Biological Molecules: 
Structure and Methods of Analysis

By Rich Griner, 2003

This handout is to be used with your textbook: Biology, 7th ed., Campbell & Reece.

I. Carbohydrates

Carbohydrates are also called **saccharides**, which is derived from the Greek word *sakcharon* meaning sugar. The simplest carbohydrates are called **monosaccharides**, and the biologically important monosaccharides contain from 3 to 7 carbon atoms. The monosaccharides with 5 carbons (the **pentoses**) and with 6 carbons (the **hexoses**) may take on different configurations in an aqueous solution, randomly alternating between a linear and a ring form (Fig. 5.4).

The pentoses are important in several biochemical pathways, but the two most common pentoses, **ribose** and **deoxyribose**, are critical structural components of nucleic acids as described later.

\[
\begin{align*}
\text{Ribose} & \quad \text{Deoxyribose} & \quad \text{Glucose} & \quad \text{Fructose} & \quad \text{Galactose} & \quad \text{Mannose} \\
\end{align*}
\]

The four most important hexoses are **glucose**, **fructose**, **galactose**, and **mannose**. Hexoses have important structural roles in cells, but they also are the preferred energy source of almost all cell types. As the preferred energy source, hexoses are required in great abundance. Thus, in order to maintain adequate levels of hexoses, cells store them by combining monosaccharides into **disaccharides** and **polysaccharides** via dehydration synthesis (Fig. 5.5).

When more than two monosaccharides are linked together, the structure is called an **oligosaccharide**. Oligosaccharides containing more than a dozen residues are commonly synthesized, and many of the residues are often mannose. Such oligosaccharides are typically linked to proteins to form **glycoproteins**.

Large quantities of hexoses are stored most efficiently when they are combined to form much larger polysaccharides. The major storage polysaccharide in plant cells is called **starch**, which is composed entirely of glucose. There are two major forms of starch, **amylose** and **amylopectin**, which differ in the way the glucose molecules are linked together. Amylose is a linear chain of several thousand glucose molecules each linked between carbons 1 and 4. These chains are free to twist and turn, and they typically assume a coiled conformation. Amylopectin may contain up to one million glucose molecules, and it is a more complex structure in which the glucose chain may branch every 25-30 molecules. Most of the glucose residues of amylopectin are linked between carbons 1 and 4 as in amylose, but the branches are due to linkages between carbons 1 and 6. The overall structure of amylopectin resembles a tree in that the branches reach out in many different directions (Fig. 5.6). The major storage polysaccharide in animal cells is called **glycogen**, and just like starch it is composed entirely of \(\alpha\)-glucose. Glycogen is very similar in structure to amylopectin, but it branches more frequently, every 8-12 molecules (Fig. 5.6).
Plants make another important polysaccharide from glucose, but instead of a storage form for future energy needs it is for structural purposes. This polysaccharide is called **cellulose**, and it is composed of a linear chain of thousands of glucose molecules linked between carbons 1 and 4. These chains are long and straight, unlike the coiled chains of amylose. Over a dozen of these long straight chains can form hydrogen bonds between the OH and H groups of neighboring chains leading to the formation of microfibrils. These microfibrils of cellulose are very strong, and they are the major component of plant cell walls (Fig. 5.8). Animals produce the enzymes necessary to digest (breakdown) starch, and to synthesize and breakdown glycogen, but they do not produce the enzymes necessary to digest cellulose. This is because the ring form of glucose used to make starch or glycogen is different from the form used to make cellulose. When glucose converts from a linear to a ring form there is a 50% chance that it will form either an α or a β configuration (Fig. 5.7). The difference between the α and β configurations is the orientation of the OH and H on the #1 carbon of a six-sided ring or the #2 carbon of a five-sided ring. Specifically, when the OH is in the down position the configuration is α and when the OH is in the up position the configuration is β. Thus, animals synthesize an enzyme capable of breaking down the bonds between α-glucose molecules found in starch and glycogen, but they do not make an enzyme capable of breaking the bonds between the β-glucose molecules found in cellulose.

