its concentration. Hence evolution has favored a similar molecule, a monomeric protein which is more compact. Thus the life systems have utilized the heme group for two different functions (i) for transporting oxygen and (ii) for storing oxygen. This has achieved by combining the heme group with different proteins ie the globin part.

Also it is interesting to see, why there is a requirement of a protein part ie amino acids attached with heme group. When molecular oxygen encounters an isolated heme molecule, it rapidly converts the Fe(II) to Fe(III). The oxidized heme binds oxygen very poorly. Obviously, if this happened to the Fe(II) groups of hemoglobin and myoglobin, the proteins would be less useful as oxygen carriers. Oxidation of the heme iron is prevented by the presence of the **distal histidine** side chain, which prevents the O₂ from forming a linear Fe-O-O bond. The bond between Fe and O₂ is bent, meaning that this bond is not as strong as it might be. *Weaker oxygen binding means easier oxygen release.*

Chlorophyll

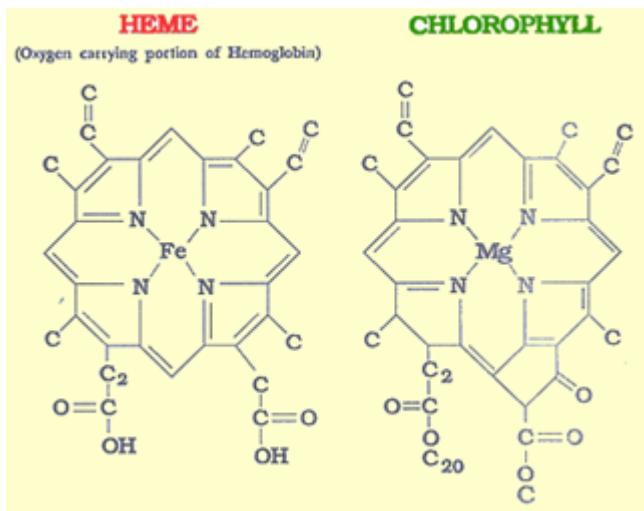


Figure 16-Structure comparison of heme and chlorophyll

Chlorophyll is the molecule that traps this 'most elusive of all powers' - and is called a photoreceptor. It is found in the chloroplasts of green plants, and is what makes green plants, green. The basic structure of a chlorophyll molecule is a porphyrin ring (similar to haeme) coordinated to a central atom. This is very similar in structure to the heme group found in hemoglobin, except that in heme the central atom is iron, whereas in chlorophyll it is magnesium.

Haemocyanin

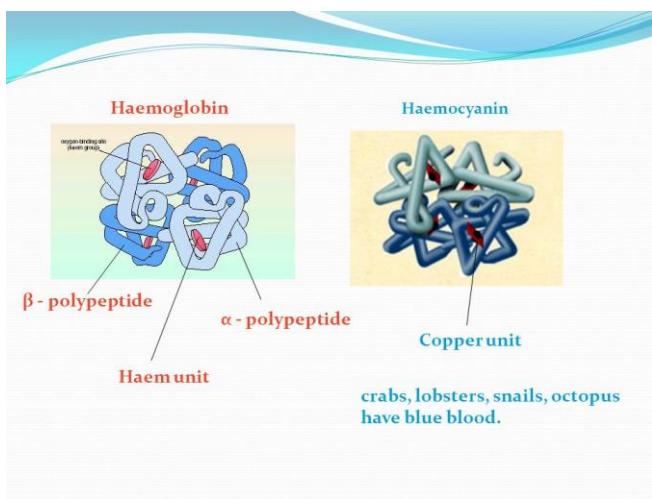


Figure 17- Structural comparison of haemoglobin and haemocyanin

Haemocyanin is another evolved compound from the porphyrin ring. Instead of Fe, haemocyanin has Cu in it. This also serves as an oxygen carrier in some animals.

Leghaemoglobin

This is the oxygen carrier found in the root nodules of legumes. It is synthesized in a symbiotic interaction with nitrogen fixing bacteria in the root nodules. Similar to haemoglobin it is red in color and carries oxygen for the growing bacteria in the root nodule.

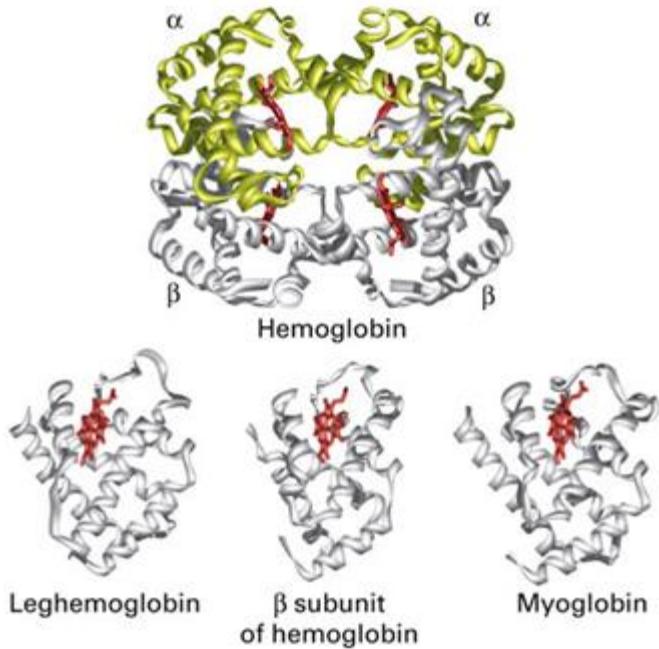


Figure 18- Structural comparison of hemoglobin, leghemoglobin, myoglobin and beta subunit of hemoglobin

Now it is easy for us to understand that among the molecules discussed such as haemoglobin, myoglobin, haemocyanin and leghaemoglobin, there is a similarity in the structure. Just by modifying a small part such as replacement of the central atom (Fe /Cu/Mg) of the porphyrin ring these molecules are modified (evolved) to perform diverse functions. This is nature's chemical engineering and are products of evolution.

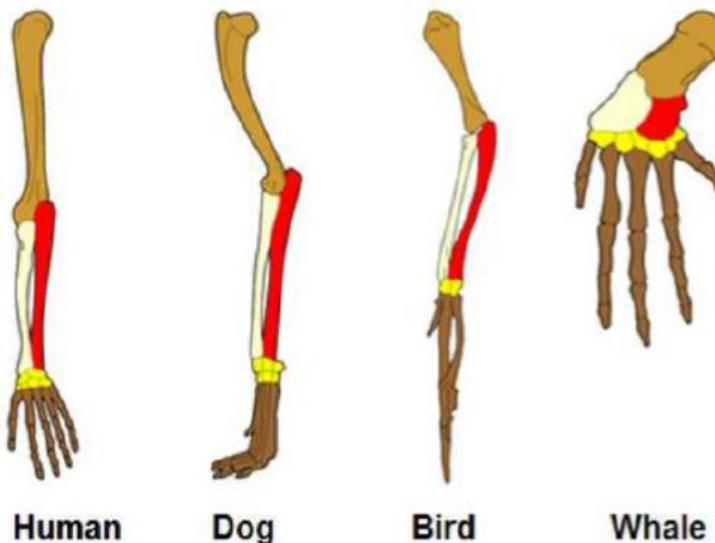


Figure 19- The forelimbs of vertebrates evolved from an ancestral pattern. Even vertebrates as dissimilar as whales and bats have the same basic arrangement of bones. The shades are based on homologous bones.

4. Form and function evidence: The **function** of a biological structure can be **inferred** from its **shape**. If we take our own bodies to illustrate, we know that ear is a canal with a funnel shaped external ear. We know that the external ear is funnel shaped for focusing sounds from the atmosphere. It is possible to explain that why many animals have more sound perception than humans based on the architecture of external ear. Vestigial organs are remains of a structure that was functional in some ancestor but is no longer functional in the organism in question. For example, humans have a tail bone (the coccyx) but no tail.

-----2 hours-----

The mechanisms responsible for evolution are:

1. Mutation

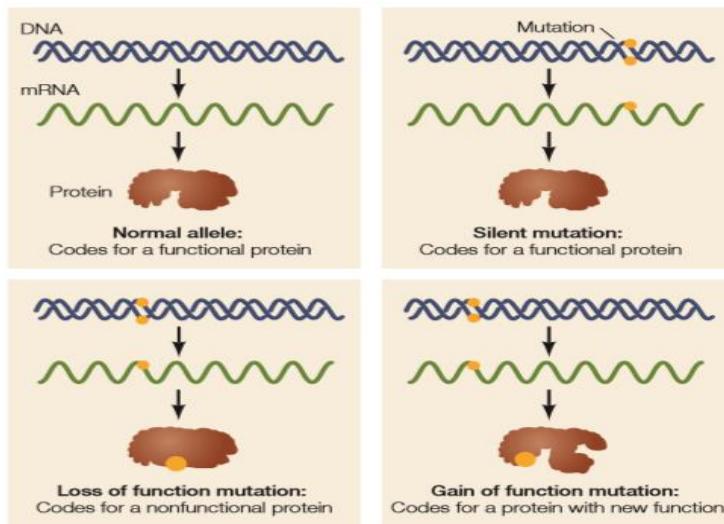


Figure 20- Mutations may or may not affect the protein phenotype. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

The origin of genetic variation is mutation. Accurate DNA replication, transcription, and translation all depend on the reliable pairing of complementary bases. Errors occur, though infrequently, in all three processes—least often in DNA replication. But, the consequences of DNA errors are the most severe because only they are heritable. Mutations are heritable changes in genetic information, which may occur spontaneously or may be induced.

Mutations have different phenotypic effects phenotypically, we can understand mutations in terms

of their effects on proteins and their function:

All mutations are alterations in the nucleotide sequence of DNA. At the molecular level, we can divide mutations into two categories:

- A **point mutation** results from the gain, loss, or substitution of a single nucleotide. After DNA replication, the altered nucleotide becomes a mutant base pair. If a point mutation occurs within a gene (rather than in a noncoding DNA sequence), then one allele of that gene (usually dominant) becomes another allele (usually recessive).

- **Chromosomal mutations** are more extensive than point mutations. They may change the position or

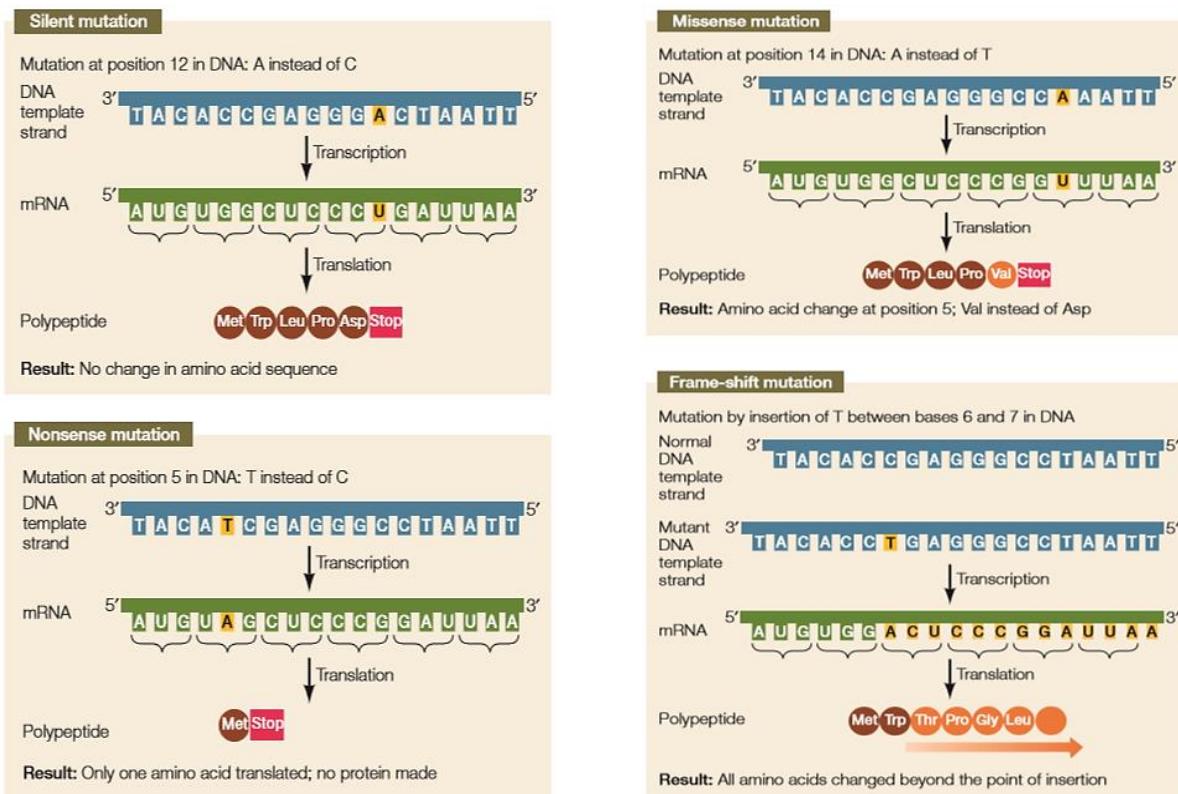
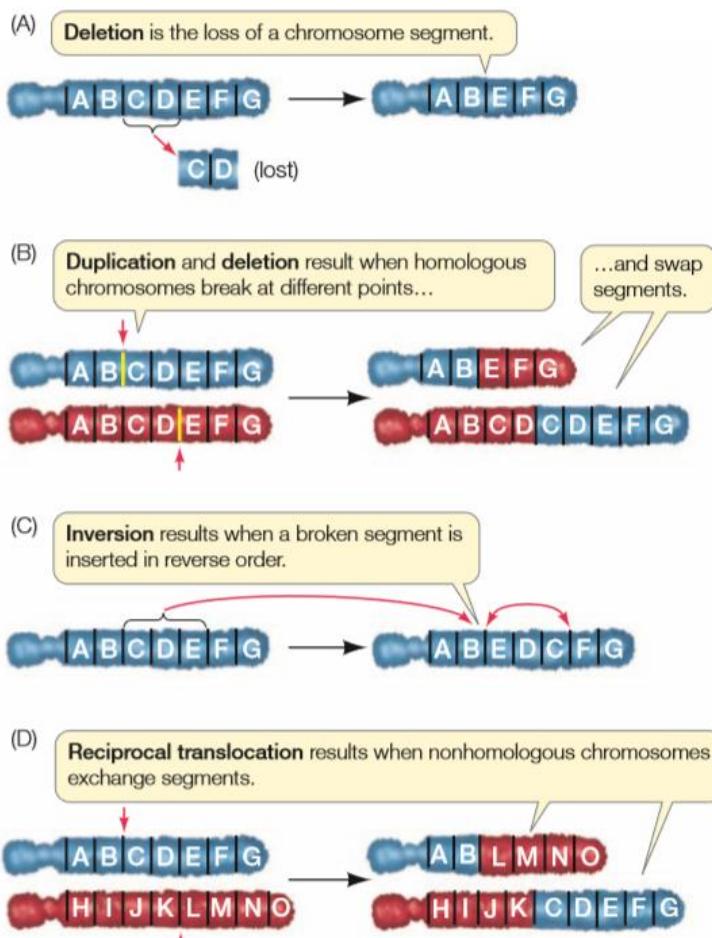


Figure 21- Point mutations. Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

occur in the genetic material. Whole DNA molecules can break and rejoin, grossly disrupting the sequence of genetic information. There are four types of such chromosomal mutations: deletions, duplications, inversions and translocations. These mutations can be caused by severe damage to chromosomes resulting from mutagens or by drastic errors in chromosome replication.

orientation of a DNA segment without actually removing any genetic information, or they may cause a segment of DNA to be duplicated or irretrievably lost. Changes in single nucleotides are not the most dramatic changes that can



Effects of mutation:

Mutations provide the scope for evolution: A mutation in somatic cells may benefit the organism immediately. Second, a mutation in germ line cells may have no immediate selective advantage to the organism but may cause a phenotypic change in offspring. If the environment changes in a later generation, that mutation may be advantageous and thus selected for under these conditions.

