]

BIOCHEMISTRY REVIEW

Overview of Biomolecules

Deborah W. Louda, Associate Professor
Charles E. Schmidt College of Medicine
Florida Atlantic University

Table of Contents

Accompanying power point slides are indicated by **(PP #)**.

Chapter 1: Introduction to Biomolecules

Biochemistry is the study of the chemistry of cells and organisms. Thus it is concerned with the types of molecules found in biological systems, their structure, and their chemical properties. Biochemistry also deals with the function of these molecules, how they interact, and what reactions they undergo.

I. Properties of Biomolecules

A. General Properties

Biomolecules are organic molecules, not fundamentally different from other, typical organic molecules. They are the same types of molecules, react in the same ways, and obey the same physical laws.

B. Composition and Structure

Biomolecules contain mainly carbon, which behaves as it always does in organic compounds, forming 4 bonds, usually with a tetrahedral arrangement. **(PP 2)** The carbon skeleton can be linear, branched, cyclic, or aromatic. Other important elements are H, O, N, P and S. About 30 elements are required by biological systems, including iodine and many metals, though most of these are needed in only trace amounts. **(PP 3)**

Biomolecules contain the same types of functional groups as do organic molecules, including hydroxyl groups, amino groups, carbonyl groups, carboxyl groups, etc. **(PP 4-5)** However, many biomolecules are polyfunctional, containing two or more different functional groups which can influence each other's reactivity. **(PP 6)** Biomolecules tend to be larger than typical organic molecules. Small biomolecules have molecular weights over 100, while most biomolecules have molecular weights in the thousands, millions, or even billions. Because of their large size, the majority of biomolecules have specific 3-dimensional shapes. The atoms of a biomolecule are arranged in space in a precise way, and proper arrangement is usually needed for proper function. The 3-dimensional shape is maintained by numerous non-covalent bonds between atoms in the molecule. **(PP 7)** Because of the weak nature of most non-covalent bonds and because of interactions between the biomolecule and the solvent, the biomolecule's structure is flexible rather than static.

C. Stereochemistry

As is common with organic compounds, many biomolecules exhibit stereochemistry. When four different types of atoms or functional groups are bonded to one carbon atom, the carbon is stereogenic (or chiral or asymmetric) and the

compound can exist in two different isomeric forms that have different configurations in space. The two configurations are mirror images of each other and are not superimposable. **(PP 8)** When two compounds are mirror images of each other they are called enantiomers or optical isomers, a subclass of stereoisomers. Enantiomers usually have identical chemical properties, and differ only in the way they rotate plane-polarized light or interact with other chiral compounds. Most biomolecules have several or many asymmetric carbons and so may have many diastereomers, a subclass of stereoisomers that are non-mirror images and have different properties. **(PP 9)** Stereochemistry is important because biological systems usually use only one specific isomer of a given compound.

II. Types of Biomolecules

Biomolecules can be divided into several major classes and a few minor classes.

A. Amino Acids and Proteins

Amino acids are relatively small molecules with molecular weights around 100-200. **(PP 10)** They are used to produce energy, to synthesize other molecules like hormones, and to make proteins. Proteins are polymers of amino acids. **(PP 11)** They fold into specific shapes and range in molecular weight from several thousand to over a million. **(PP 12)** Proteins function as enzymes (which catalyze reactions), structural elements, transport molecules, antibodies, etc.

B. Carbohydrates (sugars & starches)

The smallest carbohydrates are the monosaccharides with molecular weights of around 100-200. **(PP 13)** They are a major source of energy for biological systems. Polysaccharides are polymers of monosaccharides with molecular weights often in the millions. **(PP 14)** Polysaccharides also have definite shapes and serve as structural elements or as stored metabolic energy. **(PP 15)**

C. Lipids (fats & oils)

Lipids are relatively small water-insoluble molecules with molecular weights of up to 750-1500. **(PP 16)** Because they are defined by their water-insolubility, they are chemically more diverse than the other classes of biomolecules, with about half a dozen major types. Lipids are used for energy production and storage, hormones, structural elements of cell membranes, and vitamins. Lipids do not polymerize to form macromolecules, but they can aggregate non-covalently to form very large structures. **(PP 17)**

D. Nucleotides and Nucleic Acids

Nucleotides are relatively small molecules with molecular weights in the hundreds. **(PP 18)** They function in transferring energy and in helping enzymes to catalyze reactions. Nucleic acids (DNA and RNA) are large polymers of nucleotides, with molecular weights up into the billions. **(PP 19)** They form structures like the double helix, and they function in storing, transmitting, and utilizing genetic information. **(PP 20)**

E. Small Organic Molecules

In addition to the major classes of biomolecules, there are many relatively small organic molecules required by cells for very specific functions; these molecules do not fall neatly into one of the above major categories. These molecules can be precursors of biomolecules that help enzymes function (often related to vitamins), or can be intermediates in metabolic pathways, etc. **(PP 21)**

F. Inorganic Ions

Though not actually biomolecules, many inorganic ions are required by cells, often in trace amounts. These include calcium, sodium, iron, magnesium, potassium, chlorine, etc. Inorganic ions perform a variety of functions such as structural elements (calcium in bone), regulation of osmotic pressure and transport (sodium), and components of proteins and enzymes (iron).

G. Combinations of Biomolecules

Sometimes one biomolecule can contain components from two of the major classes, such as a lipoprotein (lipid plus protein) or a glycoprotein (carbohydrate plus protein).

Chapter 2: Amino Acids

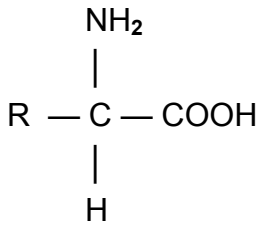
I. Introduction

The major function of amino acids is to act as the building blocks of proteins. Amino acids themselves can be used by the cell to produce energy and are the starting point for making many nitrogen-containing compounds.

II. General Structure

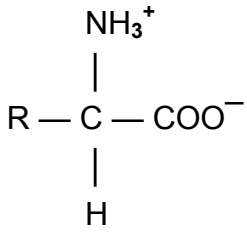
A. Formula

As the name implies, amino acids contain two functional groups, a carboxylic acid group and an amino group. The common amino acids are α-amino acids where both functional groups are attached to the same carbon atom.



Also attached to the central carbon are a hydrogen atom and an R group, which is different in each amino acid.

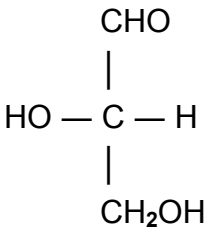
The form above is called the non-ionic form. Both the amino group and carboxyl group are capable of ionizing. At neutral pH, which is normal for biological systems, both groups are ionized. **(PP 2)**



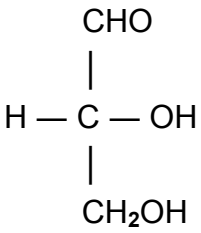
This doubly ionized form is called the zwitterion (hybrid ion with one positive charge and one negative charge) and overall has a zero charge. **(PP 3-6)** Crystalline amino acids have this structure, and the electrostatic forces between molecules explain the higher-than-expected melting points of amino acids.

B. Stereochemistry

If the R group is something other than a hydrogen atom, then the central carbon is asymmetric and there will be two enantiomers (mirror images). **(PP 7)** The compound glyceraldehyde is used as a reference compound for distinguishing stereoisomers. **(PP 8)**

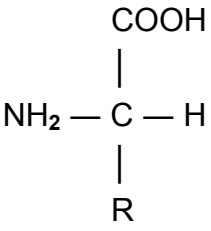


L – glyceraldehyde

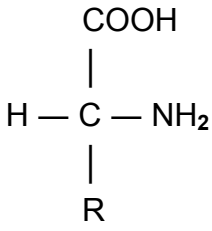


D - glyceraldehyde

The prefixes L and D stand for levo (rotates light to left) and dextro (rotates light to right). For amino acids



L - amino acid



D - amino acid

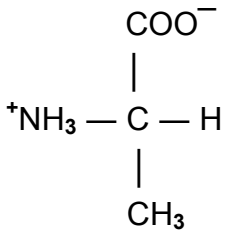
It is the L-amino acids that are biologically important, with very few exceptions. Amino acids found in proteins are normally L-isomers.

C. Classes

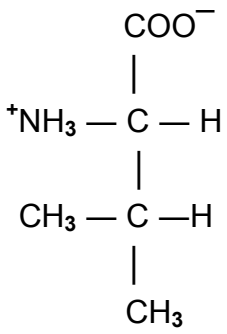
There are 20 common or major amino acids that are found in proteins. They are divided into groups based on the nature of the R group. However, not every amino acid falls neatly into a category, so there can be variations in how amino acids are classified. For instance, the glycine R-group is sometimes classified as hydrophilic and sometimes as hydrophobic. Each amino acid can be designated by a three-letter abbreviation, or by a one-letter abbreviation. **(PP 9)**

1. Nonpolar aliphatic R groups

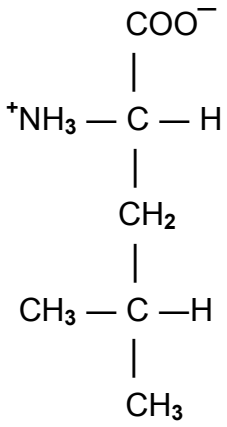
The R group of these amino acids is hydrophobic, but not the entire amino acid. The R groups are mainly hydrocarbon in nature. **(PP 10)**



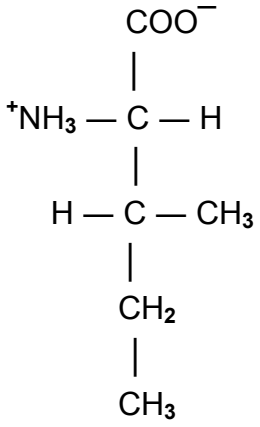
Alanine - Ala, A



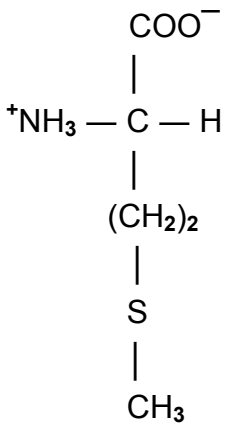
Valine - Val, V



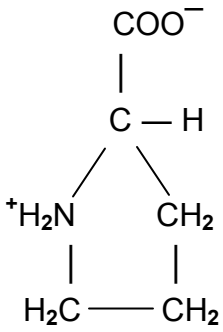
Leucine - Leu, L



Isoleucine - Ile, I
(There are 4 possible isomers, only one of which is biologically important.)



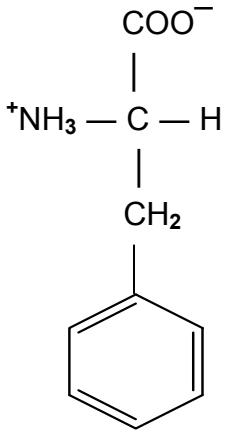
Methionine - Met, M
(The R group contains sulfur.)



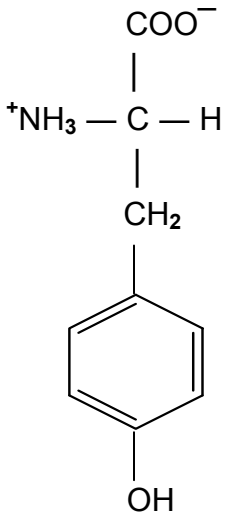
Proline - Pro,P
(This is an imino acid and less flexible than the others.)

2. Aromatic R groups

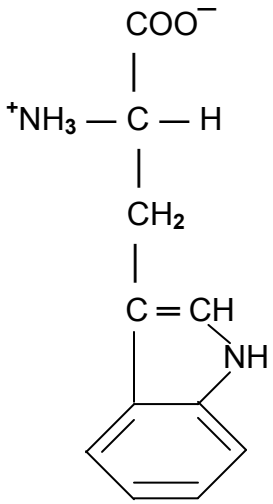
The R groups of these amino acids are aromatic and will absorb UV light at 280 nm. **(PP 11)**



Phenylalanine - Phe, F
(The R group is a phenyl group, and is non-polar.)



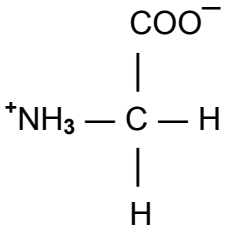
Tyrosine -Tyr, Y
(The R group is a phenolic group, more polar than Phe.)



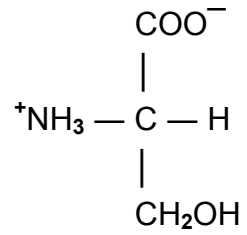
Tryptophan -Trp, W
(The R group is an indole group, more polar than Phe.)

3. Polar, uncharged R groups (at pH = 7.0)

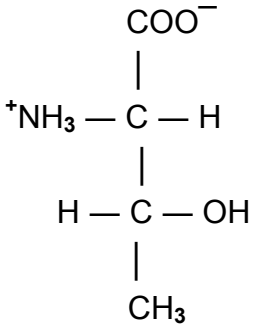
The R groups of these amino acids are hydrophilic. **(PP 12)**



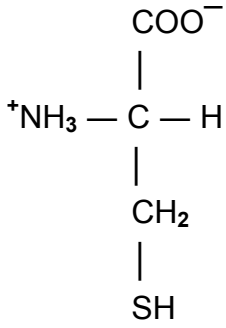
Glycine - Gly, G
(This is the only amino acid without D and L forms. The hydrogen R group is not really hydrophilic, but this amino acid has similar properties to the others in this class.)



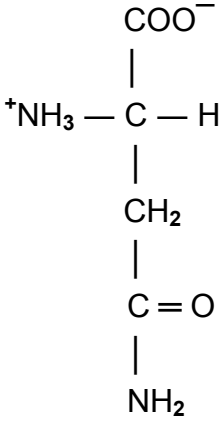
Serine - Ser, S
(The R group is a hydroxyl group.)



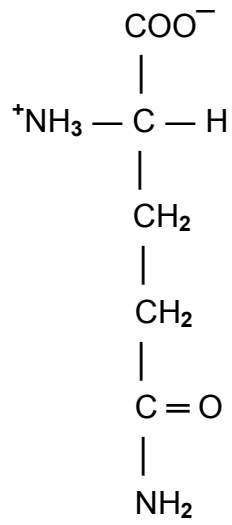
Threonine -Thr, T
(The R group is a hydroxyl group. There are 4 possible isomers, only one of which is biologically important.)



Cysteine - Cys, C
(The R group is a thiol or sulfydryl group.)



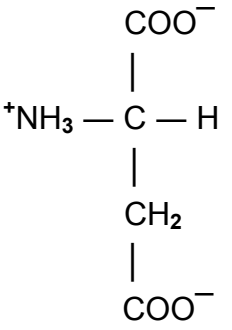
Asparagine - Asn, N
(The R group is an amide group.)



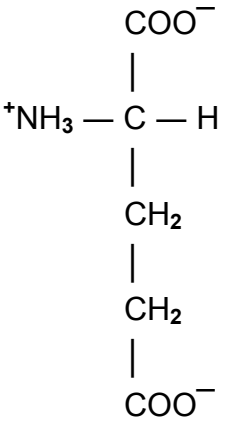
Glutamine - Gln, Q
(The R group is an amide group.)

4. Negatively-charged R groups (at pH 7.0)

These amino acids are acidic and contain an extra carboxyl group. **(PP 13)**



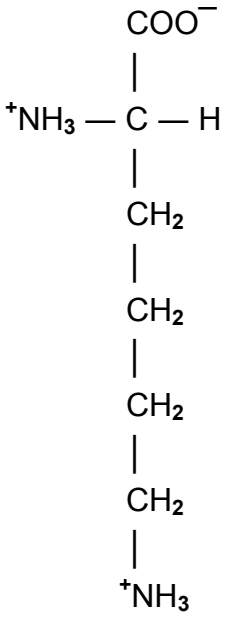
Aspartate - Aspartic Acid - Asp, D



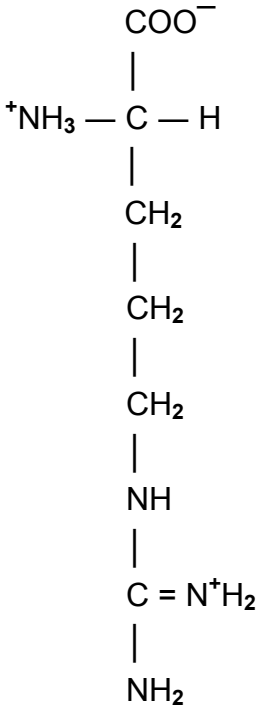
Glutamate - Glutamic Acid - Glu, E

5. Positively-charged R groups (at pH 7.0)

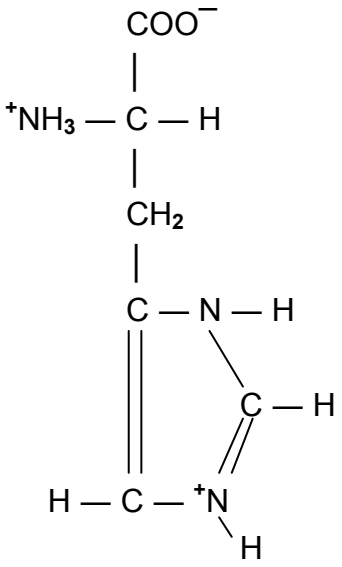
These amino acids are basic and contain an extra basic group. **(PP 14)**



Lysine - Lys, K
(The R group is an amino group.)



Arginine - Arg, R
(The R group is a guanidino group.)



Histidine - His, H
(The R group is an imidazole group. This is the only amino acid with an R group that ionizes around neutral pH, creating two forms of the amino acid.)

6. Amino acid properties

The nature of the R-group determines the behavior of amino acids. **(PP 15-18)**

7. Other amino acids

In addition to the 20 standard amino acids found in proteins, some proteins contain modified amino acids, such as 4-hydroxyproline or ε-N-methyllysine. Another 300 amino acids have been found in biological systems that have functions other than as protein components. They have a variety of functions as precursors in biosynthesis and intermediates in metabolic pathways. **(PP 19)**

III. Acid-Base Properties

Both the carboxyl group and the amino group (and sometimes the R-group) have acid-base properties, so amino acids have complicated acid-base behavior. These properties are crucial in determining the behavior of both individual amino acids and the proteins that contain them.

A. Acid-Base Theory

An acid is a proton donor relative to H_2O. Weak acids do not dissociate completely, while strong acids do. Weak acids reach an equilibrium between dissociated and undissociated forms (generally ionizing less than 1%).

$$HA + H_2O \leftrightarrow H_3O^+ + A^-$$

The strength of the acid is given by its K_a, the acid dissociation constant. **(PP 20)**

$$K_a = \frac{[H_3O^+][A^-]}{[HA]} \approx 10^{-5} \text{ for organic acids}$$

A base is a proton acceptor relative to H_2O.

$$B + H_2O \leftrightarrow BH^+ + OH^-$$

The strength of a base is given by its K_b, the base dissociation constant. **(PP 21)**

$$K_b = \frac{[BH^+]\,[OH^-]}{[B]} \approx 10^{-4} \text{ for organic bases}$$

In any of these reactions, an acid and a base react to form a conjugate base and a conjugate acid. Rather than work with negative exponents, pK_a and pK_b are used.

$$pK_a = -\log K_a \approx 5 \text{ for organic acids}$$

$$pK_b = -\log K_b \approx 4 \text{ for organic bases}$$

To find the relationship between pK_a and pK_b , consider one acid-base pair. **(PP 22)**

$$HA + H_2O \leftrightarrow H_3O^+ + A^- \quad \text{and} \quad A^- + H_2O \leftrightarrow HA + OH^-$$

$$K_a = \frac{[H_3O^+]\,[A^-]}{[HA]} \qquad K_b = \frac{[HA]\,[OH^-]}{[A^-]}$$

$$\text{so} \quad K_a \cdot K_b = \frac{[H_3O^+]\,[A^-]}{[HA]} \cdot \frac{[HA]\,[OH^-]}{[A^-]}$$

$$K_a \cdot K_b = [H_3O^+]\,[OH^-] = K_w = 10^{-14} \text{ at } 25^{\circ} \text{ C (water dissociation constant)}$$

$$\text{and} \quad -\log K_a - \log K_b = -\log K_w = 14$$

Therefore, $pK_a + pK_b = 14$

So rather than use pK_b , only pK_a is used. The strength of a base is described by the pK_a of its conjugate acid where $pK_a = 14 - pK_b \approx 10$ for organic bases. Thus compounds with low pK_a values are good acids while compounds with high pK_a values are good bases. **(PP 23-24)**

B. Titration Curves

One of the best ways to find the pK_a of a substance is to determine its titration curve. A weak acid can be titrated with a strong base, and the pH of the resulting solution plotted as a function of the amount of base added. **(PP 25)**

$$HA + NaOH \rightarrow Na^+ + A^- + H_2O$$

The titration curve is expressed by the Henderson-Hasselbalch equation, derived as follows. **(PP 26)**

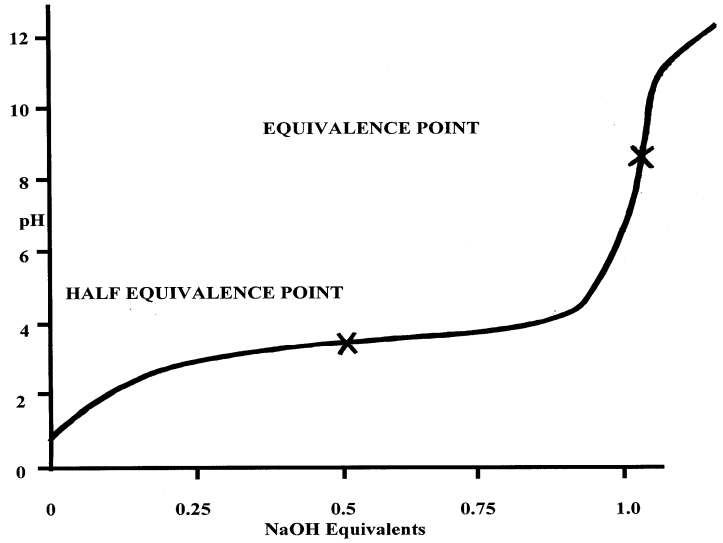
$$K_a = \frac{[H_3O^+][A^-]}{[HA]}$$

$$-\log K_a = -\log [H_3O^+] - \log [A^-] / [HA]$$

$$pK_a = pH - \log [A^-] / [HA]$$

and $$pH = pK_a + \log [A^-] / [HA]$$

Thus at any point on a titration curve (as shown below), the pH is determined first by the strength of the acid HA being titrated (pK_a) and second by the relative amounts of acid (HA) and conjugate base (A^-), which is determined by the amount of NaOH added.