A simple biochemical test using **Benedict’s reagent** can establish the presence of monosaccharides and most disaccharides. Benedict’s reagent is composed of copper (II) ions, Cu$^{+2}$, and citrate ions in a buffer with a basic pH. The Cu$^{+2}$ ions in this reagent have weak oxidizing activity and can react with the reducing group C = O which is only found in the linear form of sugars. Such sugars are termed reducing sugars. When Benedict’s reagent is in the presence of reducing sugars, a reduction/oxidation reaction occurs in which the Cu$^{+2}$ ions are reduced to copper (I) ions, Cu$^{+1}$. The Cu$^{+1}$ ions then form Cu$_2$O which is a red solid. As the Benedict’s reagent is exposed to increasing quantities of reducing sugars, its color changes from blue to green, yellow, orange, or red.

The C = O group is found at the #1 carbon position of glucose, mannose, and galactose, but at the #2 position of fructose (Fig. 5.3). If this carbon participates in the formation of a glycosidic bond, then the sugar is locked into its ring structure, and it is no longer a reducing sugar. The glycosidic bond of sucrose is formed using the #1 carbon of glucose and the #2 carbon of fructose. When this bond is formed neither glucose nor fructose is free to interconvert to its linear form, and neither can act as a reducing sugar. However, the glycosidic bonds of maltose and lactose involve the #1 carbon of glucose and galactose, respectively, and the #4 carbon of glucose. Thus, when either of these disaccharides is formed a glucose residue has a #1 carbon free (i.e. not participating in a glycosidic bond). This glucose is able to interconvert between its linear and ring forms, and thus, it is a reducing sugar even though it is part of a disaccharide. Likewise, one or more of the terminal glucose residues of polysaccharide molecules can reduce the Cu$^{+2}$ ions of Benedict’s reagent, but the low number of these residues compared to the overall mass of the polysaccharide prevents starch from giving a positive reaction with this test.

Because Benedict’s reagent is not specific for any one type of sugar, it cannot be used to make accurate quantitative determinations of the amount or concentration of carbohydrate in a sample. This test is best used for making relative comparisons of the amount of carbohydrate among a group of samples. To test for the presence and quantity of specific polysaccharides, enzyme-based assays are used. Why do you think that an enzyme-based assay would provide specificity in a carbohydrate assay?

A test for the presence of starch uses **Lugol’s reagent**. Lugol’s reagent, which contains iodine, I$_2$, and potassium iodide, KI, is an amber colored solution that turns a blue-black in the presence of starch. The color is produced when the I$_2$ molecules become trapped within the helices of the starch molecule. The I$_2$ will also become trapped within the helices of the glycogen molecule but to a lesser degree because of the higher number of branches in glycogen versus starch. Therefore, the color change produced when Lugol’s reagent reacts with glycogen is less than when it reacts with starch. Like the Benedict’s test, a Lugol’s test for starch is not specific for a particular type of polysaccharide, and it is not very accurate as a quantitative assay. To test for the presence and quantity of specific polysaccharides, enzyme-based assays are used. Would cellulose give a positive test with Lugol’s reagent? Why or why not?
II. Lipids

The simplest lipids are fatty acids. Structurally, fatty acids are characterized by a carboxylic acid group on one end attached to a string of hydrocarbons known as an acyl chain. Organisms synthesize fatty acids in a variety of lengths with a range of 14 – 20 carbons long.

Fatty acids are also synthesized with various numbers of double bonds, and each double bond produces a 30 degree bend in the acyl chain (Fig. 5.11). Fatty acids having no double bonds in their acyl chain are called saturated because the carbons are saturated with hydrogens. Fatty acids that have double bonds are unsaturated because the two carbons that form a double bond are each bound to one fewer hydrogen that the other carbons of the chain. The terms monounsaturated and polyunsaturated refer to the presence of one double bond and greater than one double bond, respectively.

Fatty acids are identified both with a unique name and a numerical designation that indicates the number of carbons and the number of double bonds in the molecule. For example, palmitic acid (16:0) and stearic acid (18:0) are saturated, oleic acid (18:1) is monounsaturated, and linoleic (18:2) is polyunsaturated.