Germ line and somatic mutations can be harmful: Mutations in germ line cells that get carried to the next generation are often deleterious, especially if the offspring are homozygous for a harmful recessive allele. In their extreme form, such mutations produce phenotypes that are lethal. Lethal mutations can kill an organism during early development, or the organism may die before maturity and reproduction.

Figure 22- Chromosomal Mutations Chromosomes may break during replication, and parts of chromosomes may then rejoin incorrectly. The letters on these chromosome illustrations represent large segments of the chromosomes. Each segment may include anywhere from zero to hundreds or thousands of genes. . Image courtesy- Sadava et al, Life: The science of Biology, 9th edition.

2. Gene flow may change allele frequencies

Frequency of appearance of an allele also depends on the migration of individuals and movements of gametes between populations. This is called as gene flow. If the arriving individuals survive and reproduce in their new location, they may add new alleles to the population's gene pool, or they may change the frequencies of alleles already present if they come from a population with different allele frequencies.

3. Genetic drift:

In small populations, random changes in allele frequencies from one generation to the next may produce large changes in allele frequencies over time. This is called as genetic drift.

During population bottlenecks, genetic variation can be reduced by genetic drift. In this case, disasters such as floods and fires can drastically reduce the size of the population, leaving by chance, individuals that are not necessarily representative of the original population. A population forced through a bottleneck is likely

to lose much of its genetic variation. Genetic drift can have similar effects when a few pioneering individuals colonize a new region. Because of its small size, the colonizing population is unlikely to have all the alleles found among members of its source population. The resulting change in genetic variation, called a founder effect, is equivalent to that in a large population reduced by a bottleneck.

4. Nonrandom mating can change genotype frequencies

Mating patterns may alter genotype frequencies if individuals in a population do not choose mates at random.

1 hour

Constraints in evolution:

All life forms are constrained by various factors. For example, the growth in cell size is constrained by the stringencies of surface area-to-volume ratios. Folding of protein is limited by the bonding capacities of their constituent molecules. And the energy transfers that fuel life must operate within the laws of thermodynamics. Evolution works within the boundaries of these universal constraints.

The additional constraints which will be placed on evolution are:

1. Development process constraints evolution:

Evolutionary changes cannot start from scratch happen in this way. Current phenotypes of organisms are constrained by historical conditions and past selective pressures. The new species is formed or evolved from an existing species by incremental change. The old species is not discarded but it is changed. Thus evolution occurs from one form to another.

Metamorphosis of the tadpole larva into an adult frog is one of the most striking transformations in biology. In amphibians, metamorphosis is initiated by hormones from the tadpole, and these changes prepare an aquatic organism for a terrestrial existence.

The developmental constraints prevent the organism from producing a change which is extremely different from the ancestors.

2. Trade-offs constraint evolution

Adaptations frequently impose both fitness costs and benefits. For an adaptation to evolve, the fitness benefits it confers must exceed the fitness costs it imposes—in other words, the tradeoff must be worthwhile. For example, there are metabolic costs associated with developing and maintaining certain conspicuous features (such as antlers or horns) that males use to compete with other males for access to females. The fact that these features are common in many species suggests that the benefits derived from possessing them must outweigh the costs.

3. Short-term and long-term evolutionary outcomes sometimes differ

The short-term changes in allele frequencies within populations can be observed directly, they can be manipulated experimentally, and they demonstrate the actual processes by which evolution occurs. By themselves, however, they do not enable us to predict long-term evolutionary changes. Long-term patterns

of evolutionary change can be strongly influenced by events that occur so infrequently (a meteorite impact, for example) or so slowly (continental drift) that they are unlikely to be observed during short-term studies. The ways in which evolutionary processes act may change over time with changing environmental conditions.

30 min

References:

- 1) Taylor, D.J., Green, N.P., Stout, G.W. and Soper, R., 1997. *Biological science* (Vol. 983). Cambridge, United Kingdom: Cambridge University Press.
- 2) Sadava, D.E., Hillis, D.M., Heller, H.C. and Berenbaum, M., 2009. *Life: the science of biology* (Vol. 2). Macmillan.
- 3) Hall, B. and Strickberger, M.W., 2008. *Strickberger's evolution*. Jones & Bartlett Learning.

Historical overview of vaccine development:

The first recorded attempts to deliberately induce immunity were performed by the Chinese and Turks in the fifteenth century. They were attempting to prevent smallpox, a disease that is fatal in about 30% of cases and that leaves survivors disfigured for life. Reports suggest that the dried crusts derived from smallpox pustules were either inhaled or inserted into small cuts in the skin (a technique called *variolation*) in order to prevent this dreaded disease.

1718- Lady Mary Wortley Montagu, the wife of the British ambassador in Constantinople, observed the positive effects of *variolation* on the native Turkish population and had the technique performed on her own children.

1798- Edward Jenner observed milkmaids who had contracted the mild disease cowpox were subsequently immune to the much more severe smallpox, Jenner reasoned that introducing fluid from a cowpox pustule into people might protect them from smallpox. To test this idea, he inoculated an eight-year-old boy with fluid from a cowpox pustule and later intentionally infected the child with smallpox. As predicted, the child did not develop smallpox. However, this method is ethically incorrect.

1870- Louis Pasteur had succeeded in growing the bacterium that causes fowl cholera in culture, and confirmed this by injecting it into chickens that then developed fatal cholera.

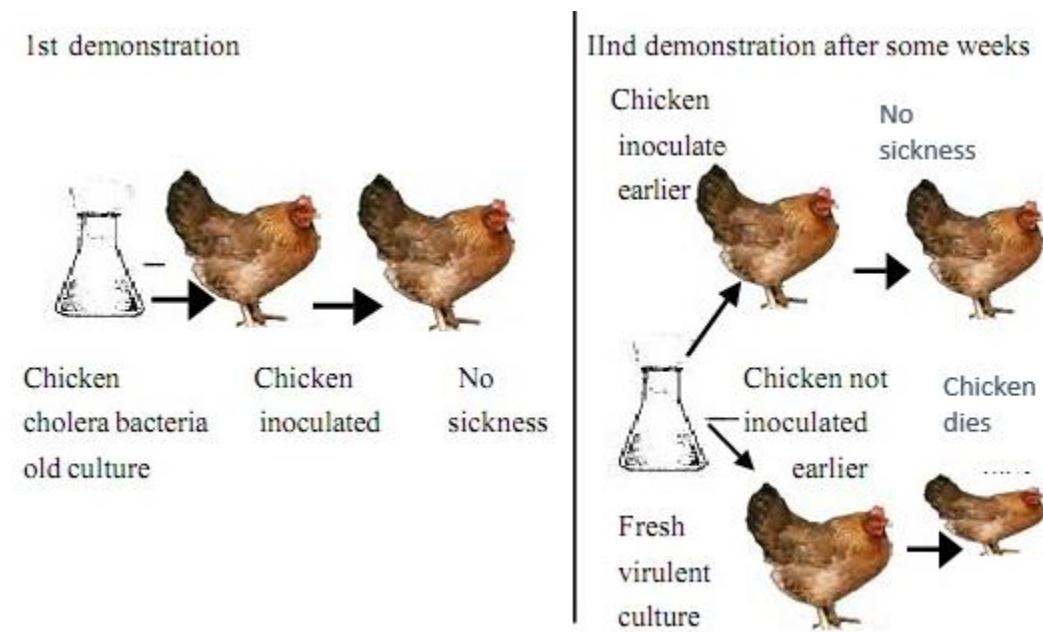


Figure 1- After returning from a summer vacation, he and colleagues resumed their experiments. 1st set of experiments involved old cholera bacterial culture, which caused the chickens to develop the disease on inoculation (inject) with the bacterial culture. However, the chickens recovered from the disease. In the second set of experiment, the chickens were inoculated with a new culture. The chickens who had been inoculated earlier and had survived the first set of experiment survived this stage as well. However, those chickens that were not inoculated earlier, developed the disease and died.

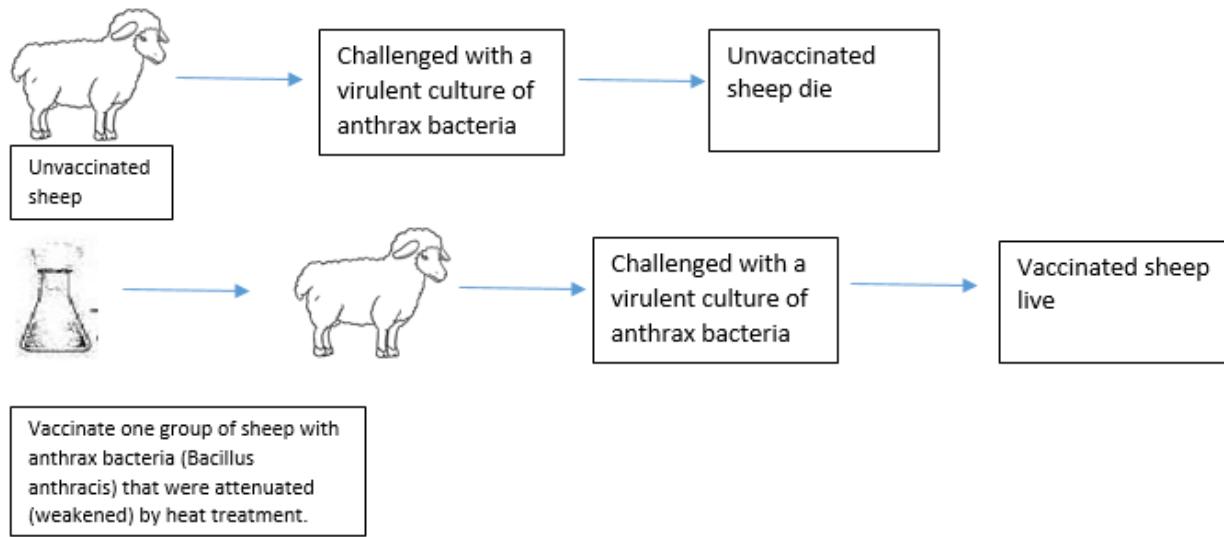


Figure 2- Pasteur extended these findings to other diseases, demonstrating that it was possible to attenuate a pathogen and administer the attenuated strain as a vaccine. In a now classic experiment performed in the small village of Pouilly-le-Fort in 1881, which marked the beginnings of the discipline of immunology.

In 1885, Pasteur administered his first vaccine to a human, a young boy who had been bitten repeatedly by a rabid dog. The boy, Joseph Meister, was inoculated with a series of attenuated (weakened) rabies virus preparations. The rabies vaccine is one of very few that can be successful when administered shortly after exposure. Joseph lived, and later became a caretaker at the Pasteur Institute, which was opened in 1887 to treat the many rabies victims that began to flood in when word of Pasteur's success spread; it remains to this day an institute dedicated to the prevention and treatment of infectious disease.

Introduction:

A vaccine is a biological preparation that improves immunity to a particular disease. A vaccine typically contains an agent that resembles a disease-causing microorganism and is often made from weakened or killed forms of the microbe, its toxins or one of its surface proteins. The agent stimulates the body's immune system to recognize the agent as foreign, destroy it, and keep a record of it, so that the immune system can more easily recognize and destroy any of these microorganisms that it later encounters.

