uc3m | Universidad Carlos III de Madrid

OpenCourseWare (2023)

CHEMISTRY II

Verónica San Miguel Arnanz

Teresa Pérez Prior

Berna Serrano Prieto

Department of Materials Science and Engineering and Chemical Engineering

Structure and Properties of Macromolecules

1. CARBOHYDRATES

Carbohydrates, also known as saccharides, represent the most abundant class of biological molecules on Earth when considering mass. While all organisms have the capacity to synthesize carbohydrates, a significant portion is generated by photosynthetic entities such as bacteria, algae, and plants. These organisms capture solar energy through photosynthesis, converting it into stored carbohydrates. Consequently, carbohydrates serve as the metabolic building blocks for nearly all other biomolecules. The breakdown of carbohydrates releases energy crucial for sustaining animal life. Moreover, carbohydrates form covalent bonds with various molecules. Carbohydrates, identified as polyhydroxy aldehydes or ketones, or compounds yielding such substances upon hydrolysis, comprise a versatile molecular class. Their covalent linkage with lipids results in glycolipids, integral components of biological membranes. Proteins with covalently attached carbohydrates are termed glycoproteins.

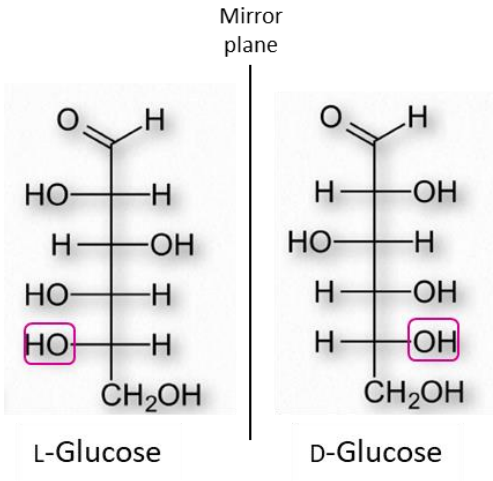
The characterization of carbohydrates is based on the number of monomeric units they encompass. Monosaccharides, the smallest structural units, defy further breakdown under mild conditions. The term "carbohydrate," derived from "hydrate of carbon," corresponds to the empirical formula with n typically being 5 or 6, though it can extend to 9. Oligosaccharides, polymer chains containing 2 to around 20 monosaccharide residues, frequently manifest as disaccharides or trisaccharides. Oligosaccharides, often covalently linked to molecules like glycoproteins, exhibit characteristic glycosidic bonds between monosaccharide units. Polysaccharides, comprising more than 20 monosaccharide residues, differ from oligosaccharides because water is eliminated during their polymer formation. They are versatile polymers, existing as linear or branched structures, and can consist of hundreds to thousands of monosaccharide units. Molecular weights for polysaccharides can extend up to a million or more.

1.1. Chemical and Physical Properties

Monosaccharides

Monosaccharides exhibit characteristics of colorless crystalline solids, displaying high solubility in water while remaining insoluble in nonpolar solvents. The fundamental structure of common monosaccharides comprises unbranched carbon chains, wherein single bonds interconnect all carbon atoms. In the open-chain configuration, one carbon atom forms a carbonyl group through a double bond with an oxygen atom, while a hydroxyl group accompanies each remaining carbon. The positioning of the carbonyl group determines the classification of the monosaccharide: if at the end of the carbon chain, it is an *aldose*, and if at any other position, it is a *ketose*.

Aldoses, featuring a minimum of three carbons, and ketoses, with a minimum of four carbons, possess chiral centers. Carbon atom numbering starts from the aldehydic carbon, designated as 1 (Figure 1). Conventionally, sugars adopt the D configuration when the chiral carbon with the highest number, farthest from the carbonyl carbon, has the hydroxyl group on the right (dextro) in a projection formula with the carbonyl carbon at the top. Conversely, the L isomer is identified when the hydroxyl group is on the left (levo) in the same projection formula.

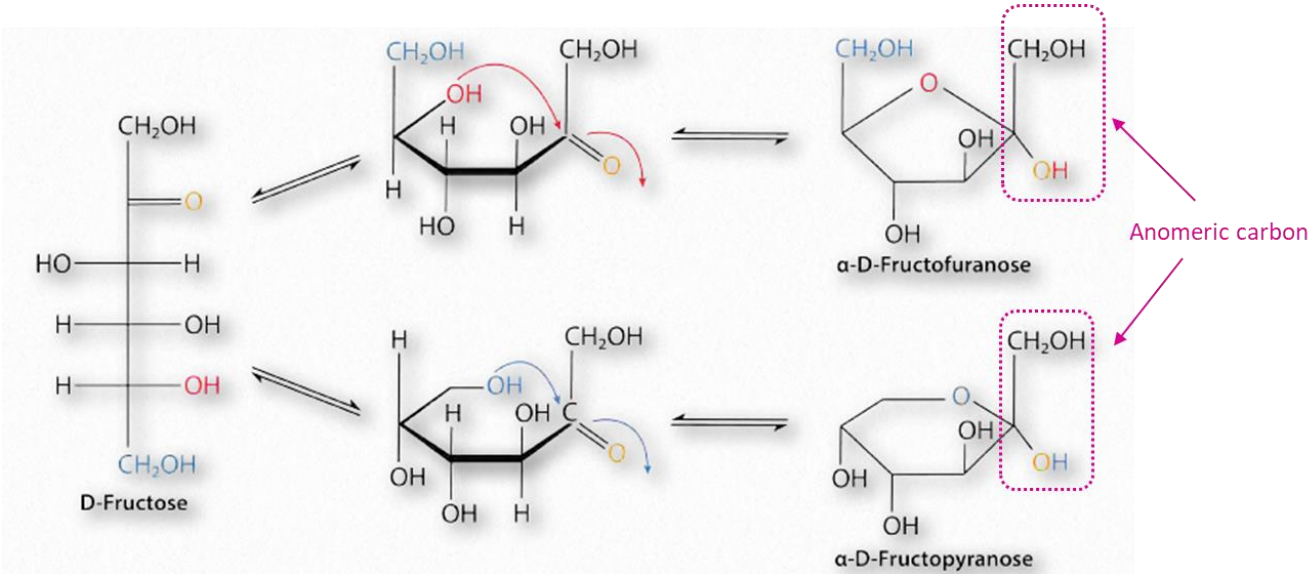


Source: Biochemistry: Free For All, 2018. Kevin Ahern, Indira Rajagopal, and Taralyn Tan, Oregon State University. https://open.umn.edu/opentextbooks/textbooks/866.

Figure 1. Fischer projection of the aldose Glucose.

Monosaccharides with five and six carbons readily undergo cyclization, creating a new asymmetric carbon known as the *anomeric carbon*. Like other asymmetric carbons in sugars, this anomeric carbon can exhibit the hydroxyl group in two positions referred to as α and β. Sugars such as α-D-glucose and β-D-glucose, differing solely in the configuration of the anomeric carbon, are termed anomers. Rings formed with five atoms are termed furanoses (named for furan), while those with six atoms, like glucose, are referred to as pyranoses (named for pyran). The carbonyl carbon in the ring structure becomes the anomeric carbon, binding to the oxygen of a hydroxyl elsewhere in the chain. α- and β- forms of a given sugar can readily interconvert in solution, provided the anomeric hydroxyl is free. Most pentoses and hexoses can adopt both furanose and pyranose structures. The α and β anomers of D-glucose undergo mutarotation in aqueous solution, briefly opening into the linear form and then closing again to produce the β anomer. Consequently, a solution of β-D-glucose and a solution of α-D-glucose eventually form identical equilibrium mixtures with identical optical properties.

Similarly, ketones can react with alcohols to form hemiketals. The intramolecular reaction of a ketose sugar, such as fructose, results in a cyclic hemiketal (Figure 2). The equilibrium mixture in an aqueous solution predominantly consists of the cyclic pyranose or furanose forms, with the linear aldehyde or ketone structure being a minor component. A furanose refers to carbohydrates with a five-membered ring system, and its configuration, either D or L, is determined by the highest numbered chiral carbon. The furanose ring may exhibit α or β configuration, depending on the direction of the anomeric hydroxy group, and undergoes mutarotation in solution, resulting in an equilibrium mixture of α-β configurations.



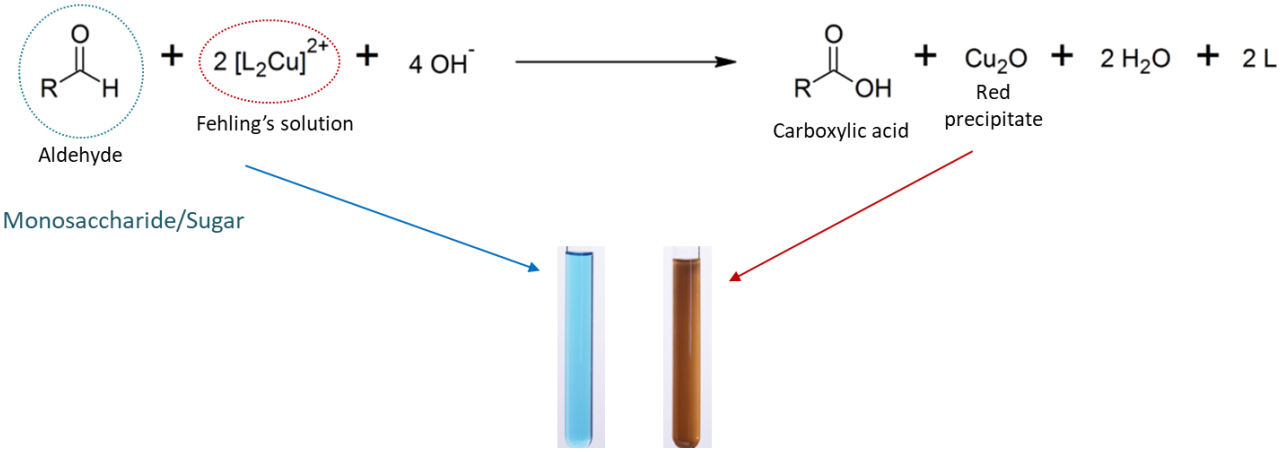
Source: Biochemistry: Free For All, 2018. Kevin Ahern, Indira Rajagopal, and Taralyn Tan, Oregon State University. https://open.umn.edu/opentextbooks/textbooks/866.

Figure 2. Cyclation of a ketose.

Monosaccharides undergo oxidation through the action of mild oxidizing agents, such as cupric (Cu^{2+}) ions. This oxidation process transforms the carbonyl carbon into a carboxyl group. Sugars, including glucose, capable of reducing cupric ions are classified as reducing sugars. This oxidation results in a complex mixture of carboxylic acids.