The synthesis of unsaturated fatty acids requires specific enzymes called desaturases that produce the double bonds between specific carbons in the acyl chains. Animals do not have a desaturase necessary to produce a double bond below the 10th carbon in an acyl chain. Nevertheless, animals require several unsaturated fatty acids containing double bonds below the 10th carbon such as linoleic acid, which has double bonds between the 12th and 13th carbons. However, the desaturases necessary for the synthesis of these molecules are present in plants; therefore, animals obtain these fatty acids from plants they consume as food. Such fatty acids are referred to as essential, because it is essential that they are part of the animal’s diet.

The length of the fatty acid chain and the number of double bonds has a direct impact on the physical properties of a fatty acid. Remember that the chemical structure of a molecule determines its physical properties. For fatty acids, an important physical property that is affected by the length of the acyl chains and the number of double bonds is the melting point. The melting point is the temperature at which a substance changes from a solid to a liquid. Molecules in a solid state are packed together in an orderly fashion with very much movement, while molecules in a liquid state are moving around in a random pattern termed Brownian motion. Therefore, the melting point is affected by anything that affects the ability of molecules to pack together or anything that increases Brownian motion.

Just as smaller molecules diffuse faster than larger molecules, fatty acids with short acyl chains exhibit more Brownian motion at a given temperature than fatty acids with long acyl chains. Therefore, a shorter fatty acid will have a lower melting point than a longer fatty acid. Furthermore, the bent acyl chains of unsaturated fatty acids interfere with the ability of fatty acids to pack together. Therefore, an unsaturated fatty acid will have a lower melting point than a saturated fatty acid of equal acyl chain length.

Organisms use fatty acids to synthesize more complex lipid structures. For example, three fatty acids may be combined via dehydration synthesis with one molecule of glycerol to form a neutral lipid called triacylglycerol or a triglyceride (Fig. 5.10). Triglycerides are storage form of lipids. The adipose cells of animals synthesize and store triglycerides when food intake exceeds its requirement for energy and growth. Plants also synthesize and store triglycerides, but the triglycerides synthesized by plants differ significantly from those synthesized by animals. Specifically, they differ in the type of fatty acids that they contain. Triglycerides from plants typically contain shorter more unsaturated fatty acids, while triglycerides form animals typically contain longer less unsaturated fatty acids. Therefore, the triglycerides from each type of organism differ significantly in their physical properties. Triglycerides from plants are usually liquid at room temperature and are classified as oils, while those from animals are usually solid at room temperature and are classified as fats (Fig. 5.11).

Even more complex lipid structures are synthesized from triglycerides. When one of the fatty acids chains is replaced by phosphate, a phospholipid is formed (Fig. 5.12). Other charged or polar molecules are usually linked to the phosphate. Together, the phosphate and the additional molecule form the “head group” of the phospholipid. The most common additions to the phosphate are choline, serine, ethanolamine, and inositol, and the resulting phospholipids are named phosphatidylcholine, phosphatidylserine, phosphatidylethanolamine, and phosphatidylinositol.

The addition of a charged and/or polar head group makes that end of the phospholipid hydrophilic, while the acyl chains form a tail that is hydrophobic. Any molecule that is hydrophobic on one end and
hydrophilic on the other is termed amphipathic, and it the amphipathic property of phospholipids that allows them to form the multi-molecular lipid bilayers which compose all cellular membranes (Fig. 5.13).

A separate class of lipids is the steroids. The steroids all have a common ring structure, but they differ in the number and types of substitutions attached to the ring. The prototypical steroid, cholesterol, is synthesized in great quantities by the liver. Cholesterol serves as a precursor for the synthesis of the steroids hormones such as testosterone, estrogen, aldosterone, and vitamin D. Despite the presence of a few polar groups, such as the hydroxyl on cholesterol, these molecules are considered completely hydrophobic.