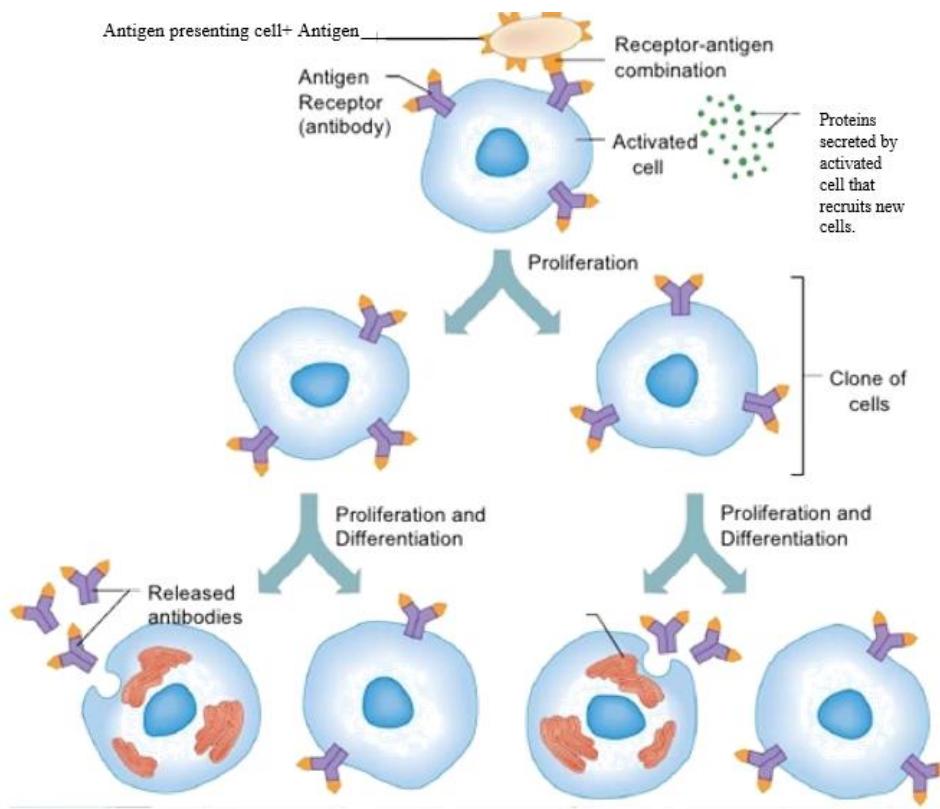


Figure 3- Cascade of immune reactions. On contact with an antigen, the Antigen presenting cell (APC) present the antigen to an effector cell. The effector cell recognizes the antigen that the APC holds at its cell surface and releases proteins that calls more immune cells to destroy the antigen. The effector cell has a second function of division (proliferation)

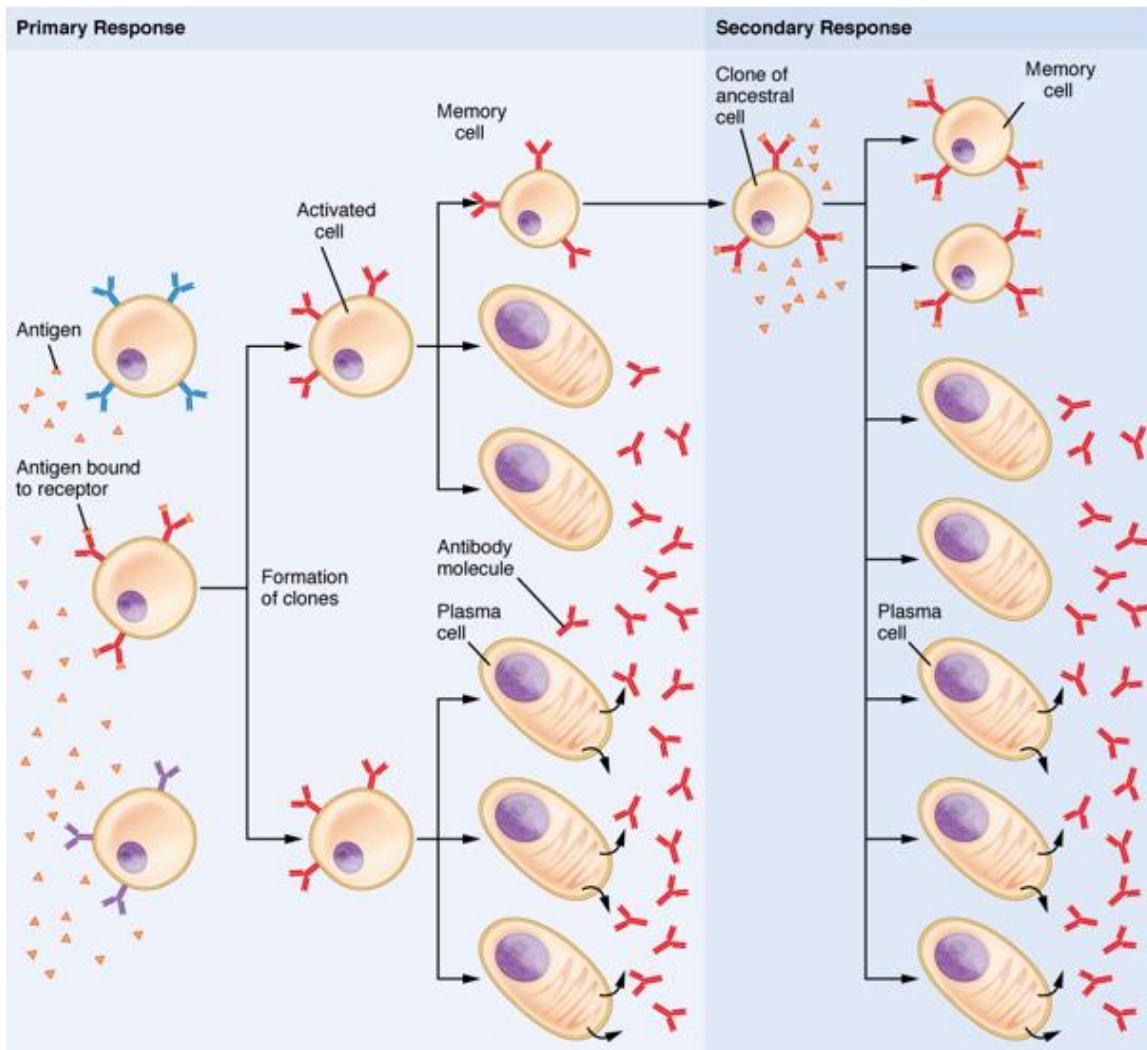


Figure 4- Primary response- When an antigen is encountered by the immune system for the first time, the body takes more time to recruit and activate all the cells and destroy the antigen. The result of primary response is destruction of the antigen and memory cells. Memory cells remember the properties of the antigen and in case of a second attack by the same antigen, cause a more rapid and more vigorous response. This is called the secondary response.

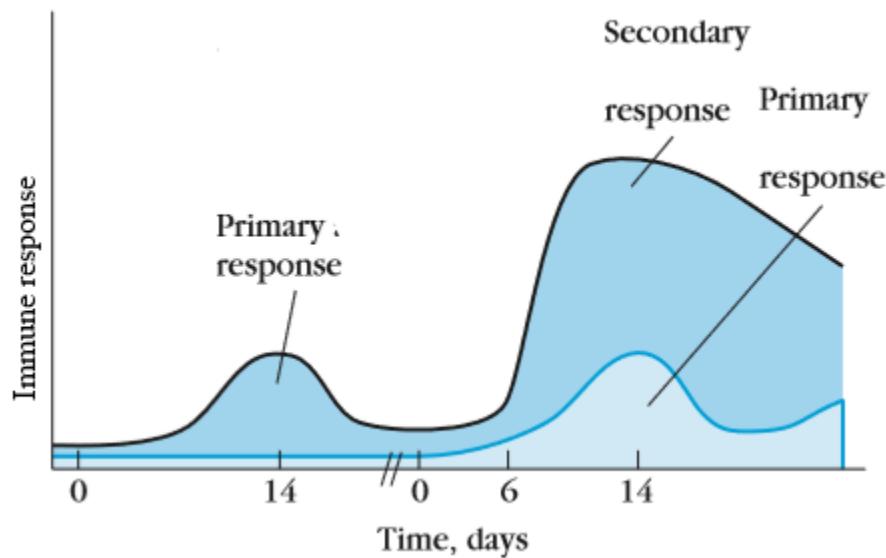


Figure 5- Primary response of antigen indicated in blue line. Primary and secondary response to antigens indicated by black line

Vaccine mode of action:

Immunization is the process of eliciting a long-lived state of protective immunity against a disease-causing pathogen. Exposure to the live pathogen followed by recovery is one route to immunization. Vaccination, or intentional exposure to forms of a pathogen that do not cause disease (a **vaccine**), is another. In an ideal world, both engage antigen-specific cells and result in the generation of memory cells, providing long-lived protection. However, vaccination does not ensure immunity, and a state of immune protection can be achieved by vaccination.

A state of immunity can be induced by passive or active immunization. Short-term passive immunization is induced by the transfer of preformed antibodies. Natural infection or vaccination can induce active immunization and lead to long-term immunity.

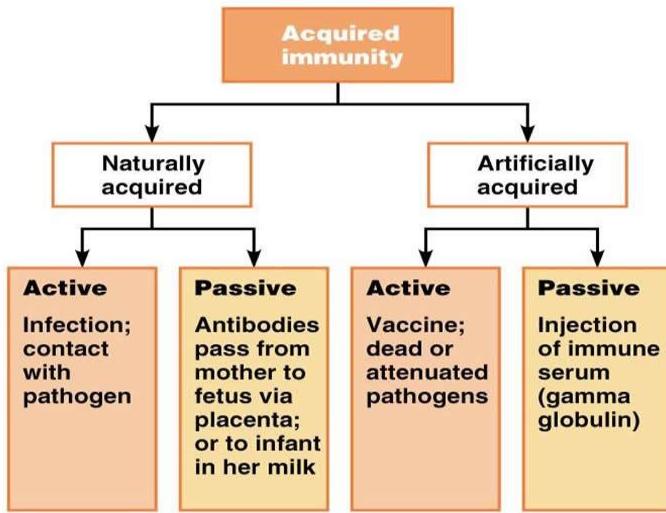


Figure 6- Immunity can be gained by passive or active way

Types of vaccines:

Five types of vaccines are currently used or under experimental consideration in humans: live, attenuated (avirulent) microorganisms; inactivated (killed) microorganisms and purified macromolecules (subunits).

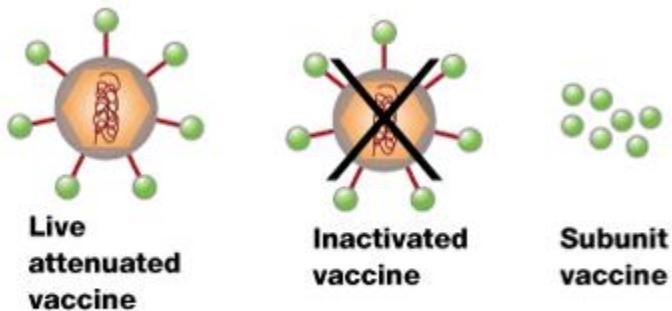


Figure 7- Types of vaccines

Live vaccines have the advantage of inducing both humoral and cell-mediated immunity, and can produce more effective overall protective immunity. However, live, attenuated vaccines carry the risk of reversion, which is not an issue with recombinant forms.

Scientists produce inactivated vaccines by killing the disease-causing microbe with chemicals, heat, or radiation. Such vaccines are more stable and safer than live vaccines: The dead microbes can't mutate back to their disease-causing state. Inactivated vaccines usually don't require refrigeration, and they can be easily stored and transported in a freeze-dried form, which makes them accessible to people in developing countries.

Isolated protein components of pathogens expressed in cell culture can be used to create effective vaccines, especially when the toxic effects of the pathogen are due to discrete protein

products. Polysaccharide and other less immunogenic vaccines may be conjugated to more immunogenic proteins to enhance or maximize the immune response.

Some vaccines show effect on a pathogen, however this effect is lost in the offsprings of the pathogen: Example of influenza vaccines

The virions are surrounded by an outer envelope, Inserted into the envelope are two glycoproteins, hemagglutinin (HA) and neuraminidase (NA).

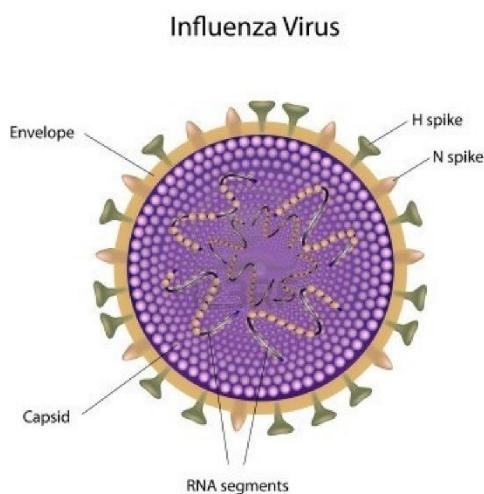


Figure 8- Influenza virus particle.

The first time a human influenza virus was isolated was in 1934; this virus was given the subtype designation HON1 (where H is hemagglutinin and N is neuraminidase). The hemagglutinin trimer binds to sialic acid groups on host-cell glycoproteins and glycolipids. Neuraminidase, as its name indicates, cleaves N-acetylneurameric (sialic) acid from nascent viral glycoproteins and host-cell membrane glycoproteins, an activity that presumably facilitates viral budding from the infected host cell.

A number of vaccines have been designed that target hemagglutinin and neuraminidase. However, the virus still is capable of surviving against the vaccines because of two processes that the virus is capable of:

1. Producing a series of spontaneous point mutations that occur gradually, resulting in minor changes in HA and NA.
2. Producing sudden emergence of a new subtype of influenza, who's HA and possibly also NA are considerably different from that of the virus present in a preceding epidemic.

Evolution of the immune system:

The innate immune system acts as the first line of defense and exists to provide early defense against pathogen attack, and to alert the adaptive immune system to the fact that pathogen invasion has begun. Many attributes of the innate immune system cells like macrophage cells are very similar to the movement of the most primitive organism amoeba. However, at such a basic level of organism, a stronger defense mechanism is not present. Innate immunity in eukaryotes can be thought of as arising from the need of a unicellular microorganism such as an amoeba to

discriminate between food and other amoebas. Innate immunity in eukaryotes allows the initial recognition of a foreign pathogen. However, this initial defense system does not provide the memory response required to identify a second round of attack by the pathogen. Therefore, a vaccine cannot be designed for such an organism that can only raise the innate response.

Genomic analysis of plants and animals provides evidence that a sophisticated mechanism of host defense was in existence by the time the ancestors of plants and animals diverged.

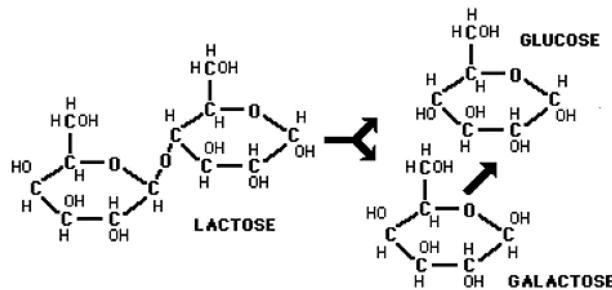
It has been known for at least 50 years that all jawed fish can mount an adaptive immune response. On the other hand, hagfish and lampreys, which are jawless vertebrates, lack all signs of an adaptive immune system: they do not have organized lymphoid tissue, they lack primary immune responses, and most importantly, they do not exhibit immunological memory. It was only in 1998 that the answers to these questions began to become apparent.

Adaptive immune system is the second line of defense of the immune system. In jawed fish and all 'higher' vertebrates, adaptive immunity is possible because a transposable element invaded a stretch of DNA, presumably a gene that was similar to an immunoglobulin gene. On the other hand, hagfish and lampreys, which are jawless vertebrates, lack all signs of an adaptive immune system: they do not have immune organs, they lack primary immune responses, and most importantly, they do not exhibit immunological memory. By contrast, even cartilaginous fish, the earliest jawed fish to survive to the present day, have organized immune organs, albeit primitive, and the ability to mount adaptive immune responses.

Thus the ability of mounting an adaptive immune response allows the organism to enlist cells for clearing out the invading pathogen and most important, produce memory cells for the specific pathogen (launch a secondary immune response). Such organisms which can mount an adaptive immune response and a secondary immune response can be vaccinated.