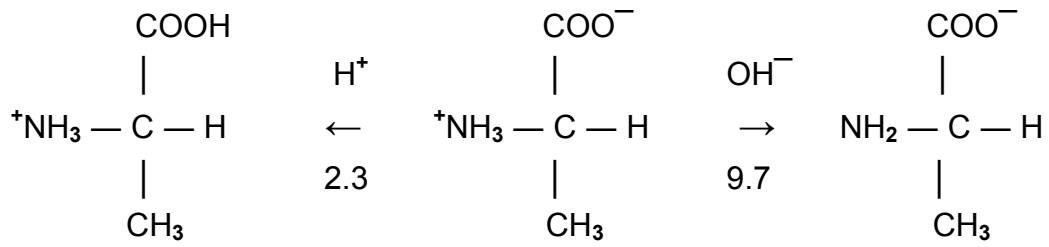
There are two important points on this curve. When there are equal amounts of HA and NaOH present (one equivalent), all the HA has been neutralized and only A^- remains. This is called the equivalence point, and it occurs at a basic pH since A^- is a base. When half as much NaOH is present as HA, then half of the HA has been neutralized to A^-, and the other remains as HA, creating equal amounts of acid and conjugate base. This is the half-equivalence point. The pH here can be calculated from the Henderson-Hasselbalch equation.

$$pH = pK_a + \log [A^-] / [HA]$$

Since $[A^-] = [HA]$, the ratio is 1. Since log 1 = 0, the pH = pK_a. Thus the pK_a can be found from the half-equivalence point on the titration curve. **(PP 27-28)**

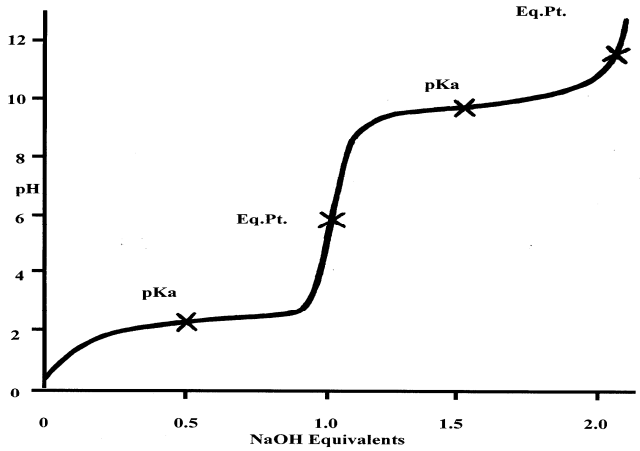
C. Amino Acid Titration

Amino acids can act as both acids and bases and so are called amphoteric. Starting with the zwitterion, the COO^- group can accept a proton as the pH is lowered. The $^+NH_3$ group can lose a proton as the pH raised. Using alanine as an example,



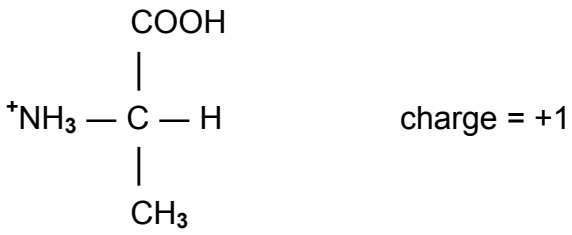
fully protonated zwitterion fully deprotonated

The pK_a of the carboxyl group is 1.8 – 2.4, depending upon the amino acid. The pK_a of the amino group is 8.8 –11.0 (usually 9 – 10). The titration curve of the amino acid will show both pK_a values.



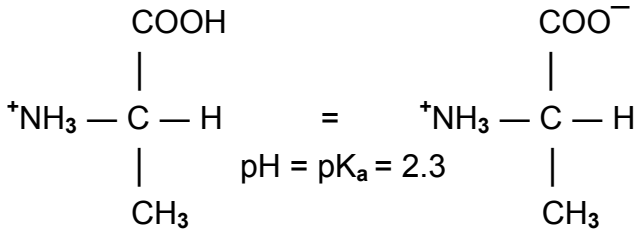
Equivalence points occur in steep areas, while pK_a values occur in flat areas.

The form(s) of the amino acid present depends on the pH. At very low pH = 1, the amino acid will be in the fully protonated form.

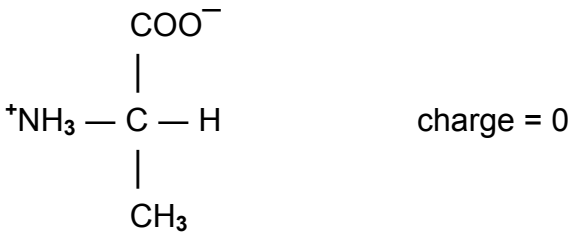


As base is added, the COOH group will be titrated to COO^-. As the pK_a is approached, some of the amino acid will lose its proton. The exact amounts of COOH and COO^- are given by the Henderson-Hasselbalch equation.

When the half-equivalence point has been reached, half the COOH has been converted to COO^-.



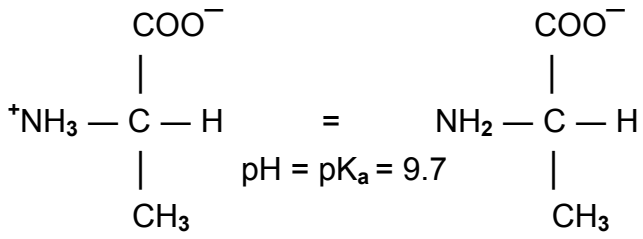
As more base is added, the rest of the COOH is neutralized. All of it is neutralized when the equivalence point is reached.



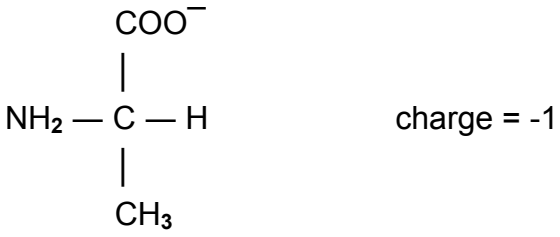
All the amino acid is in the zwitterion form at this point. The amino acid has no net charge, so this is called the isoelectric point or isoionic point (pI). **(PP 29)**

$$pI = \frac{pK_{a1} + pK_{a2}}{2} = \frac{2.3 + 9.7}{2} = 6.0$$

As more base is added, the $^{+}NH_3$ group is titrated to NH_2. When half this group has been titrated, the second half-equivalence point is reached at the pK_a of this group.

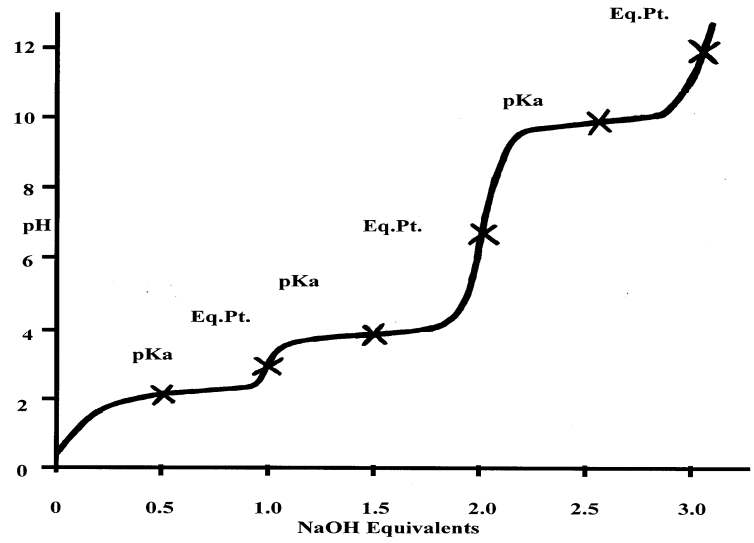


Following the addition of more base, another equivalence point is reached where all the amino acid is in the fully deprotonated form.

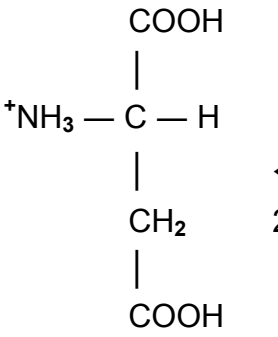
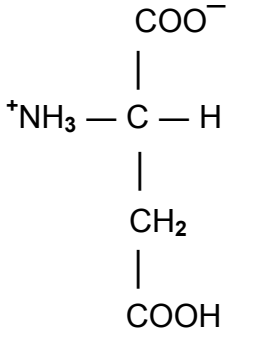
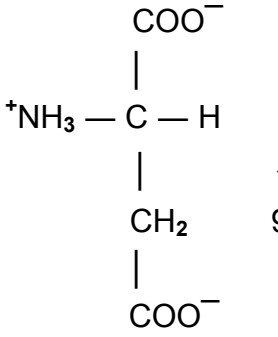
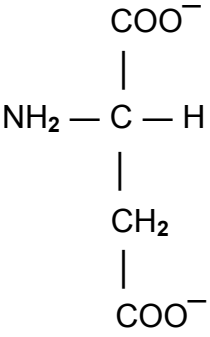


Several points are important. First, all 20 amino acids will have one pK_a for the carboxyl group in the 1.8 – 2.4 range, and another pK_a for the amino group in the 9 – 10 range. The titration of other amino acids with 2 pKa values, such as glycine, would be similar. **(PP 30)** Second, the titration curve has two distinct parts. The carboxyl group is titrated first and then the amino group, so there is no overlap. The carboxyl group reacts with NaOH first because it is the stronger acid. Third, the deprotonation of a group does not all occur abruptly at the pK_a. The Henderson-Hasselbalch equation will give the relative amounts of HA and A^{-} at any pH. Significant amounts of the two forms will be present within one pH unit of the pK_a. Fourth, there will be one form of the amino acid present at each equivalence point. In contrast, there will be a 50/50 mixture of two forms at each half-equivalence point, which is also a pK_a value. Fifth, the titration curve does not have to start at low pH. The amino acid could start at a high pH and be titrated with strong acid (HCl), or start at neutral pH and be titrated with both NaOH and HCl. The exact same titration curve will result. Sixth, the pK_a values of amino acids are somewhat lower than those of other organic acids due to the proximity and influence of the amino group which creates a stronger carboxyl group. **(PP 31-32)**

Some amino acids have an R group with acid-base properties. If the R group is ionizable, then there will be a third pK_a, the value of which depends upon the nature of the group. For example, aspartate has pK_a values of 2.1, 9.8, and 3.9.



The curve has three distinct areas, each corresponding to the titration of one group on the amino acid. The pK_a values can be assigned to each group, with 2.1 for the α-carboxyl group, 9.8 for the amino group, and 3.9 for the R group. The following forms of the amino acid occur throughout the titration, with one form present at each equivalence point and a mixture of forms present at each pK_a.

One form at pH < 1.0		One form at pH = 3.0		One form at pH = 6.85		One form at pH > 11.0
	↔ 2.1		↔ 3.9		↔ 9.8	
	Two forms at pH = pK_a = 2.1		Two forms at pH = pK_a = 3.9		Two forms at pH = pK_a = 9.8	

Other amino acids with 3 pK_a values are: **(PP 33)**

Glutamate – 2.2, 4.3 (R), 9.7

At pH = 4.3, the acidic R group ionizes- $COOH \leftrightarrow COO^-$

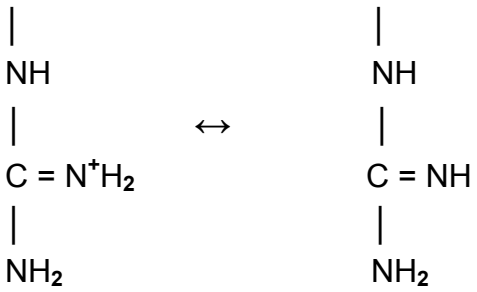
The titration curve of glutamate would be similar to that of aspartate. **(PP 34)**

Lysine – 2.2, 9.0, 10.5 (R)

At pH = 10.5, the basic R group ionizes- $^+NH_3 \leftrightarrow NH_2$

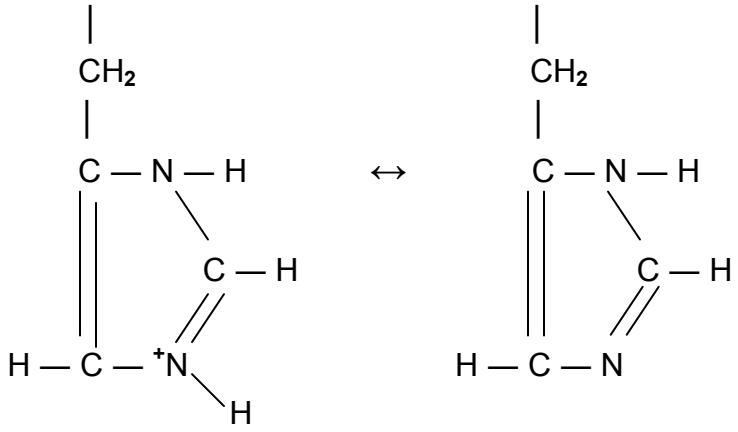
Arginine – 2.2, 9.0, 12.5 (R)

At pH = 12.5, the basic R group ionizes-



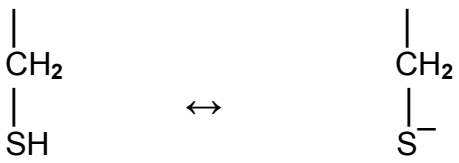
Histidine – 1.8, 6.0 (R), 9.2

At pH = 6.0, the basic R group ionizes-



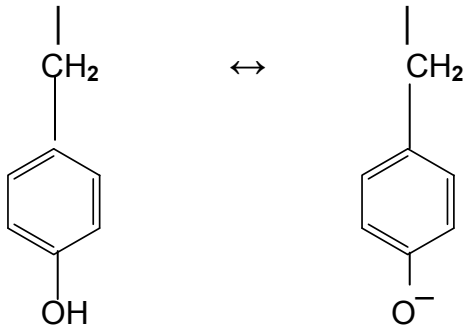
Cysteine – 1.7, 8.3 (R), 10.8

At pH = 8.3, the R group ionizes-



Tyrosine – 2.2, 9.1, 10.1 (R)

At pH = 10.1, the R group ionizes-



The R-groups of serine and threonine can ionize but only above pH= 13, so these pKa values are often ignored. For each amino acid, the charge will vary with pH depending upon the groups present and their pKa values. **(PP 35-38)**

IV. Reactions of Amino Acids

Amino acids will undergo all the reactions typical of acids and amines. However, some specialized reactions are used in biochemistry to detect and quantify amino acids.

A. Ninhydrin reacts with all amino acids having α-amino groups to give a purple-blue color. Proline yields a different yellow product because it is an imino acid. The color of the product can be quantified to find the amino acid concentration. **(PP 39)**

B. Sanger's reagent (1-fluoro-2,4-dinitrobenzene) reacts with amino groups to give a yellow product. Unlike ninhydrin, the R group of the amino acid is part of the product, so the different amino acids can be distinguished as well as quantified. **(PP 40)**

C. Dansyl chloride reacts with amino groups to give a fluorescent product. Very small amounts of amino acids can be detected and quantified. Different amino acids can be distinguished. **(PP 41)**

D. Other groups

The carboxyl group does not undergo convenient, color-producing reactions. A few R groups (including Cys, Tyr, and Arg) can undergo specific modification. Some of these reactions with R groups are important in protein studies, such as reactions with Cys, while other such reactions are not widely used. Sometimes the reagents used for these reactions produce side reactions or yield results that can be difficult to distinguish.

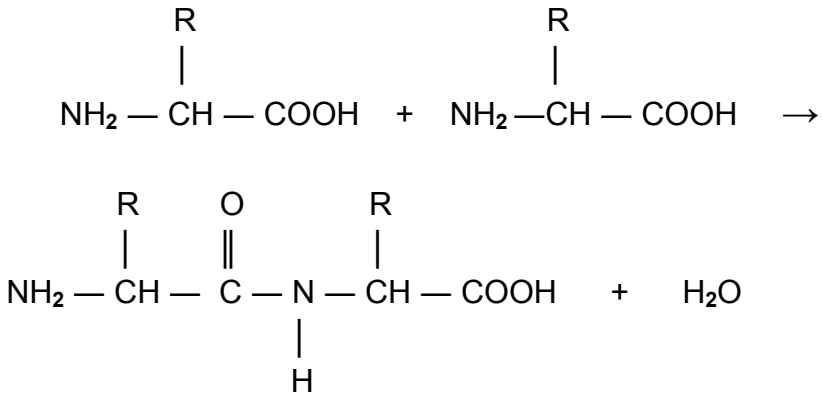
Chapter 3: Peptides

I. Introduction

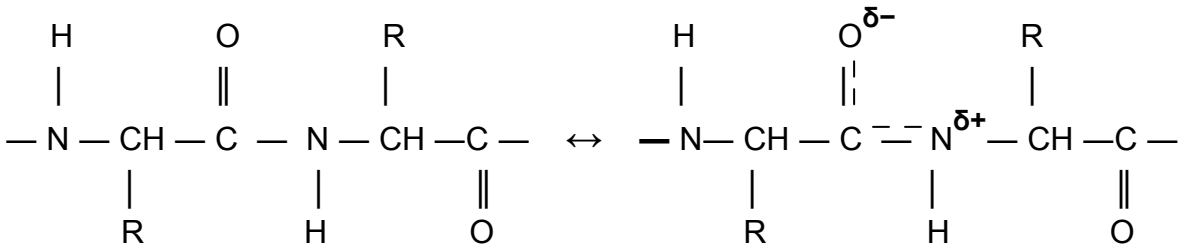
Amino acids can join together to form polymers. When a few amino acids are joined together (2 - ~50), the molecule is called a peptide. When more amino acids are joined together, the molecule is called a protein.

II. Peptide Bonds

The covalent linkage between amino acids is called a peptide bond, which is chemically an amide bond. **(PP 2)**



Additional amino acids can be added onto both ends. The bond is mostly planar and rigid, with partial double bond character due to resonance forms. **(PP 3-4)**



There is restricted rotation around the peptide bond. The configuration around a peptide bond is normally trans. **(PP 5)**

III. Properties of Peptides

A. Naming

By convention, the amino acid with the free α-amino group is written on the left and is the amino-terminal or N-terminal of the peptide. The amino acid with the free

carboxyl group is written on the right and is called the carboxyl-terminal or C-terminal. **(PP 6)** Peptides are named according to the amino acid components, beginning with the N-terminal. **(PP 7)**

Ala - Gln - Tyr - Gly - Ser - Phe - Lys

N-terminal C-Terminal

B. Chemical Properties

Peptide bonds are stable under biological conditions. Peptide bonds can be hydrolyzed by heating in strong acid or strong base. Peptide bonds can be broken by a group of enzymes known as proteases (exo vs. endopeptidase).

C. Optical Properties

Peptides are composed of L-amino acids, so they will be optically active. For very short peptides, the optical activity is the sum of the optical activities of the component amino acids. For longer peptides, it is not.

D. Acid-Base Properties

The -NH- and C = O in a peptide bond have no significant tendency to ionize or protonate and so do not have acid-base properties. Peptides do have acid-base properties caused by the N-terminal α-amino group, the C-terminal α-carboxyl group, and any ionizable R-groups present. Thus a peptide may have several to many pK_a values. **(PP 8)** The exact pK_a values may differ from their value in the free amino acids. Peptides can be titrated, but it may be impossible to pick out individual pK_a values if several are similar and overlap. There will be some pH, however, where the negative charges balance the positive charges, and the peptide has no net charge. This point will be the peptide's isoelectric point. **(PP 9-12)**

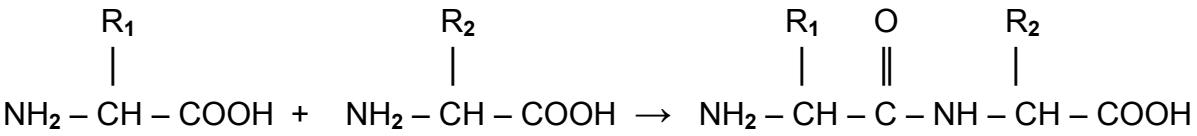
E. Biological Activity

Many peptides are biologically active. Some hormones are peptides, such as insulin (51 amino acids) and glucagon (29 amino acids). Some toxins and antibiotics are peptides.

IV. Peptide Synthesis

Peptides can be made very efficiently by the cell. A 100-amino acid protein with a specific amino acid sequence can be made by the cell in about five seconds.

Peptide synthesis can be done in the laboratory with much more difficulty. The problem is to join together the amino acids in a specific sequence. **(PP 13)**

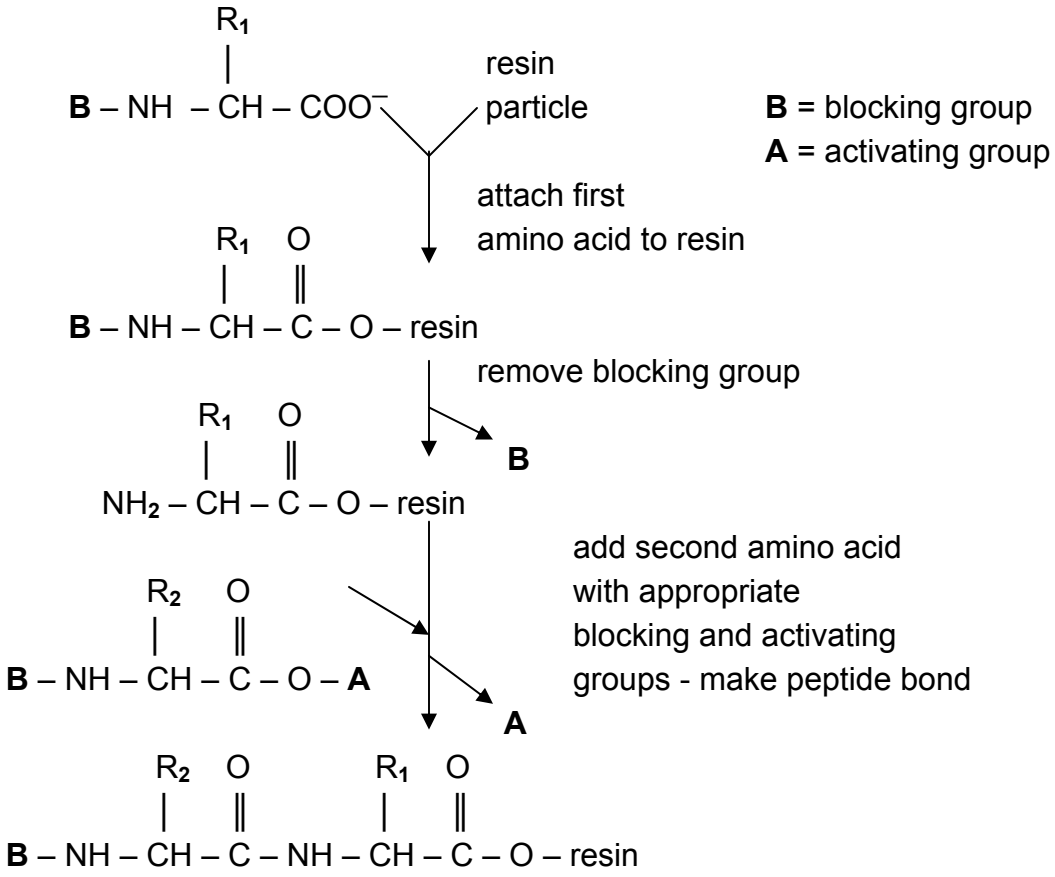


These two amino acids could react to give amino acid 2 as the N-terminal, or two molecules of amino acid 1 or two molecules of amino acid 2 could join together. As more amino acids are joined, the problem escalates.

To synthesize a peptide, the following things must be done. The amino groups that should not form a peptide bond must be reacted with a blocking reagent. Certain R-groups (amine, carboxyl, sulfydryl groups) may also have to be protected. The appropriate carboxyl group must be activated so it will react to form a peptide bond under laboratory conditions (rather than a salt). Once the peptide bond is made, protecting groups need to be removed. Then the process must be repeated for adding the third amino acid.

Thus making even a small peptide involves a large number of reactions. Since the yield from any reaction is never 100%, the major limitation is the yield of the peptide. Even if the yield from each reaction is high, by the time 50 amino acids have been joined together, the overall yield is low.