Fehling's reaction serves as the foundation for a semiquantitative test designed to detect the presence of reducing sugars. This test has been employed for many years to identify and measure elevated glucose levels in individuals with diabetes mellitus. Sugars possessing free anomeric carbon atoms exhibit notable reducing agent properties, leading to the reduction of various oxidizing agents, including hydrogen peroxide, ferricyanide, certain metals such as Cu and Ag, and others. These reactions lead to the conversion of the sugar into a sugar acid. For instance, adding alkaline $CuSO_4$ to an aldose sugar induces the formation of a red cuprous oxide (Cu_2O) precipitate. It converts the aldose into an aldonic acid, such as gluconic acid. The appearance of a red Cu_2O precipitate constitutes a positive test for the presence of an aldehyde (Figure 3).

Carbohydrates exhibiting the ability to reduce oxidizing agents in this manner are termed reducing sugars. By quantifying the amount of oxidizing agent reduced by a sugar solution, one can accurately determine the concentration of the sugar.



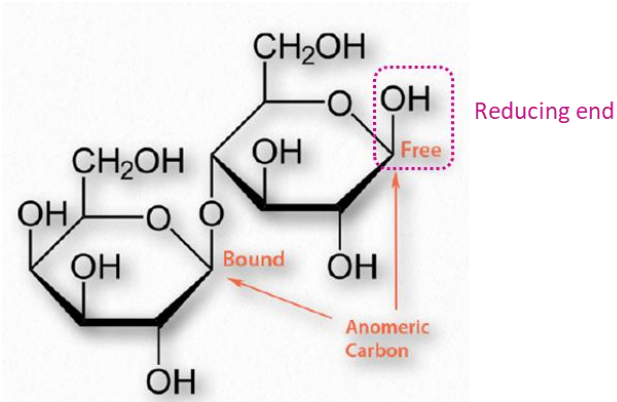
Sources:

- Fehling's test reaction: Srikeit, CC BY-SA 3.0, https://commons.wikimedia.org/wiki/File:Fehling_test.png.
- Fehling's test image: Science Photo Library, CC BY-NC 4.0, https://ndla.no/subject:1:721307df-c384-4a7f-ad69-668853c766c6/topic:9:6db04aa9-8de7-4bc3-8981-04daf4335635/resource:79a46804-ffb3-4882-bb1c-59c606d046fd.

Figure 3. Fehling reaction: monosaccharide is oxidized and cupric ion is reduced.

Disaccharides

Disaccharides form when the anomeric carbon of one sugar molecule interacts with one of the hydroxyl groups in another sugar molecule. Consequently, in the case of disaccharides and other carbohydrate polymers, it is essential to consider both the types of monosaccharide residues present and the atoms involved in glycosidic bonds. The disaccharide lactose depicted in the Figure 4 is a mixed acetal, featuring one intramolecularly provided hydroxyl group and one hydroxyl from the other monosaccharide.



Source: Principles of Biology: Biology 211, 212, and 213, 2017. Lisa Bartee, Walter Shriner, and Catherine Creech, Open Oregon Educational Resources, https://open.umn.edu/opentextbooks/textbooks/989.

Figure 4. Disaccharide Lactose formed from glucose and galactose.

The oxidation of a sugar by cupric ion, defining a reducing sugar, only occurs with the linear form, which exists in equilibrium with the cyclic form(s). When the anomeric carbon is part of a glycosidic bond (as in a full acetal or ketal), the easy interconversion between linear and cyclic forms is impeded. Since the carbonyl carbon can undergo oxidation solely when the sugar is in its linear form, forming a glycosidic bond renders a sugar nonreducing. In the description of disaccharides or polysaccharides, the end of a chain with a free anomeric carbon (not involved in a glycosidic bond) is commonly referred to as the reducing end.

Polysaccharides

Polysaccharides, also known as glycans, exhibit variations in their monosaccharide unit composition, chain length, types of bonds linking the units, and degree of branching. They are broadly categorized into two classes. Homoglycans, or homopolysaccharides, are polymers composed of residues of only one type of monosaccharide. On the other hand, heteroglycans, or heteropolysaccharides, are polymers containing residues of more than one type of monosaccharide.

Furthermore, most polysaccharides can be categorized based on their biological functions, falling into either storage or structural roles. For instance, starch and glycogen are storage polysaccharides, while cellulose and chitin serve as structural polysaccharides. Storage polysaccharides act as readily metabolizable food, providing energy reserves for cells. On the other hand, structural polysaccharides offer robust support for the skeletons of arthropods and green plants, respectively.

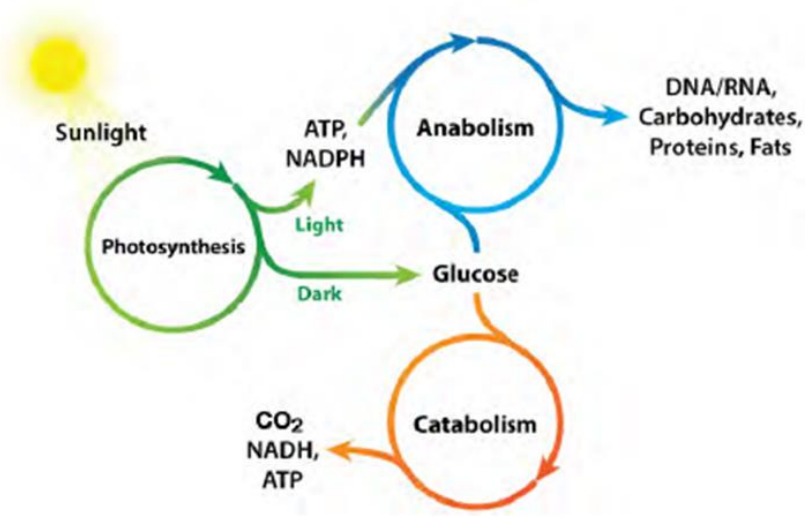
1.2. Metabolism

Metabolic regulation, a focal point in biochemistry, stands out as one of the most intriguing characteristics of living organisms. Among the many enzyme-catalyzed reactions possible within a cell, every single one likely undergoes some form of regulation. The imperative to regulate every facet of cellular metabolism becomes evident when delving into the intricacies of metabolic reaction sequences.

Metabolism encompasses the entire network of chemical reactions carried out by living cells. Metabolites, small molecules that serve as intermediates in the degradation or biosynthesis of biopolymers, play a crucial role. The term "intermediary metabolism" describes reactions involving these low-molecular-weight molecules. It is practical to distinguish between reactions that synthesize molecules (anabolic reactions) and those that degrade molecules (catabolic reactions). Anabolic reactions synthesize compounds necessary for cell maintenance, growth, and reproduction. These biosynthesis reactions generate simple metabolites like amino acids, carbohydrates, coenzymes, nucleotides, and fatty acids, as well as larger molecules such as proteins, polysaccharides, nucleic acids, and complex lipids. In some species, all complex molecules constituting a cell are synthesized from inorganic precursors (carbon dioxide, ammonia, inorganic phosphates, etc.). Certain species derive energy from these inorganic molecules or from establishing a membrane potential. Photosynthetic organisms utilize light energy to drive biosynthesis reactions.

On the other hand, catabolic reactions break down large molecules to release smaller molecules and energy. These reactions also degrade small molecules to inorganic products. While all cells perform degradation reactions as part of normal cell metabolism, some species rely on them as their primary energy source. For instance, animals require organic molecules as food, and the ultimate source of these organic molecules is a biosynthetic pathway in another species. It is crucial to note that all catabolic reactions involve the breakdown of compounds synthesized by a living cell—whether it's the same cell, a different cell in the same individual, or a cell in a different organism. In addition to the energy needed for biosynthesis, organisms require energy for other cellular activities such as transport and movement.

The Figure 5 illustrates anabolism and catabolism. Anabolic reactions utilize small molecules and chemical energy to synthesize organic molecules and for cellular work. Solar energy is a significant metabolic energy source in photosynthetic bacteria and plants. Some molecules, including those obtained from food, undergo catabolism to release energy and produce either monomeric building blocks or waste products. Blue arrows represent biosynthesis pathways, and red arrows represent catabolism pathways.



Source: Biochemistry: Free For All, 2018. Kevin Ahern, Indira Rajagopal, and Taralyn Tan, Oregon State University. https://open.umn.edu/opentextbooks/textbooks/866.

Figure 5. Anabolic and Catabolic reactions.

Glycolysis

Glucose plays a pivotal role in the metabolism of plants, animals, and numerous microorganisms due to its relatively high potential energy, making it an excellent fuel. Beyond being a reliable fuel source, glucose is a remarkably versatile precursor, providing a diverse array of metabolic intermediates for various biosynthetic reactions.

Glycolysis, the first elucidated metabolic pathway and likely the best understood, is nearly universal as a central pathway in glucose catabolism, exhibiting the largest carbon flux in most cells. In glycolysis, a molecule of glucose undergoes degradation in a series of enzyme-catalyzed reactions, ultimately yielding two molecules of the three-carbon compound pyruvate. Throughout the sequential reactions of glycolysis, some of the free energy released from glucose is conserved in the form of ATP and NADH.

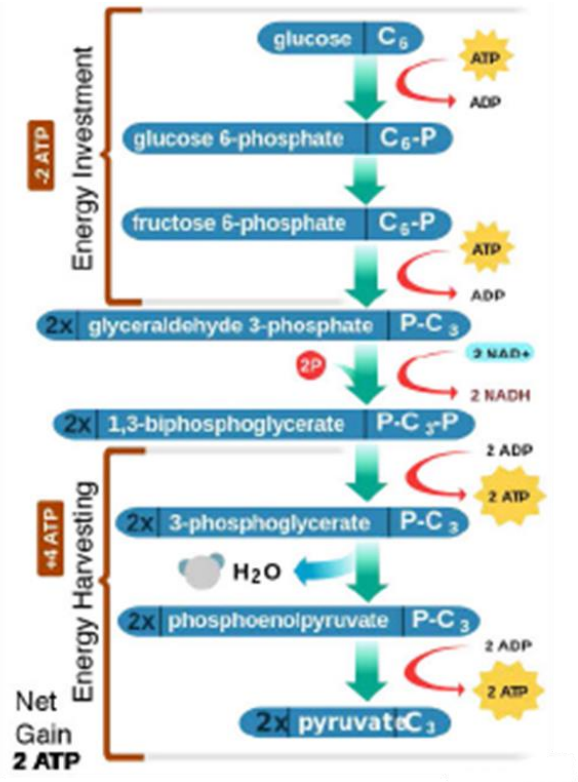
Glycolysis occurs in two phases: the preparatory phase and the payoff phase (Figure 6). The preparatory phase involves the breakdown of glucose into glyceraldehyde 3-phosphate, and the energy gain occurs in the subsequent payoff phase. For each glucose molecule, two ATP are consumed in the preparatory phase, and four ATP are produced in the payoff phase, resulting in a net yield of two ATP per glucose molecule converted to pyruvate. This energy is largely conserved through the phosphorylation of four molecules of ADP to ATP. Additionally, during glycolysis, two molecules of the electron carrier NADH are formed per glucose molecule.