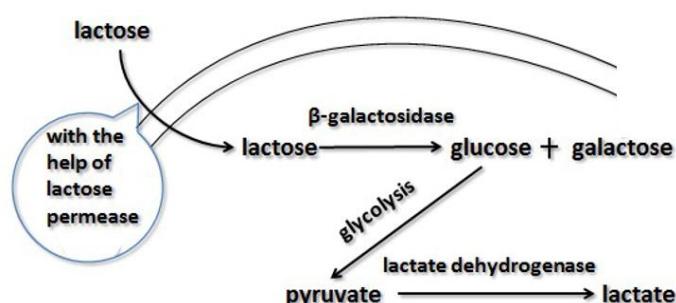
Operons

An operon is a group of genes transcribed at the same time. It is **only** found in prokaryotes. They regulate biochemical process. Each operon consists of a promoter, an operator and a regulator whose products function in a common pathway.



The lactose operon or the lac operon is one of the many operons that operates in bacteria. The lac operon encodes a set of genes that are involved in the metabolism of a simple sugar, lactose.

Figure 1- Lactose is a disaccharide composed of two sugars (galactose and glucose) with a β -linkage between carbon 1 of galactose and carbon 4 of glucose



The lac operon encodes three proteins (i) β -galactosidase - the product of the *lacZ* gene (ii) lactose permease - the product of the *lacY* gene and (iii) lactose transacetylase -the product of the *lacA* gene. All these genes are transcribed from a common promoter site.

Figure 2- Activity of β galactosidase and lactose permease in the metabolism of lactose

Escherichia coli use either glucose or lactose. There can be four options like:

1. When glucose is present and lactose is absent the E. coli does not produce β galactosidase.
2. When glucose is present and lactose is present the E. coli does not produce β galactosidase.
3. When glucose is absent and lactose is absent the E. coli does not produce β galactosidase.
4. When glucose is absent and lactose is present the E. coli does produce β galactosidase

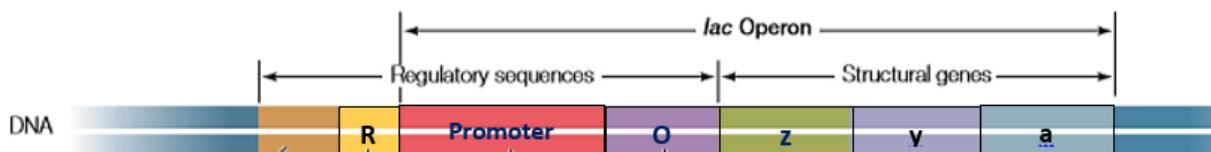
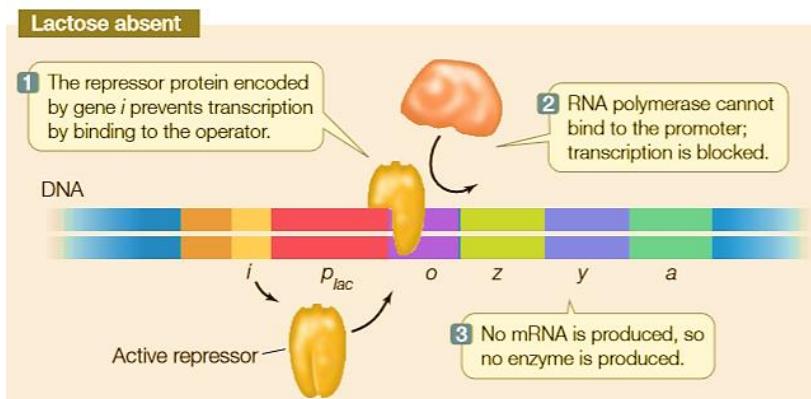


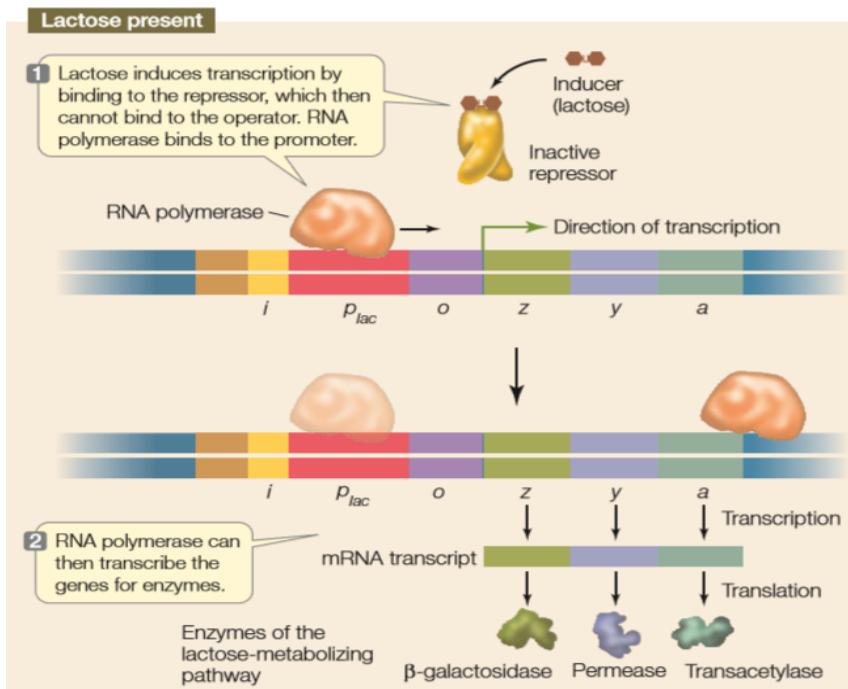
Figure 3- Lac operon (R- Regulatory gene (i) codes for repressor protein; O- operator)

The regulation of lac operon



repressor protein is synthesized from regulator genes

Figure 4- Repression of lac operon expression in absence of lactose



When glucose and lactose are present RNA polymerase can sit on the promoter site but it is unstable and it keeps falling off. For stabilizing RNA polymerase another protein is needed known as the activator protein. The activator protein only works when glucose is absent. In this way E. coli only makes enzymes to metabolize other sugars in the absence of glucose.

When lactose is absent, a repressor protein is continuously synthesized. It binds to the operator site of the lac operon. The operator site is located just in front of the three genes of the operon. As a result of repressor binding the RNA polymerase is not able to move forward to the three genes. Therefore the transcription of the genes are blocked. The

When lactose is present, a small amount of a sugar allolactose is formed within the bacterial cell. This fits onto the repressor protein at another active site (allosteric site). This causes the repressor protein to change its shape (a conformational change). It can no longer sit on the operator site. RNA polymerase can now reach its promoter site and transcribes the gene lac Z, lac Y and lac A. As a result, the lactose can be metabolized.

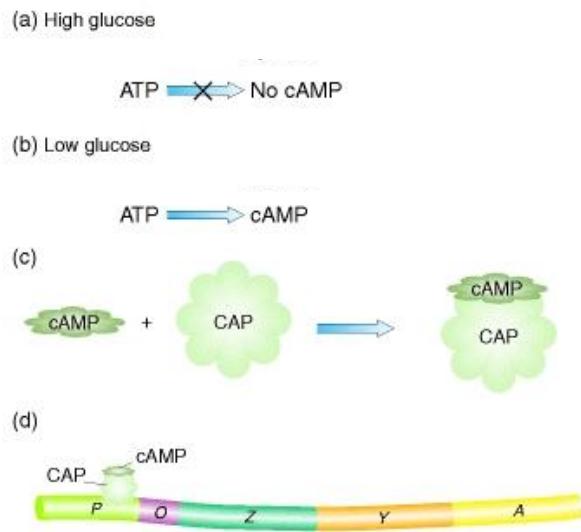
Figure 5- Activation of lac operon in presence of lactose

When glucose and lactose are

Carbohydrates	Activator protein	Repressor protein	RNA polymerase	<i>lac</i> Operon
+ GLUCOSE + LACTOSE	Not bound to DNA	Lifted off operator site	Keeps falling off promoter site	No transcription
+ GLUCOSE - LACTOSE	Not bound to DNA	Bound to operator site	Blocked by the repressor	No transcription
- GLUCOSE - LACTOSE	Bound to DNA	Bound to operator site	Blocked by the repressor	No transcription
- GLUCOSE + LACTOSE	Bound to DNA	Lifted off operator site	Sits on the promoter site	Transcription

Catabolite Repression of the *lac* Operon: Choosing the Best Sugar to Metabolize:

An additional control system is superimposed on the repressor–operator system. This control system is thought to have evolved because the cell can capture more energy from the breakdown of glucose than it can from the breakdown of other sugars. If both lactose and glucose are present, the synthesis of β -galactosidase is not induced until all the glucose has been utilized. Thus, the cell conserves its energy pool used, for example, to synthesize the Lac enzymes by utilizing any existing glucose before going through the energy-expensive process of creating new machinery to metabolize lactose.



The glucose breakdown product modulates the level of an important cellular constituent—cyclic adenosine monophosphate (cAMP). When glucose is present in high concentrations, the cell's cAMP concentration is low; as the glucose concentration decreases, the cellular concentration of cAMP increases correspondingly. The high concentration of cAMP is necessary for activation of the lac operon. cAMP binds with CAP (catabolite activator protein) to form the cAMP-CAP complex. This complex then binds to the CAP binding site on the lac operon and increases the affinity of RNA polymerase for the lac promoter. In this way, the catabolite repression system contributes to the selective activation of the lac operon.

Figure 6- The operon is inducible by lactose to the maximal levels when cAMP and CAP form a complex. (a) Under conditions of high glucose, a glucose breakdown product inhibits an enzyme, preventing the conversion of ATP into cAMP. (b) Under conditions of low glucose, there is no breakdown product, and therefore the enzyme is active and cAMP is formed. (c) When cAMP is present, it complexes with CAP. (d) The cAMP–CAP complex acts as an activator of lac operon transcription by binding to a region within the lac promoter.

Introduction:

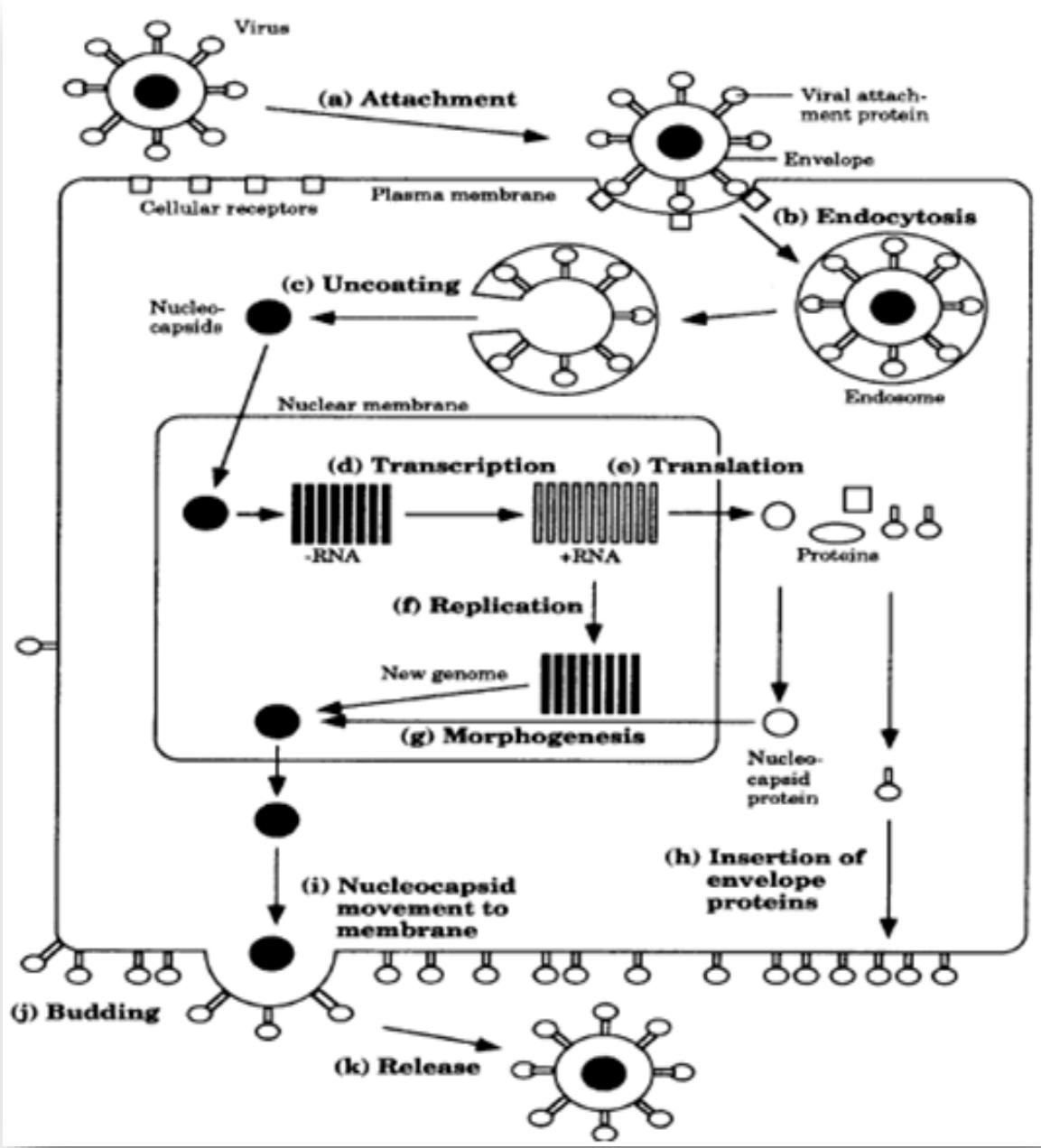
Viruses are the simplest form of life and they exist on the borderline between the living and the inanimate, non-biological world. They are mostly DNA/RNA that happen to be wrapped in a coating (termed a *capsid*). The capsid affords protection for the viral genes and allows viral genes to gain entrance to appropriate host cells. They are obligate parasites in the sense that they can only replicate after they have invaded and parasitized a host cell. However, the distinction of viruses from living organisms is drawn on basis of following observable facts about viruses:

- Virus particles are produced from the assembly of preformed components, whereas other agents grow from an increase in the integrated sum of their components and reproduce by division.
- Virus particles themselves do not grow or undergo division.
- Viruses lack the genetic information that encodes apparatus necessary for the generation of metabolic energy or for protein synthesis.