A general test for the presence of lipids uses a lipophilic dye such as Sudan IV. Because Sudan IV is lipophilic, it associates with any lipid or hydrophobic molecules. When Sudan IV, which is red, is added to a sample containing a mixture of lipids and other molecules, the dye will partition with the lipids allowing them to be visualized.

This test is limited because it is not specific for any particular type of lipid, and thus, it cannot differentiate between fatty acids, triglycerides, phospholipids, or steroids. It is also not specific for lipids in general. Its lipophilic nature allows it to partition into any group of hydrophobic molecules, such as alcohols or organic solvents like chloroform. Therefore, this test is a general indicator only. A more specific test for an individual type of lipid requires the separation of a group of lipids by thin-layer chromatography (TLC), and visualization using other dyes. This method always requires that a standard (known sample) of the lipid in question be included in the test for comparative purposes.

IV. Nucleic Acids

Nucleic acids have a number of roles in cells ranging from temporary carriers of energy to permanent storage of the information necessary to make proteins, but they are all made from the same three types of molecular components: a pentose sugar, phosphate, and a base (Fig. 5.29). The pentose sugars are always either ribose or deoxyribose, and the bases are based on either a purine or pyrimidine ring. (Note: these structures contain nitrogen; what other cellular molecule contains nitrogen?) The names of the two purines found in nucleic acids are adenine (A) and guanine (G), and the names of the three pyrimidines found in nucleic acids are thymine (T), cytosine (C), and uracil (U).

These bases are joined to the #1 carbon of ribose or deoxyribose to form structures called nucleosides, and the addition of one or more phosphates to the pentose at the #6 carbon produces structures called nucleotides.

The polynucleotides ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) are linear strands of nucleotides linked by their sugars and phosphates leaving the bases protruding out to one side (Fig. 5.29). In DNA, the bases between two strands interact via hydrogen bonds to form base pairs resulting in double stranded DNA. The formation of base pairs follows two very simple rules. 1. Adenine and thymine always form two hydrogen bonds with each other, and 2. cytosine and guanine always form three hydrogen bonds with each other. Neither adenine nor thymine ever forms hydrogen bonds with cytosine or guanine. Because base pairing is always complementary (i.e. A only pairs with T and G only pairs with C), the entire strands of double stranded DNA are complementary to each other. If the sequence of bases on one strand is known, the sequence of bases on its complementary strand can be intuitively inferred. If one strand has the sequence GCTACTGTAAACTG, what would be the orientation and sequence of the complementary strand?

The triphosphate forms of nucleotides are molecules that are very rich in energy, and ATP, specifically, supplies energy to drive many endergonic biochemical reactions (Fig. 6.8). This is because the bonds between the phosphates (phosphodiester bonds) are “high energy” bonds, and when these bonds are broken that energy is released and utilized by other reactions. Biochemical reactions that break bonds like the phosphodiester bond of ATP require the addition of water. What are these reactions called?

Several fluorescent dyes are used to quantitatively assay DNA and RNA, but the most common method for qualitative analysis of DNA and RNA is via electrophoresis through a gel made of agarose (a polysaccharide from seaweed). Agarose electrophoresis separates these nucleic acids by size. The agarose gel also contains ethidium bromide which binds to the DNA or RNA and fluoresces under uv light. Thus, after the nucleic acids have been separated by electrophoresis, the gel is placed under uv light to
visualize the DNA or RNA. The locations of the DNA or RNA appear as bands within the gel, and these are compared to DNA standards (known sizes) which are also present in the gel. This allows you to identify the size of the DNA or RNA in your samples, but it does not reveal anything about the sequence of bases. Other techniques such as Southern blotting for DNA, northern blotting for RNA, or direct biochemical sequencing are also common techniques that you will learn more about in future biology classes.

III. Proteins

Proteins are long chains of several hundred to several thousand amino acids. These proteins are assembled using 20 different amino acid structures.