Preparatory phase

Phosphorylation of glucose and its conversion to glyceraldehyde 3-phosphate

Payoff phase

Oxidative conversion of glyceraldehyde 3-phosphate to pyruvate and the coupled formation of ATP and NADH



Source: Survey of Biochemistry and Biotechnology, A. Goodman, California Polytechnic State University San Luis Obispo, https://commons.libretexts.org/book/chem-347272.

Figure 6. Glucose catabolism, Glycolysis.

The overall equation for glycolysis reveals that for each glucose molecule degraded to pyruvate, two molecules of ATP are generated from ADP and Pi, and two molecules of NADH are produced by the reduction of NAD^+. The hydrogen acceptor in this reaction is NAD^+, which undergoes reduction to form the reduced coenzyme NADH.

Overall reaction:

$$Glucose + 2\ NAD^+ + 2\ ADP + 2\ P_i \rightarrow 2\ pyruvate + 2\ NADH + 2\ H^+ + 2\ ATP + 2\ H_2O$$

$$Glucose + 2\ NAD^+ \rightarrow 2\ pyruvate + 2\ NADH + 2\ H^+ \ \Rightarrow\ \Delta G_1'^0 = -146\ kJ/mol$$

$$2\ ADP + 2\ P_i \rightarrow 2\ ATP + 2\ H_2O \ \Rightarrow\ \Delta G_2'^0 = 2(30.5\ kJ/mol) = +61.0\ kJ/mol$$

$$\Delta G_{sum}'^0 = \Delta G_1'^0 + \Delta G_2'^0 = -85\ kJ/mol$$

Gluconeogenesis

Gluconeogenesis, the process of converting pyruvate and related three- and four-carbon compounds into glucose, occurs universally in animals, plants, fungi, and microorganisms. Although the reactions are essentially the same in all tissues and species, the metabolic context and regulation of the pathway vary between species and tissues.

All species have the ability to synthesize glucose from simple two-carbon and three-carbon precursors through gluconeogenesis, which means the formation of new glucose. Certain species,

particularly photosynthetic organisms, can generate these precursors by fixing carbon dioxide, resulting in the net synthesis of glucose from inorganic compounds. It's important to note in our discussion of gluconeogenesis that every glucose molecule used in glycolysis had to be synthesized in some species. The pathway for gluconeogenesis shares several steps with glycolysis, the pathway for glucose degradation. However, four reactions specific to the gluconeogenic pathway replace the three metabolically irreversible reactions of glycolysis. This opposing set of reactions exemplifies separate, regulated pathways for synthesis and degradation.

The availability of glucose is controlled by regulating the uptake and synthesis of glucose and related molecules, as well as the synthesis and degradation of storage polysaccharides composed of glucose residues, such as glycogen in bacteria and animals and starch in plants. Glycogen and starch can be intracellularly degraded to release glucose monomers, which can either fuel energy production through glycolysis or serve as precursors in biosynthetic reactions.

Gluconeogenesis serves as the anabolic counterpart to glycolysis and primarily occurs in the cells of the liver and kidneys, with little involvement in other cells in the body. While seven of the eleven reactions in gluconeogenesis use the same enzymes as glycolysis, the reaction directions are reversed. The ΔG values of these reactions in the cell are typically near zero, allowing their direction to be controlled by small changes in substrate and product concentrations. Notably, the three regulated enzymes of glycolysis catalyze reactions whose cellular ΔG values are not close to zero, making the manipulation of reaction direction non-trivial. To favor gluconeogenesis when necessary, cells employ "work-around" reactions catalyzed by different enzymes.

The synthesis of one glucose molecule from two pyruvate molecules requires four ATP and two GTP molecules (in animals), along with two molecules of NADH. For each glucose molecule formed, six high-energy phosphate groups are required—four from ATP, two from GTP, and two molecules of NADH are needed to reduce two molecules of 1,3-bisphosphoglycerate. This process is not a simple reversal of glycolysis, which would only require two ATP molecules. Despite its high energy cost, gluconeogenesis is necessary to ensure its irreversibility.

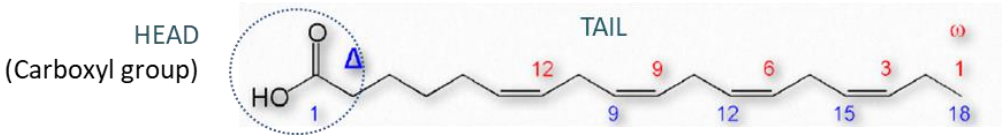
$$2\,Pyruvate + 2\,NADH + 4\,ATP + 2\,GTP + 6\,H_2O + 2\,H^+ \rightarrow Glucose + 2\,NAD^+ + 4\,ADP + 2\,GDP + 6\,P_i$$

2. LIPIDS

2.1. Classification

Fatty Acids

A fatty acid consists of a lengthy hydrocarbon chain (referred to as the "tail") and a terminal carboxyl group (or "head") (Figure 7). Fatty acids are carboxylic acids featuring hydrocarbon chains spanning from 4 to 36 carbons long. The hydrocarbon chain in some fatty acids is unbranched and fully saturated (containing no double bonds); in others, it has one or more double bonds. Variations among fatty acids include the length of their hydrocarbon tails, the number and positions of carbon–carbon double bonds, and the presence of branches.

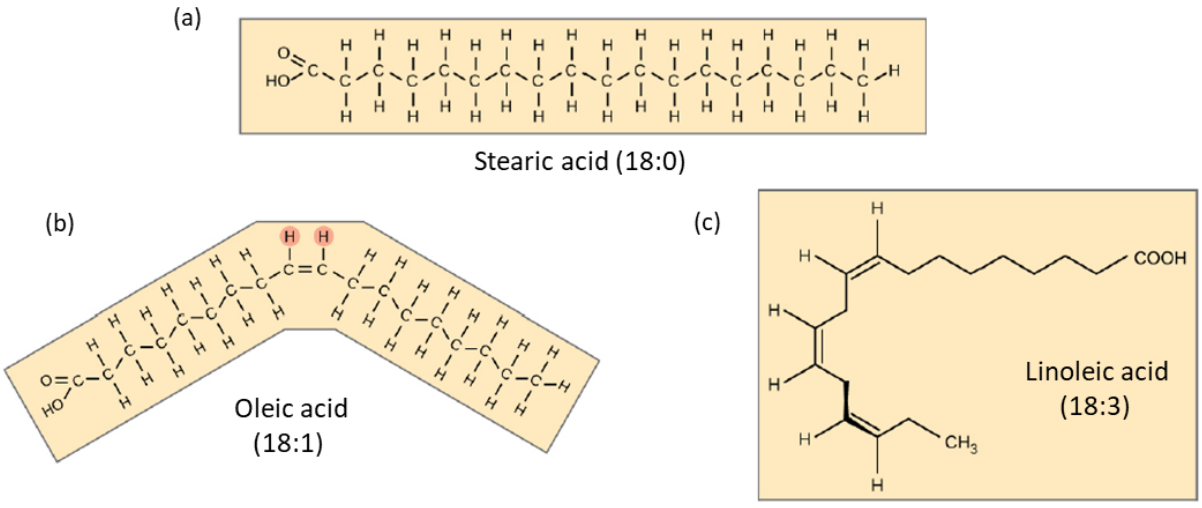


Source: Survey of Biochemistry and Biotechnology, A. Goodman, California Polytechnic State University San Luis Obispo, https://commons.libretexts.org/book/chem-347272.

Figure 7. Fatty acid structure.

Fatty acids lacking carbon–carbon double bonds are termed saturated, whereas those with at least one such bond are termed unsaturated. Monounsaturated fatty acids have a single carbon–carbon double bond, and polyunsaturated ones have two or more. Most fatty acids have a pKa of about 4.5 to 5.0, ionizing them at physiological pH. Acting as detergents due to their long hydrophobic tail and polar head, fatty acids exhibit low concentrations in cells to avoid membrane disruption. Detergents, classified as surfactants, can reduce water surface tension, facilitating the mixing of hydrophobic compounds (like oil and grease) with water.

Saturated and unsaturated fatty acids exhibit distinct physical properties. Typically, saturated fatty acids are solid at room temperature (22°C), resembling waxy substances, while unsaturated fatty acids are liquid under the same conditions. The length and degree of unsaturation in a fatty acid's hydrocarbon chain influence its melting point. As the hydrocarbon tail lengthens, the melting points of saturated fatty acids rise due to increased van der Waals interactions among neighboring tails, demanding more energy to disrupt these interactions. A comparison of stearate (18:0), oleate (18:1), and linolenate (18:3) structures in Figure 8 reveal their differences. The flexible saturated hydrocarbon tail of stearate allows rotation around every carbon–carbon bond, promoting close packing in a stearic acid crystal. Oleate and linolenate, with *cis* double bonds causing bends in their hydrocarbon chains, hinder rotation and prevent tight packing. As a result, cis-unsaturated fatty acids have lower melting points than saturated ones. With increased unsaturation, fatty acids become more fluid. For instance, stearic acid (melting point 70 °C) is solid at body temperature, while oleic acid (melting point 13 °C) and linolenic acid (melting point –17 °C) are both liquids.

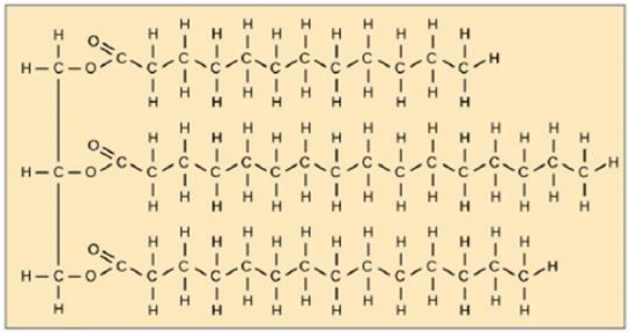


Source: Biology 2e, 2018. Mary Ann Clark, Matthew Douglas, Jung Choi, Rice University, https://openstax.org/details/books/biology-2e.

Figure 8. Saturated and unsaturated fatty acids structures.

Triacylglycerols

As implied by their name, triacylglycerols consist of three fatty acyl residues esterified to glycerol, a three-carbon alcohol (refer to Figure 9). Owing to their long hydrocarbon chains, triacylglycerols are highly hydrophobic. Consequently, they can be stored in cells in an anhydrous form, meaning the molecules are not solvated by water. This characteristic avoids taking up space and adding mass, ensuring energy storage efficiency.