Life cycle of viruses:

The following steps represent the generalized productive life cycle for animal viruses:

- A)** The virus attaches to host receptor on the plasma membrane through its many attachment proteins.
- B)** The host plasma membrane surrounds the virus and draws it into the cell, a process called endocytosis.
- C)** Viral envelope fuse with endosomal membrane releasing nucleocapsids
- D)** Nucleic Acid (NA) of virus covered by capsid proteins
- E)** Viral NA enters cytoplasm
- F)** Viral NA serves as a template for new viral NA
- G)** Nucleocapsid enters the nucleus and begins the process of morphogenesis
- H)** Envelope proteins pass through the ER and golgi body
- I)** Nucleocapsid movement to membrane
- J)** Budding process
- K)** Viral particle release



Classification of eukaryotic Viruses:

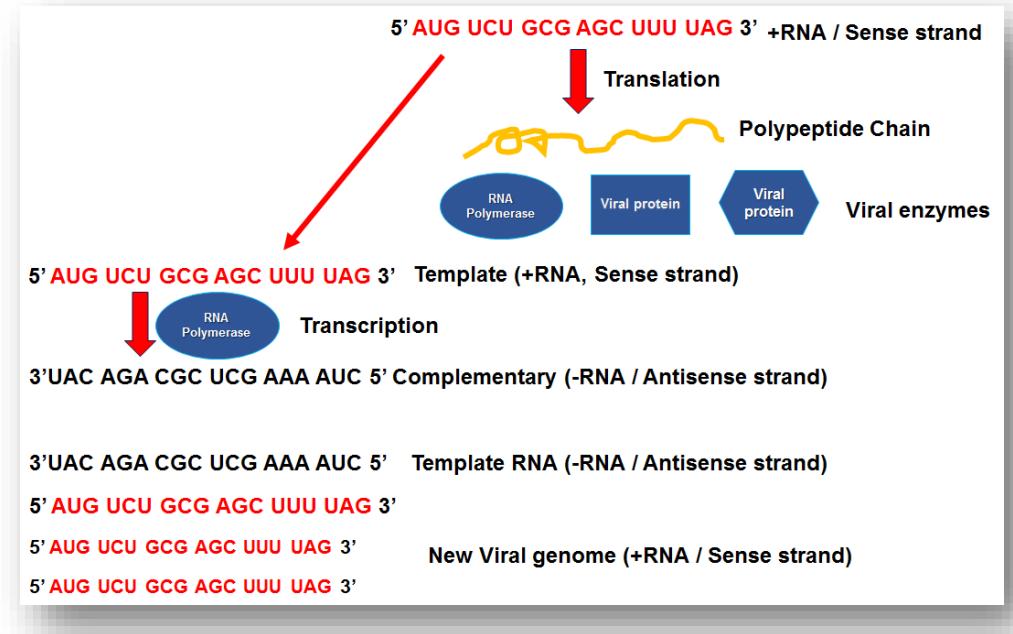
The primary characteristics used to differentiate eukaryotic viruses are associated with their nucleic acid. First, viruses may be separated based on whether they are DNA viruses or RNA viruses. The nucleic acid may be single-stranded (ss) or double-stranded (ds), depending on the species. If the ssRNA is able to function as mRNA it is referred to as plus strand RNA (+RNA); if it is the equivalent to antisense RNA it is known as minus strand RNA (-RNA). Virion shape is also used to differentiate among the viruses since they have a number of distinctive forms: cylindrical or helical, spherical, icosahedral, bulletshaped, or even more complex shapes.

Replication strategies:

This genetic material is the source for coding all the proteins needed for the virus to survive and propagate in a host cell. After the process of replication or protein synthesis stages are completed, new virus particles are produced by the association of the newly synthesized protein coats and the newly synthesized virus genetic material. The replication of the nucleic acid is extremely diverse, however, some general features can be outlined.

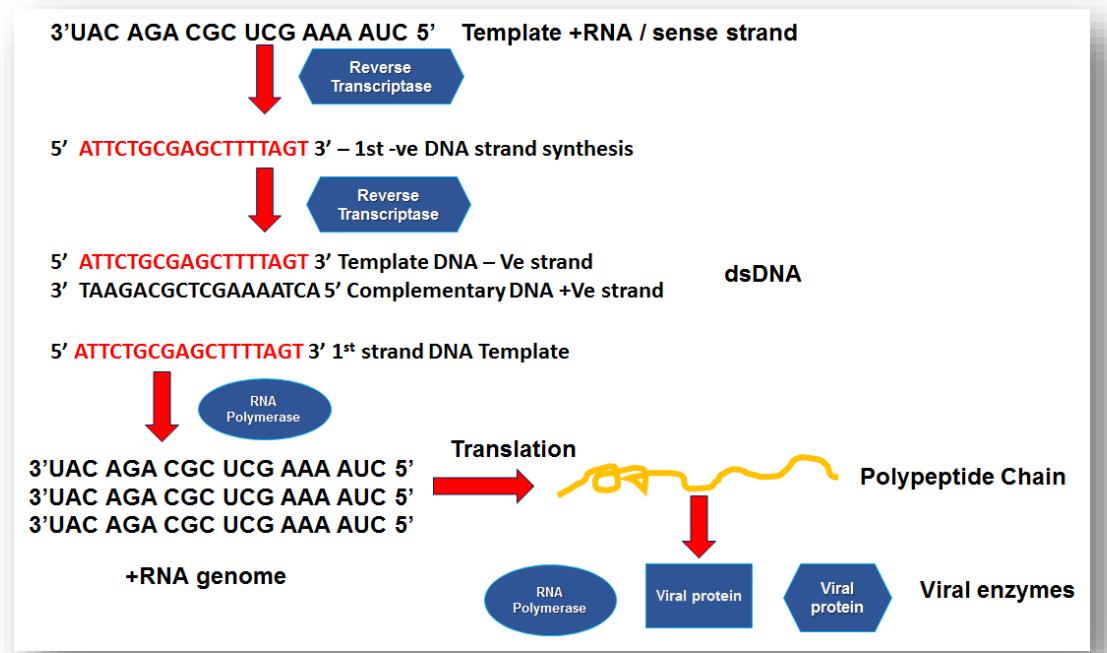
+ ssRNA Virus genome replication

Single-stranded +RNA viruses, such as picornviruses and togaviruses, maybe directly translated. Aviral encoded RNA-dependent RNA polymerase is produced early on and catalyzes the transcription of intermediate complementary -RNA strands. The minus strands then serve as templates for the synthesis of genomic +RNAs.



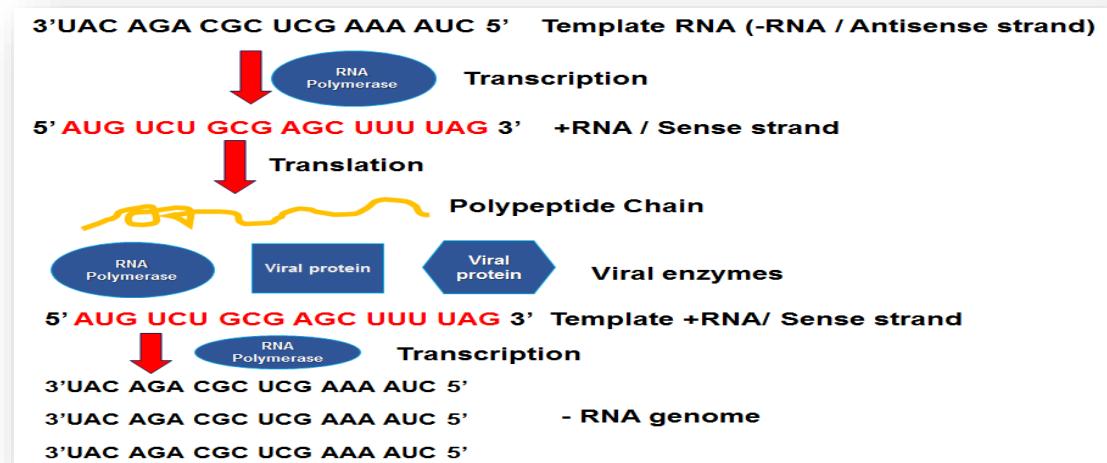
RT directed +ssRNA virus genome replication

Retroviruses are an unusual group of +RNA viruses in that they synthesize new +RNA using a DNA template. The plus genome is converted in a step-wise manner to a dsDNA molecule by reverse transcriptase, which is carried by the virus. Reverse transcriptase subsequently synthesizes a DNA strand complementary to the first and the resulting dsDNA integrates into the host genome. Finally +ssRNA are created using the dsDNA as template



- ssRNA virus genome replication

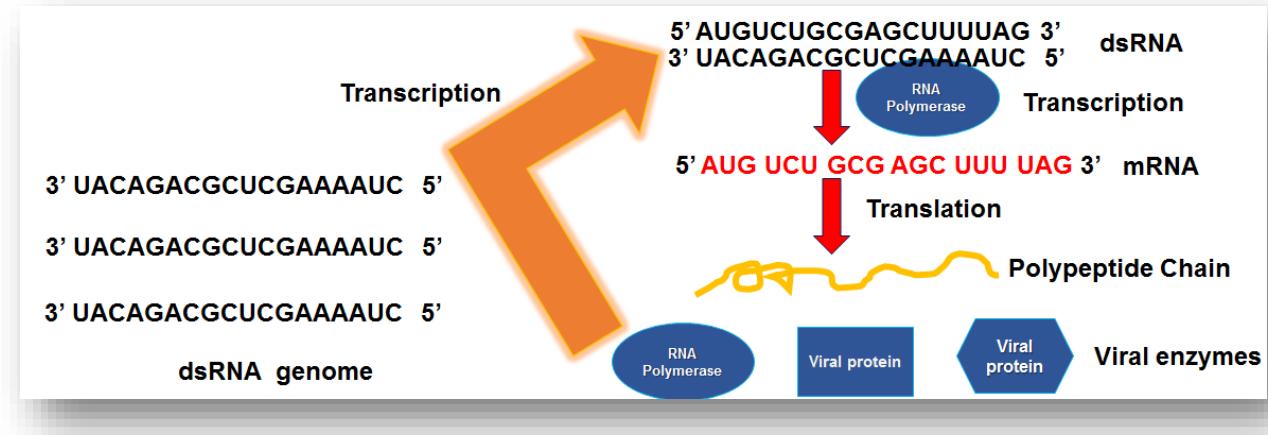
Single-stranded –RNA viruses are those with nucleic acids (genomes) that cannot be translated. The –RNA strand serves as a template for the synthesis of +RNA that functions as mRNA. Viral proteins translated from the mRNA promote the synthesis of full-length +RNA strands that function as templates for the synthesis of full-length –RNA genome strands. Usually the virus brings into the cell an RNA-dependent RNA polymerase for making viral mRNA.



ds RNA virus

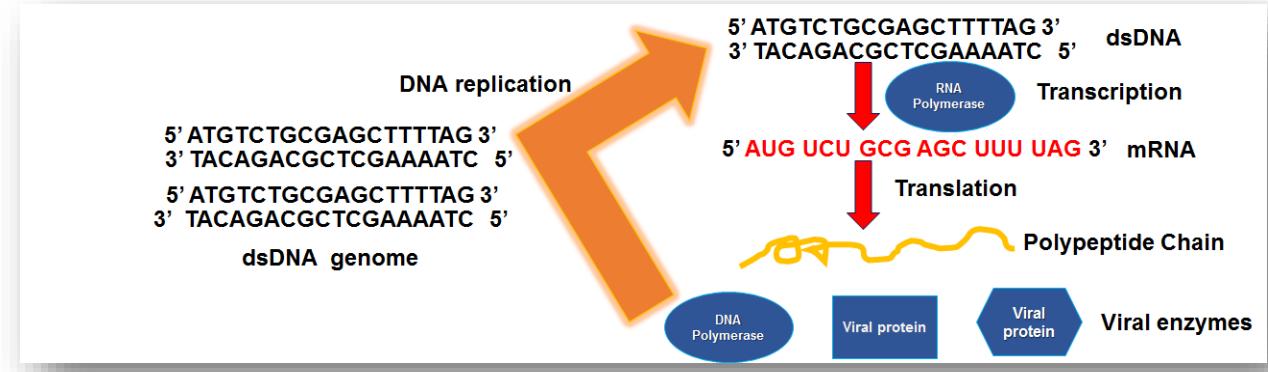
Double-stranded RNA viruses, such as reoviruses, are segmented and carry a dsRNA-dependent RNA polymerase which they use to transcribe their genomes. The RNA segments and the mRNA molecules specified by them only encode for one protein each. These viruses replicate their dsRNA

genomes by producing many copies of plus strand RNA that are not translated, but serve as templates for the synthesis of complementary minus strands.



ds DNA virus genome replication

The simplest viruses to understand are those with genomes of double-stranded DNA (dsDNA). Once the nucleocapsid of this type of virus enters the cell, it proceeds to the nucleus where it mimics the genome of the host cell. The viral genome is replicated using the host cell DNA polymerase, and the viral genome is transcribed by the host cell RNA polymerase. The resulting transcripts carrying information encoding viral proteins are then transported to the cytoplasm and viral proteins are produced with the host ribosomes. These newly synthesized viral proteins are used as the protein capsid around newly replicated viral DNA molecules. These new virions are released from the cell and trigger new rounds of replication. But some viruses produce their own DNA polymerase and use them for the next round of replications.



Making the most of a limited genome: In a continuous arrangement of one nucleotide after the other, it is possible that a second reading frame could be formed in which nucleotide number 2 to 4 forms the first codon, nucleotide number 5 to 7 forms second codon and so on. It is easy to imagine that on similar lines, a third reading frame is also possible. So we can imagine that depending on whichever reading frame is used during translation, the same gene may produce two or three entirely different polypeptide or proteins from the same gene.

5' ATTCTGCGAGCTTTAGT 3' - Coding DNA
3' TAAGACGCTCGAAAATCA 5' - Template DNA

5' AUU CUG CGA GCU UUU AGU 3' - mRNA

Ile-leu-Arg-Arg-Phe-Ser - one reading frame AA chain

5' A UUC UGC GAG CUU UUA GU 3' - mRNA

Phe-Cys-Glu-Leu-Leu - Second reading frame AA chain

5' AU UCU GCG AGC UUU UAG U 3'

Ser-Ala-Ser-Phe-Stop - Third reading frame AA chain

The above information showing how the same nucleotide sequence of an mRNA be read as three different proteins due to alterations of the reading frame. While in the first reading frame, the mRNA sequence shall be translated as “Ile-leu-Arg-Arg-Phe-Ser” amino acid residues; in the second the same mRNA will be translated into “Phe-Cys-Glu-Leu-Leu”. In the third reading frame, the translation will be termed prematurely as a stop codon gets created.