Instead of doing all reactions in solution, an improvement was made by joining one end of the growing peptide to an insoluble resin solid support which could then be isolated by filtration or centrifugation. **(PP 14)** In this solid-phase synthesis, reagents are added and the reactions occur, while the yield of the peptide remains high because it is attached to a resin particle. **(PP 15)** This procedure has been automated and can make a 100 - amino acid protein in a few days.



This process is repeated for each additional amino acid. When completed, the peptide is detached from the resin particle.

Chapter 4: Protein Sequence

I. Introduction

Proteins are the same as peptides, only larger with more amino acids. The amino acids are still joined in a long, linear chain by peptide bonds.

A. Size and Structure

Proteins can range in size from around 50 amino acids (MW = 6000) to around 8000 amino acids (MW = 1,000,000). One exceptionally large protein has 34,000 amino acids and a molecular weight of 3.8 million. **(PP 2)** Some proteins consist of a single polypeptide chain **(PP 3)**, but many of the large proteins consist of two or more polypeptide chains associated together. **(PP 4)** Such proteins are called multimeric or multisubunit proteins, and the individual polypeptide chains are called subunits. The subunits may be identical or they may be of different types. The number of subunits may range from two to 50 or more. The amino acids are almost always the 20 standard amino acids, but sometimes these are slightly modified (hydroxyproline). **(PP 5-6)**

B. Prosthetic Groups

Many proteins consist only of amino acids, but some contain a non-amino-acid component called a prosthetic group. **(PP 7)** Some prosthetic groups are small organic molecules related to vitamins, some are lipids (lipoproteins), some are carbohydrates (glycoproteins), and some are metal atoms or ions (metalloproteins). **(PP 8)** The prosthetic group is usually important for the protein to function properly.

C. Function

Proteins perform many different functions in biological systems. **(PP 9)**

1. Enzymes catalyze almost all the chemical reactions that occur in the cell. For example, DNA polymerase is a protein that makes DNA.
2. Transport proteins carry specific molecules in the blood or through cell membranes. Hemoglobin carries oxygen.
3. Storage proteins contain nutrients or metabolic energy. Ovalbumin in egg white provides nutrition to the embryo.
4. Motile proteins enable a cell or organism to move. Actin is found in muscle.
5. Structural proteins provide strength or protection. Hair is made of keratin.
6. Defense proteins include antibodies and venoms.
7. Regulatory proteins include hormones and proteins which mediate hormonal effects.

D. Protein Characterization

Proteins can be separated, purified, and characterized using a variety of methods. A fundamental property of a protein is structure. Because proteins are large with complicated 3-dimensional shapes, protein structure is broken down into several different levels. These levels are called primary, secondary, tertiary, and quaternary structure. **(PP 10)**

II. Primary Structure

A. Definition

Primary structure refers to the amino acid sequence of a protein. It means not merely what amino acids are in the protein, but their specific order. To know the primary structure is to be able to list the amino acids, starting with the N-terminal amino acid all the way through the C-terminal amino acid. **(PP 11)**

Asp	-	Lys	-	Ser	-	Thr	-	-	-	-	-	-	-	-	Ala	-	Ile	-	Leu
1		2		3		4									159		160		161

The primary sequence is most important because the sequence is what makes a protein a specific protein. All molecules of the protein lysozyme have the same sequence (and therefore the same number of amino acids and the same molecular weight). All molecules of the protein ovalbumin will have the same sequence, but the ovalbumin sequence will be different from the lysozyme sequence. The sequence is of crucial importance, for one change in the amino acid sequence can make the protein non-functional. Some variation in sequences can occur. For instance, 30% of the proteins in humans have some sequence differences. The same protein from different species will usually show some variation. **(PP 12-13)**

Primary structure is routinely determined. It requires that the protein be pure. Usually some preliminary experiments are done before the actual sequencing.

1. Amino acid composition

The numbers and types of amino acids in a protein (but not the sequence) can be determined by hydrolyzing the protein with strong acid or base or breaking it down enzymatically into its amino acids. The 20 different amino acids are then separated by various techniques and quantitated. The result is that the protein is found to have 10 His, 12 Ala, 9 Lys, etc. **(PP 14)** Alternatively, each amino acid can be expressed as a percentage of the total. This technique is now automated.

The problem is that no breakdown method is completely satisfactory. Strong acid completely destroys Trp and partially destroys Ser, Thr, and Tyr. Amounts of these last three can be estimated by measuring their disappearance over time and extrapolating back to time = 0. In addition, Asn is converted to Asp and Gln is converted to Glu. Base hydrolysis destroys Cys, Ser, Thr, Arg and modifies others, but can be used to measure Trp. Since individual enzymes break only certain peptide bonds, a mixture of enzymes must be used to ensure complete breakdown. Of course, the enzymes digest themselves and contaminate the results, but amounts of Trp, Asn, and Gln can be determined. **(PP 15)**

2. End group analysis

Determining N and C-terminal amino acids gives two reference points in the amino acid sequence. It can also reveal (if two amino acids show up in one test) that the protein is multimeric with different types of subunits.

The N-terminal can be found by reacting the intact protein with Sanger reagent, dansyl chloride, etc., followed by acid hydrolysis. Only the N-terminal has an α-amino group, so it will be modified by the reagent. After the protein is broken down, the modified amino acid can be separated and identified. **(PP 16)**

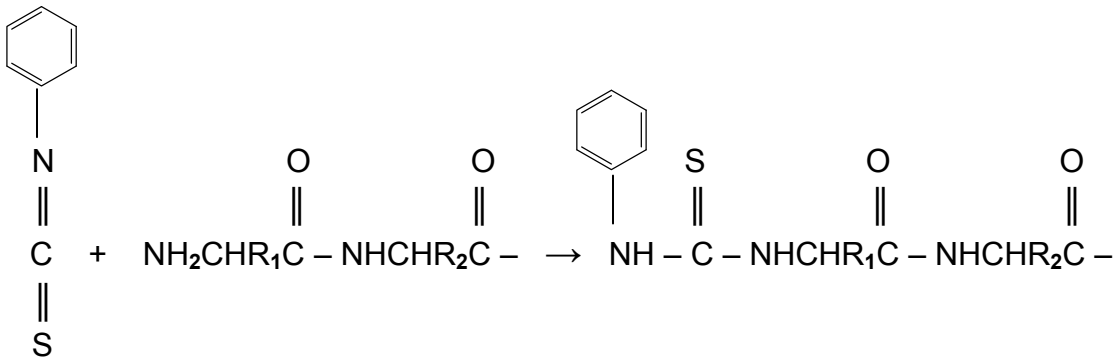
The C-terminal does not undergo a similar reaction. The C-terminal amino acid can be reduced to the amino alcohol using lithium borohydride. The protein is then degraded and the amino alcohol identified, but this can be difficult since the amino alcohol is not colored, fluorescent, etc. Another method is to break down the protein with hydrazine, creating amino acyl hydrazides of all the amino acids except the C-terminal which can then be separated and identified. However, many side reactions occur which makes interpreting the results difficult. **(PP 17)**

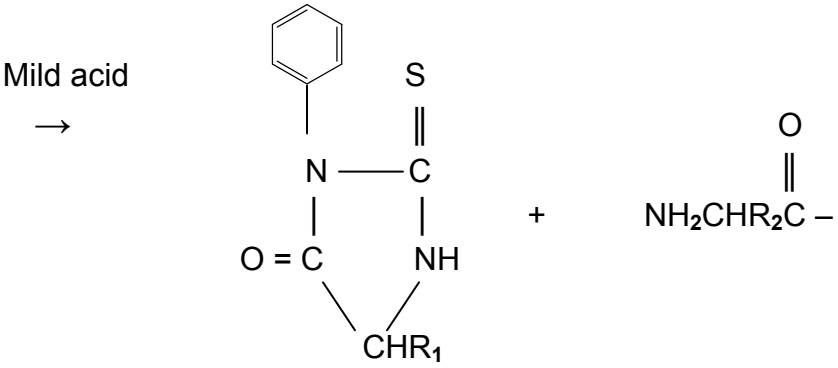
The best method is to use enzymes which degrade the protein from the C-terminal end. Such enzymes are called carboxypeptidases. **(PP 18)** However, each enzyme has different specificity for the amino acids in the peptide bonds it cleaves. For instance, several will not work on peptide bonds involving Pro. In addition, not all peptide bonds are cleaved at the same rate, and once the last amino acid is cleaved off, the next-to-the-last amino acid is susceptible to cleavage. If a mixture of carboxypeptidases is used and the timing is correct, in most cases the C-terminal amino acid is cleaved off and can be identified. **(PP 19-20)**

3. Sequencing

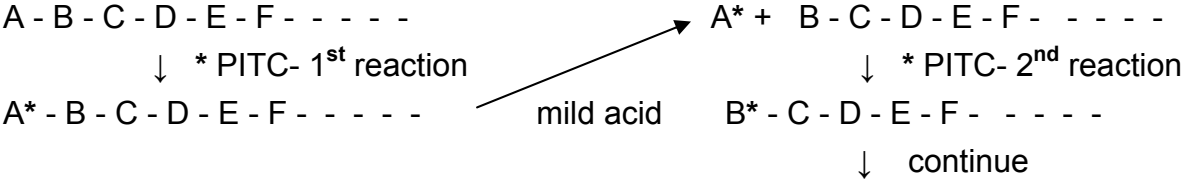
The sequencing itself can be done by the Edman degradation. The Edman reagent is phenylisothiocyanate which reacts with the N-terminal amino acid. Mild acid will then cleave off this modified amino acid leaving the rest of the protein chain intact.

This is unlike other reagents that modify the N-terminal amino acid and then require strong acid to remove the modified amino acid, destroying the rest of the protein. **(PP 21)**





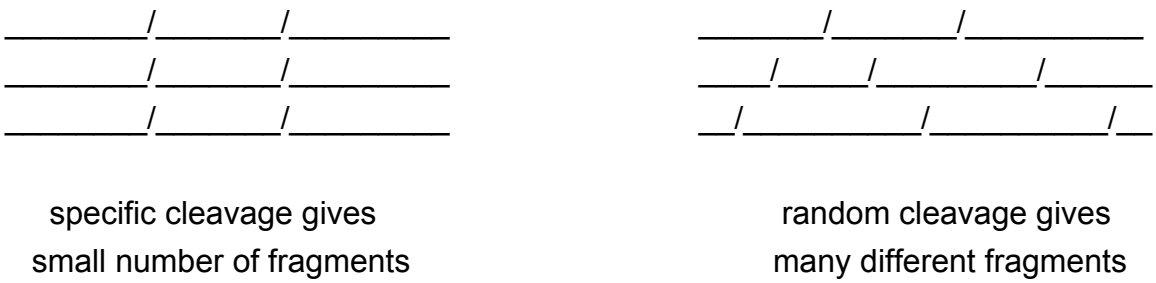
The remainder of the protein chain is separated from the modified amino acid, which is identified. The remainder of the chain is reacted again with phenylisothiocyanate; the second amino acid (the new N-terminal) now reacts, can be cleaved off, and identified. The procedure can be repeated for 50-100 amino acids. **(PP 22)** Since each step requires several reactions, none of which is 100% complete, and also requires a separation where yields are never 100%, there is a limit as to how many times the cycle can be repeated before the results become equivocal. The process has been automated. **(PP 23-24)**



4. Specific cleavage

Since most proteins have more than 100 amino acids, they cannot be directly sequenced in their entirety. In such a case, the large protein is broken down into smaller fragments of 50-100 amino acids. The fragments can then be separated and each can be sequenced individually.

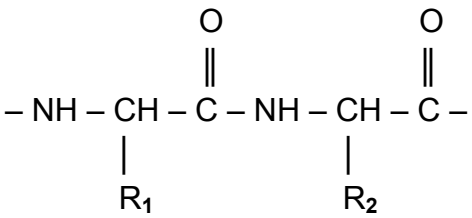
The problem is to break specific peptide bonds in the protein so that each protein molecule is broken into the exact same fragments. If every protein molecule is broken randomly, as by acid hydrolysis, then the fragment mixture would be impossible to sequence. **(PP 25)**



specific cleavage gives small number of fragments

random cleavage gives many different fragments

The best way to specifically cleave the protein chain is by using enzymes. Some enzymes will break a protein chain at specific amino acids.



Trypsin will cleave a protein chain when R_1 is Lys or Arg. **(PP 26)** Chymotrypsin will cleave when R_1 is Phe, Tyr, Trp. Staphylococcal V8 protease cleaves when R_1 is Asp, Glu. Some enzymes are specific for R_2 rather than R_1 . One chemical reagent which cleaves specifically is cyanogen bromide, which cleaves when R_1 is Met. **(PP 27-32)**

Following cleavage and sequencing of each fragment, the problem is to order the fragments. Suppose the following fragments are found for a peptide that contains 15 amino acids, A-O. (This small peptide could be sequenced directly by the Edman degradation without needing specific cleavage, but it illustrates the issue of ordering the specific cleavage fragments.)

A B C D E F G H I J K L M N O

If A has been identified as the N-terminal amino acid and O is the C-terminal amino acid then there are two possible sequences.

A B C D E F G H I J K L M N O

or

A B C D E H I J F G K L M N O

To establish the order a second specific cleavage is done with a second reagent that will generate a second set of fragments.

A B C D E H I J F G K L M N O

Based on these fragments, again two orders are possible.

A B C D E H I J F G K L M N O

or

A B C J F G K D E H I L M N O

Only one order is possible from both sets of fragments. **(PP 33)**

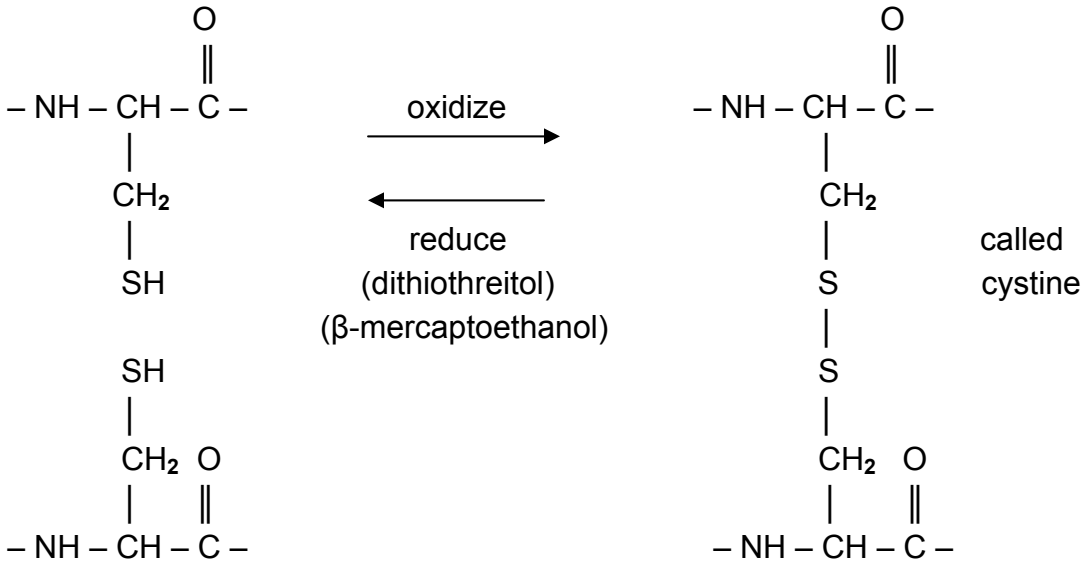
5. Multimeric Proteins

If the subunits of a multimeric protein are of different types, then they must be separated before primary structure is determined. The subunits must be purified and each type sequenced separately.

B. Disulfide Bonds

In addition to peptide bonds, there is another type of covalent linkage which occurs in proteins. This other type of bond is a disulfide bridge. Primary structure can be defined as all the covalent linkages in a protein, which would include knowing the sequence of amino acids joined by peptide bonds as well as knowing the position of disulfide bonds. However, disulfide bonds function to maintain the 3-dimensional shape of the protein, so they are sometimes considered a part of the higher levels of protein structure.

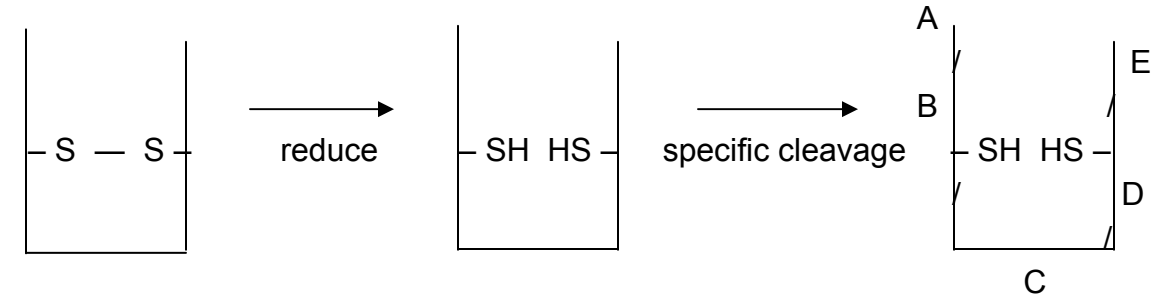
Disulfide bonds occur when two cysteines are close to each other in a protein molecule. **(PP 34)**

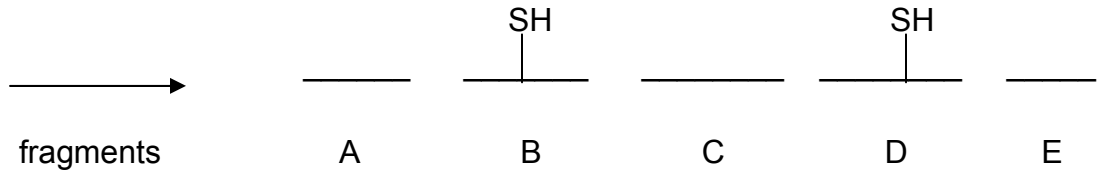


Disulfide bonds can occur within one polypeptide chain **(PP 35)**, or can occur between subunits in a multimeric protein. **(PP 36)**

Disulfide bonds interfere with sequencing and so must be cleaved before sequencing. However, once the sequence is found, the position of disulfide bonds must also be determined. This is done by fragmenting with the disulfide bonds in place to see what parts of the protein are linked by disulfide bonds.

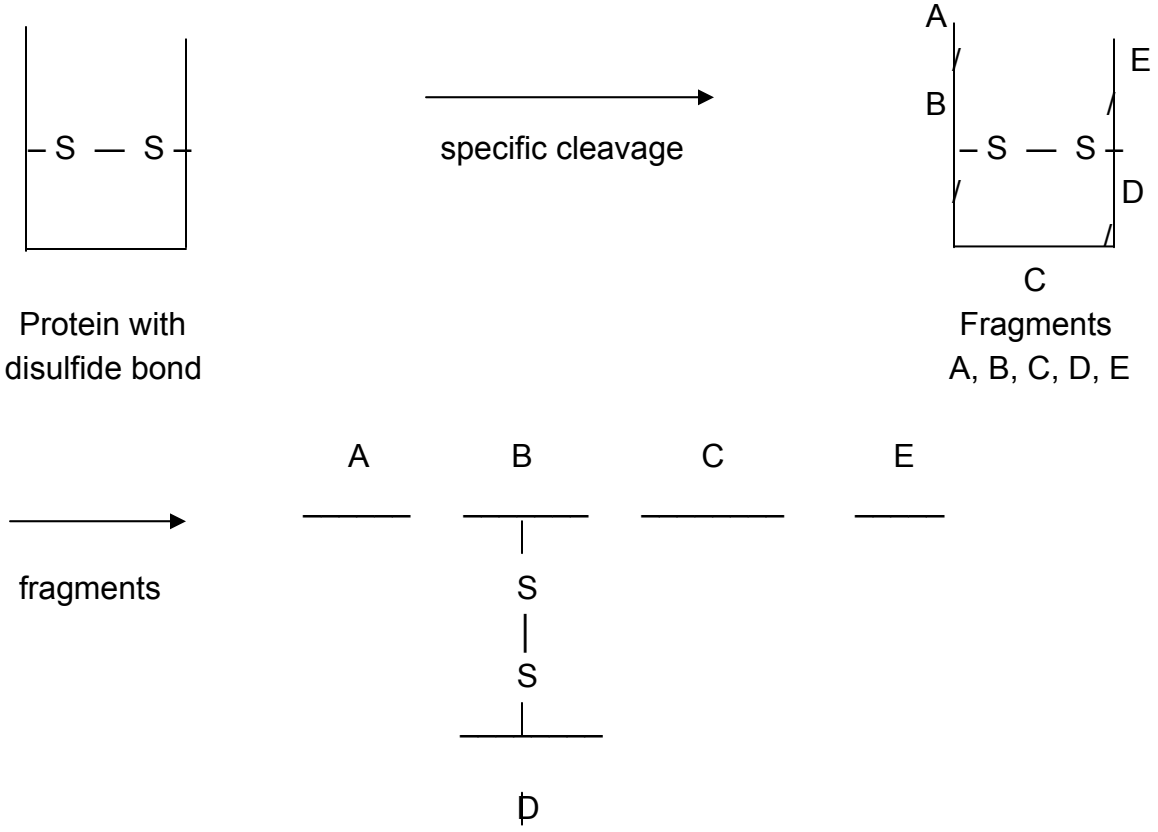
To find the positions of disulfide bonds, the protein is first subjected to conditions that break the disulfide bonds, followed by specific cleavage and sequencing of the fragments.





Each fragment is sequenced.

Then, the protein is specifically cleaved with the disulfide bonds in place.



Fragments connected by disulfide bonds can now be isolated and analyzed to determine the exact position of the disulfide linkages. **(PP 37-38)**

While the Edman degradation is the oldest and most common method for sequencing a protein, mass spectrometry is a more recent technique that can also be used. The precise molecular weights for many fragments from a protein are found, and this information can be compiled to determine a sequence.

C. Information From Sequencing

Sequencing can tell how closely related proteins are by comparing sequences. Certain proteases, for instance, have similar sequences and so belong to the same family of proteases. Trypsin, chymotrypsin and elastase share a significant amount of amino acid sequence, even though they cleave proteins differently. Since sequence information is now available for many proteins, a new protein can often be assigned to a family of proteins based on its sequence and this gives important information about the protein's structure and function. Likewise, the same protein from different species can be compared. The more similar the sequences, the more closely the two species are

related. As an example, cytochrome c is a small protein of about 100 amino acids. The sequence in humans and in chimpanzees is identical. However, cytochrome c in sheep has 10 different amino acids, cytochrome c in fish has 18 different amino acids, and cytochrome c in insects has 31 different amino acids. The evolutionary relationship of the species is reflected in the protein.

The critical amino acids for a protein's function can be found because they are the ones that will be present in all related proteins. In cytochrome c, 28 amino acids are invariant and therefore essential for proper function. Less crucial amino acids can vary without affecting function.

Sometimes a protein's primary structure provides clues about the shape of the protein and the higher levels of the protein's structure. For instance, certain amino acids are more likely to be found on the surface of a protein while others are usually found in the interior. Understanding a protein's sequence and its relationship to other proteins can also help in determining a protein's function.

Chapter 5: Protein Conformation

I. Introduction

Primary structure includes all the covalent bonds in a protein, both peptide bonds and disulfide bonds. However, rotation is possible around many of these covalent bonds, so there are a large number of possible 3-dimensional shapes that a protein can assume. The spatial arrangement of atoms in a protein is called its conformation. Despite the almost limitless number of possible conformations of a protein, each protein will have a specific, unique 3-dimensional structure. **(PP 2)** The higher levels of protein structure describe this conformation with increasing complexity. **(PP 3-5)** There are several important points about conformation.