Source: Principles of Biology: Biology 211, 212, and 213, 2017. Lisa Bartee, Walter Shriner, and Catherine Creech, Open Oregon Educational Resources, https://open.umn.edu/opentextbooks/textbooks/989.

Figure 9. Triacylglycerol structure.

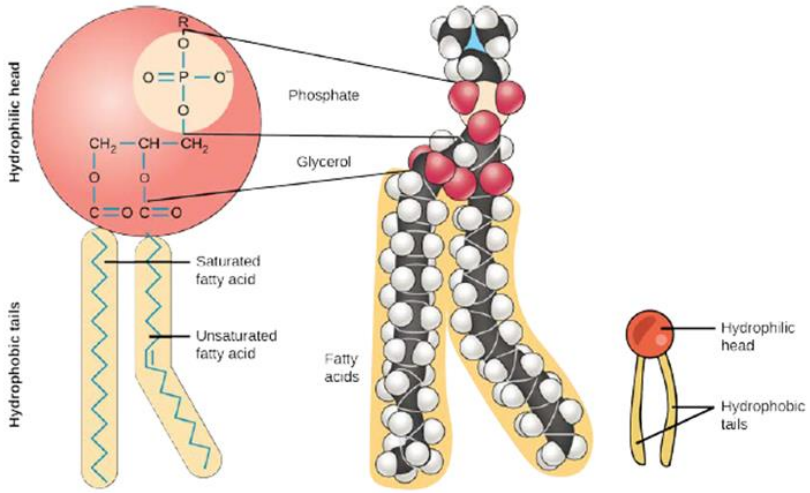
Triacylglycerols can exist in either solid (fats) or liquid (oils) states, depending on their fatty acid compositions and temperature conditions. Those containing only saturated long-chain fatty acyl groups tend to be solids at body temperature, while those with unsaturated or short-chain fatty acyl groups tend to be liquids.

Hydrolysis of acylglycerols can occur through heating with acid or base or treatment with lipases (enzyme). Alkali-induced hydrolysis, termed saponification, yields free fatty acids and glycerol salts. This process was historically used to produce soap, involving the use of potassium hydroxide (potash) leached from wood ashes to hydrolyze animal fat, predominantly triacylglycerols.

Triacylglycerols coalesce into fat droplets within most cells, often observed near mitochondria in cells relying on fatty acids for metabolic energy. In mammals, adipose tissue, consisting of specialized adipocytes, is the primary fat storage site. Each adipocyte contains a large fat droplet that occupies most of the cell volume. Adipose tissue is distributed throughout the bodies of mammals, with significant deposits just under the skin and in the abdominal cavity. Subcutaneous fat, especially prevalent in aquatic mammals, functions as both an energy storage depot and thermal insulation.

Glycerophospholipids

Generally, glycerophospholipids exhibit a structural pattern where saturated fatty acids are esterified to C-1, while unsaturated fatty acids are esterified to C-2 of the glycerol backbone. The defining characteristic is including a phosphate group at C-3 of the glycerol structure (Figure 10). These lipids play pivotal roles as integral components of cell membranes and are present in modest concentrations in various cellular compartments.



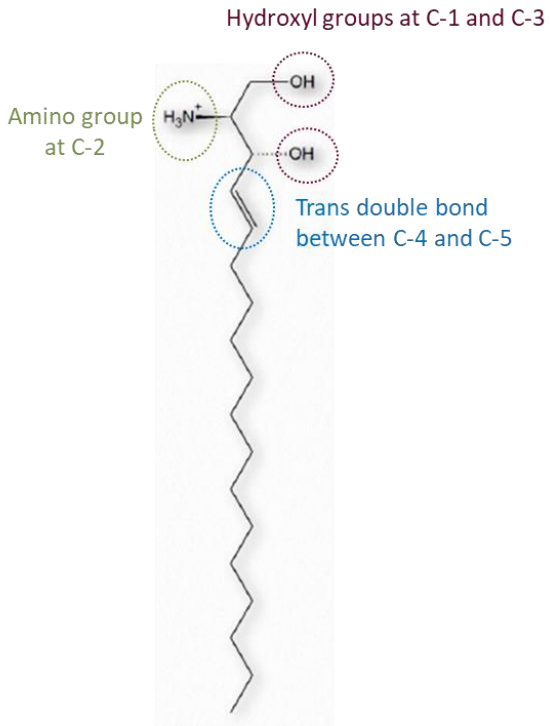
Source: Biology 2e, 2018. Mary Ann Clark, Matthew Douglas, Jung Choi, Rice University, https://openstax.org/details/books/biology-2e.

Figure 10. Glycerophospholipid structure.

Following glycerophospholipids in abundance within plant and animal membranes are sphingolipids.

Sphingolipids

In mammals, sphingolipids exhibit pronounced abundance in central nervous system tissues. It's worth noting that most bacteria lack sphingolipids. Sphingolipids feature sphingosine (trans-4-sphingenine) as their structural backbone—a linear alcohol with 18 carbons, a trans double bond between C-4 and C-5, an amino group at C-2, and hydroxyl groups at C-1 and C-3 (refer to Figure 11).



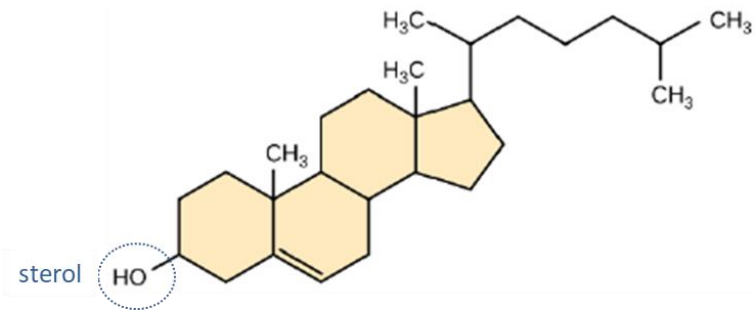
Source: Biochemistry: Free For All, 2018. Kevin Ahern, Indira Rajagopal, and Taralyn Tan, Oregon State University. https://open.umn.edu/opentextbooks/textbooks/866.

Figure 11. Sphingolipid structure.

Despite their relatively low presence in most membranes, glycosphingolipids, a subtype of sphingolipids, play crucial roles in various cellular functions. Particularly at cell surfaces, glycosphingolipids contribute to determining tissue and organ specificity elements. Cell–cell recognition and tissue immunity seem to rely on, in part, on specific glycosphingolipids.

Steroids

Steroids boast a distinctive structure characterized by four fused rings—three six-carbon rings labeled A, B, and C, along with a five-carbon D ring. Among steroids, cholesterol holds particular significance as a crucial component of animal plasma membranes. Cholesterol, categorized as a sterol, features a hydroxyl group at C-3 (Figure 12).



Source: Biology 2e, 2018. Mary Ann Clark, Matthew Douglas, Jung Choi, Rice University, https://openstax.org/details/books/biology-2e.

Figure 12. Cholesterol structure.

Notably, cholesterol tends to accumulate in lipid deposits (plaques) on the walls of blood vessels, and these plaques have been linked to cardiovascular disease, which can lead to heart attacks or strokes. Despite its association with cardiovascular issues, cholesterol remains indispensable in mammalian biochemistry. It not only serves as a component of specific membranes but also acts as an essential precursor for the synthesis of steroid hormones and bile salts.

2.2. Oxidation of Fatty Acids

Fatty acids undergo a complex oxidation process that occurs in the mitochondria of cells and consists of multiple stages. Here is a summary of the critical stages of the fatty acid oxidation process:

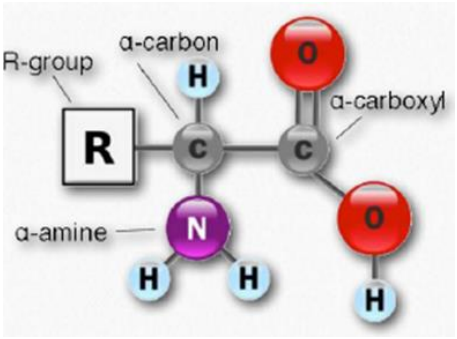
- *Activation of Fatty Acid*: The fatty acid is activated by adding a coenzyme A (CoA) molecule to form a compound called Acyl-CoA. This process consumes one molecule of ATP.
- *Transport into the Mitochondria*: Acyl-CoA is transported into the mitochondrial matrix through a specialized transporter.
- *Beta-Oxidation*: Beta-oxidation is the pivotal stage of fatty acid oxidation. In this stage, Acyl-CoA is cleaved into two-carbon units (acetyl-CoA) through a series of reactions involving oxidation, hydration, and oxidation once again. This process produces one molecule of $FADH_2$ and one molecule of NADH for each complete beta-oxidation cycle.

- *Energy Production*: The acetyl-CoA produced by beta-oxidation undergoes oxidation in the Krebs cycle, also known as the citric acid cycle, producing more molecules of $FADH_2$ and NADH. These molecules are utilized in the electron transport chain to generate ATP through oxidative phosphorylation.

3. PROTEINS

Proteins serve as essential agents in biological functions, with amino acids acting as the fundamental building blocks. The remarkable diversity observed in the multitude of natural proteins stems from the intrinsic properties of only 20 commonly occurring amino acids. These properties encompass the ability to polymerize, distinctive acid–base characteristics, diverse structure and chemical functionality in the amino acid side chains, and chirality.

The structure of a typical amino acid is illustrated in the Figure 13. At the core of this structure lies the tetrahedral alpha carbon, covalently linked to the amino and carboxyl groups. Additionally, a hydrogen and a variable side chain are bonded to this alpha carbon. The identity of each amino acid is determined by the side chain, often referred to as the R group.

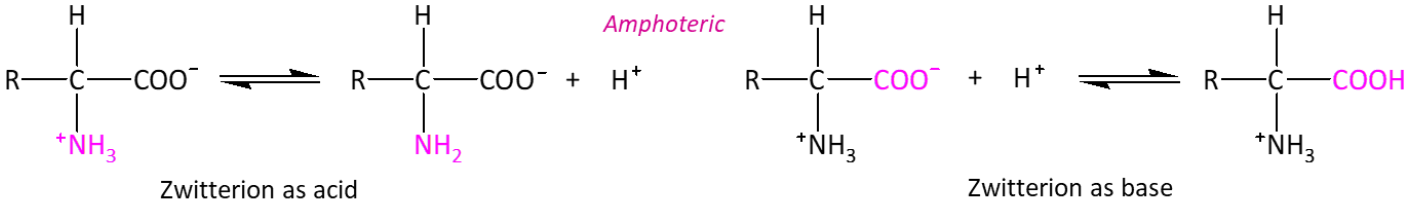


Source: Survey of Biochemistry and Biotechnology, A. Goodman, California Polytechnic State University San Luis Obispo, https://commons.libretexts.org/book/chem-347272.