Overlap of genes within a genome

Though in most organisms, a gene normally encodes only one polypeptide or protein as translation always begins on the same reading frame, in several viruses, the same DNA base sequence codes for two different proteins by employing two different reading frames giving raise to overlapping genes. For example, virus or bacteriophage has two overlapping genes in its genome. Overlapping genes are also observed in the genomes of virus or phage. It has been suggested that because of the fixed, small size of viral capsid, viruses require economical use of a limited amount of DNA to code for the variety of proteins needed to infect a host cell and replicate within it. Overlapping genes provide an opportunity to achieve this. Normally, overlap of genes is considered as a primitive characteristics of a genome.

What is recombinant DNA technology?

Molecular biology technique borrows many of its principles based on observation of the microbial world. In the early 1960s, before the advent of gene cloning, studies of genes often relied on indirect or fortuitous discoveries, such as the ability of bacteriophages to incorporate bacterial genes into their genomes. A huge contribution to recombinant DNA technology is based on the products and processes followed by the microbes.

The cornerstone of most molecular biology technologies is the gene. To facilitate the study of genes, they can be isolated and amplified. One method of isolation and amplification of a gene of interest is to clone the gene by inserting it into another DNA molecule that serves as a vehicle or vector that can be replicated in living cells. When these two DNAs of different origin are combined, the result is a recombinant DNA molecule. Although genetic processes such as crossing-over technically produce recombinant DNA, the term is generally reserved for DNA molecules produced by joining segments derived from different biological sources. The recombinant DNA molecule is placed in a host cell, either prokaryotic or eukaryotic. The host cell then replicates (producing a clone), and the vector with its foreign piece of DNA also replicates. The foreign DNA thus becomes amplified in number, and following its amplification can be purified for further analysis.

Insights from bacteriophage lambda (λ) cohesive sites

In 1962, Allan Campbell noted that the linear genome of bacteriophage λ forms a circle upon entering the host bacterial cell, and a recombination (breaking and rejoicing) event inserts the phage DNA into the host chromosome. Reversal of the recombination event leads to normal excision of the phage DNA. Further analysis revealed that phage λ had short regions of single-stranded DNA whose base sequences were complementary to each other at each end of its linear genome. These single-stranded regions were called “cohesive” (cos) sites.

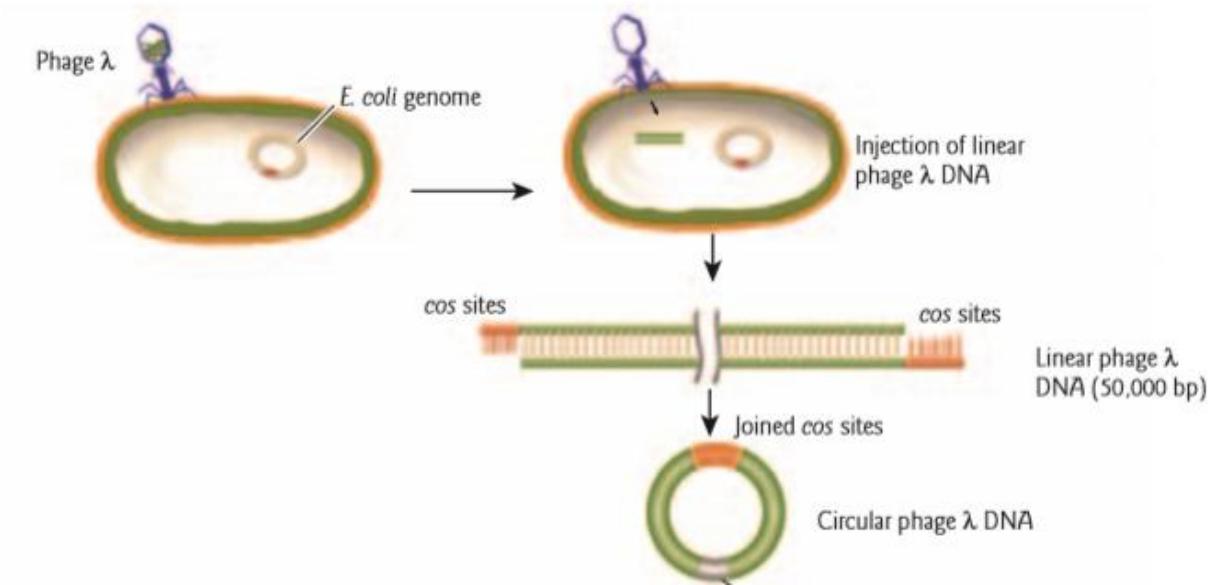


Figure 1- Cloning (insertion) of viral DNA into bacterial (host) cell. The cohesive ends are complementary to each other and thus gain a circular form in the host after insertion.

Complementary base pairing of the cos sites allowed the linear genome to become a circle within the host bacterium. The idea of joining DNA segments by “cohesive sites” became the guiding principle for the development of genetic engineering. With the molecular characterization of restriction and modification systems in bacteria, it soon became apparent that the ideal engineering tools for making cohesive sites on specific DNA pieces were already available in the form of restriction endonucleases. Hamilton Smith and co-workers demonstrated unequivocally that restriction endonucleases cleave a specific DNA sequence. The activity of cutting by restriction endonucleases, mostly, produced cohesive ends. Restriction endonucleases (also referred to simply as restriction enzymes) received their name because they restrict or prevent viral infection by degrading the invading nucleic acid.

Modification system.

At the time, it was known that methyl groups were added to bacterial DNA at a limited number of sites. Most importantly, the location of methyl groups varied among bacterial species. Arber and colleagues were able to demonstrate that modification consisted of the addition of methyl groups to protect those sites in DNA sensitive to attack by a restriction endonuclease. Methyl-modified target sites are no longer recognized by restriction endonucleases and the DNA is no longer degraded. Once established, methylation patterns are maintained during replication. In contrast, foreign DNA that is unmethylated or has a different pattern of methylation than the host cell DNA is degraded by restriction endonucleases.

Cutting and joining DNA

Two major categories of enzymes are important tools in the isolation of DNA and the preparation of recombinant DNA: restriction endonucleases and DNA ligases. Restriction endonucleases recognize a specific, rather short, nucleotide sequence on a double-stranded DNA molecule, called a restriction site, and cleave the DNA at this recognition site or elsewhere, depending on the type of enzyme. DNA ligase joins two pieces of DNA by forming phosphodiester bonds. To increase the efficiency of the reaction, researchers often use the enzyme terminal deoxynucleotidyl transferase to modify the blunt ends (no cohesive ends). For example, if a single-stranded poly(dA) tail is added to DNA fragments from one source, and a single stranded poly(dT) tail is added to DNA from another source, the complementary tails can hydrogen bond. Recombinant DNA molecules can then be created by ligation.

It should be appreciated that the entire framework of molecular cloning rests heavily on the contributions made by the microbial world.

Molecular cloning

The basic procedure of molecular cloning involves a series of steps:

First, the DNA fragments to be cloned are generated by using restriction endonucleases.

Second, the fragments produced by digestion with restriction enzymes are ligated to other DNA molecules that serve as vectors. Vectors can replicate autonomously (independent of host genome replication) in host cells and facilitate the manipulation of the newly created recombinant DNA molecule.

Third, the recombinant DNA molecule is transferred to a host cell. Within this cell, the recombinant DNA molecule replicates, producing dozens of identical copies known as clones. As the host cells replicate, the recombinant DNA is passed on to all progeny cells, creating a population of identical cells, all carrying the cloned sequence.

Finally, the cloned DNA segments can be recovered from the host cell, purified, and analyzed in various ways.

Vector DNA

Cloning vectors are carrier DNA molecules. Four important features of all cloning vectors are that they: (i) can independently replicate themselves and the foreign DNA segments they carry; (ii) contain a number of unique restriction endonuclease cleavage sites that are present only once in the vector; (iii) carry a selectable marker (usually in the form of antibiotic resistance genes or genes for enzymes missing in the host cell) to distinguish host cells that carry vectors from host

cells that do not contain a vector; and (iv) are relatively easy to recover from the host cell. There are many possible choices of vector depending on the purpose of cloning.

Plasmid DNA as a vector

Plasmids are naturally occurring extrachromosomal, double-stranded circular DNA molecules that carry an origin of replication and replicate autonomously within bacterial cells. The plasmid vector pBR322, constructed in 1974, was one of the first genetically engineered plasmids to be used in recombinant DNA. Plasmids are named with a system of uppercase letters and numbers, where the lowercase “p” stands for “plasmid.” In the case of pBR322, the BR identifies the original constructors of the vector (Bolivar and Rodriguez), and 322 is the identification number of the specific plasmid. pUC18 is a derivative of pBR322.

Cloning

Plasmid vectors are modified to contain a specific antibiotic resistance gene and a multiple cloning site (also called the polylinker region) which has a number of unique target sites for restriction endonucleases. Cutting the circular plasmid vector with one of these enzymes results in a single cut, creating a linear plasmid. A foreign DNA molecule, referred to as the “insert,” cut with the same enzyme, can then be joined to the vector in a ligation reaction. Ligations of the insert to vector are not 100% productive, because the two ends of a plasmid vector can be readily ligated together, which is called self-ligation. But, if the vector is joined with a foreign insert, the 5'-phosphate is provided by the foreign DNA. Another strategy involves using two different restriction endonuclease cutting sites with noncomplementary sticky ends. This inhibits self-ligation and promotes annealing of the foreign DNA in the desired orientation within the vector.

Transformation: transfer of recombinant plasmid DNA to a bacterial host

The ligation reaction mixture of recombinant and nonrecombinant DNA described in the preceding section is introduced into bacterial cells in a process called transformation. The permeable “competent” cells are then mixed with DNA to allow entry of the DNA into the bacterial cell. Successfully transformed bacteria will carry either recombinant or nonrecombinant plasmid DNA. Multiplication of the plasmid DNA occurs within each transformed bacterium. A single bacterial cell placed on a solid surface (agar plate) containing nutrients can multiply to form a visible colony made of millions of identical cells. As the host cell divides, the plasmid vectors are passed on to progeny, where they continue to replicate. Numerous cell divisions of a single transformed bacteria result in a clone of cells (visible as a bacterial colony) from a single parental cell. This step is where “cloning” got its name. The cloned DNA can then be isolated from the clone of bacterial cells.

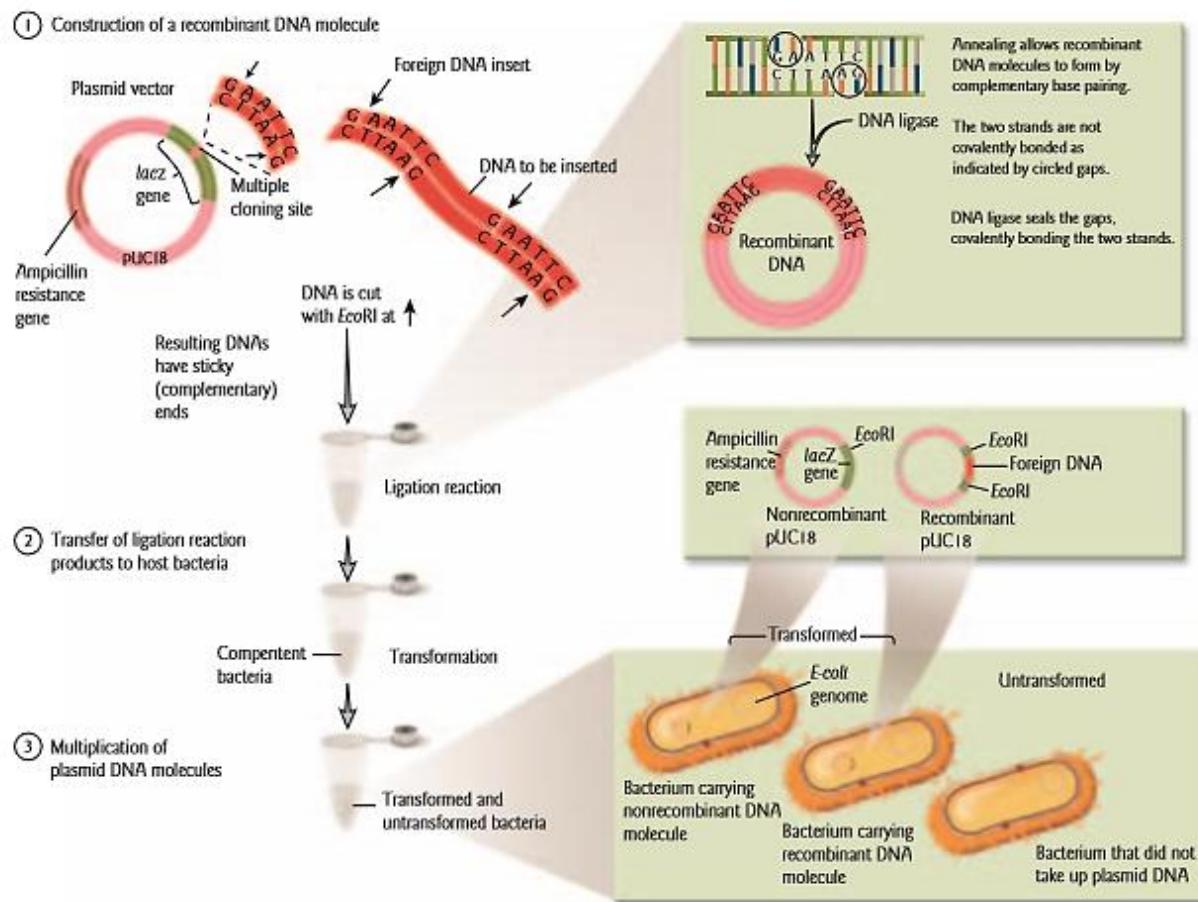


Figure 2- Production of recombinants and transformation

Recombinant selection

What needs to be included in the medium for plating cells so that nontransformed bacterial cells are not able to grow at all? The answer depends on the particular vector, but in the case of pUC18, the vector carries a selectable marker gene for resistance to the antibiotic ampicillin. Ampicillin, a derivative of penicillin, blocks synthesis of the peptidoglycan layer that lies between the inner and outer cell membranes of *E. coli*. Ampicillin does not affect existing cells with intact cell envelopes but kills dividing cells as they synthesize new peptidoglycan. The ampicillin resistance genes carried by the recombinant plasmids protects against ampicillin. Nontransformed cells contain no pUC18 DNA, therefore they will not be antibiotic-resistant, and their growth will be inhibited on agar containing ampicillin. Transformed bacterial cells may contain either nonrecombinant pUC18 DNA (selfligated vector only) or recombinant pUC18 DNA (vector containing foreign DNA insert). Both types of transformed bacterial cells will be ampicillin-

resistant. Further screening may be done to separate out the recombinant from non-recombinant types.