A. Function of a protein depends upon its conformation. Protein molecules that lose their proper shape will not be able to function properly.

B. The proper conformation for a protein is often (but not always) the one which is the most thermodynamically stable. **(PP 6)** What shape is most stable depends on what amino acids are in the protein. **(PP 7)** Thus conformation depends upon the amino acid sequence. **(PP 8-9)** It is a goal of protein research to be able to deduce protein conformation from a protein's primary structure. Unfortunately, given the large number of possible conformations and the difficulty of estimating energies and stabilities, this cannot be done at present. Limited predictions about the structure of regions in proteins can currently be made.

C. Conformation is maintained and held together mainly by non-covalent forces. (Disulfide bonds and other covalent crosslinks also help to maintain protein shape.) The various non-covalent forces include the following types. **(PP 10)**

1. Hydrogen bonds can form in many ways between certain R groups of amino acids, between portions of the peptide backbone, and between polar amino acids and the surrounding water molecules. **(PP 11)** Many polar amino acids tend to cluster on the outside of a protein molecule where they can interact with the solvent.

2. Ionic interactions involve attractions between opposite charges. Two oppositely charged amino acids can form an ionic bond known as a salt bridge. **(PP 12)** Other electrostatic interactions involve induced or permanent dipoles.

3. Van der Waals forces are weak attractions between close, uncharged atoms. A random variation in the electron cloud of one atom creates a momentary dipole that induces an opposite dipole in another atom and causes an attraction. **(PP 13)**

4. Hydrophobic interactions occur when two or more hydrophobic groups cluster together and so avoid interaction with water. **(PP 14)** There is no actual attraction

between the non-polar groups, but rather the stability comes from the thermodynamic favorability of keeping these groups from water where they cause the water to assume a highly structured solvation layer. Non-polar amino acids tend to cluster in the interior of the protein where water is excluded.

While non-covalent forces are weak, there are a large number of them in a given protein. The amino acid sequence will dictate how the chain must be spatially arranged in order to maximize these forces. **(PP 15-19)**

D. Types

1. Secondary structure refers to the arrangement of neighboring amino acids, which often occurs in a regular, repeating structure.

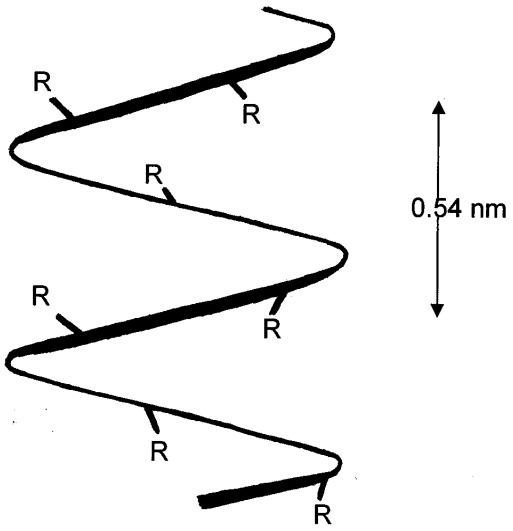
2. Tertiary structure refers to the complete 3-dimensional structure of the polypeptide chain.

3. Quaternary structure occurs in multimeric proteins and refers to the spatial arrangement of the subunits.

II. Secondary Structure

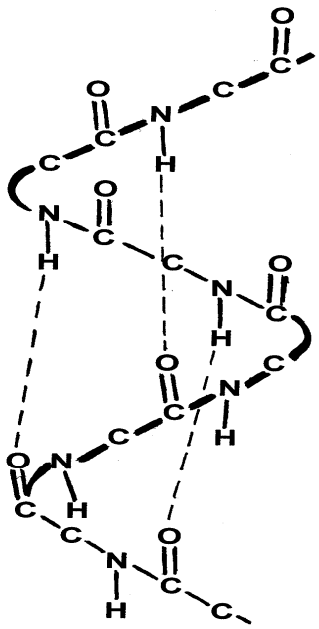
Secondary structure focuses on neighboring amino acids (20-40). The major types of 2° structure are regular, repeating arrangements.

A. α-Helix Structure. The amino acids are arranged into a spiral shape with the peptide backbone in a helix and the R-groups pointing out. **(PP 20)**



One turn of the helix is 0.54 nm and contains 3.6 amino acids. Almost all known α-helices are right-handed (clockwise spiral) while left-handed helices (counterclockwise spiral) are rarely observed. **(PP 21)**

What holds the α-helix together is hydrogen bonding between peptide bond atoms in the backbone. Hydrogen bonds form between an amino acid and the fourth amino acid further up the chain. The accumulation of many hydrogen bonds makes the structure very stable since each N – H and C = O can form a hydrogen bond. **(PP 22-24)**

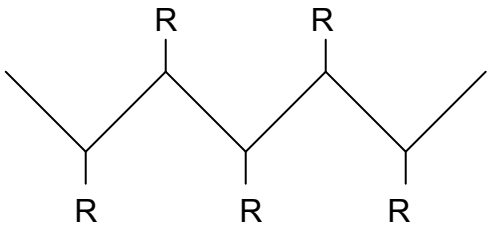


A protein can contain several segments of α-helix. **(PP 25).** However, not all amino acid sequences can form a stable α-helix. Too many amino acids with the same charged R group (acidic or basic) will disrupt an α-helix. Too many bulky R groups will cause a problem (Leu, Thr). Proline cannot conform to the α-helix shape. When present, it causes a bend in the direction of the helix. **(PP 26-27)**

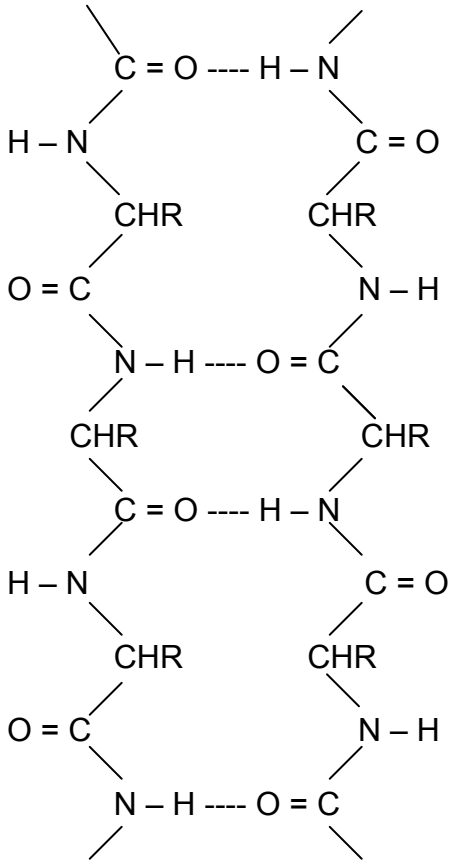
The first R-group and the third or fourth one will interact. Often one will be positively charged and the other negatively charged, or both will be hydrophobic.

B. β-Pleated Sheet Structure. A second type of regular 2° structure is the β-pleated sheet. It is more extended than an α-helix and the chain is arranged in a zig-zag. Several zig-zag chains line up to form a pleated sheet. **(PP 28)**

side view

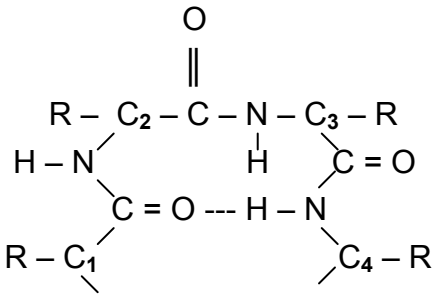


The neighboring chains are H-bonded to each other. **(PP 29)**



More than 2 chains (or different parts of the same chain) can align in this manner. If the peptide bonds run in opposite directions the β-sheet is called antiparallel. **(PP 30-31)** If they run in the same direction it is called parallel. **(PP 32-33)** The structure often contains relatively small R groups like Gly and Ala. The sheet can twist. **(PP 34)** Sometimes there will be one amino acid in one chain that is not H-bonded, creating a β-bulge. **(PP 35-36)**

C. Bends in the Polypeptide Chain. A number of different types of bends occur where the polypeptide chain must reverse direction. Tight turns involve four amino acids, often specific ones, that form a hydrogen bond. Gly and Pro are often found in bends. **(PP 37)**



D. Irregular 2° Structure. In many areas of a protein, the amino acids may not assume a regular, repeating structure. Irregular structure can occur. This does not mean random. The spatial arrangement is still specific, dictated by the amino acids.

E. Secondary Structure in Proteins

Different amino acids occur with different frequencies in the various types of secondary structure. **(PP 38)**

Proteins can be divided into two categories with regard to 2° structure.

1. Globular proteins have a very compact 3-dimensional structure, within which exist different areas of 2° structure. Different globular proteins have very different amounts of 2° structure types. **(PP 39)**

	α	β	irregular
myoglobin	78%	0%	22%
chymotrypsin	14%	45%	41%
lysozyme	40%	12%	48%

2. Fibrous proteins have long, extended structures consisting often of just one main type of 2° structure. There are several categories of fibrous proteins. **(PP 40)**

a. α-keratins

α-keratins make up hair, wool, feathers, nails, skin, horns, etc. The structure of α-keratins is a right-handed α-helix. They contain many hydrophobic residues, making them insoluble and water-repellent. Adjacent chains of α-helix are linked by disulfide bonds. The more disulfide bonds, the stronger the structure is. In hair, two α-helices interact to form coils, two coils combine to make a protofilament, and four protofilaments make one filament. These are packaged and wound to form macrofibrils, many of which make up a hair strand. Disulfide cross-links occur at several levels, producing a strong structure. In the case of hair, the structure is flexible but does not stretch. **(PP 41-42)**

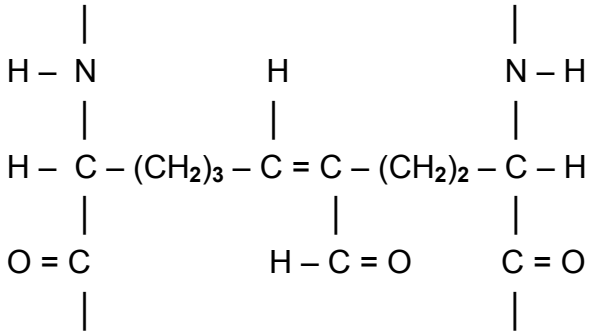
b. β-keratins

β-keratins include silk and spider webs that are flexible but do not stretch. The structure is an antiparallel β-sheet with a high percentage of small amino acids (Gly, Ala). There are no disulfide bonds. The small R groups allow pairs of the β-sheets to stack together, forming an ordered array. **(PP 43)**

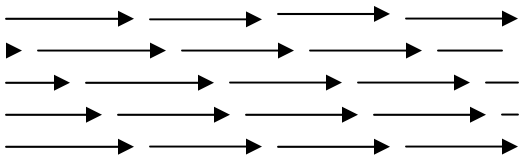
c. Collagen

Collagen (existing in several different types) is found in tendons, cartilage, etc. Collagen has a unique left-handed helix with 3.3 amino acids/ turn. The helix is more extended than an α-helix. Collagen contains 35% Gly and 21% proline plus

hydroxyproline, usually in a repeating sequence of Gly - X - Pro (Hyp). Three such chains (~ 1000 amino acids) are wound together in a right-handed triple helix held together by hydrogen bonds and covalent crosslinks between oxidized lysines. **(PP 44)**



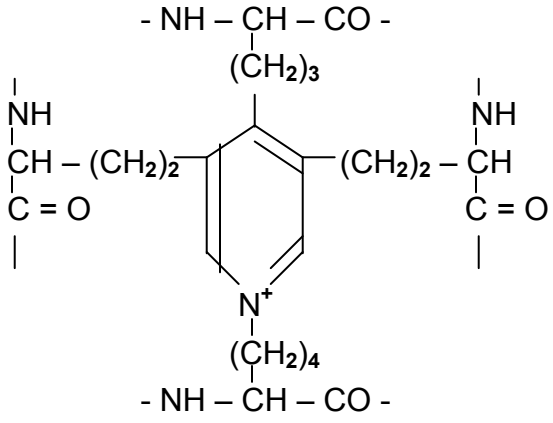
A collagen fibril consists of many triple helices (tropocollagen molecules) arranged head to tail in a staggered array. Covalent cross-links produce a fibril that does not stretch. **(PP 45)**



repeats every fourth row

d. Elastin

Elastin is also found in connective tissue and it will stretch. The basic unit is tropoelastin which is 800 amino acids, probably arranged in some kind of coil or helix. It is rich in Gly, Ala, and Lys, but low in Pro. Lysines form covalent cross-links where four lysines join to form desmosine, which stretches reversibly. **(PP 46-47)**



III. Tertiary Structure

Tertiary structure refers to the overall three-dimensional shape of a protein, also called conformation.

A. General Features

1. Tertiary structure is specific for each protein. All molecules of myoglobin will have the same tertiary structure, while all molecules of lysozyme will share a different tertiary structure. **(PP 48)**

2. Tertiary structure is determined by primary structure. Amino acid chains will fold into a shape that maximizes favorable interactions, and this will vary with amino acid sequence. Proteins with similar sequences have similar tertiary structures. **(PP 49-50)**

3. Tertiary structure is maintained mainly by non-covalent forces (H – bonds, salt bridges, hydrophobic interactions). Non-polar R groups will tend to be in the interior of the protein while polar and charged amino acids are found mainly on the exterior. Disulfide bonds also play a role.

4. Globular proteins have very compact structures. A protein with 600 amino acids will be 200 nm long if arranged in a β-sheet, 90 nm long if arranged in an α-helix, but is actually 13 nm long as a globular protein. The fraction of space occupied by atoms in a globular protein is about 0.75, the same as for solids. **(PP 51)**

5. Tertiary structures are not rigid. There is some flexibility and fluctuation in their structure. **(PP 52-53)**

B. Determination of Tertiary Structure

Normal chemical methods of analysis are not useful in determining higher levels of protein structure because the forces in 2°, 3°, and 4° structure are largely non-covalent. Thus they are easily disrupted and difficult to study.

The major method for studying higher levels of protein structure is X-ray diffraction (X-ray crystallography). Crystals of the protein are subjected to X-rays. **(PP 54)** Just as light waves diffract around an object in a microscope (and can produce an image), X-rays will diffract around the protein's atoms in a specific way if the atoms are arranged in a regular array (crystals). **(PP 55)** The diffraction pattern yields thousands of spots where X-rays have diffracted and positively reinforced. **(PP 56)** By measuring the position and intensity of the spots, it is then possible to use a complex mathematical calculation (a Fourier transform) to generate an image of the protein molecule. The analysis is difficult and the resolution of the image depends on the quality of the crystals and diffraction pattern. For thousands of proteins, the position of every atom is now known. For other proteins, only the outline of the shape is known. **(PP 57)**

NMR is also a technique which can be used to find the tertiary structure of small proteins. The interaction of atoms close to each other in the tertiary structure can be seen in a 2-dimensional NMR spectrum and used to calculate a structure. **(PP 58-60)**

C. Example of Tertiary Structure

Myoglobin was the first protein studied by X-ray crystallography. It is a small protein with 153 amino acids and a molecular weight of 16,700. It is found in muscle where it binds and transports O_2 for use when the muscle is working. It contains an iron-porphyrin group called heme which binds O_2. **(PP 61-62)**

X-ray analysis shows myoglobin is 78% α-helix in 8 segments ranging from 7 to 23 amino acids. Most of the hydrophobic R groups are in the interior. All but two of the polar R groups are on the surface. All peptide bonds are planar and trans. Prolines occur at bends. Other bends contain Ser, Thr, and Asn which tend to disrupt α-helices when close to each other. The heme group rests in a crevice. The iron binds to a His.

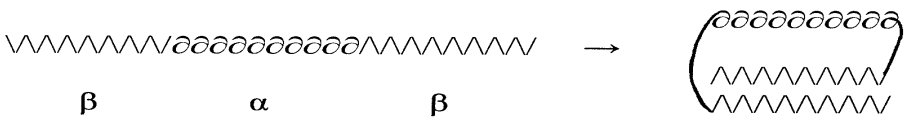
D. Common Tertiary Structures

Certain patterns in tertiary structure are seen in many different proteins. Since the proteins often have very different sequences and function, these patterns may have unusual stability and so recur. Such patterns include an even number of β strands arranged in a barrel shape. Another arrangement is four α-helices connected by peptide loops. Several other patterns also appear commonly. **(PP 63-64)**

E. Protein Folding

An interesting aspect of tertiary structure is how a protein finds the right tertiary structure. A protein is made and folds properly in about 5 seconds in a cell. It would take a protein 10^{50} years to find its proper structure by chance, trying out all possible 3^{o} structures. Thus protein folding is not random. **(PP 65)**

The principles of protein folding are not well-understood. It is thought that local secondary structures form first. Secondary structures then interact to form super-secondary structures, such as a βαβ loop where hydrophobic amino acids of each section would interact. **(PP 66-67)**



Supersecondary structures interact to form domains, which interact to produce the overall tertiary structure. Different proteins may have different mechanisms of folding. Some may fold in steps as described. Others may collapse into a folded state mediated by hydrophobic interactions. **(PP 68-69)**

Not all proteins fold spontaneously. Polypeptide chain binding proteins have been found that help some proteins fold properly by preventing non-specific aggregation and guiding the assembly of complex proteins. These helper proteins are called chaperones. **(PP 70)**

F. Denaturation

Denaturation refers to the loss of proper tertiary structure caused by breaking non-covalent bonds (but not covalent peptide bonds) in a protein. Proteins can be denatured by heat, pH changes, and certain chemicals, any of which will disrupt H-bonds, salt bridges, or hydrophobic interactions. To completely denature a protein, any disulfide bonds must also be broken. It is not necessary to disrupt all the non-covalent forces in order to disrupt conformation and destroy function. **(PP 71)**

Different proteins have different stabilities. Some are relatively difficult to denature and others are relatively simple. With some proteins, denaturation is irreversible and the protein is permanently damaged. With other proteins, the denaturation process can be reversed if the denaturing agent is removed. This reversal is called renaturation. The renatured protein is fully functional. Renaturation is consistent with amino acid sequence determining 3^{o} structure and with a precise pathway for protein folding. **(PP 72-73)**

IV. Quaternary Structure

When proteins contain multiple subunits (which may be identical or different), the 3-dimensional arrangement of the subunits is called quaternary structure.

A. General Features

1. Not all proteins have quaternary structure. A protein with a single amino acid chain cannot have quaternary structure.

2. Multimeric proteins vary from proteins with two identical subunits, to proteins with ten different subunits, to very large protein complexes with 102 subunits of three different types.

3. Non-covalent forces (H-bonds, salt bridges, hydrophobic interactions) are most important in maintaining 4^{o} structure. However, covalent disulfide bridges can also form between subunits.

4. Quaternary structure can be determined by X-ray crystallography.

B. Example of Quaternary Structure

Hemoglobin is a well-studied protein with multiple subunits. Hemoglobin (MW = 64,500) transports oxygen in blood. It has four subunits of two different types ($\alpha_2 \beta_2$). **(PP 74)** The α subunit has 141 amino acids while the β subunit has 146 amino acids. Each subunit has one heme group which can bind an O_2 molecule. **(PP 75)** The α and

β chains both contain considerable α-helix and have 1°,2°, and 3° structures similar to myoglobin. This is not surprising since all three protein chains have the same function and are probably related through evolution.

The hemoglobin molecule is roughly spherical with the four subunits in a tetrahedral shape. There are many contact points between the α and β chains, but few between the two α chains or the two β chains. Hydrophobic amino acids make most of the contact points along with a few salt bridges.

Hemoglobin has a property known as cooperative binding. The binding of one O_2 to one of the four subunits enhances the chances that O_2 will bind to the other three subunits by a factor of 500. This makes it an efficient oxygen carrier. This means that the binding of one O_2 is somehow transmitted to the other subunits. The structures of oxyhemoglobin (with O_2's) and deoxyhemoglobin (without O_2's) are slightly different. The tertiary structure of one subunit changes as the first O_2 binds **(PP 76)**, and this change is transmitted throughout the protein with the subunits undergoing small changes in their relative positions. **(PP 77-78)** As a result of the changes as the first O_2 binds, the other subunits find it progressively easier to bind O_2. Specifically, certain salt bridges must be broken as O_2 binds. As the first O_2 binds the most salt bridges must be broken, so binding the first O_2 is relatively hard (harder than for myoglobin). The remaining O_2's require that fewer salt bridges be broken, so their binding becomes easier and easier, producing positive cooperativity. (There are also examples of negative cooperativity.)

Myoglobin, being a single chain, does not display cooperativity. It binds oxygen tightly, which makes it well-designed to store oxygen, especially in muscle where the O_2 concentration is relatively low. Hemoglobin, in contrast, must pick up O_2 in the lungs (where the O_2 concentration is high) and release it in the peripheral tissues (where the O_2 concentration is low). **(PP 79)** The cooperative O_2 binding ensures that as hemoglobin leaves the lungs it will be fully and efficiently oxygenated. In the tissues with lower O_2 levels, hemoglobin will off-load a significant amount of its O_2. Thus, it too is well-designed. **(PP 80-81)**

Hemoglobin also demonstrates the importance of 1° structure. Sickle-cell anemia is caused by a single mutation in hemoglobin. Amino acid 6 of the β chain is changed from the normal Glu to Val. This creates a hydrophobic sticky spot on the outside of the hemoglobin molecule. **(PP 82)** Hemoglobin molecules can then polymerize into long chains. **(PP 83)** This distorts red blood cells from their normal disc shape into an elongated sickle shape. **(PP 84-85)** This distorted cell is more fragile and can break, causing anemia. Distorted cells can also clog capillaries, causing pain and tissue death. Not all amino acid changes in hemoglobin are so detrimental. Over 300 variant hemoglobins are known and most function with only minor problems, if any.

Chapter 6: Enzymes

I. Introduction

In order for a cell or organism to stay alive, thousands of chemical reactions must occur to produce energy, make biomolecules, etc. These reactions would take place very slowly unless catalyzed. Enzymes are biological catalysts.

A. Function - Enzymes speed up reactions 10^6 - 10^{16} times. **(PP 2)** They are true catalysts required in only small amounts and fully recoverable in their original form at the end of a reaction.

B. Structure - Most enzymes are globular proteins, although a few are nucleic acids. Enzymes that are proteins can be multimeric. The specific tertiary and/or quaternary structure of an enzyme is crucial in its ability to function.

C. Naming - Most enzymes end with -ase and the name indicates the function as well as the molecule it works on (substrate), such as DNA polymerase or glucose-6-phosphatase.

D. Prosthetic Groups and Cofactors - Many enzymes contain prosthetic groups needed for their function. Other enzymes do not have prosthetic groups, but do require a helper molecule in order to catalyze a reaction. This helper, known as a cofactor, associates with the enzyme during the reaction. Cofactors are often metal ions or organic molecules (coenzymes). A cofactor that remains permanently with the enzyme becomes a prosthetic group. **(PP 3)**

E. Specificity - Enzymes are very specific regarding their substrates, working on just one specific molecule (often one isomer) or on a group of closely related molecules. Most enzymes will also catalyze just one reaction, so a large number of different enzymes is needed. An enzyme will generally work on just one substrate molecule at a time, although it can convert 10^5 substrate molecules into product per second.

II. Factors in Enzyme Catalysis

In order for a chemical reaction to occur, the reacting molecules must be brought together in the proper orientation and with sufficient energy for the reaction to take place. Enzymes affect all these factors.