Figure 13. Amino acid structure.

Amino acids exhibit chirality, molecules with four groups attached to the alpha carbon. This asymmetric -carbon gives rise to two non-identical mirror image isomers or enantiomers.

The amino and carboxyl groups of amino acids and the ionizable R groups of certain amino acids act as weak acids and bases. When an amino acid lacking an ionizable R group is dissolved in water at neutral pH, it takes on the form of a dipolar ion, or zwitterion (derived from the German term for "hybrid ion"), which is electrically neutral but possesses formal positive and negative charges on different atoms, capable of functioning as both an acid and a base (Figure 14). Substances with this dual (acid-base) nature are termed amphoteric and are commonly referred to as ampholytes, signifying their amphoteric electrolytic properties.

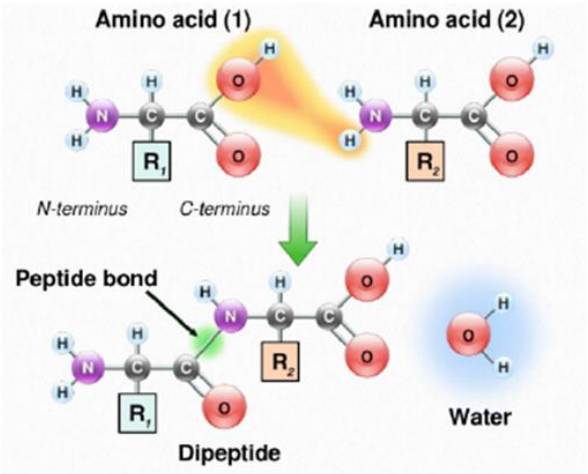


Source: Image made by the authors.

Figure 14. Amphoteric character of amino acids.

The physical characteristics of amino acids are affected by the ionic states of the amino and carboxyl groups, as well as any ionizable groups in the side chains. Each ionizable group is associated with a specific pK_a value, indicating the pH at which the concentrations of the protonated and unprotonated forms are equal. When the pH is below the pK_a, the protonated form predominates, and the amino acid acts as a true acid, capable of donating a proton. Conversely, when the pH exceeds the pK_a of the ionizable group, the unprotonated form prevails, and the amino acid exists as the conjugate base, acting as a proton acceptor. Each amino acid possesses at least two pK_a values corresponding to the ionization of the alpha-carboxyl and alpha-amino groups. Seven common amino acids also feature ionizable side chains with distinct pK_a values, leading to varying net charges at a given pH.

The bond formed between amino acids is an amide bond, commonly known as a *peptide bond*. This linkage arises from the condensation of the carboxyl group of one amino acid with the amino group of another, accompanied by the loss of a water molecule (Figure 15). Notably, the groups involved in peptide bonds do not carry ionic charges, unlike the free amino and carboxyl groups of amino acids in solution. The uncharged amino and carboxyl groups at the ends of a peptide chain are termed the N-terminus (amino terminus) and the C-terminus (carboxyl terminus), respectively. At neutral pH, each terminus bears an ionic charge. Conventionally, amino acid residues in a peptide chain are numbered from the N-terminus to the C-terminus, written from left to right, reflecting the direction of protein synthesis. The synthesis commences with the N-terminal amino acid, usually methionine, and proceeds sequentially toward the C-terminus by adding one residue at a time.



Source: Survey of Biochemistry and Biotechnology, A. Goodman, California Polytechnic State University San Luis Obispo, https://commons.libretexts.org/book/chem-347272.

Figure 15. Amphoteric character of amino acids.

The linear sequence of amino acids in a polypeptide chain constitutes the primary structure of a protein, while higher structural levels are denoted as secondary, tertiary, and quaternary structures. Primary structure defines the linear arrangement of amino acid residues in a protein.

3.1. Structure of Proteins

Nearly all biological processes rely on the specialized functions of one or more protein molecules. Proteins play crucial roles in producing other proteins, controlling cellular metabolism, regulating the movement of various molecular and ionic species across membranes, converting and storing cellular energy, and performing numerous other activities. The structure of proteins, however, is not merely linear; instead, polypeptide chains are intricately folded into compact shapes featuring coils, zigzags, turns, and loops. Over the past five decades, researchers have determined over a thousand proteins' three-dimensional shapes, or conformations. Conformations refer to spatial arrangements of atoms that depend on the rotation of one or more bonds. Unlike configurations, which require breaking and re-forming covalent bonds, conformations can change without disrupting covalent.

Although each amino acid residue has multiple potential conformations, and proteins have an astronomical number of possible conformations, under physiological conditions, each protein adopts a single stable shape known as its *native conformation*. This native, folded structure is crucial for the biological function of a protein and is influenced by factors such as interactions with solvent molecules (typically water), the pH and ionic composition of the solvent, and, most significantly, the protein's sequence.

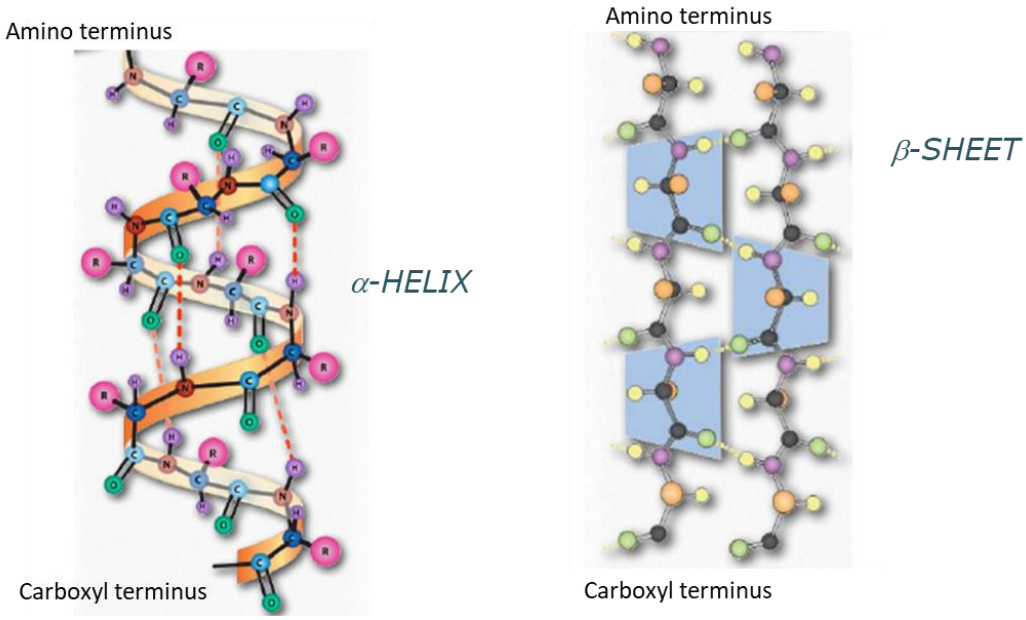
Up to four levels of structure can describe proteins. The primary structure delineates the linear sequence of amino acid residues. The three-dimensional structure of a protein encompasses secondary, tertiary, and quaternary structures. Noncovalent forces primarily stabilize these structural levels:

- *Hydrogen Bonds*: Commonly formed within a protein's structure, especially between atoms of the peptide backbone. Side chains with hydrogen bond-forming capabilities are usually located on the protein surface and engage with water molecules.
- *Hydrophobic Interactions*: Nonpolar side chains of amino acids and other nonpolar solutes tend to cluster in a nonpolar environment, minimizing their interaction with water. This clustering is entropically driven, and the hydrophobic interior of a protein primarily contains hydrophobic residues.
- *Ionic Interactions*: Arise from electrostatic attractions or repulsions between positively and negatively charged amino acid side chains. Charged residues are typically located on the protein surface, optimizing interaction with water.
- *Van der Waals Interactions*: Include both attractive forces due to instantaneous dipole-induced dipole interactions and repulsive forces. While individual interactions are weak, their sheer numbers contribute significantly to protein stability.

Environmental changes or chemical treatments may disrupt the native conformation, resulting in denaturation—a loss of biological activity. The energy required for denaturation is often relatively small, equivalent to disrupting a few hydrogen bonds.

Understanding protein structure begins with examining the peptide bonds that link amino acids. Peptide bonds exhibit partial double-bond character, limiting their free rotation. The rigid nature of peptide bonds restricts the range of conformations possible for a polypeptide chain, which can be envisioned as a series of rigid planes with common rotation points at C_α atoms.

The term *secondary structure* refers to a chosen segment of a polypeptide chain, describing the local spatial arrangement of its main-chain atoms without considering side chains or its relationship to other segments. Some particularly stable secondary structures widely found in proteins are the α-helix, β conformation, and the β turn (Figure 16). The alpha-helix involves a tightly wound polypeptide backbone, forming hydrogen bonds between carbonyl oxygens and amide protons, with a dipole moment along the helix axis. Beta-pleated sheets resemble zigzagging patterns and are formed by hydrogen-bonded peptide strands.



Source: Survey of Biochemistry and Biotechnology, A. Goodman, California Polytechnic State University San Luis Obispo, https://commons.libretexts.org/book/chem-347272.

Figure 16. Secondary structure of proteins.

The three-dimensional arrangement of all atoms in a protein is termed *tertiary structure*, resulting from folding a polypeptide into a closely packed structure. The tertiary structure involves interactions among amino acid residues that are distant from the primary structure. Noncovalent interactions, such as the hydrophobic effect and disulfide bridges, are crucial in stabilizing tertiary structure. Supersecondary structures, or motifs, are recognizable combinations of alpha-helices, beta-strands, and loops that appear in different proteins, sometimes associated with specific functions.

Quaternary structure pertains to the organization and arrangement of subunits in proteins with multiple subunits. These subunits are separate polypeptide chains, and proteins with quaternary structures

are termed oligomers. The subunits can be identical or different, forming dimers, tetramers, etc., with weak noncovalent interactions holding them together.

In summary, as proteins fold into their native conformations, primary structures' individual components interact, forming secondary structures. Tertiary structures involve interactions among distant residues, with noncovalent forces playing a significant role. Quaternary structures arise from interactions between multiple polypeptides, forming oligomeric proteins with distinct subunit arrangements. The sequence of amino acids ultimately dictates the properties and functions of a protein.