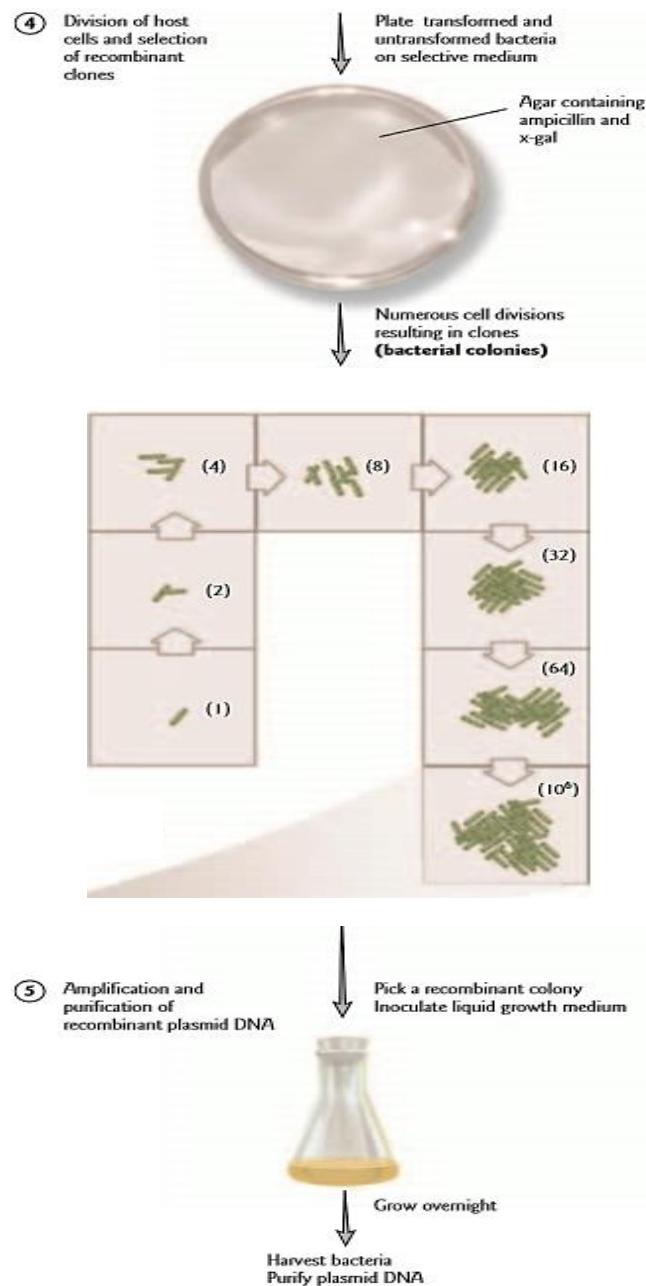


Figure 3- Recombinant selection on a selective media

The final step is purification of the plasmid DNA and can be further used for various purposes.

Application of molecular cloning

Molecular cloning provides scientists with an essentially unlimited quantity of any individual DNA segments derived from any genome. This material can be used for a wide range of purposes,

including those in both basic and applied biological science. A few of the more important applications are i) To study the genome organization and gene expression; ii) Production of recombinant proteins; iii) Transgenic organisms/genetically modified organisms (GMOs); Although most GMOs are generated for purposes of basic biological research (transgenic mouse), a number of GMOs have been developed for commercial use, ranging from animals and plants that produce pharmaceuticals or other compounds (pharming), herbicide-resistant crop plants, and fluorescent tropical fish (GloFish) for home entertainment. iv) Gene therapy; Gene therapy involves supplying a functional gene to cells lacking that function, with the aim of correcting a genetic disorder or acquired disease. Gene therapy can be broadly divided into two categories. The first is alteration of germ cells, that is, sperm or eggs, which results in a permanent genetic change for the whole organism and subsequent generations. This “germ line gene therapy” is considered by many to be unethical in human beings. The second type of gene therapy, “somatic cell gene therapy”, is analogous to an organ transplant. In this case, one or more specific tissues are targeted by direct treatment or by removal of the tissue, addition of the therapeutic gene or genes in the laboratory, and return of the treated cells to the patient. Clinical trials of somatic cell gene therapy began in the late 1990s, mostly for the treatment of cancers and blood, liver, and lung disorders.

Human Insulin Production

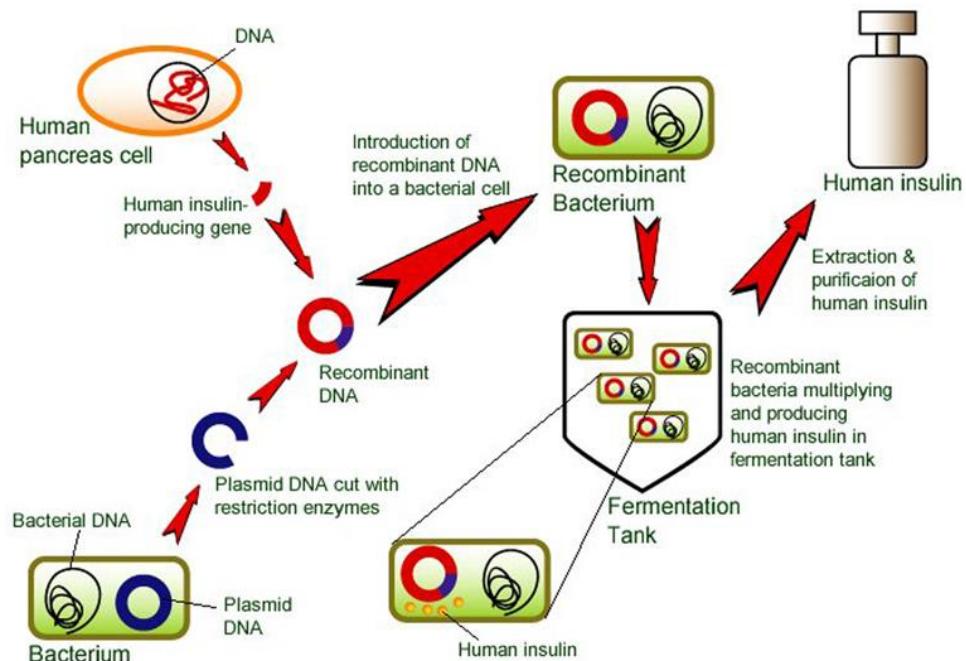


Figure 4- Human insulin production using molecular cloning. The production of human insulin chains in bacterium is a classic example of the application of molecular cloning for improving human life. Initially, bovine and porcine sources were used for the purpose of human insulin production. However, this gave rise to various ethical and medical problems. These problems are successfully overcome by recombinant DNA technology to produce insulin in bacteria.

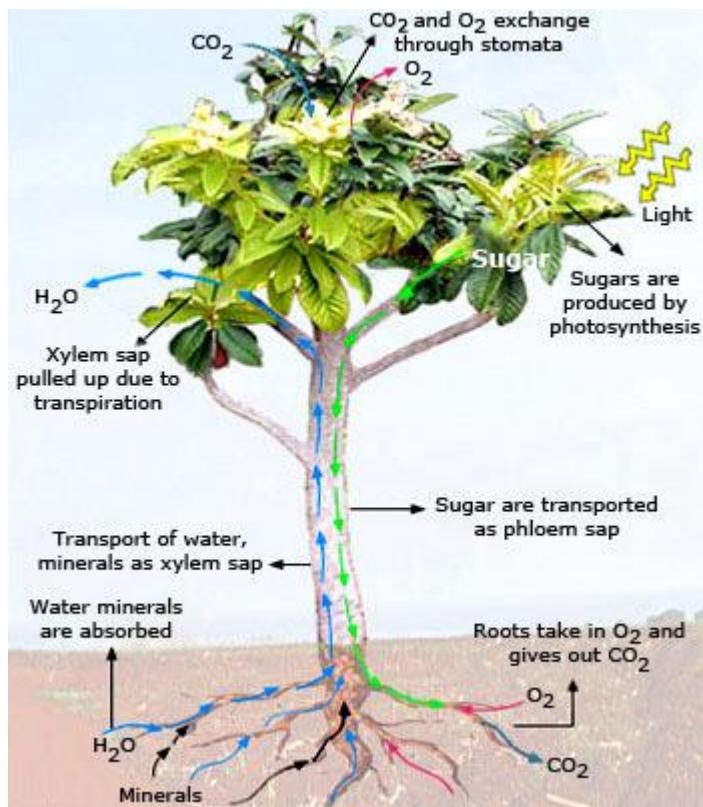
CASE STUDY

WATER ABSORPTION AND TRANSPORT IN PLANTS

Presentation of the problem: We need a pump to uplift water to a tank. The requirement of a pump is to pull water against the gravity. Look at the human body. We do also have a pump (heart) to circulate blood. Now imagine tall tree. In plants, there are no standing tanks, pumps (hearts), or valves that can move water up trees. The leaves at the tip of the tree must get water to survive and for preparation of food. Do a tree have a pump to pull water against the gravitational force? How much water can a tree process? How do they prevent loss of water? Today we know that a big tree can process around 3000 liters of water a day. What are the motive forces of water movement in plants? What will be the speed of transport?

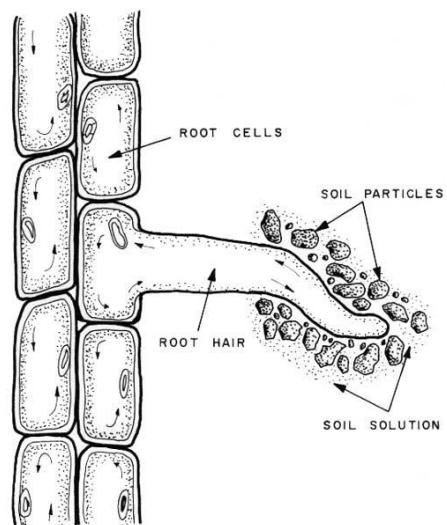
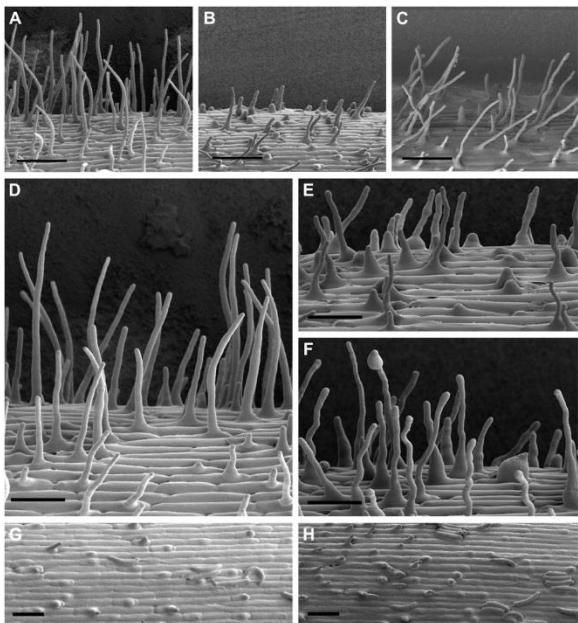
The route map of water transport in plants

Root Hair → Cortex cells → Xylem vessels of root → Xylem vessels of stem → Xylem vessels of leaf → Stomata (Escapes in the form of water vapor—Transpiration)



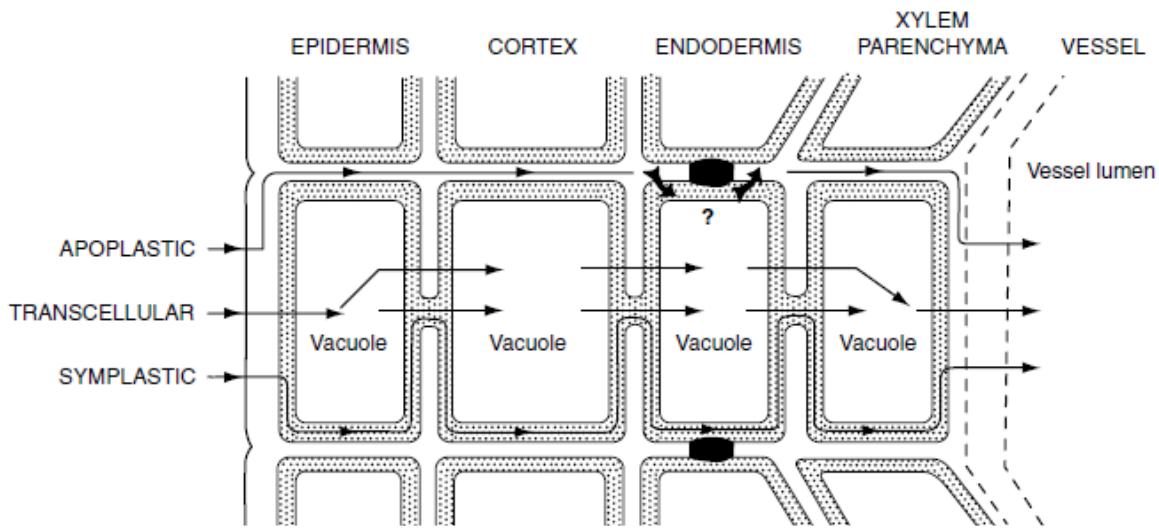
The hurdles of water relations in plants: (a) Water absorption (b) Water transport in xylem upwards against gravitational pull (c) Transport of sap in Phloem

Water absorption: The water absorbing cell of the root is the root hair. Root hair cell is a root epidermal cell with a hair like projection. This is an adaptation to increase the water absorbing area. Another advantage of this **root hair architecture** is to maintain contact with the soil by bending into the soil particle and penetrating into the crevices. If you take a root, you will find a root hair zone where the frequency of the root hairs are high. This is towards the root tip. Hence all part of the root are not absorbing water. Because root is an organ which is **multipurpose**: (i) To absorb water and mineral ions (ii) To support the plant

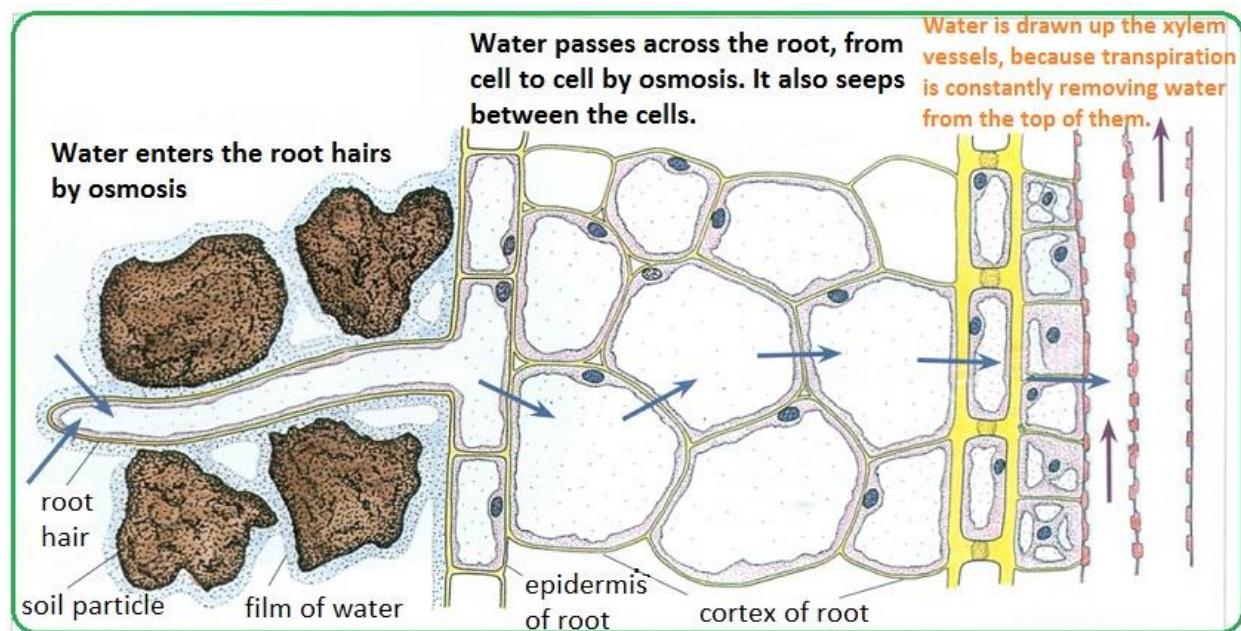


SEM pictures of root hairs in Rice

The cell wall is made up of cellulose. It is hydrophilic in nature. Root hair cell imbibe water. From here water enters the cell sap through osmosis. The water movement till the xylem cell is through osmosis. Then it enters xylem cell by the **transpiration pull**. So the water absorbed by the root has to move **radially** to reach the xylem. Here is the three possible ways of radial transport till the xylem vessel.