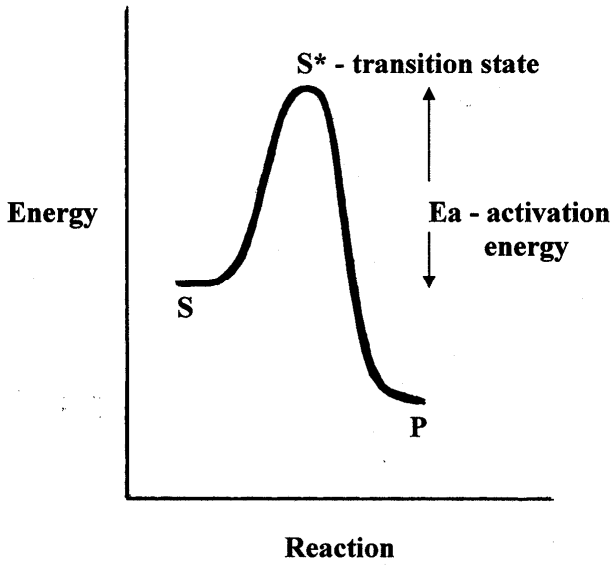
A. Orientation

During an enzyme-catalyzed reaction, enzyme and substrate bind together. This binding automatically orients the substrate in the proper way for the reaction to occur. Without an enzyme, proper orientation is a matter of chance during random collisions, and only a small percentage of molecules would collide with the proper orientation.

With an enzyme, all substrate molecules are properly oriented which increases the rate of the reaction. **(PP 4)**

B. Activation Energy

An energy diagram shows the energy involved in a chemical reaction.



Substrate and product have intrinsic energy levels. In order for S to become P, it must be able to climb the energy barrier called the activation energy. **(PP 5)**

This energy barrier exists because S and P both have stable electron configurations. To convert S into P, the stable structure of S must be disrupted so bonds can break, reform, etc., and the electrons rearrange to form P. At the top of the energy barrier, an extremely unstable and short-lived chemical species exists called the transition state.

The rate of a reaction is controlled by the activation energy. With a high E_a, very few S molecules in the population will have enough energy to react, so the reaction takes place slowly. **(PP 6)** The rate can be increased by increasing the energy of the S molecules (heating them up) but this is not practical in biological systems where many molecules are heat-sensitive.

Alternatively, the activation energy can be lowered so now many S molecules have the required energy. **(PP 7)** Enzymes participate in the reaction or change its mechanism so that the activation energy is lowered and the reaction rate increases. Enzymes bind to the substrate during the reaction, forming an enzyme-substrate complex, and this results in a lower E_a.

Enzymes do not alter the equilibrium of the reaction which is governed by the relative energy levels of S and P. (The energy level of P can be higher, lower, or the same as S.) By speeding up the reaction, enzymes let the reaction reach equilibrium more quickly. **(PP 8-9)**

III. Enzyme Reactions

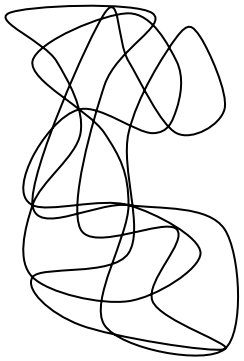
The simplest scheme for an enzyme-catalyzed reaction is given below.

$$E + S \leftrightarrow E S \rightarrow E + P$$

The substrate binds reversibly to form an enzyme-substrate complex. Formation of product (it can be reversible) is the rate-limiting step.

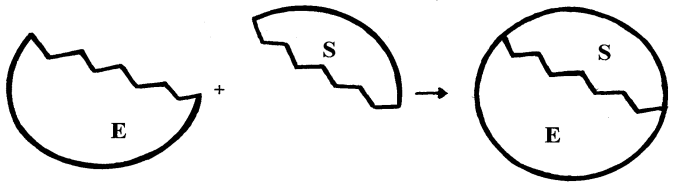
A. Enzyme-Substrate Binding

The substrate binds to a specific place on the enzyme called the active site. The active site is only a small portion of the enzyme (~5%). The active site has a specific 3-dimensional structure, usually a crevice on the enzyme surface. **(PP 10)** The amino acids forming and lining the crevice make up the active site.



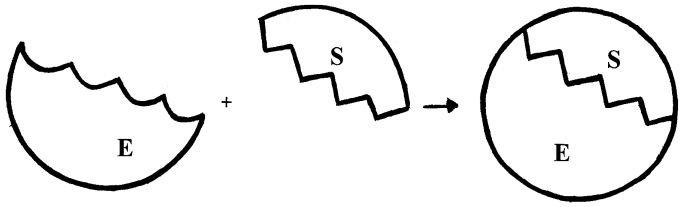
Binding between enzyme and substrate is non-covalent. The forces involved are hydrogen bonds, ionic attractions, hydrophobic interactions, and van der Waals forces.

The enzyme can pick out and bind its specific substrate from the thousands of other molecules present in the cell. This is because the substrate and active site have complementary sizes and shapes. Two models demonstrate this idea. The first is the lock and key model, where the active site is viewed as rigid and can bind only the complementary substrate, fitting like a lock and key. **(PP 11)**



enzyme substrate enzyme-substrate complex

While this describes some enzymes, the enzyme is usually not so rigid. Most enzymes are more flexible and are explained by the induced fit model. The enzyme's active site starts out largely complementary to the substrate, but changes slightly as the substrate binds to become even more complementary and produce a tight fit. **(PP 12)**



enzyme substrate enzyme-substrate complex

The enzyme is looking for several characteristics of the substrate. First, the substrate must have the right size. **(PP 13)** Too large a molecule will not be able to get into the active site and too small a molecule will not bind well. The substrate must fit snugly. Second, the substrate must have the right shape, or else the substrate will not bind well to the entire active site. Third, the substrate must have the right chemical groups in the right places. **(PP 14)** An amino acid in the active site may form an H-bond with a group on the substrate. A positively charged amino acid might interact with a negative charge on the substrate. A hydrophobic R group on an amino acid might interact with a hydrophobic group on the substrate. Unless a molecule has all the right features, it cannot bind to the active site, and this is how an enzyme is so specific as to its substrate. In some cases, an enzyme is looking primarily for a certain group, such as phosphate, and will remove phosphate from several different but related molecules. In other cases, the enzyme looks at the entire molecule and can pick out one isomer of one compound.

The importance of 3° structure is evident. If the active site does not have the proper 3-dimensional structure, it will not be able to bind the substrate or catalyze a reaction. **(PP 15-18)**

B. Catalysis

Exactly how enzymes bring about a reaction is not completely understood. The goal is to understand, step-by-step at the molecular level, exactly what events take place during the reaction. There are very few enzymes which are understood to this extent. What is known are some general features of enzyme catalysis.

1. Orientation - All enzymes properly orient the substrate as they bind it so that bonds to be broken are in close proximity to the amino acids doing the reaction. These amino acids are called the catalytic site (or active site).

2. Strain - Enzymes can slightly strain or distort the substrate molecule, making it easier to break certain bonds.

3. Transition State - An enzyme's active site is often more complementary to the transition state than to the original substrate. By stabilizing the transition state with more favorable binding, the enzyme facilitates the reaction. **(PP 19)**

4. Covalent Intermediates - Some enzymes form an unstable covalent intermediate with the substrate at some point during the reaction. This changes the reaction mechanism and lowers the activation energy. The enzyme is back in its original form by the end of the reaction. **(PP 20)**

5. Acid-Base Catalysis - The enzyme can act as an acid, donating protons, or as a base, accepting protons from the substrate. This transfer of protons can stabilize a charged intermediate to form a species that breaks down more readily into the products. **(PP 21)**

6. Metal Ion Catalysis - Metal ions (Fe, Cu, Zn, Mn) function in two main ways. They can participate in oxidation-reduction reactions, or they can be used to stabilize or shield charges that develop on the substrate during the reaction. **(PP 22)**

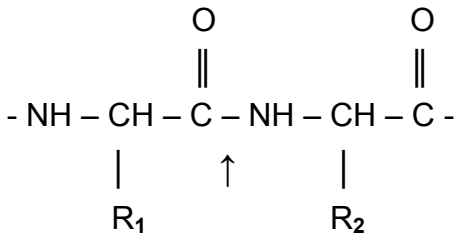
7. Electrostatic Catalysis - Any charged group in the enzyme can stabilize an opposite charge that develops on the substrate during the reaction.

Most enzymes use a combination of these factors to bring about a reaction. **(PP 23-26)**

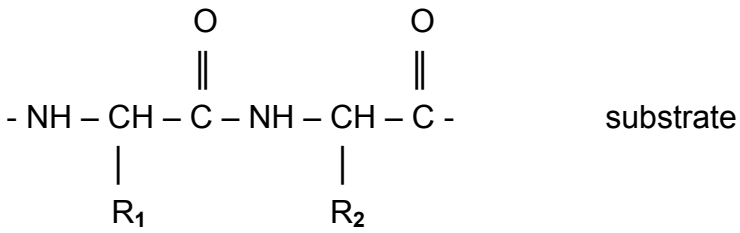
C. Chymotrypsin

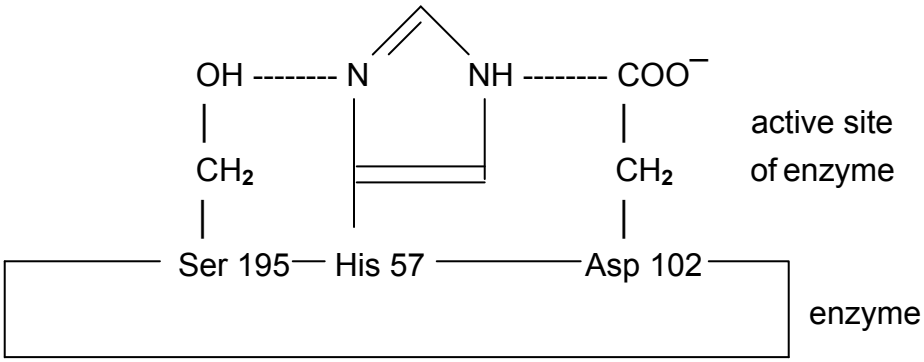
Chymotrypsin is a well-understood enzyme. It illustrates several of the general principles of enzymes. **(PP 27)**

Chymotrypsin is a proteolytic enzyme. It cleaves peptide bonds when R_1 is aromatic (Trp, Tyr, Phe). **(PP 28)**

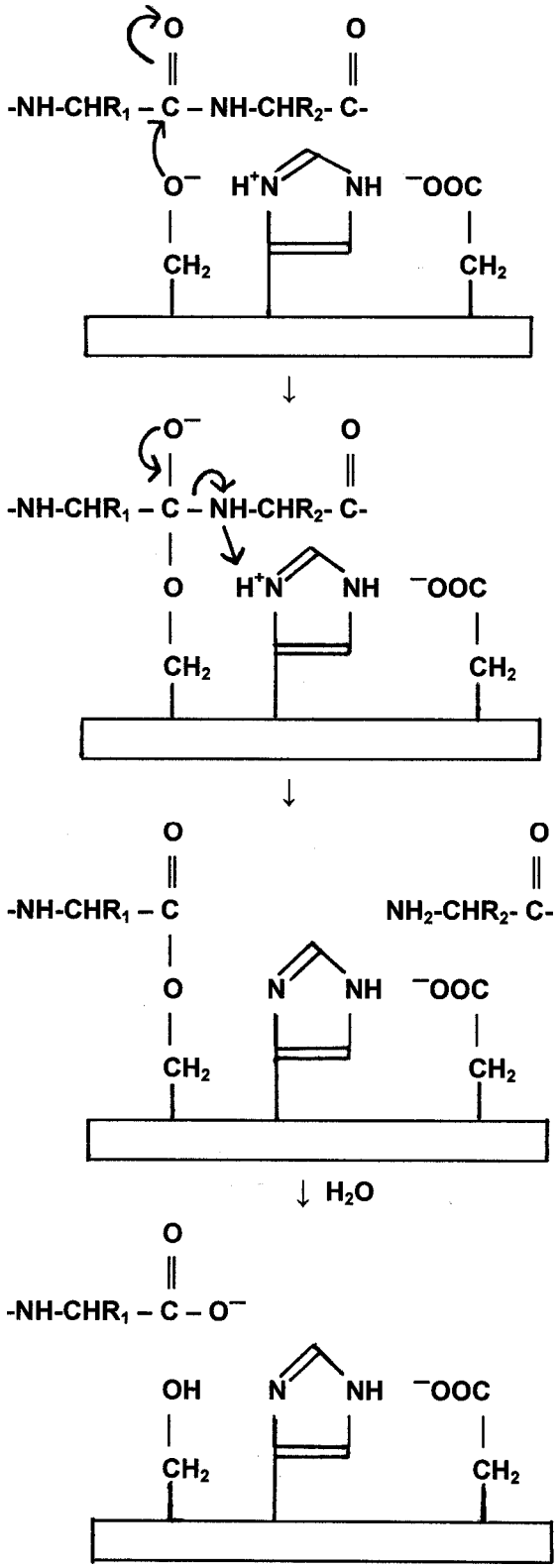


Chymotrypsin has a hydrophobic pocket which selects and binds the large hydrophobic R group, and positions the peptide bond to be broken. Three amino acids are important in the catalytic site, Ser 195, His 57, and Asp 102, which lie next to each other in the tertiary structure. The Ser is H-bonded to the His which is H-bonded to Asp. **(PP 29)**





The Ser oxygen attacks the electropositive carbon of the peptide bond. The Ser proton is picked up by the His, and its positive charge is stabilized by the negative charge of Asp 102. Attack of the Ser oxygen produces a tetrahedral intermediate which rearranges to break the peptide bond and form an unstable covalent intermediate between enzyme and substrate. **(PP 30-32)**


-NH-CHR1 – C – NH-CHR2- C-
O
O
O⁻
H⁺N
NH
⁻OOC
CH2
CH2
-NH-CHR1 – C – NH-CHR2- C-
O⁻
O
O
H⁺N
NH
⁻OOC
CH2
CH2
-NH-CHR1 – C
NH2-CHR2- C-
O
O
N
NH
⁻OOC
CH2
CH2
↓ H2O
-NH-CHR1 – C – O⁻
O
OH
N
NH
⁻OOC
CH2
CH2

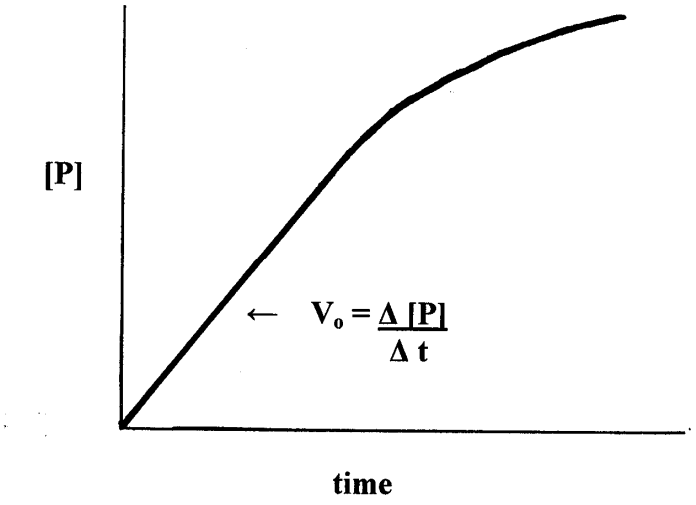
Water hydrolyzes the ester bond. **(PP 33-34)** The repulsion between the product's carboxyl group and Asp 102 causes the product to leave the enzyme. **(PP 35)**

Chymotrypsin uses acid/base catalysis and forms a covalent intermediate. Also, the active site is more complementary to the tetrahedral transition state (extra H-bonds form with the peptide bond oxygen) than to the original substrate.

IV. Enzyme Kinetics

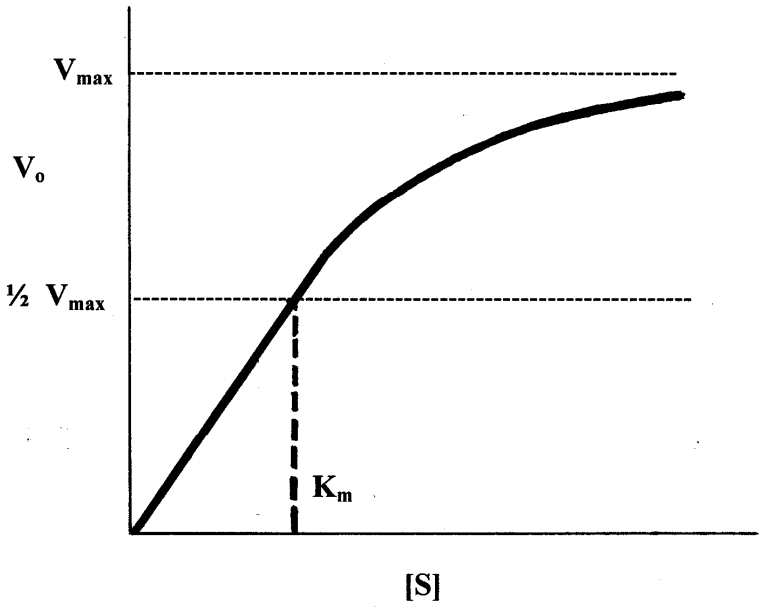
One way to explore an enzyme's mechanism is to study its action quantitatively, using enzyme kinetics. Enzyme kinetics studies the rate of the reaction in response to changing reaction conditions. One of the most fundamental conditions is the substrate concentration.

When studying the reaction S → P, either the appearance of P or the disappearance of S can be monitored, whichever is more convenient experimentally. If P can be measured by spectroscopy, then the concentration of P is monitored. The following results are seen over time.



Enzyme-catalyzed reactions typically slow down with time due to enzyme lability, inhibition by the product, or reversibility of the reaction. For this reason, V_o is always measured, which is the initial velocity at the beginning of the reaction. **(PP 36)**

When the substrate concentration is varied, a V_o value is measured for each value of [S]. When V_o is plotted against [S], a hyperbolic curve is produced. **(PP 37)**



At low substrate concentrations, initial velocity is directly proportional to [S]. At high substrate concentrations, initial velocity is independent of [S]. This is because, at high substrate concentrations, the enzyme is saturated with substrate and is already working as fast as it can.

The equation for a hyperbola is

$$y = \frac{ax}{b + x}$$

For this curve, the constant a = V_{max}, which is the maximum rate at high substrate concentration. The constant b = K_m, which is the substrate concentration needed to produce a rate of ½ V_{max}. So the equation which describes this curve and the variation of rate with substrate concentration becomes

$$V_o = \frac{V_{max}\,[S]}{K_m + [S]}$$

This is called the Michaelis-Menten equation. **(PP 38)** After this equation was found empirically, it was then derived by a theoretical consideration of how enzymes work.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} E + P$$

The theoretical treatment demonstrated that $V_{max} = k_2 e_o$ where k_2 is the rate constant for the rate limiting step and e_o is the enzyme concentration. In addition, it was found that $K_m = (k_{-1} + k_2) / k_1$ and can indicate the strength of the enzyme-substrate binding. **(PP 39)** A small K_m reflects a high affinity of the enzyme for its substrate, and means that an enzyme will be functioning at a significant level even when a small amount of substrate is present. An enzyme with a high K_m requires higher amounts of substrate in order to work well. **(PP 40)**

V_{max} and K_m are therefore important properties of an enzyme and are normally determined in a study of enzyme kinetics. However, these values are difficult to find precisely using a hyperbolic curve, so there are several algebraic rearrangements of the Michaelis-Menten equation that yield a straight line and are used for accurate determination of these kinetic values.

Once the basic kinetic properties of an enzyme are established, further kinetic studies can reveal important information about the enzyme. Substrate analogs can be used to see if the enzyme will react with them and to what extent. Such comparisons can help determine how the enzyme binds the substrate and reacts with it. The optimum pH for enzyme can be found, and measuring the effect of pH on enzyme activity can indicate what groups are important in the enzyme's active site. Using specific inhibitors of an enzyme can also help to define the crucial characteristics of the substrate and the mechanism of the enzyme. **(PP 41-42)**

V. Enzyme Regulation

If the cell is to perform efficiently, then its metabolism must be regulated. For instance, a cell may have metabolic pathways for synthesizing alanine and also for breaking down alanine for energy production. If the amino acid alanine is available in large amounts from the environment, then there is no need for the cell to use its resources making alanine. Instead, excess alanine can be broken down to produce energy. If external levels of alanine are low, then the cell needs to make alanine and

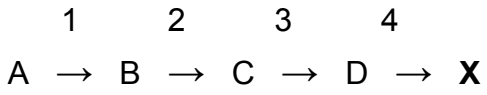
should not break it down. Thus the pathways for making alanine and for breaking down alanine should not function at the same time.

Regulation of reactions and metabolic pathways is achieved in many ways through regulating one or more enzymes in a metabolic pathway. Enzyme activity can be regulated in some straightforward ways, such as controlling the amount of substrate or cofactor that is available. Changes in pH or temperature can change enzyme activity, though these methods are rarely used inside cells. Another way of regulating reactions is to control the presence of enzymes. Each enzyme is coded for by a gene in the cell's DNA. Genes can be turned on and used to make the corresponding enzyme, or turned off so that the enzyme cannot be made. If the enzyme is not made, then the reaction cannot occur and the pathway using that enzyme will be shut down. This saves energy for the cell (making an enzyme consumes energy), but it is a relatively long-term method of control since it takes time for the cell to make the enzyme when it is needed.

Another method of regulating enzymes is to control the activity of existing enzyme molecules. Enzyme activity can be raised or lowered to quickly respond to changing conditions. This is a relatively short-term method of control, but it does require an investment in energy to make the enzyme molecules and have them present. It is not necessary to regulate every enzyme in a metabolic pathway. It is usually the first enzyme in a pathway which is regulated so that the entire pathway can be turned on and off as needed. There are two major methods for regulating enzyme activity.

A. Allostery

Allosteric enzymes have their activity changed through the non-covalent binding of a molecule known as an effector or modulator. In the following metabolic pathway which converts compound A to compound X using enzymes 1-4, enzyme 1 is often an allosteric enzyme inhibited by the end product X. **(PP 43)**



When levels of X are high and no more X is needed, the pathway will be shut down as X binds to enzyme 1. When levels of X are low, X will not be bound to enzyme 1, so the enzyme is active and X will be produced. The binding of the effector is non-covalent and reversible, so that when the effector leaves, the enzyme is restored to its original form and function.

Effectors can be either positive (which activate the enzyme) or negative (which inhibit the enzyme). Many allosteric enzymes have several effectors, some positive and some negative. A wide variety of different compounds can act as effectors, including substrates, products, cofactors, and metabolic intermediates. Each effector will bind to a specific site on the enzyme known as the regulatory or allosteric site. If there are multiple effectors, there can be multiple allosteric sites. Each site will be specific for its effector, similar to the way an active site is specific for its substrate.

Effectors work by changing the conformation of the enzyme. **(PP 44)** When the effector binds, it causes a change in the tertiary and/or quaternary structure of the enzyme which changes the active site. As a result, the enzyme works better (in response to a positive effector) or works less well or not at all (in response to a negative effector). **(PP 45)** Most allosteric enzymes are large and often have multiple subunits. Often one type of subunit is catalytic, binding the substrate and carrying out the reaction, while the other type of subunit is regulatory and binds the effectors.

How much enzyme is affected depend upon how much enzyme is bound to the effector, and that in turn depends upon the concentration of the effector. An equilibrium exists between the free enzyme, the effector, and the enzyme-effector complex.

$$\text{Enzyme} + \text{Effector} \leftrightarrow \text{Enzyme-Effector Complex}$$

The higher the concentration of the effector, the more is bound to the enzyme and the more the enzyme will be regulated. This allows the overall activity of the enzyme to be adjusted to whatever level is needed by the cell.