3.2. Function of Proteins

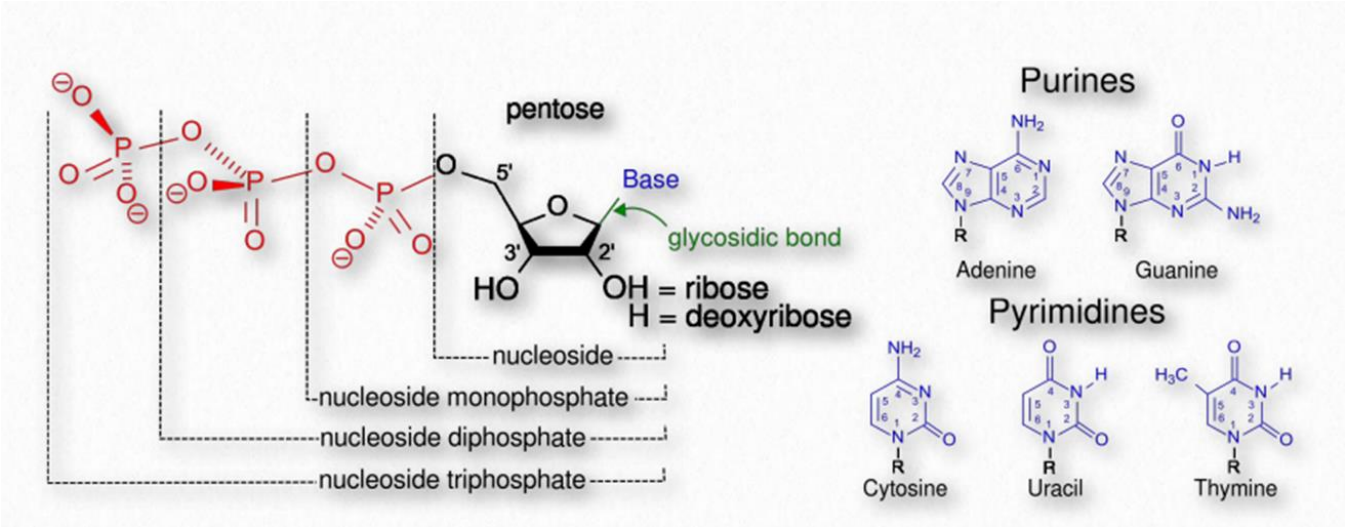
Proteins are essential agents in biological functions, playing pivotal roles in virtually every cellular activity. To categorize the vast array of proteins, a convenient approach is to classify them based on their specific biological roles.

- *Enzymes*: These proteins act as catalysts, accelerating biological reactions with remarkable specificity. Each enzyme operates in a particular metabolic pathway, catalyzing reactions vital for cellular functions.
- *Regulatory Proteins*: This class of proteins doesn't undergo chemical transformations but exerts control over the physiological functions of other proteins. Examples include insulin, a hormone regulating glucose metabolism.
- *Transport Proteins*: Dedicated to ferrying specific substances from one location to another, transport proteins perform critical roles. Hemoglobin, for instance, transports oxygen from the lungs to tissues, while serum albumin facilitates the transport of fatty acids. Membrane transport proteins facilitate the movement of metabolites across permeability barriers.
- *Storage Proteins*: These proteins serve as reservoirs for essential nutrients. Notably, casein in milk is a nitrogen source for mammalian infants during times of need, contributing to growth.
- *Structural Proteins*: Vital for creating and maintaining biological structures, structural proteins provide strength and protection to cells and tissues. Collagen, a fibrous protein, forms inelastic fibrils in bones, connective tissues, and tendons, while elastin contributes elasticity, particularly in ligaments.
- *Protective or Exploitive Proteins*: This diverse group includes cell defense, protection, or exploitation proteins. Immunoglobulins, such as antibodies produced by vertebrate lymphocytes, exemplify this category. Antibodies possess the remarkable ability to recognize and neutralize foreign molecules while ignoring intrinsic host molecules, safeguarding against infections.

4. NUCLEIC ACIDS

In 1944, Oswald Avery, Colin MacLeod, and Maclyn McCarty conclusively demonstrated that DNA carries genetic information. Despite limited knowledge about its structure, subsequent years witnessed the determination of nucleotide structures, culminating in James D. Watson and Francis H. C. Crick's proposal of the double-stranded DNA model in 1953.

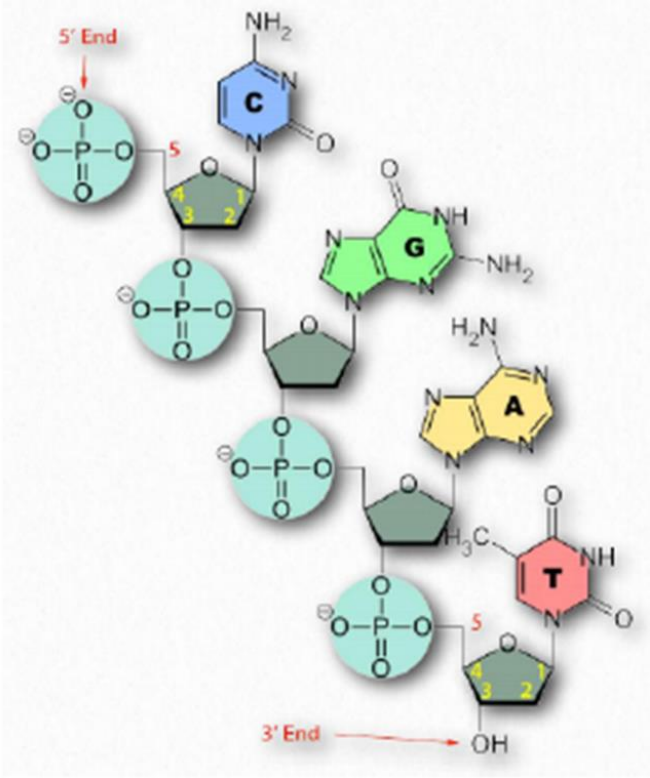
Nucleic acids constitute the fourth major macromolecule class, akin to proteins and polysaccharides. This discussion delves into the structure of nucleic acids and their cellular packaging, exploring enzymes that interact with DNA and RNA. Numerous proteins and enzymes collaborate to ensure accurate interpretation of genetic information. Nucleic acids, specifically polynucleotides, are composed of nucleotides with three essential components: a nitrogenous base, a pentose, and one or more phosphates (Figure 17). The absence of a phosphate group turns a nucleotide into a nucleoside. Nitrogenous bases, derived from pyrimidine and purine, are planar and absorb ultraviolet light due to conjugated double bonds. The sugar components, ribose or deoxyribose, form the stable beta-configuration of the furanose ring in nucleotides. Nucleosides, the N-glycosides of these sugars, connect to heterocyclic bases through b-N-glycosidic bonds. Nucleotides, phosphorylated derivatives of nucleosides, can be ribonucleotides or deoxyribonucleotides, depending on the sugar present.



Source: Survey of Biochemistry and Biotechnology, A. Goodman, California Polytechnic State University San Luis Obispo, https://commons.libretexts.org/book/chem-347272.

Figure 17. Nucleotide structure.

Nucleic acids exist as linear polymers of nucleotides, linked by phosphodiester bridges in a 3 to 5 direction. DNA and RNA represent the two major classes of nucleic acids, with DNA holding a central role in preserving and transcribing genetic information. In the directional sense of polynucleotide chains, the 5' end lacks an attached residue, while the 3' end has no residue attached to its 3'-carbon. Like polypeptide chains, polynucleotide chains exhibit directionality, and the DNA strand direction is conventionally defined by reading across the sugar residue atoms (Figure 18). Polynucleotides have common nitrogenous bases, except one, thymine for DNA and uracil for RNA.



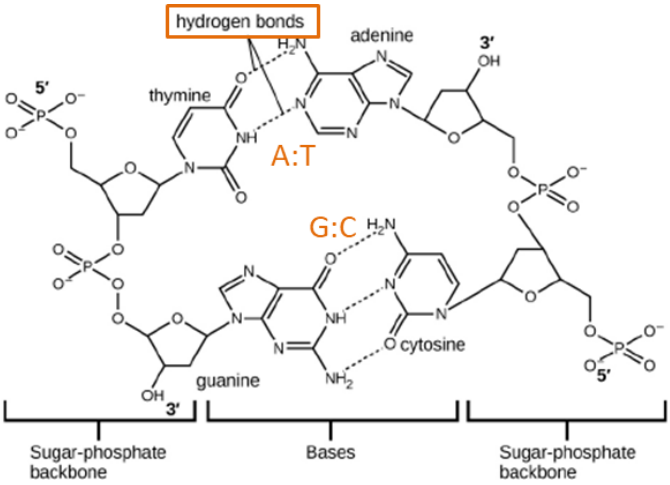
Source: Survey of Biochemistry and Biotechnology, A. Goodman, California Polytechnic State University San Luis Obispo, https://commons.libretexts.org/book/chem-347272.

Figure 18. Polynucleotide structure (Deoxyribonucleic acid, DNA).

4.1. DNA Structure

In 1953, James Watson and Francis Crick proposed a comprehensive three-dimensional model of DNA structure that aligned with all available data. Most DNA molecules comprise two polynucleotide strands, with bases on one strand forming hydrogen bonds with the opposite strand's bases. The resultant right-handed double helix features hydrophilic backbones of alternating deoxyribose and phosphate groups on the exterior, adopting a C-2' endo conformation. Inside the double helix, the hydrophobic and nearly planar ring structures of purine and pyrimidine bases stack closely and perpendicularly to the long axis.

The sequence of bases in one strand maintains a complementary relationship with the other strand, ensuring the preservation of information across both sequences. Several factors contribute to the stability of the DNA double helix. Internal and external hydrogen bonds play a pivotal role, with A:T pairs forming two hydrogen bonds and G:C pairs forming three (Figure 19). Negatively charged phosphate groups on the helix's exterior exhibit minimal interference and engage in electrostatic interactions with cations like Mg^{2+} in the solution. Additionally, the helix's core, comprising base pairs, undergoes stacking through hydrophobic interactions and van der Waals forces, contributing significantly to the overall stabilization energy.