Each of the 20 amino acids has a common core structure, consisting of a central carbon atom called the α-carbon, which is covalently bonded to an amino group (NH$_2$), a carboxylic acid group (COOH), and a hydrogen (H). Each amino acid differs in that the α-carbon is covalently bonded to a fourth group (R) that is different for each amino acid. The groups occupying the R location differ in their physical properties. Specifically, the different physical properties of R groups are nonpolar (hydrophobic), polar (hydrophilic), acidic (anions), or basic (cations) (Fig. 5.15).

Amino acids are joined together through a dehydration synthesis/condensation reaction that links the amino group of one amino acid to the carboxyl acid group of another amino acid (Fig. 5.16). The resulting bond between the amino and carboxyl groups is called a peptide bond. Multiple amino acids are linked in this manner to form polypeptide chains. At one end of each chain is a free amino group known as the amino terminus (N-terminus), while at the other end is a free carboxyl group known as the carboxyl terminus (C-terminus).

N-terminus

\[
\begin{array}{c}
\text{H} \\
\text{R} \\
\text{C-C-N-C-C-N-C-C-N-C-C-N-C-C-N-C-C-N-C-C} \\
\text{H} \\
\text{O} \\
\text{H} \\
\text{H} \\
\text{O} \\
\text{H} \\
\text{O} \\
\text{H} \\
\text{O} \\
\text{H} \\
\text{O} \\
\text{H} \\
\text{O} \\
\text{H} \\
\text{H} \\
\text{H}_2\text{N} \\
\end{array}
\]

C-terminus

Locate the peptide bonds in this short peptide.

The term polypeptide, or simply, peptide, is typically used for chains containing no more than 50 amino acids, while the term protein applies to molecules containing more than 50 amino acids. However, in practice any amino acid chain containing no more than a few hundred amino acids can accurately be described as a peptide.

The specific sequence of amino acids in a protein is the primary (1º) structure of that protein, and this structure may be recited by naming each amino acid beginning on the N-terminus (Fig. 5.18).

What is the primary structure of this short peptide?

Of course, proteins twist, turn, and fold into complex three dimensional shapes. These twists and turns are possible because there is free rotation of the bonds around the α-carbon of each amino acid. This leads to the formation of structures such as the α-helix and the β-sheet, which are examples of secondary
(2\textsuperscript{nd}) structures of proteins. Both of these types of secondary structure are very common and multiple \(\alpha\)-helices and \(\beta\)-sheets may be found in almost all proteins such as the enzyme lysozyme shown in Fig. 5.20.

The \(\alpha\)-helix is a spiral or corkscrew shape that a peptide chain can form due to hydrogen bonding between an N-H and a C = O within the peptide backbone. Specifically, hydrogen bonds form between an N-H and a C = O exactly 4 amino acids away. The only way for a hydrogen bond to form between two atoms separated by 4 amino acids is for the peptide chain to assume a helical conformation (Fig. 5.20). The R groups of the amino acids, which are not shown in the figure, all project outward, away from the helix and are not involved in forming the hydrogen bonds that form the helix.

The \(\beta\)-sheet is a planar structure that forms when the linear peptide chain turns or folds back and runs parallel to another region of linear peptide chain. This allows hydrogen bonds to form between N-H and C = O groups within the backbone of the chain just as in the \(\alpha\)-helix except that the amino acids forming these hydrogen bonds may be separated by many amino acids within the primary sequence (Fig. 5.20).

The third level of protein structure, tertiary \(3\textsuperscript{rd}) structure, is the more complex three dimensional shape that the protein assumes. The formation of this tertiary conformation is due to the physical characteristics of the R groups of the amino acids in the protein. Remember that proteins are composed of four different classifications of amino acids. While the polar and charged amino acids are hydrophilic and interact readily with the water molecules of the aqueous environment of the cell, the nonpolar amino acids are hydrophobic and thus, minimize their interaction with the aqueous environment of the cell. In order to move away from the aqueous environment, nonpolar amino acids are typically found clustered together within the tertiary structure and away from the surface of the protein. Meanwhile, primarily hydrophilic amino acids are found on the exterior surface of proteins.