B. Covalent Modification

A second way of regulating enzymes is through the covalent joining of a regulating group, called a covalent modifier, to a specific site on the enzyme. This changes the enzyme conformation and so changes enzyme activity. For some enzymes, covalent modification makes them more active. For other enzymes, covalent modification makes them less active. The regulation is reversible since the covalent modifier can be removed. **(PP 46)**

The major differences of covalent modification compared to allostery include the following:

1. Binding of a covalent modifier is covalent while the binding of an effector is non-covalent.
2. A few covalent modifiers, mainly phosphate, can regulate many different enzymes while each allosteric enzyme has its own effector(s). **(PP 47)**

3. Because covalent bonds are being made and broken in covalent modification, chemical reactions are occurring which require catalysis by additional enzymes. Generally there is one extra enzyme to attach the covalent modifier to the regulated enzyme, and a second extra enzyme to remove the covalent modifier. Allostery requires nothing extra because the binding of the effector is non-covalent. The extra enzymes involved in covalent modification must themselves be regulated in some way (through allostery, covalent modification, hormones, etc.) The result is that the extent of covalent modification will depend upon the activity of the extra enzymes. An advantage of covalent modification is that many enzymes can be coordinately regulated by simultaneous covalent modification. An enzyme can be both allosteric and covalently modified.

C. Other mechanisms of enzyme regulation also exist, although they are generally less common.

1. Enzymes generally work better at some pH values than others, depending upon the pH of their normal environment. For instance, pepsin works well in the acidic pH of the stomach where it helps digest proteins. **(PP 48)**

2. Some enzymes are made as inactive precursors called zymogens. The zymogen must first be cleaved before the enzyme becomes active. This regulation is irreversible, and occurs with certain enzymes such as digestive enzymes. These enzymes should not be active until they reach the digestive tract or else they would damage tissues. **(PP 49)**

Chapter 7: Carbohydrates

I. Introduction

Carbohydrates are sugars and starches. They are the major source of energy in many organisms, serve to store energy, and are structural components in some organisms. Because of their wide distribution, carbohydrates are the most abundant type of biomolecule.

A. Formula and Structure

Most carbohydrates have the general formula $(CH_2O)n$. This suggested that they were hydrates of carbon, hence the name. However, they are not hydrates of carbon, nor do all carbohydrates conform to this formula. Carbohydrates are fundamentally polyhydroxy aldehydes and ketones, and some contain nitrogen, phosphorus, or sulfur.

B. Classes

There are three major classes of carbohydrates based on size. **(PP 2)**

1. Monosaccharides are the simplest sugars, containing 3-7 carbons and one aldehyde or ketone group.
2. Oligosaccharides consist of ~2-10 monosaccharide units joined together. Most are disaccharides. Some are joined to lipids, or to proteins as prosthetic groups.
3. Polysaccharides are very large, with hundreds or thousands of monosaccharide units joined together.

II. Monosaccharides

A. General Properties **(PP 3)**

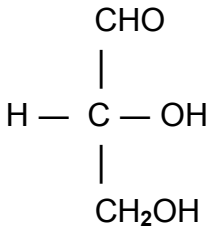
1. They are white solids, water-soluble (polar), and often have a sweet taste.
2. The formula is $(CH_2O)n$ where n = 3-7. The most common number of carbons is 5 or 6.
3. The carbon skeleton is unbranched, connected by single bonds. One carbon contains a carbonyl oxygen (aldehyde or ketone). All other carbons contain a hydroxyl group.
4. Monosaccharides are named with the suffix - ose. Every monosaccharide can be classified as an aldose or ketose, depending upon the functional group. **(PP 4-5)** They can also be classified according to the number of carbon atoms: 3 carbons = a triose; 4 C = a tetrose; 5 C = a pentose; 6 C = a hexose; 7 C = a heptose. The two classifications are often combined, such as aldopentose or ketohexose. **(PP 6-7)**

B. Structure and Stereochemistry

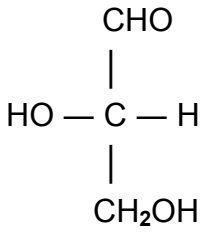
The most important classification is aldose vs. ketose.

1. Aldoses

The simplest aldose contains three carbons with one aldehyde and two hydroxyl groups. The name for this compound is glyceraldehyde.



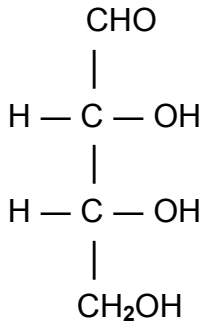
D-glyceraldehyde



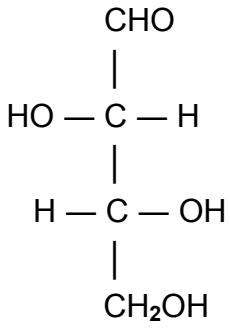
L-glyceraldehyde

Glyceraldehyde contains a chiral or asymmetric carbon, also referred to as a stereogenic or chirality center. The central carbon is attached to four different substituents, giving two forms that are mirror images of each other but are not superimposable. They are designated D and L. **(PP 8-9)** All other monosaccharides are classified as D or L, depending on the configuration of the asymmetric carbon farthest from the aldehyde group.

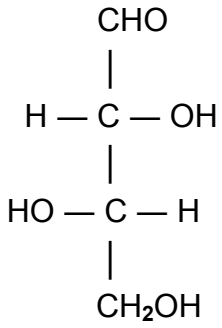
A molecule with n chiral centers will have 2^n stereoisomers. Thus 4-carbon aldoses have two stereogenic centers and four stereoisomers. **(PP 10-11)**



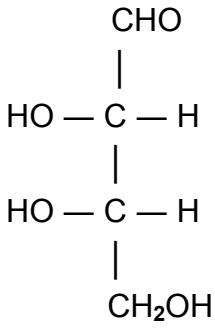
D-erythrose



D-threose



L-threose

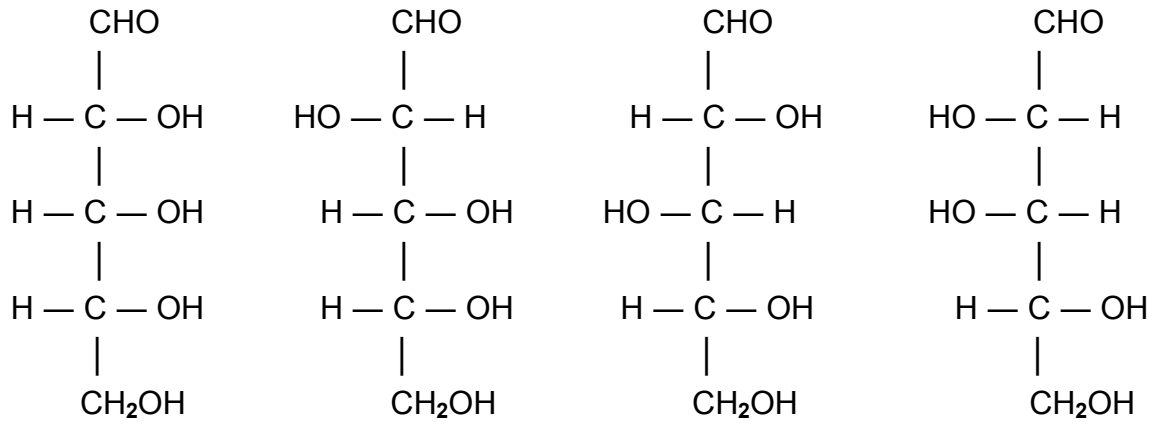


L-erythrose

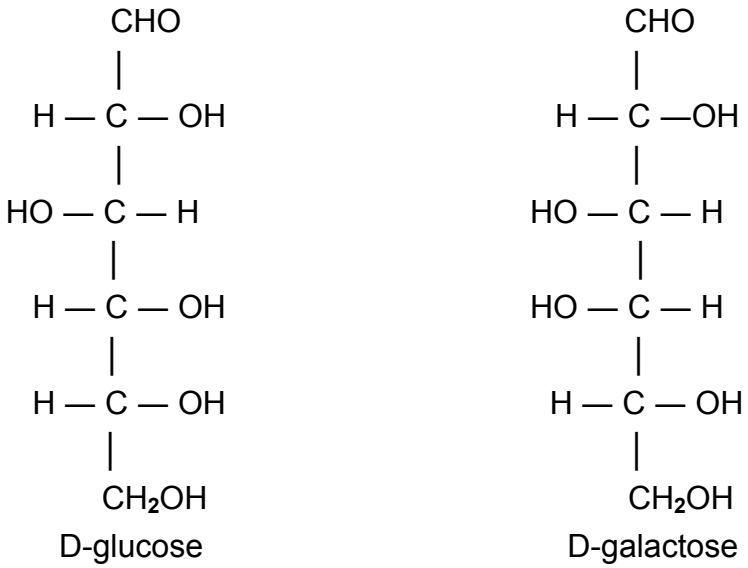
The first two structures are D-sugars because the configuration around the third carbon is like that of D-glyceraldehyde. The last two structures are L-sugars because they resemble L-glyceraldehyde. D-threose and L-threose are mirror images

(enantiomers) and will have all the same properties except their behavior with plane-polarized light and their interaction with other chiral compounds. D-erythrose and D-threose are not mirror images (are diastereomers) and are given different names because they are different compounds with different properties.

The D-sugars are the biologically important ones. There are four D-aldopentoses and eight D-aldohexoses (and the same number of L-isomers). **(PP 12-14)**



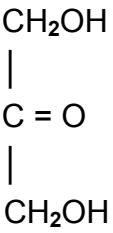
The most important aldoses and the most common in biological systems are D-ribose (an aldopentose) and two aldohexoses, D-glucose and D-galactose.



Sugars that differ in configuration around one carbon are called epimers. D-glucose and D-galactose are C-4 epimers. **(PP 15-19)**

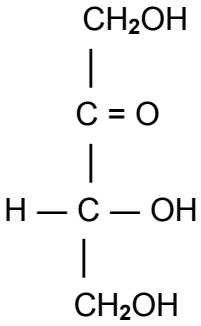
2. Ketoses

The simplest ketose will have three carbons, one ketone group, and two hydroxyl groups. **(PP 20)**

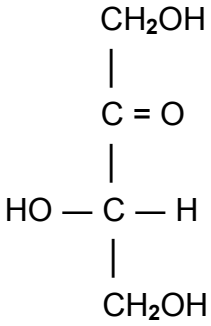


dihydroxyacetone

This compound does not have a chiral carbon and so does not have D and L forms. When a fourth carbon is added, a stereogenic center is created. **(PP 21)**



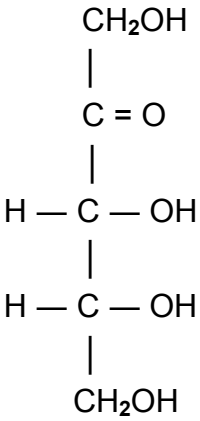
D-erythrulose



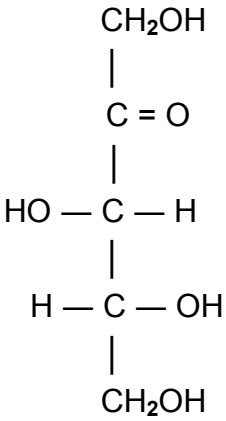
L-erythrulose

D and L designations are based on glyceraldehyde. If the OH on the last chiral carbon points right, it is a D isomer. If the OH points left it is an L-isomer.

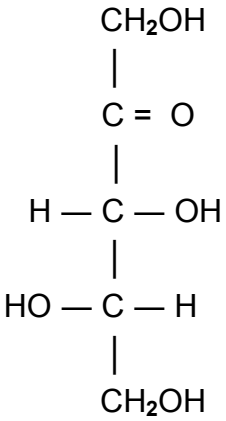
For ketopentoses, there are two chiral carbons and four stereoisomers. **(PP 22-23)**



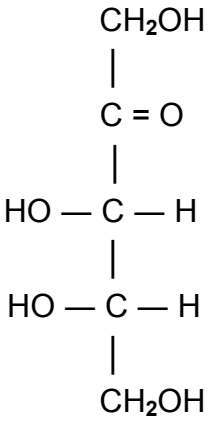
D-ribulose



D-xylulose

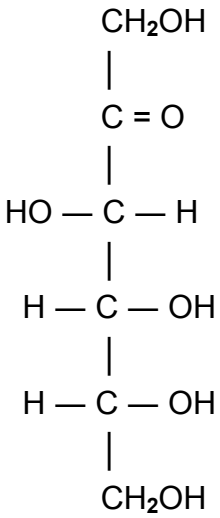


L-xylulose



L-ribulose

The various stereoisomers again form pairs of enantiomers or diastereomers. The D-isomers are the biologically important ones. Ketoses are generally named by inserting ‘ul’ into the name of the corresponding aldose. (ribose-ribulose). The most important ketose is D-fructose, which is a ketohexose.



When there are five or six carbons present, other ketoses are possible where the ketone group is on carbon 3. However, all the biologically important ketoses have the ketone group on carbon 2. **(PP 24-25)**

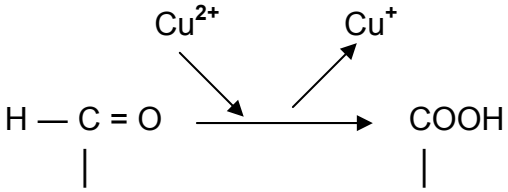
C. Reactions of Monosaccharides

Monosaccharides will undergo reactions characteristic of the functional groups. Some are useful in detecting and identifying monosaccharides.

1. Oxidation

a. Fehling’s or Benedict’s reaction

Fehling’s solution is alkaline cupric ion complexed with tartrate ion, while Benedict’s solution is complexed with citrate ion. When reacted with a sugar, the sugar is oxidized and the copper is reduced. **(PP 26)**



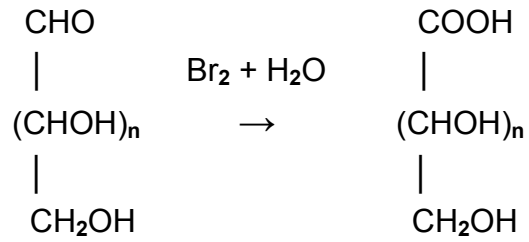
If the sugar is an aldose, the aldehyde is directly oxidized to the acid. If the sugar is a ketose, an acid is formed as the carbon chain breaks. In both cases, the blue Cu^{2+} is

reduced to Cu^+, which precipitates out as rust-colored Cu_2O and indicates a positive test. Any carbohydrate which reacts in this test is called a reducing sugar. All monosaccharides are reducing sugars.

A similar test is Tollen's test which uses alkaline Ag^+. The sugar is oxidized while Ag^+ is reduced to Ag, which precipitates out as a silver film. **(PP 27)**

b. Bromine water

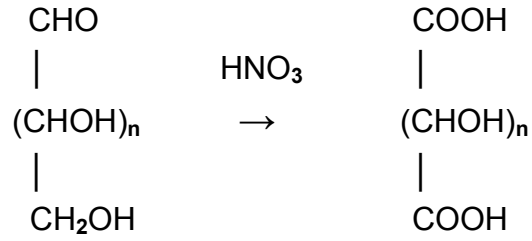
$Br_2 + H_2O$ is a weaker oxidizing agent than Benedict's reagent and so will oxidize aldoses, but not ketoses. **(PP 28)** Thus, it can distinguish the two types of sugars.



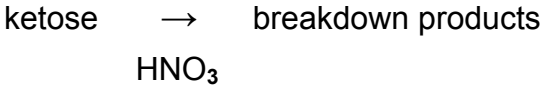
The acid formed is given the general name of an aldonic acid or glyconic acid with specific names like ribonic acid or gluconic acid.

c. HNO_3

Nitric acid is a strong oxidizing agent that reacts with both aldoses and ketoses. For aldoses, both the aldehyde group and primary alcohol group are oxidized. **(PP 29)**

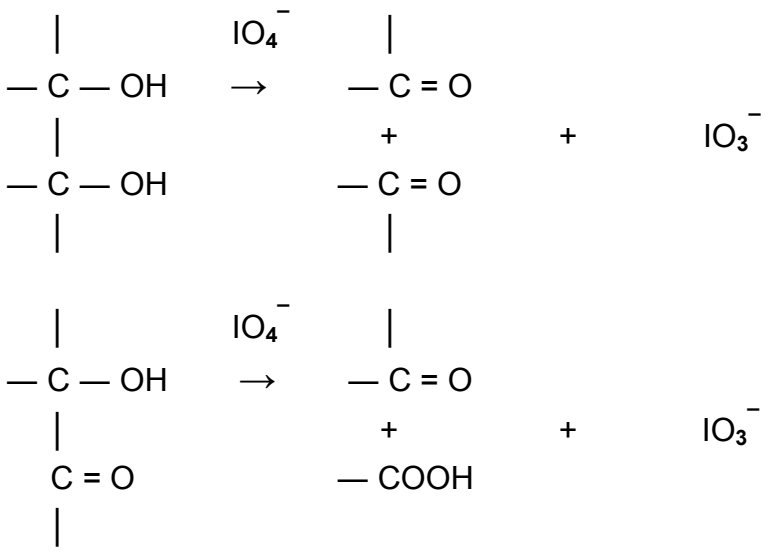


The general name for the product is an aldaric acid or a glycaric acid with specific names like ribaric acid or glucaric acid. For ketoses, oxidation occurs with chain breakage.



d. HIO_4

Periodic acid oxidizes both aldoses and ketoses, but it does so in such a specific way that it is used for structural determination and identification. IO_4^- cleaves a C-C bond when both carbons carry oxidizable groups such as hydroxyls or carbonyls. **(PP 30)** One molecule of HIO_4 is reduced to iodate, IO_3^-, and both carbons are oxidized to the next highest oxidation state.



The reaction of groups is

— C — OH → — C = O (carbonyl)

— C = O → — COOH (acid)

— COOH → CO_2

When dealing with monosaccharides, every carbon has an oxidizable group and so all C-C bonds will be broken. The result is a mixture of one-carbon compounds, and the following rules apply. **(PP 31)**

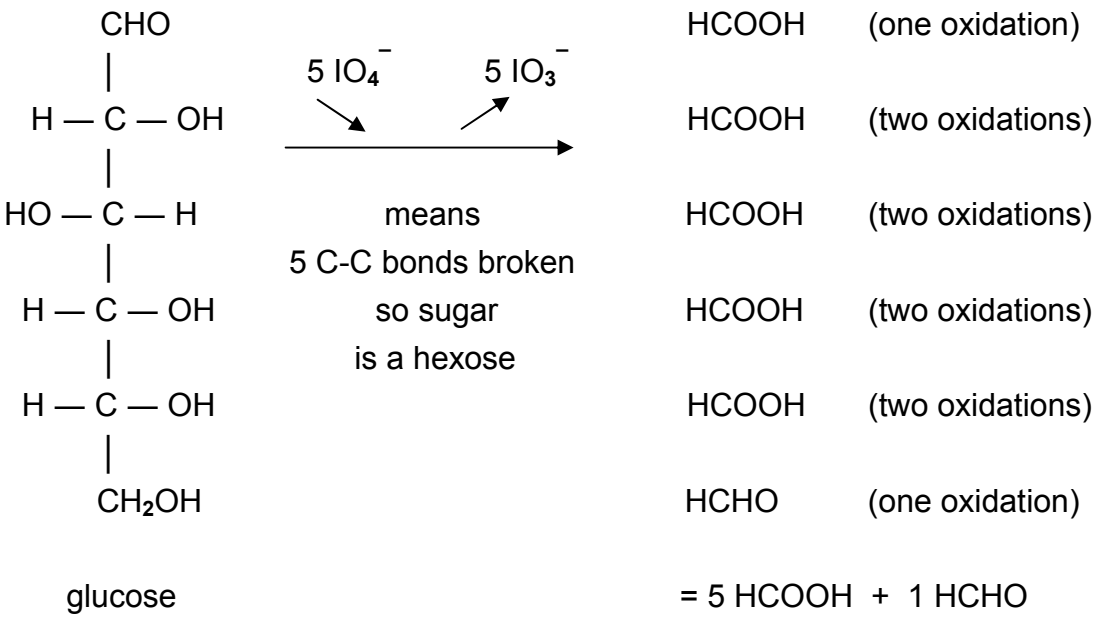
one mole of HIO_4 is used per mole of C – C bonds broken

one mole of HCHO is formed per mole of 1° alcohol oxidized

one mole of HCOOH is formed per mole of 2° alcohol oxidized

one mole of HCOOH is formed per mole of aldehyde oxidized

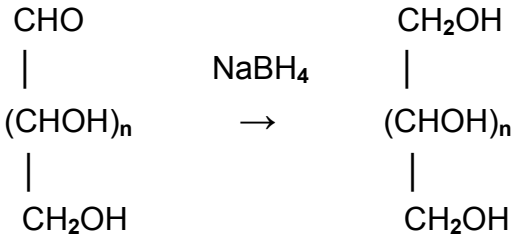
one mole CO_2 is formed per mole of ketone oxidized



From the products and the amount of IO_4^- used, much about the sugar structure can be deduced. **(PP 32)**

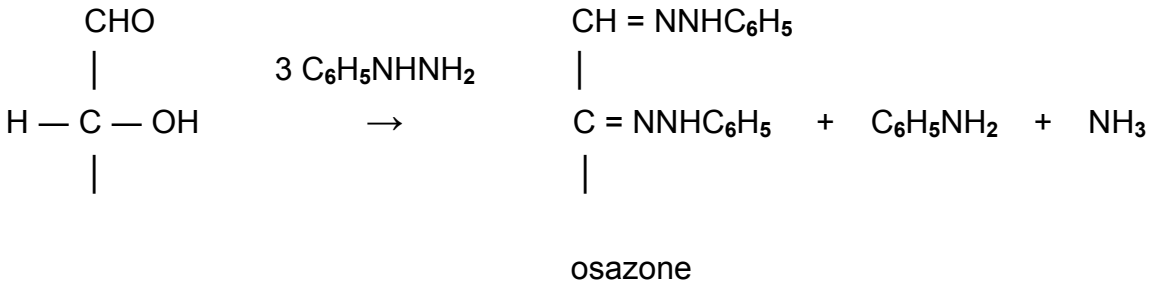
2. Reduction

Aldoses can be reduced to alditols (aldehyde → alcohol) using reducing agents like borohydride. **(PP 33-35)**



3. Osazones

Aldoses react with phenylhydrazine to form phenylhydrazones. If taken to completion, the product is an osazone. **(PP 36)**



Different carbohydrates form osazones with different melting points and crystal structures allowing for identification of the initial aldose. C-2 epimers give the same osazone. Ketoses usually react more slowly, if at all.

4. Other Reactions

Monosaccharides are stable in dilute acids. Strong acids can dehydrate or break chains. Dilute bases can cause rearrangements around certain carbons. Strong bases can fragment the chain.

Monosaccharides can be interconverted using several different reactions. Epimerization around carbon-2 of an aldose can take place using pyridine. An aldose can be shortened by one carbon using a series of reactions called the Ruff degradation, or extended by one carbon using the Kiliani-Fischer synthesis. However, these methods involve several steps and can create a mixture of products, so their usefulness is limited.

D. Haworth Structures

There is a further complication in the structure of monosaccharides.

1. Unusual Properties

Certain properties of monosaccharides are inconsistent with the presence of a normal carbonyl group.