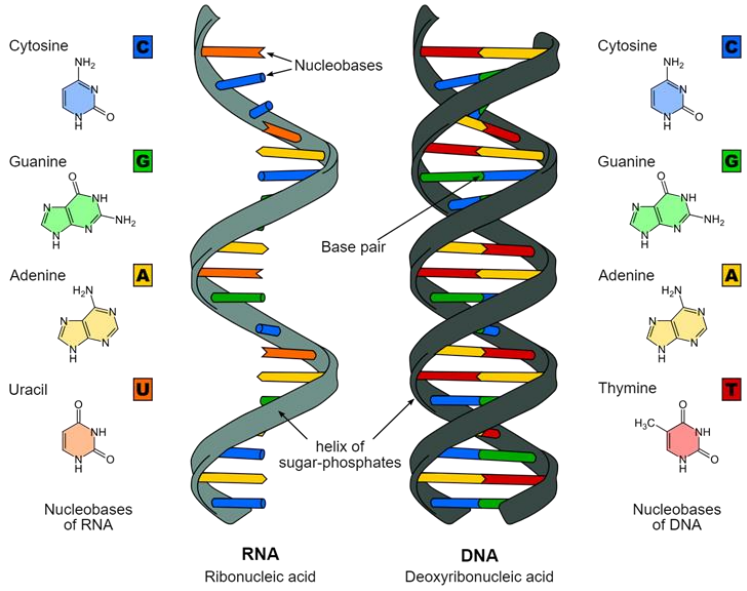
Source: Survey of Biochemistry and Biotechnology, A. Goodman, California Polytechnic State University San Luis Obispo, https://commons.libretexts.org/book/chem-347272.

Figure 19. Sequence of bases in DNA structure.

4.2. RNA structure

The ribose sugar in RNA adopts a cyclic structure, encompassing five carbon atoms and one oxygen atom. Notably, the presence of a chemically reactive hydroxyl (−OH) group attached to the second carbon in the ribose sugar renders RNA susceptible to hydrolysis. This chemical vulnerability of RNA, distinct from DNA, which lacks a reactive −OH group in the equivalent position on the sugar moiety (deoxyribose), is believed to contribute to the evolutionary preference for DNA as the primary carrier of genetic information in most organisms.

RNA typically exists as a single-stranded biopolymer (Figure 20). However, self-complementary sequences within the RNA strand induce intrachain base-pairing, folding the ribonucleotide chain into intricate structural forms comprising bulges and helices. The three-dimensional configuration of RNA plays a pivotal role in its stability and functionality. This structural flexibility allows cellular enzymes to modify the ribose sugar and nitrogenous bases by attaching chemical groups (e.g., methyl groups). Such modifications facilitate the formation of chemical bonds between distant regions in the RNA strand, leading to intricate contortions in the RNA chain that further enhance its stability.



Source: Difference RNA and DNA structures: Difference DNA RNA-DE.svg: Sponk / *translation: Sponk, CC BY-SA 3.0, https://commons.wikimedia.org/wiki/File:Difference_DNA_RNA.svg.

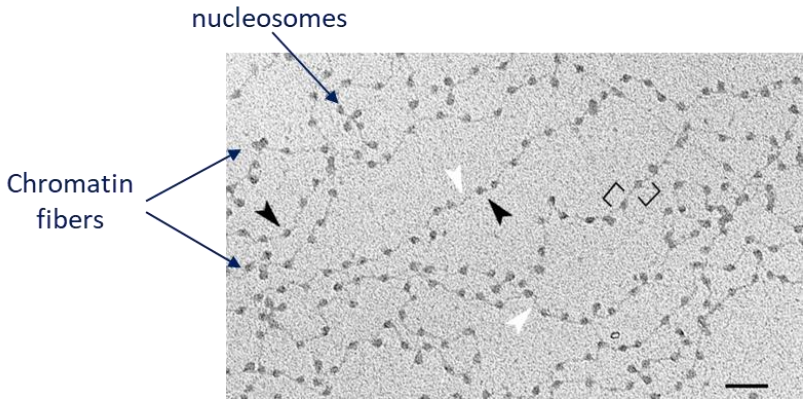
Figure 20. Comparation between DNA and RNA structures.

4.3. Genome Organization

The hereditary material, DNA (deoxyribonucleic acid), is organized into a sequence of four nucleotides, encoding information based on their specific order. Genomic organization encompasses the linear arrangement of DNA elements and their division into chromosomes. Additionally, it can refer to the three-dimensional structure of chromosomes and the spatial arrangement of DNA sequences within the nucleus.

Chromatin, a complex of DNA and various proteins, is crucial in packaging DNA into a more compact form. DNA is also associated with proteins in prokaryotes, forming condensed structures distinct from eukaryotes' chromatin. Within a resting cell, chromatin appears as long, slender threads, approximately 30 nm in diameter, known as 30 nm fibers. Human nuclei accommodate 46 such chromatin fibers, representing chromosomes. The principal proteins in chromatin are called histones.

In Figure 21 is shown chromatin unfolds after treatment with a low ionic strength solution, revealing a "beads-on-a-string" organization in an electron micrograph. The "beads" are DNA–histone complexes called nucleosomes, with the "string" being double-stranded DNA. Efficient condensation of over 2 meters of DNA into a nucleus about 5 micrometers in diameter is achieved by wrapping DNA around nucleosomes and organizing them into helical filaments, forming loops associated with the nuclear matrix.



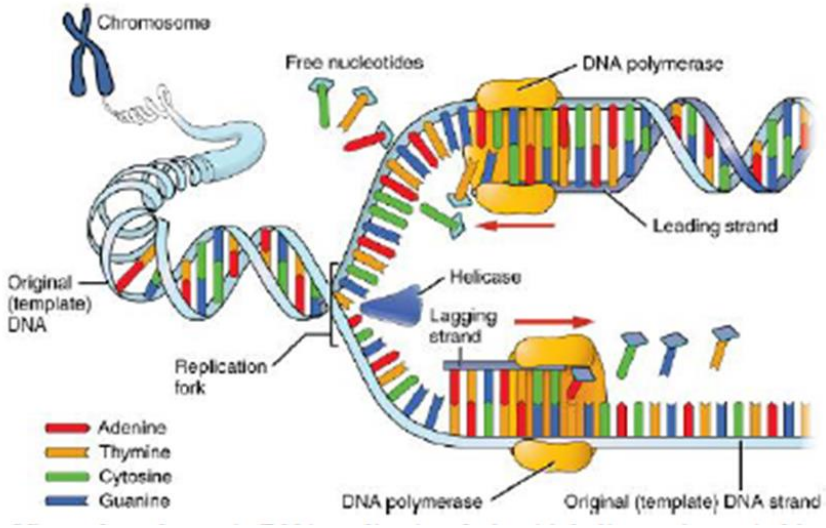
Source: Chromatin fibers: Chromatin_nucleofilaments.png: Chris Woodcockderivative work: Gouttegd, CC BY-SA 3.0, https://commons.wikimedia.org/wiki/File:Chromatin_nucleofilaments_(detail).png.

Figure 21. Electron micrograph of chromatin fibers.

Human genomic DNA, identified by Friedrich Miescher in 1869, consists of 46 chromosomes, including autosomes and sex chromosomes. With over 6 billion letters, the human genome necessitates extensive packaging to fit inside the micrometer-sized nucleus. Chromatin, with histones as the predominant proteins, undergoes various levels of packaging, shortening chromosome length by seven-fold.

4.4. DNA Replication

As genetic information resides in DNA, transferring this information from a parental cell to two daughter cells necessitates precise duplication of DNA, a process known as *DNA replication*. Due to the complementary nature of the two strands in double-helical DNA, the nucleotide sequence of one automatically dictates the sequence of the other. Watson and Crick proposed that the helix's two strands unwind during DNA replication, and each DNA strand serves as a template for synthesizing a complementary strand (Figure 22). This results in the production of two double-stranded daughter molecules, each comprising one parental strand and one newly synthesized strand. Termed semiconservative replication, this process ensures the conservation of one strand from the parental DNA in each daughter molecule.



Source: Survey of Biochemistry and Biotechnology, A. Goodman, California Polytechnic State University San Luis Obispo, https://commons.libretexts.org/book/chem-347272.

Figure 22. DNA replication.

To maintain genetic information across generations, DNA replication must be rapid (replicating the entire DNA complement before each cell division) and accurate. Chromosome replication initiates at a specific site called the origin of replication, progressing bidirectionally until two replication complexes converge at the termination site, signaling the end of replication. The replisome, the protein machine conducting the polymerization reaction, contains various proteins crucial for swift and precise DNA replication. Each replication fork, where parental DNA unwinds, hosts one replisome in bidirectional DNA replication. The termination site marks the point where the two double-stranded DNA molecules separate. Notably, each daughter molecule includes one parental and one newly synthesized strand.

4.5. DNA Repair

DNA stands as the sole cellular macromolecule subject to repair, a testament to the organism's investment in preserving genetic information, considering the disproportionate cost of mutated or damaged DNA compared to the energy expended in repair processes. DNA repair systems, intricately designed, underscore the cellular imperative to safeguard the integrity of genetic information. A myriad of processes, both spontaneous and catalyzed by environmental agents, can induce DNA damage. Even during replication, errors leading to mismatched base pairs, like G paired with T, can potentially compromise the information content in DNA.

The chemistry of DNA damage is intricate and diverse, prompting a broad enzymatic response that catalyzes fascinating chemical transformations within DNA metabolism. Damage to DNA manifests in various forms, encompassing base modifications, nucleotide deletions or insertions, cross-linking of DNA strands, and breakage of the phosphodiester backbone. While environmental agents such as chemicals or radiation contribute to some DNA damage, errors in normal cellular processes, like incorporating an incorrect nucleotide during replication, can also result in damage. Although severe damage may be lethal, a substantial portion of in vivo damage undergoes repair. Permanent alterations in the nucleotide sequence, termed mutations, can involve substitution, insertion, or deletion of one or more base pairs. In mammals, an accumulation of mutations correlates strongly with cancer.

DNA repair mechanisms play a dual role, protecting individual cells and ensuring the genetic integrity of subsequent generations. In single-celled organisms, both prokaryotic and eukaryotic, unrepaired DNA damage may transform into mutations transmitted directly to daughter cells after DNA replication and cell division. In multicellular organisms, mutations only pass on to the next generation if they occur in the germ line. Germ-line mutations might not manifest noticeable effects on the host organism but can profoundly impact progeny, particularly if the mutated genes play vital roles in development.

Repair systems can be classified:

- *Direct repair*, a process circumventing the need to break the phosphodiester backbone of DNA, rectifies some lesions. Certain compounds, like acids and oxidizing agents, can modify DNA through alkylation, methylation, or deamination.
- *Excision repair* involves the removal of a damaged DNA strand segment followed by DNA synthesis, utilizing the undamaged strand as a template. Excision repair, a general term encompassing various methodologies, excises the damaged DNA portion, initiates repair synthesis, and concludes with ligation to restore continuity to the repaired strand.

4.6. DNA Recombination

Recombination constitutes any event leading to the transfer or exchange of DNA fragments between chromosomes or within a single chromosome. Genetic recombination events can be broadly classified into three categories. *Homologous genetic recombination*, also known as general recombination, facilitates exchanges between any two DNA molecules or segments sharing an extensive region of nearly identical sequence. The specific sequence of bases is inconsequential as long as it is similar in the two DNAs. *Site-specific recombination* confines exchanges to a particular DNA sequence. At the same time, *DNA transposition*, distinct from the other classes, involves a short DNA segment with the remarkable ability to relocate within a chromosome.