The tertiary structure is also affected by the interactions between the various hydrophilic amino acids (Fig. 5.22). For example, polar amino acids may form hydrogen bonds with each other; anionic and cationic amino acids may form ionic bonds; while two amino acids carrying the same charge may repel each other. These interactions contribute to the tertiary structure by drawing different regions of the protein together or pushing them apart. For example a lysine residue on one \(\alpha\)-helix in the protein may form an ionic bond with an aspartate residue in a \(\beta\)-sheet, thus drawing those two parts of the protein into closer proximity. Additional contributions to tertiary structure may come from the formation of covalent bonds. Specifically, two cysteine residues are capable of forming a covalent bond, called a disulfide bond, between their sulfhydryl groups. Disulfide bonds are usually only found in extracellular proteins, but they make a significant contribution to the tertiary structure of those proteins.

Once the tertiary level of structure has been obtained, a protein is classified as being either fibrous or globular (Fig. 5.23). Most fibrous proteins have structural roles such as keratin, which is found in hair. However, most globular proteins function as enzymes, transport proteins, or receptors. In order for any proteins to carry out its proper functional role, it must fold correctly into its proper tertiary conformation.

All of these different types of interactions leading up to tertiary structure occur within a single polypeptide chain (Fig. 5.24). However, some functional proteins are composed of more that one polypeptide chain. For example, the globular oxygen-carrying protein hemoglobin, which is found in red blood cells, is composed of four polypeptide chains, and the fibrous protein collagen is composed of three polypeptide chains. The interactions between multiple polypeptides leads to a fourth level of structure called the quarternary \(4\textsuperscript{th}) structure (Fig. 5.23).

Scientists have the choice of several very sensitive and specific assays to measure proteins. These assays are accurate both qualitatively and quantitatively. The simplest is known as the Biuret assay. The Biuret reagent is similar to Benedict’s reagent in that its primary component is Cu\(^{2+}\) ions in an alkaline solution, and it is light blue in color. The Cu\(^{2+}\) ions in Biuret reagent react with proteins because they form complexes with the nitrogen atoms that are participating in the formation of peptide bonds. The complexes that form are purple in color, and the number of complexes that form is proportional to the number of peptide bonds present (i.e. the amount of protein). Therefore, the amount of protein present is proportional to the change in color.

To accurately quantitate the color change, a spectrophotometer is used to measure the absorption of light at 550nm in samples containing protein versus samples without protein. The amount of light absorbed is proportional to the amount of protein present. To convert the amount of absorption into a specific quantity of protein, a standard curve must be run using known quantities of protein.
The Biuret assay is very specific for amino groups, and thus, it is suitable for measuring protein in most biological samples. Also, it is very linear over a wide range of proteins concentrations, but it is limited by its sensitivity. It is only able to detect protein concentrations down to 1 mg/ml, while other assays, such as the Bradford assay, are almost 1000 times more sensitive.

To test for the presence and quantity of a specific protein in a sample of many different proteins, the sample of proteins must be subjected to a western blot. This requires that the proteins be separated based on their size by electrophoresis through a polyacrylamide gel (polyacrylamide gel electrophoresis: PAGE). The separated proteins in the gel are then transferred to a membrane of nylon or nitrocellulose. This membrane is then incubated with a specific antibody that binds to the protein of interest. The antibody is linked to an enzyme, and when the substrate is added, the enzyme catalyzes a reaction that can be visualized. These enzymes act on their substrates, and either light is emitted, or a color change occurs on the blot. This allows the protein to be localized on the blot, and the intensity of the emitted light or the intensity of the color change may be used to compare the quantity of the protein to the amount of the same protein in other samples.

A common test for free amino acids uses a reagent called ninhydrin. Ninhydrin is a strong oxidizing agent that reacts with the amino group of an amino acid. The product is purple in color, and its absorbance is measured at 540nm using a spectrophotometer. Because all amino acids have amino groups, this test cannot differentiate between individual amino acids. Furthermore, ninhydrin will react with any molecule containing an amino group; therefore, it is not truly specific for amino acids. Would ninhydrin react at all with proteins?