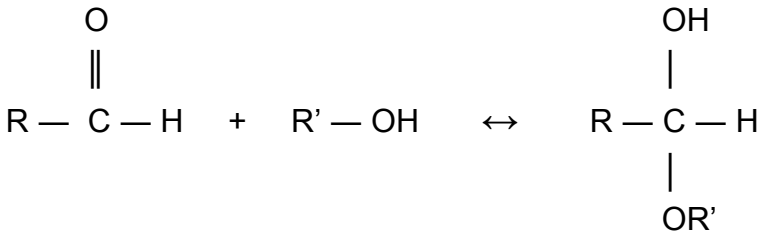
a. The addition of HCN during the Kiliani-Fischer synthesis occurs slowly with monosaccharides but rapidly with normal aldehydes.

b. Many aldoses fail to give a positive Schiff test, which normally identifies an aldehyde. (A reaction occurs with sulfur dioxide and fuchsin to give a red color.)

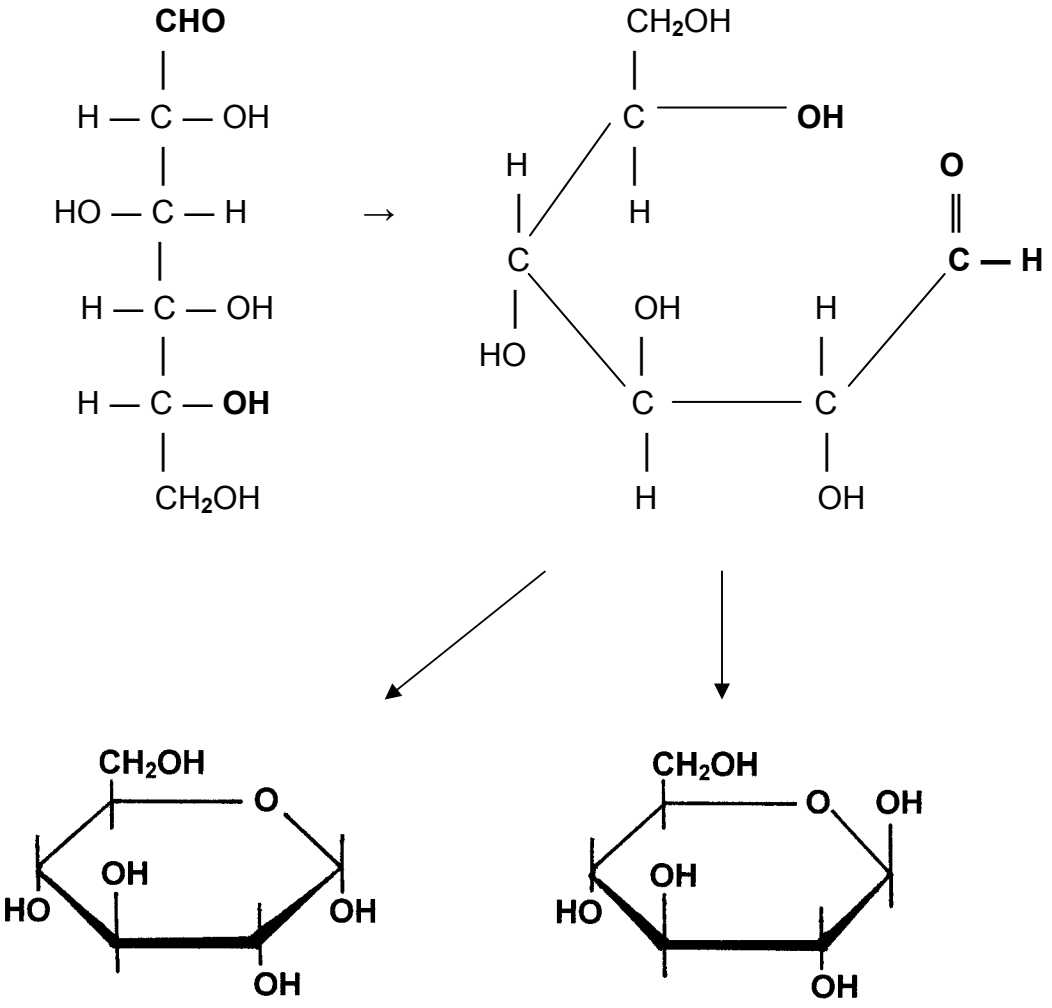
c. D-glucose can exist in two forms with different specific rotations. If D-glucose is recrystallized from water, the sugar has a specific rotation [α] = + 112° (called the α-form). If D-glucose is recrystallized from pyridine, the sugar has a specific rotation [α] = + 19° (called the β-form). If either form is dissolved in water and allowed to stand, the specific rotation changes until [α] = + 52.7°. This is called the mutarotation of glucose.

2. Hemiacetals

These properties are explained by a common reaction in organic chemistry, the formation of a hemiacetal, which occurs between an aldehyde and an alcohol. **(PP 37)**



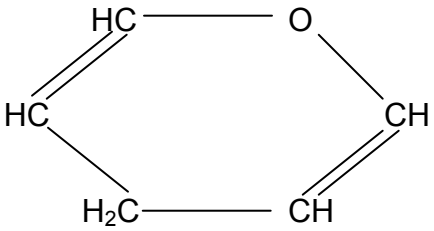
In an aldose, the hemiacetal formation is internal since both functional groups are in the same molecule. For an aldohexose, the reaction is between the C-1-aldehyde and the C-5 alcohol. **(PP 38)**

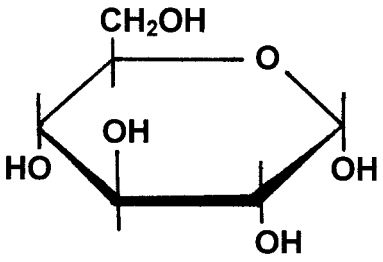


A new asymmetric carbon is created (carbon 1), with two possible configurations. The new OH group can point up or down. When the OH points down, this is α-D-glucose. When the OH points up, this is β-D-glucose. This explains the two forms of D-glucose. Since the hemiacetal is not stable and will undergo the reverse reaction readily in solution, both the α and β forms can revert to the straight-chain form (open-chain, linear form). Thus the α and β forms are interconvertible. If a solution of pure α (or pure β) is allowed to stand, some of it will convert to the straight-chain form and then some to the other cyclic form. Eventually, an equilibrium is established that is 63% β-D-glucose, 37% α-D-glucose, and a very small amount of the linear form. This explains the

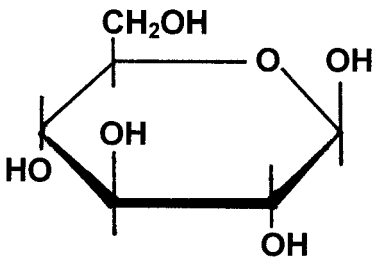
mutarotation of glucose. Also, since only a small amount of glucose has a free aldehyde group, this is why monosaccharides react slowly in reactions requiring an aldehyde group.

The two forms (α and β) are called anomers and C-1 is called the anomeric carbon. They appear to be diastereomers (non-mirror images) but are the same compound since they interconvert. Thus they are special forms of isomers, not enantiomers but not really diastereomers either. Hence there is the new designation of anomers. The ring forms are named pyranoses since they resemble the 6-membered ring compound, pyran. **(PP 39)**



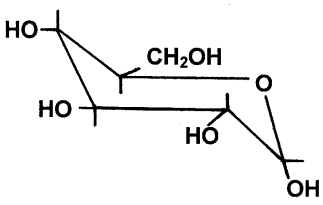


α-D-glucopyranose

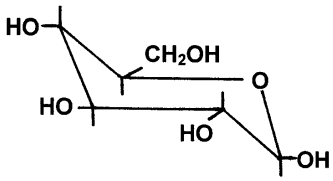


β-D-glucopyranose

The equilibrium concentrations of α and β glucose are explained by remembering that these molecules exist primarily in a chair conformation. **(PP 40)**



α

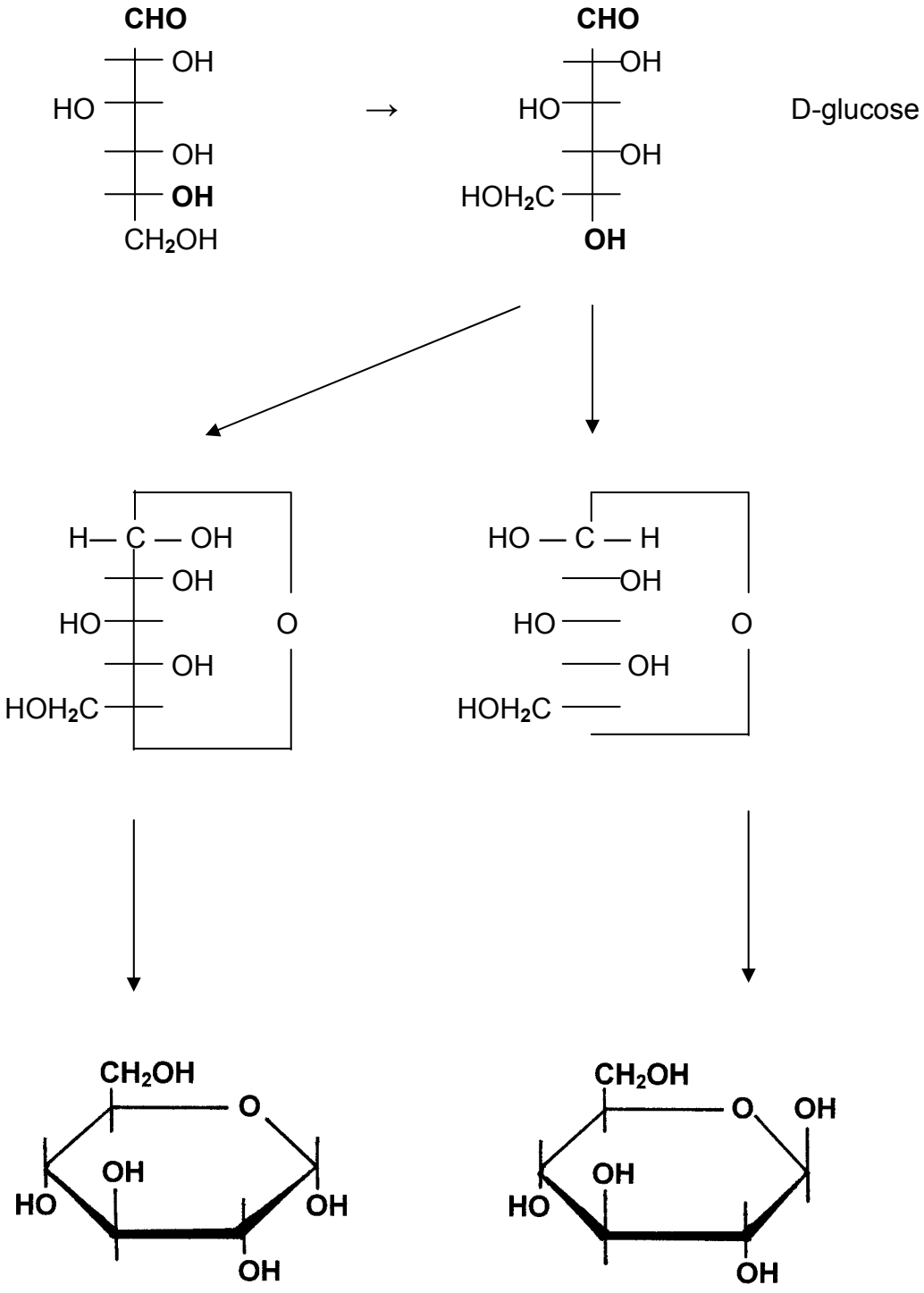


β

The α-form has four equatorial substituents and one axial substituent. The β-form has five equatorial substituents. Since equatorial substituents are more stable than axial ones, the β-form is favored.

3. Converting to Haworth structures

Any D-sugar can be transformed into a Haworth structure, starting with the straight-chain form. First, the C-C bond involving the last chiral carbon is rotated so the hydroxyl group points down. Next, the hemiacetal bond is formed, creating two pseudocyclic forms. These forms are then re-written as Haworth structures. **(PP 41)**



α β

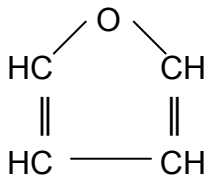
Any group on the right in the pseudocyclic form points down in the Haworth structure. Any group on the left in the pseudocyclic form points up in the Haworth structure. In the α-D pseudocyclic form, the new anomeric OH appears on the right. In the β-D form, the new OH appears on the left. **(PP 42-43)**

Other aldohexoses will behave like glucose. The rules are different for L-sugars.

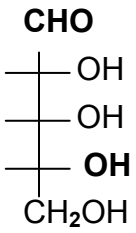
4. Other Monosaccharides

a. Aldopentoses

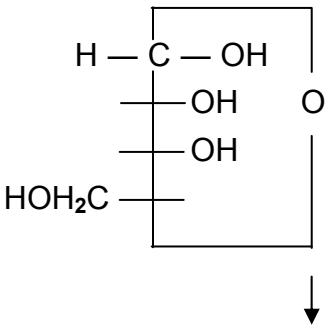
Aldopentoses undergo a similar reaction to form a 5-membered ring called a furanose, based on the compound furan. **(PP 44-45)**

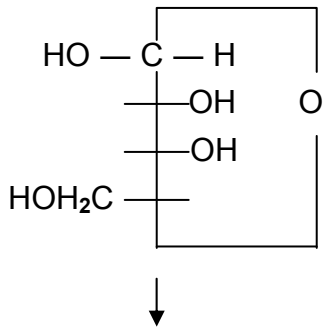


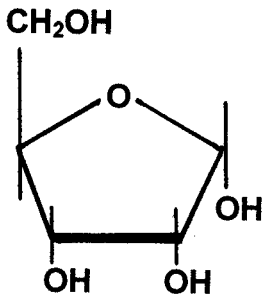
furan

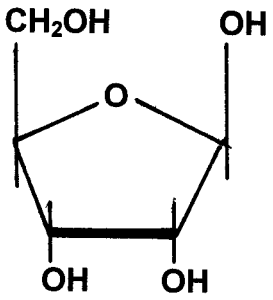


D-ribose







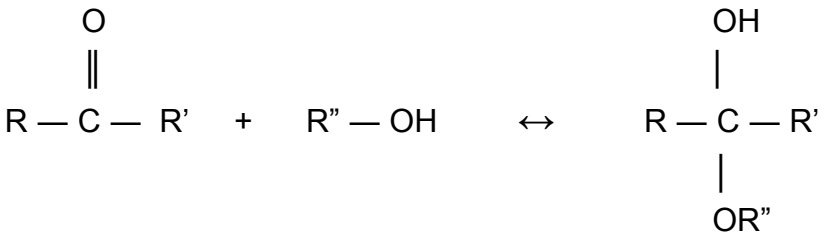


α-D-ribofuranose

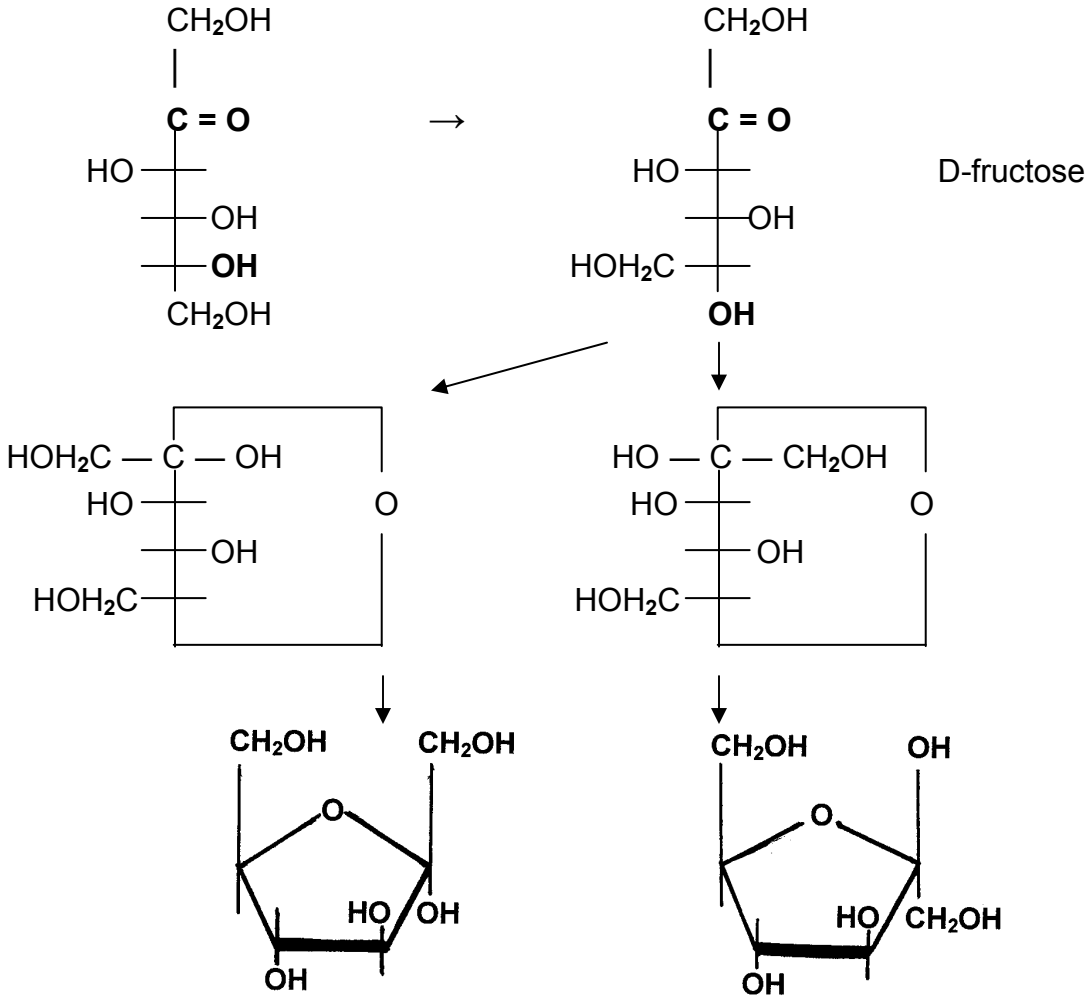
β-D-ribofuranose

b. Ketohexoses

Ketoses undergo a similar reaction to form hemiketals. **(PP 46)**



Like hemiacetals, hemiketals are unstable and this reaction is freely reversible in solution. Ketohexoses will undergo this reaction between the C-2 ketone and the C-5 hydroxyl to form furanoses. β and α forms will follow the same conventions and will mutarotate. **(PP 47)**



D-fructose

α-D-fructofuranose (OH down) β-D-fructofuranose (OH up)

The most stable ring forms are those with five or six atoms. Smaller monosaccharides will not form the cyclic structures. **(PP 48-49)**

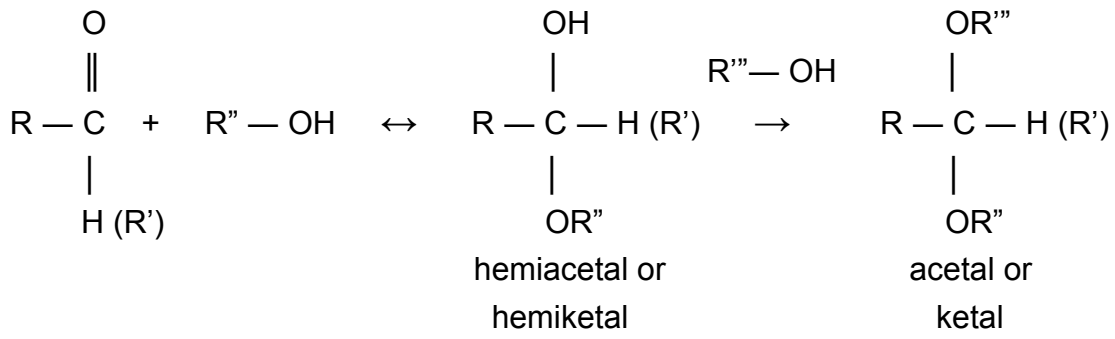
Some monosaccharides can form two types of ring structures. Ketohexoses and aldopentoses can also form pyranoses when the reaction occurs with the last OH, but the furanoses are the predominant form in biomolecules. **(PP 50)** Likewise aldohexoses can form furanoses, but the pyranoses are much more common.

E. Derivatives of Monosaccharides

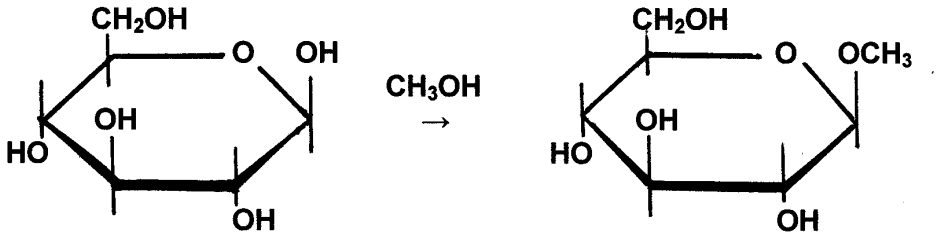
Several compounds that are related to monosaccharides have biological importance.

1. Glycosides

A hemiacetal or hemiketal can be converted to an acetal or a ketal by adding another alcohol. The acetal or ketal is stable in neutral solution, unlike the hemiacetal or hemiketal. The acetal or ketal will not be stable in acid and then the reaction can be reversed. **(PP 51)**



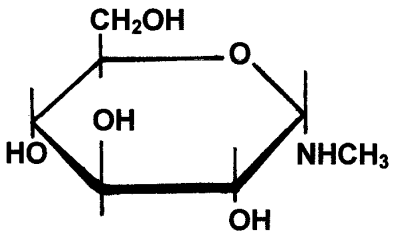
This type of reaction between a monosaccharide and any alcohol produces a glycoside.



The ring of the glycoside can no longer open. **(PP 52)** The sugar is now non-reducing, and it will not oxidize, form osazones, or mutarotate. The new bond is called a glycosidic bond and is designated α (down) or β (up). **(PP 53)** The compound is called a pyranoside if the ring is 6-membered and a furanoside if the ring is 5-membered. Glycosides are related to oligo- and polysaccharides. **(PP 54-55)**

2. N-glycosides

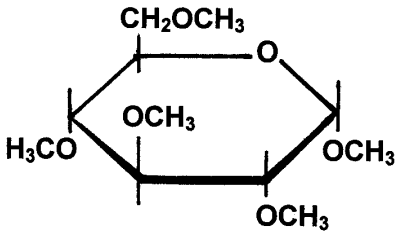
If the reaction is with an amine instead of an alcohol, a similar compound is formed called an N-glycoside. **(PP 56)**



Such compounds are related to nucleotides.

3. O-acyl derivatives

The free OH groups can be modified with reagents such as acetic anhydride or dimethyl sulfate. **(PP 57)**



These reactions are useful in structure determination.