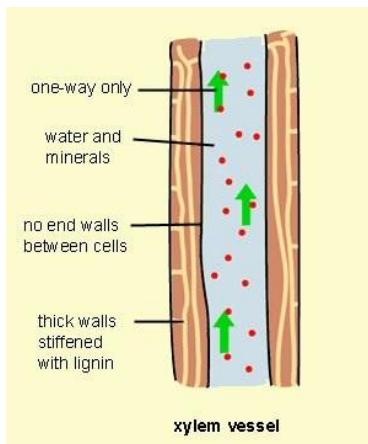
The figure showing the radial conduction of water in three different ways Apoplastic (through the non-living cell wall), symplastic (through the living parts ie cytoplasm) and transcellular (water crosses the cell wall and cell membrane and also through the vacule)



Now the question is **xylem is a dead tissue**. How it is enabled with features of water transport?

The root of water movement in the plant is through xylem vessels. It is the longest pathway of water transport. In a plant 1 m tall, more than 99.5% of the water transport pathway through the plant is within the xylem, and in tall trees the xylem represents an even greater fraction of the pathway. **Why the plant evolved with a dead tissue for water conduction?**

Xylem is dead. It means it has no membranes and no organelles. Its cross walls and contents will break down. Hence it is a **continuous system** (the lumina) from root till the leaf end. Its walls are **deposited with lignin**. This makes it **rigid**. It is **more hydrophobic**. The other advantages are (i) Being dead it can transport toxic substances (ii) Chilling does not stop water movement.



The motive forces of water movement: Capillary force, Transpiration pull (cohesion tension) and root pressure

(A) Capillary force

We know from the school days that roots absorb water and minerals from the soil. The water is transported to the leaves and is necessary for the preparation of food through photosynthesis. Look at the following figure and find the analogy. What is the force behind?

Xylem vessels are narrow and dead without cross walls making it an ideal system to develop capillary force. This facilitates the water movement upward.

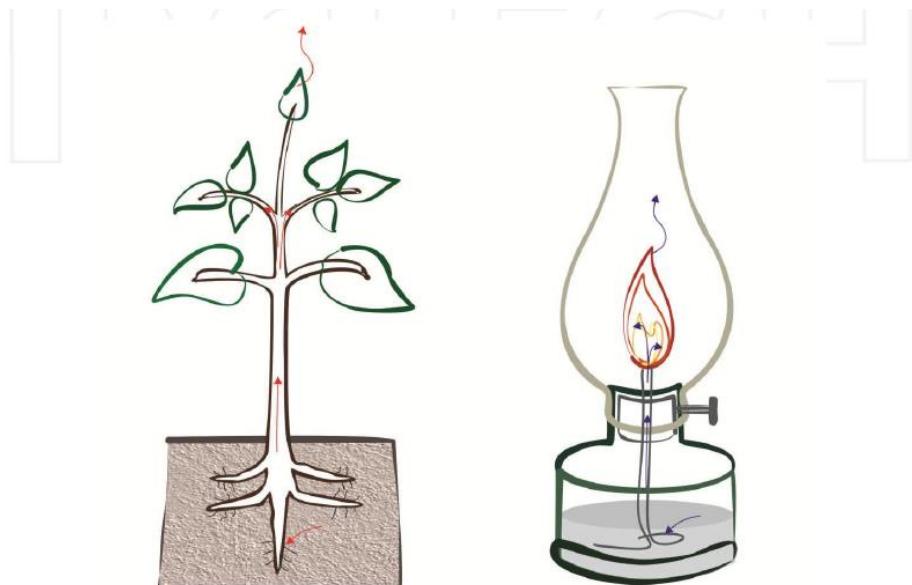
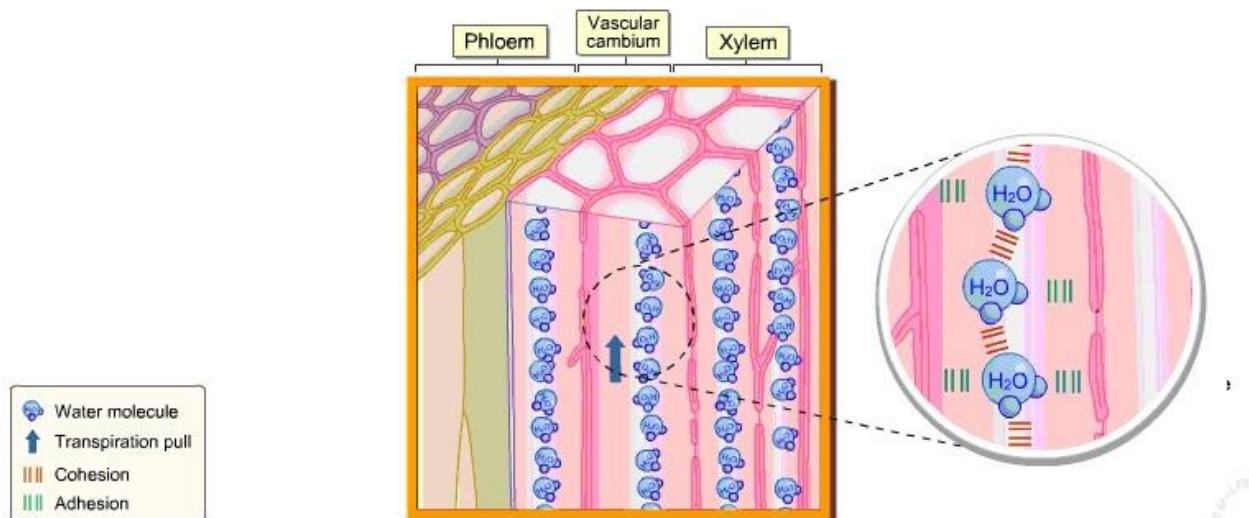


Fig. 1. Schematic representation of water flow through the plant (arrows), by analogy with the oil flow through the wick of an old fashion lampion.

The capillary force is very minimal and it is unable to lift water to very high levels as in the case of a tree. People use paper towels (and thus, capillary action) to wipe up liquid spills.

(B) Transpiration Pull (The cohesion Tension)

During transpiration water evaporates from the leaf and escape through the stomata. The water column is continuous from root hair till the stomata. As one molecule escapes, another one is pulled into the xylem. This is because of cohesive forces.



(C) Root pressure

A pressure is developed in xylem vessels when water is absorbed by root hair cells. This is the root pressure. Roots generate positive hydrostatic pressure by absorbing

ions from the soil. The absorption of ion requires energy ie ATP. Hence there is a **metabolic pump**. ie the energy through the respiration of root hair cells itself.

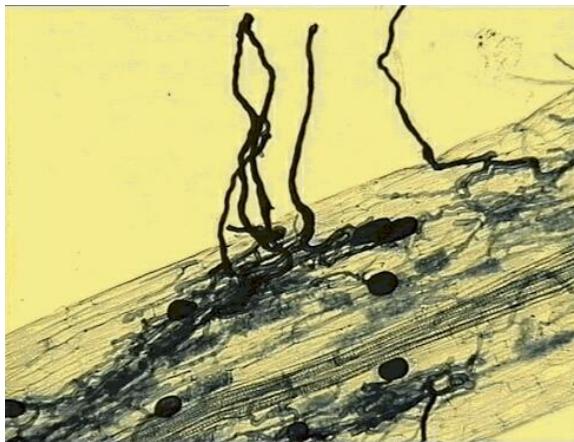
Now do you think that the from capillary water is in soil till the stomatal end is a continuous column? This is the Concept of the Soil-Plant-Atmosphere Continuum. Now look at the speed

Plant type	Speed (m h ⁻¹)
Evergreen conifers	1.2
Mediterranean sclerophylls	0.4–1.5
Deciduous diffuse-porous trees	1–6
Deciduous ring-porous trees	4–44
Herbaceous plants	10–60
Lianas	150

In plants all cells are not living. It is a system or combination both dead and living cells emerged as an efficient mechanism of water transport!

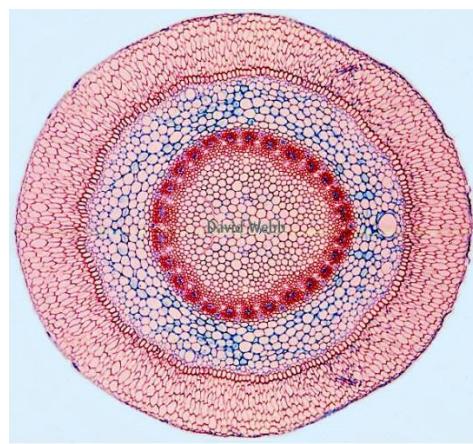
Mechanisms of enhancing efficiency of water absorption in plants

Mycorrhizae: This is a symbiotic relationship formed between root cells and fungus. The fungus helps in water absorption and mineral absorption. The fungal filaments are much finer than root hairs. Hence they can reach further areas of soil where root hair is unable to reach. In return plant cells provide food and shelter for fungus. This association can be found in most of the plants.



Velamen tissue: Many orchids are epiphytes. How do they absorb water? Orchids are equipped with special roots known as velamen roots. The outer layer of these roots are made up of **velamen tissue**. This is a dead tissue like xylem. Velamen tissue directly absorbs moisture from the atmosphere by imbibition. Now the nearby living cells absorb water from the velamen tissue.

Water absorption in hydrophytes: In these plants the function of root is merely restricted for anchorage. They do not have root hairs. Xylem is poorly developed in Hydrophytes as the water absorption takes place all over the body. In these plants stomata are inactive.

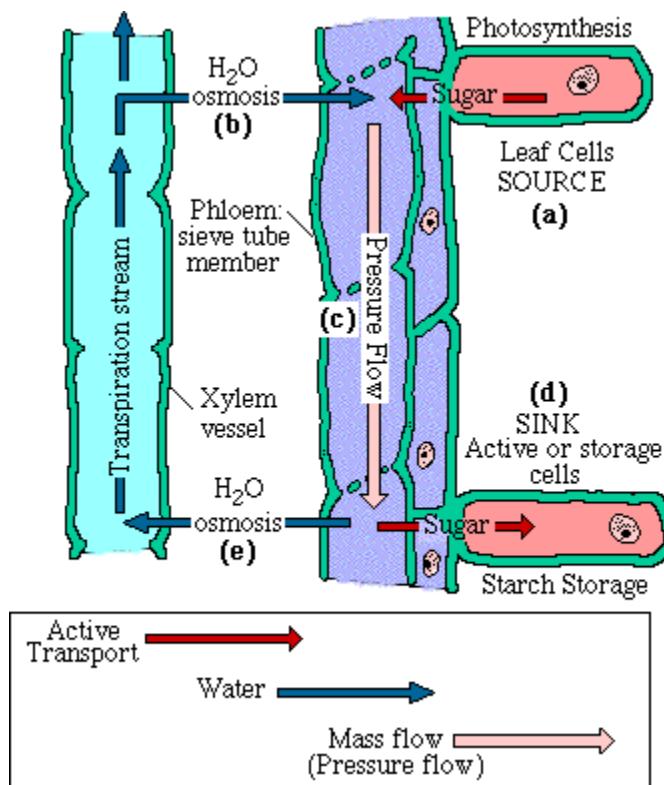


Mechanisms of phloem transport

In contrast to xylem phloem is a living tissue. It has cross walls. It transports the sap which contains mainly sugar apart from hormones. In phloem sap can be transported in both directions. How?

This can be explained on the basis of “Source sink” relationship. The leaves are the source of sugar. The sink is the consuming end may be a root which stores it. Phloem sap can be transported in any direction needed so long as there is a source of sugar and a sink able to use. The source and sink may be reversed depending on the season, or the plant's needs. Sugar stored in roots may be mobilized to become a source of food in the early spring when the buds of trees, the sink, need energy for growth and development of the photosynthetic apparatus.

The accepted mechanism needed for the translocation of sugars from source to sink is called the **pressure flow hypothesis**.



The movement of sugars in the phloem begins at the source, where (a) sugars are loaded (actively transported) into a sieve tube. Loading of the phloem sets up a water potential gradient that facilitates the movement of water into the dense

phloem sap from the neighboring xylem (b). As hydrostatic pressure in the phloem sieve tube increases, pressure flow begins (c), and the sap moves through the phloem. Meanwhile, at the sink (d), incoming sugars are actively transported out of the phloem and removed as complex carbohydrates. The loss of solute produces a high water potential in the phloem, and water passes out (e), returning eventually to the xylem.

Can xylem and phloem gets blocked? Whether air bubbles are forming there?

Xylem and phloem can be blocked due to physical blockages with secondary metabolites such as phenolics. Blockages can also be possible with viral particles or phytoplasma. This can lead to wilt of the plant and death.

Water moving up in the xylem or up and down in phloem has to be continuous channel. If any air bubbles are formed, water fails to move.