Various mechanisms underlie recombination, with many of the proteins and enzymes involved in recombination reactions also playing roles in DNA repair. Deficiencies in human genes essential for repair and recombination result in heightened sensitivity to ultraviolet light and an increased susceptibility to cancer, underscoring the shared elements and significance of these cellular processes.

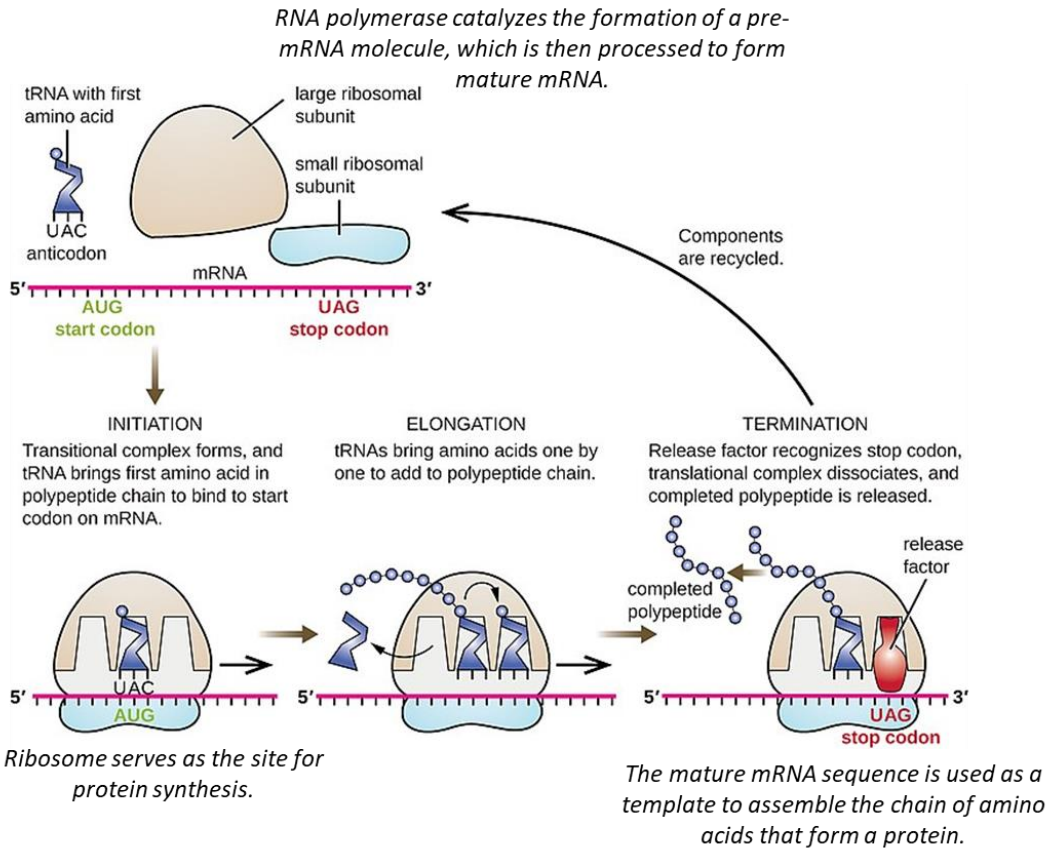
4.7. RNA Transcription and Translation

RNA molecules, excluding RNA genomes of certain viruses, derive from information permanently stored in DNA. Transcription involves an enzyme system converting genetic information in double-stranded DNA into an RNA strand with a base sequence complementary to one of the DNA strands. Three major types of RNA are produced: *Messenger RNAs* (mRNAs) encode amino acid sequences of polypeptides, *Transfer RNAs* (tRNAs) read mRNA information and transfer appropriate amino acids during protein synthesis, and *Ribosomal RNAs* (rRNAs) are components of ribosomes. Gene expression involves two main

steps: transcription, where DNA information is transferred to mRNA, and translation, where mRNA is read according to the genetic code to assemble the chain of amino acids forming a protein.

Despite the catalytic activity of small RNA molecules, ribosomal RNA comprises a large percentage of total cellular RNA, while mRNA constitutes only a small percentage. The cellular transcriptome is the sum of all RNA molecules produced in a cell under specific conditions.

The translation of mRNA information into the specific amino acid sequence of a protein involves codons and nucleotide triplets specifying amino acids (Figure 23). Adaptor molecules called tRNAs facilitate translation by recognizing codons and inserting amino acids sequentially into the polypeptide. Ribosomes execute the genetic code governing the translation of mRNA codons into proteins. It defines how codons specify the addition of amino acids during protein synthesis.



Source: Survey of Biochemistry and Biotechnology, A. Goodman, California Polytechnic State University San Luis Obispo, https://commons.libretexts.org/book/chem-347272.

Figure 23. RNA transcription and subsequent translation.

The proteome encompasses the complete set of proteins expressed in a given cell, tissue, or organism at a specific time. It represents the entire set of expressed proteins studied through proteomics. On the other hand, the genome is an organism's entire genetic material, encompassing both coding and noncoding regions of DNA or RNA. Genomics is the study of the genome.

4.8. Genetic Engineering

Recombinant DNA Technology

We have entered a biochemistry era marked by developing techniques for isolating, manipulating, and amplifying specific DNA sequences. Collectively, these methodologies form recombinant DNA technology, a powerful tool for studying the structure and function of individual genes.

Recombinant DNA molecules, constructed by combining DNA from different sources, involve joining genes into new combinations. Genomic DNA is fragmented using restriction endonucleases, and each fragment is inserted into a compatible site in a carrier molecule or vector. The resulting recombinant DNA is introduced into a living host cell, where it undergoes replication. The process of propagating host cells containing the recombinant DNA is termed cloning, even though "clone" traditionally denotes an exact genetic replica. Each clonal cell line typically harbors numerous copies of a single DNA fragment from the original genome. Most recombinant DNA experiments share six basic steps.

- Isolation and purification of DNA: Routine methods are employed to prepare both vector and target DNA molecules. In some cases, the target DNA is synthesized in vitro.
- Cleavage of DNA at particular sequences: Cleaving DNA at specific sequences is crucial. The DNA fragment of interest, termed insert DNA, is usually cleaved using commercially produced nucleases and restriction endonucleases.
- Ligation of DNA fragments: A recombinant DNA molecule is formed by cleaving the DNA of interest, yielding insert DNA, and ligating it to vector DNA. DNA fragments are typically joined using DNA ligase.
- Introduction of recombinant DNA into compatible host cells: The recombinant DNA must be introduced into a compatible host cell for propagation. The direct uptake of foreign DNA is termed genetic transformation, while recombinant DNA can also be transferred to host cells by transfection.
- Replication and expression of recombinant DNA in host cells: Cloning vectors facilitate the replication and, in some cases, expression of insert DNA in host cells. Efficient cloning and DNA expression rely on suitable vectors and hosts.
- Identification of host cells with recombinant DNA: Vectors often contain genetic markers enabling the selection of host cells with foreign DNA. Identifying a specific DNA fragment typically involves the challenging step of screening many recombinant DNA clones.

DNA Sequencing

DNA sequencing is a technique employed to ascertain the precise order of the four nucleotide bases – adenine, guanine, cytosine, and thymine – comprising a DNA strand. These bases serve as the foundational genetic information (genotype) instructing a cell's activities, destinations, and its developmental fate (phenotype). While nucleotides are not the sole determinants of phenotypes, they play

an indispensable role in their establishment. Each individual and organism possesses a unique nucleotide base sequence, representing the most fundamental gene or genome knowledge level. This sequence functions as a blueprint, containing instructions for organismal construction, and entire comprehension of genetic function or evolution necessitates obtaining this information.

Understanding DNA sequences is pivotal in various domains, from basic biological research to applied fields such as medical diagnosis, biotechnology, forensic biology, virology, and biological systematics. Comparing healthy and mutated DNA sequences aids in diagnosing different diseases, characterizing antibody repertoires, guiding patient treatment, and facilitating faster and more individualized medical care. Additionally, quick DNA sequencing expedites the identification and cataloging of diverse organisms.

The knowledge of a DNA segment's sequence holds multiple applications. Firstly, it aids in gene discovery, allowing the screening of DNA regions for characteristic features of genes that code for specific proteins or phenotypes. Secondly, it facilitates the comparison of homologous DNA sequences across different organisms, unveiling evolutionary relationships within and between species. Lastly, gene sequences can be scrutinized for functional regions by identifying common domains shared with proteins of similar function.

The advent of next-generation sequencing technologies has further expanded applications due to their cost-effectiveness and high-throughput capacity. Scientists utilize these technologies to swiftly sequence entire genomes (whole-genome sequencing), discover disease-related genes, and enhance comprehension of genomic structure and diversity among species.

Polymerase Chain Reaction

The polymerase chain reaction (PCR) is a powerful tool for amplifying minute quantities of DNA or enriching the abundance of a specific DNA sequence within a pool of mixed DNA molecules. Leveraging PCR technology eliminates the necessity of extracting large tissue samples for sequencing or cloning purposes. Moreover, PCR facilitates the generation of numerous copies of a known gene without needing prior isolation. It provides an alternative to traditional cloning for gene amplification.

In this method, denatured DNA containing the target segment acts as a template for DNA polymerase, while specific oligonucleotides function as primers for DNA synthesis. These primers, designed to complement the 3'-ends of the target DNA segment, are added in excess, significantly amplifying the concentration of the desired segment in the solution. After cycles of DNA synthesis, the DNA undergoes heating and cooling steps using thermal cyclers for dissociation, primer annealing, and further DNA synthesis. This cycling process is meticulously repeated, with controlled reaction time and temperature. The number of DNA strands with defined ends by the primers increases exponentially with each cycle, preferentially replicating the desired DNA segment until it becomes the predominant DNA in the test tube after 20 to 30 cycles. The amplified DNA can then be cloned, sequenced, or used as a probe for screening recombinant DNA libraries.

Beyond its significance in basic research, PCR technology has profound implications in medicine and forensics. Its ability to work with minute samples allows rapid determination of DNA sequences for prompt

diagnosis of infectious diseases, identification of genetic defects in fetuses or neonates, effective monitoring of cancer chemotherapy, and improved organ donor-recipient matching. PCR also plays a pivotal role in resolving paternity and criminal cases.

Furthermore, PCR has opened avenues for scientists to amplify and sequence residual DNA in ancient fossils, including those millions of years old, sourced typically from tissue embedded in amber, showcasing its versatility in various scientific disciplines.