4. Sugar alcohols and acids

The carbonyl group of a monosaccharide can be reduced to an alcohol, creating a sugar alcohol or alditol. Such compounds are found in some lipids. **(PP 58)**

CH_2OH
|
CHOH — glycerol
|
CH_2OH

Oxidation of monosaccharides will create sugar acids. There are three types.

only aldehyde is oxidized

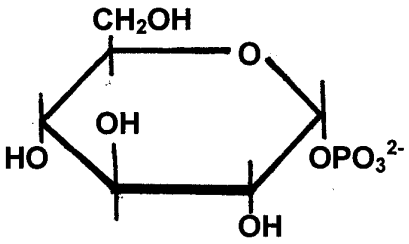
$COOH$
|
$(CHOH)_n$
|
CH_2OH

an aldonic acid - These are intermediates in metabolic pathways. **(PP 59)**

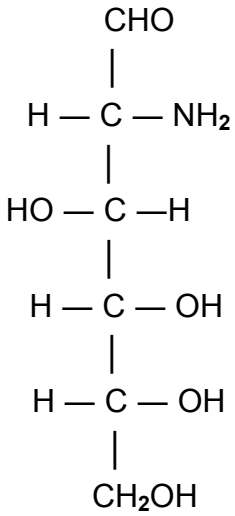
both aldehyde and 1° alcohol are oxidized	COOH | $(CHOH)_n$ | COOH	an aldaric acid - These are useful in structure determination. **(PP 60)**
only 1° alcohol is oxidized	CHO | $(CHOH)_n$ | COOH	a uronic acid - These are found in polysaccharides. **(PP 61)**

5. Sugar phosphates and amino sugars

Some monosaccharide derivatives contain other chemical groups. **(PP 62-63)**



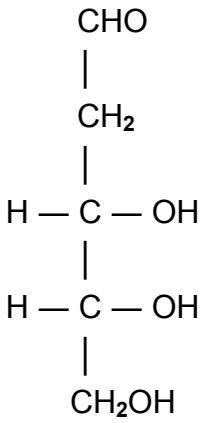
glucose-1-phosphate found in metabolic pathways



glucosamine found in polysaccharides

6. Deoxysugars

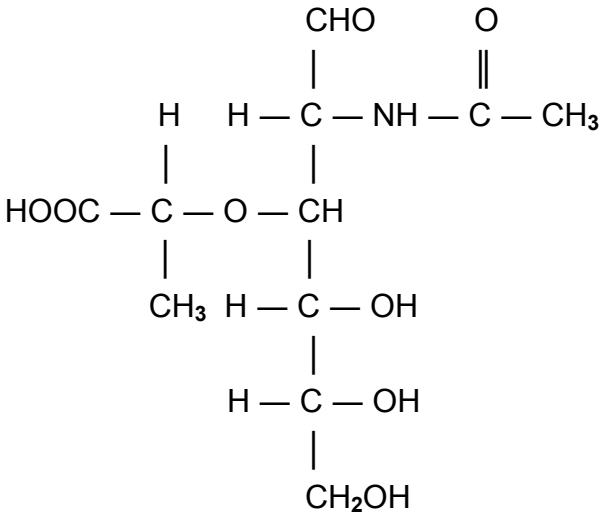
Some sugars lack an OH group. **(PP 64)**



2-deoxy-D-ribose found in nucleotides

7. Complicated derivatives

Some sugars have several extra groups. **(PP 65-67)**



N-acetylmuramic acid found in bacterial cell walls

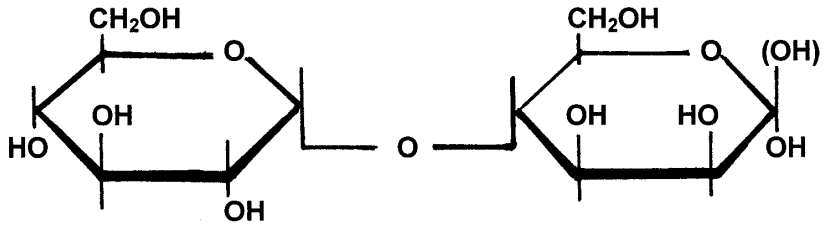
III. Oligosaccharides

A. Characteristics

Oligosaccharides contain 2-10 monosaccharides joined together. The most important are the disaccharides (two monosaccharides joined together).

The bond that holds the monosaccharides together is called a glycosidic bond. The bond (acetal or ketal) is stable at neutral and basic pH but is broken by acid hydrolysis. **(PP 68)**

There are two basic characteristics to any disaccharide. The first is the identity of the monosaccharide components. These may be the same or different. The second is the nature of the glycosidic bond. The bond is characterized by which carbons of the monosaccharides are linked and by the orientation of the bond (α or β).



In the above example, glucose is the first sugar, mannose is the second. The bond links carbon 1 of the glucose to carbon 4 of the mannose. Since the bond points down off the glucose, it is an α bond (up would be β). There is no choice about the direction of the bond from the mannose. It must be down or the sugar is not mannose. Thus this disaccharide has an α 1,4 bond.

Since a glycosidic bond is stable, there is no interconversion between α and β. These two monosaccharides joined by a β bond would make a different disaccharide. Since the aldehyde group of glucose is permanently in the acetal linkage, there can be no mutarotation or typical aldehyde reactions of the glucose unit. However, the mannose does have a hemiacetal (potential aldehyde) so it can mutarotate, and there will be two forms of this disaccharide (anomeric OH up or down) which will interconvert in solution. The disaccharide will be a reducing sugar since the mannose aldehyde group exists.

B. Examples and Analysis

Disaccharide structure can be illustrated and analyzed by looking at some specific examples.

1. Maltose

Maltose is a common disaccharide formed from the breakdown of plant and animal polysaccharides. Its formula is $C_{12}H_{22}O_{11}$, which is equivalent to $2C_6H_{12}O_6 - H_2O$, showing water is lost when a glycosidic bond is formed. Maltose has the following characteristics. **(PP 69)**

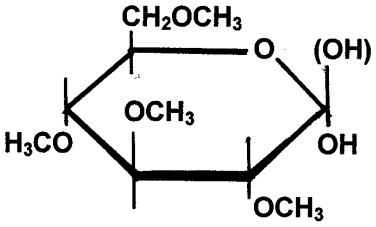
a. Acid hydrolysis produces only D-glucose, so both monosaccharides are glucose.

b. Maltose is a reducing sugar, containing a free aldehyde group and reacting with Benedict's reagent. It also exists in two forms, α-maltose with a specific rotation of +168° and β-maltose with a specific rotation of +112°, again indicating an aldehyde group which allows mutarotation.

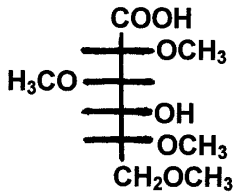
c. Maltose forms an osazone by adding two molecules of phenylhydrazine, indicating the presence of only one aldehyde group. (Two free aldehydes would add four molecules of phenylhydrazine.) This confirms that the C-1 of one of the glucoses must be involved in the glycosidic bond (which is expected since glycosidic bonds form by reacting a hemiacetal with a hydroxyl group).

d. Maltose is cleaved by the enzyme maltase, which is specific for α-bonds. Thus the bond is in the α orientation. Emulsin is an enzyme specific for β-bonds.

e. If maltose is oxidized by bromine water (converting free aldehyde to an acid), exhaustively methylated with dimethyl sulfate (so all free OHs are methylated), and then cleaved in acid, the following products result. **(PP 70)**

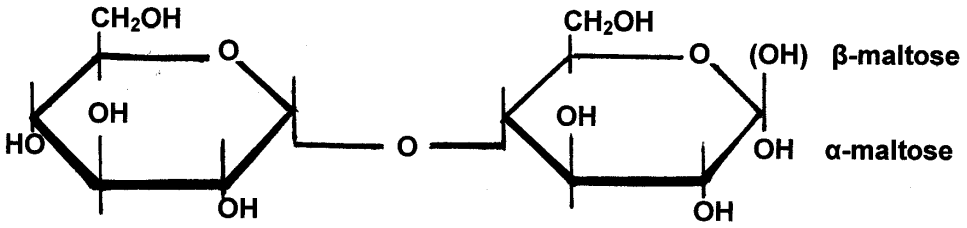


All OHs methylated except C-1, so this glucose joined through C-1



All OHs methylated except C-4, so this glucose joined through C-4

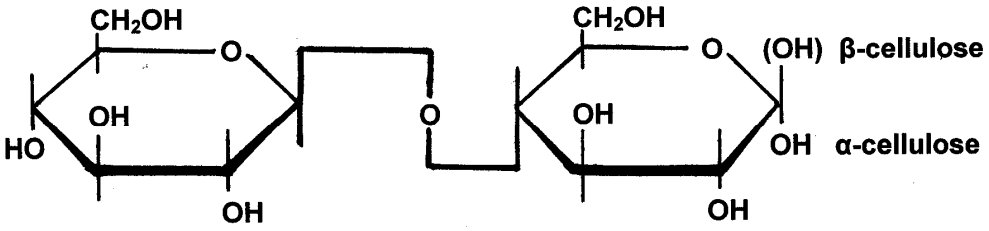
Thus the bond is from the C-1 of one glucose to the C-4 of the other. **(PP 71-72)**



non-reducing end reducing end

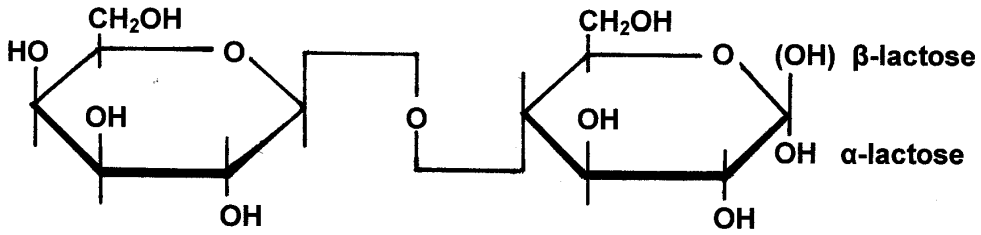
2. Cellobiose

Cellobiose is formed from the breakdown of cellulose. Acid hydrolysis yields only glucose. It is a reducing sugar that can mutarotate between α and β forms (one free reducing end). It is cleaved by emulsin but not by maltase, showing the linkage is β. Methylation shows the linkage is 1,4. **(PP 73)**



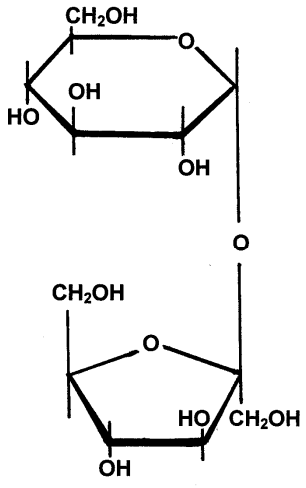
3. Lactose

Lactose, found in milk, forms D-glucose and D-galactose upon acid hydrolysis. If lactose is subjected to phenylhydrazine followed by acid hydrolysis, the products are galactose and the osazone of glucose, indicating that only the glucose has a reducing end. Therefore, the C-1 of galactose is involved in the glycosidic bond. Methylation shows the bond to be 1,4. Lactase cleaves the disaccharide, showing the bond to be β. **(PP 74-75)**



4. Sucrose

Sucrose (table sugar) is formed by plants. Acid hydrolysis yields D-glucose and D-fructose. It is non-reducing, so both anomeric carbons must be in the linkage (C-1 of glucose and C-2 of fructose). Treatment with enzymes is inconclusive regarding orientation since enzymes cleaving α and β linkages both work. X-ray analysis was needed to show that the bond is α with respect to glucose and β with respect to fructose. There is only one form of sucrose, not two anomers, and no mutarotation can occur. **(PP 76-78)**



C. Oligosaccharides in Proteins and Lipids

The other major class of oligosaccharides includes those that are covalently attached to proteins or lipids forming glycoproteins and glycolipids. Often such mixed molecules are found in the cell membrane or in proteins secreted by the cell. The function of the attached oligosaccharides is not fully understood, but several factors appear to be involved. First, attachment of hydrophilic oligosaccharides alters the polarity and solubility of proteins and lipids. Second, carbohydrates attached to proteins may help direct proper folding of the protein. A bulky oligosaccharide may prevent one interaction so another can occur. Third, groups of charged carbohydrates will repel and cause a relatively extended structure in that area of a protein, influencing 3° structure. Fourth, the oligosaccharides may protect the proteins from attack by proteases. Fifth, oligosaccharides may mediate recognition events and intercellular communication.

Oligosaccharides in glycoproteins and glycolipids tend to be very varied. They range in size from a few to ~14 monosaccharide units. There may be several different monosaccharide types including monosaccharide derivatives such as sugar acids and amino sugars. The glycosidic linkages also vary (1-2, 1-3, 1-4, 1-6, 2-3, 2-6), with some being α and others β. Only some oligosaccharides have been completely analyzed because of the difficulty in determining such a complex structure. Different glycoproteins use different oligosaccharides but one protein can contain many oligosaccharides. The carbohydrate portion can be 1-70% of the weight of a glycoprotein. **(PP 79-80)**

Many glycoproteins are found in the cell membrane where the oligosaccharides are located on the external side. Glycophorin of the erythrocyte membrane contains 16 oligosaccharides totaling 60-70 monosaccharides. The oligosaccharides are covalently attached to Ser, Thr, or Asn. Soluble glycoproteins include immunoglobulins and transport proteins such as the copper-transporting protein ceruloplasmin. Like other

soluble glycoproteins, the oligosaccharide chains of ceruloplasmin end in N-acetylneuraminic acid (a sugar acid also known as sialic acid). When these units are lost, the protein is taken up by the liver and destroyed. Thus removal of sialic acid is one mechanism for marking 'old' proteins for destruction and replacement. Within a cell, oligosaccharide attachment often marks a protein for secretion or movement to a particular cell organelle. Mannose -6-phosphate units are added to certain degradative enzymes so they are moved to lysosomes where they function in degrading old molecules.

Glycolipids are found in nerve cell membranes. Lipopolysaccharides are major components of some bacterial cell membranes.

IV. Polysaccharides

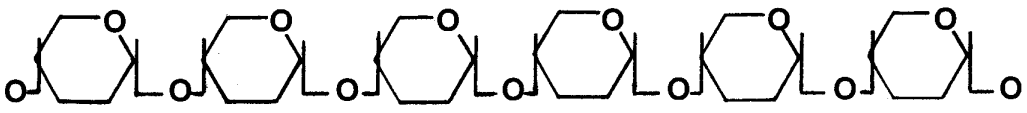
Polysaccharides, also known as glycans, have very high molecular weights. Homopolysaccharides contain one type of monosaccharide unit. Heteropolysaccharides contain two or more types of monomers. Polysaccharides do not have definite molecular weights since enzymes easily add or remove monosaccharide units. **(PP 81)**

A particular polysaccharide is characterized by its monomer types, the types of glycosidic bonds present, and the degree of branching in the carbon chain. Polysaccharides serve two main functions. First, they can serve as stores of metabolic fuel (monosaccharides). Second, they can be structural or support elements of organisms. **(PP 82-84)**

A. Starch

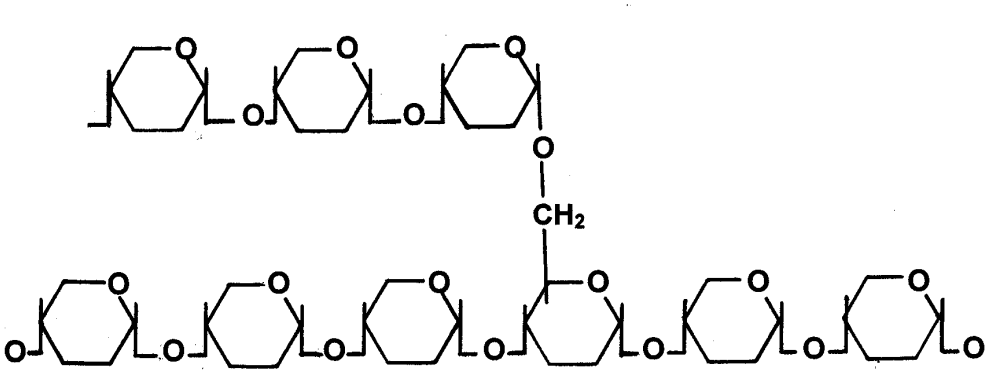
Starch is the storage polysaccharide of plants, occurring as granules inside cells, heavily hydrated with water. Starch has two components.

1. α-Amylose makes up ~20% of starch. It contains only D-glucose in an unbranched chain with the units linked by α 1,4 glycosidic bonds. Molecules have molecular weights of 150,000-600,000 which is about 1000-4000 glucose units. The molecular weight is determined by finding the percentage of C-4 atoms methylated, since such methylation can occur only at the end of a chain. **(PP 85)**



Since virtually all the anomeric carbons are in glycosidic bonds, amylose is non-reducing. It can be hydrolyzed by α-amylase to yield glucose and maltose. β-amylase hydrolyzes alternate bonds to produce maltose.

2. Amylopectin makes up 80% of starch. Molecules have molecular weights up to 100 million. It contains only glucose, but the structure is branched. Two types of glycosidic bonds are present, α 1,4 and α 1,6. Most bonds are α 1,4. When an α 1,6 bond occurs the structure branches. **(PP 86-87)**



The branch points occur every 24-30 residues (~ 5% of C-4 atoms are methylated ,and 5% of monomers are non-methylated at C-1, C-4, and C-6). It can be hydrolyzed by α-amylase to give glucose, maltose, and a limit dextran. The limit dextran is resistant to further degradation because of steric constraints. Digestion by β-amylase gives maltose and a limit dextran. Branch points can be cleaved by a debranching enzyme (α 1,6-glucosidase).

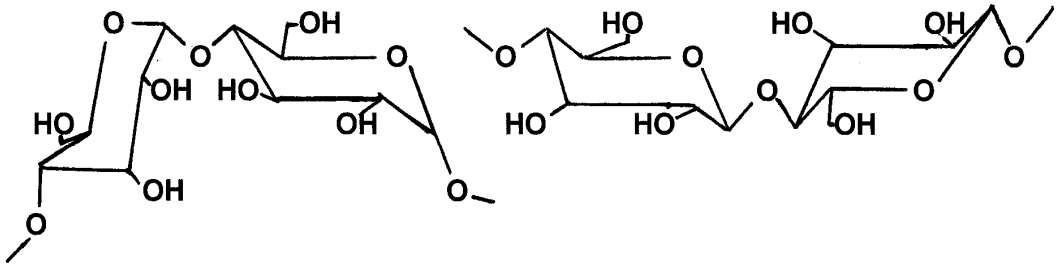
B. Glycogen

Glycogen is the storage polysaccharide of animals. It is very similar to amylopectin, but more highly branched (every 8-12 residues) and more compact. **(PP 88)** It is stored mainly in liver and skeletal muscle. Because it is branched, there are many sites at which enzymes can degrade it, allowing glucose to be released quickly when energy is needed. Glucose cannot be stored in its monomeric form because it would raise the osmolarity of the cell too high, and create such a concentration gradient (high concentration in the cell) that more glucose could not be taken up. Since glycogen is essentially insoluble, it does not cause similar problems.

C. Cellulose

Cellulose is found in the cell walls of plants where it provides support and structure. It consists entirely of D-glucose in linear chains of 10,000-15,000 monomers. However, cellulose contains β 1,4 linkages which makes it quite different from the earlier polysaccharides. **(PP 89)** Polysaccharide structure depends upon the type of covalent glycosidic bonds. That structure then tends to be stabilized by H-bonds between OH groups. When the bonds are α, the shape of the molecule is curved into a coil, which is a good compact shape for storage. This tight coil is stabilized by H-bonds. **(PP 90-91)**

When the bonds are β, such as in cellulose, the chain tends to be extended and linear, again stabilized by H-bonds. **(PP 92-94)**



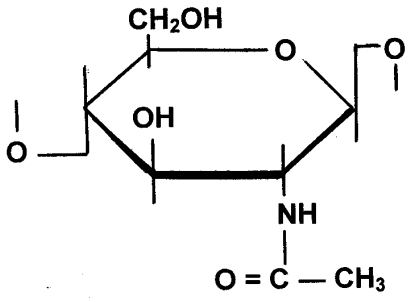
α-bonds β-bonds

Several cellulose chains lying next to each other can form a network of bonds, resulting in straight, stable, strong fibers.

Cellulose can be cleaved to glucose by the enzyme cellulase. **(PP 95-96)**

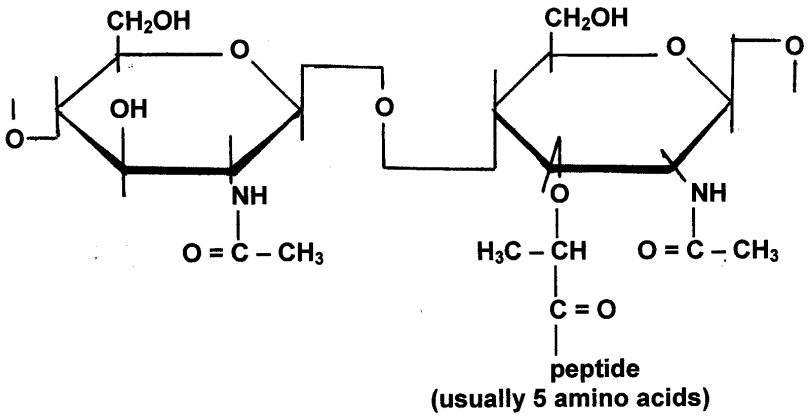
D. Chitin

Chitin forms the hard exoskeletons of arthropods (insects, lobsters, etc.). It is a homopolysaccharide of N-acetyl-D-glucosamine joined by β 1,4 linkages. Thus it will form strong, extended fibers. **(PP 97)**



E. Bacterial Cell Walls

The bacteria cell wall contains long parallel polysaccharide chains linked together by short peptides to form one enormous cage-like molecule called a peptidoglycan. The polysaccharide chains are heteropolymers of alternating N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) linked by β 1,4 bonds. **(PP 98-99)**



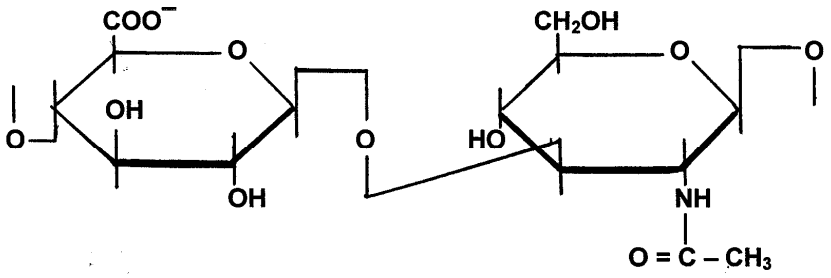
NAG NAM

The exact nature of the peptide varies with the species.

F. Extracellular Matrix

The extracellular space in animal tissue is filled with a gel-like material called the extracellular matrix. It holds cells together and provides a porous pathway for diffusion of O_2 and nutrients to cells. It is composed of a meshwork of heteropolysaccharides and fibrous proteins. The heteropolysaccharides are called glycosaminoglycans. They contain many negative charges which cause the molecule to be extended and give the solution a high viscosity.

The glycosaminoglycan hyaluronic acid or hyaluronate has alternating units of D-glucuronic acid and N-acetylglucosamine joined by β 1,3 and β 1,4 bonds with molecular weights over a million. **(PP 100)**



glucuronic acid NAG

Molecules of this type are lubricants in joints and give strength to the extracellular matrix of cartilage and tendons. Other glycosaminoglycans usually contain a uronic acid, NAG, or N-acetylgalactosamine, and some contain sulfate groups.

Proteoglycans are long glycosaminoglycans bound non-covalently to numerous protein molecules, which in turn are covalently bound to smaller glycosaminoglycan molecules such as chondroitin sulfate. The covalent bonds between carbohydrate and protein are mainly through serine residues. There can be as many as 150 polysaccharide chains per protein molecule, with about 100 protein molecules per one molecule of extended hyaluronate. This interacts with fibrous proteins like collagen and elastin, forming a cross-linked network to which cells attach (via proteins) and along which cell migration is directed.

Chapter 8: Lipids

I. Introduction

Lipids are fats and oils. They are water-insoluble substances that can be extracted from cells using organic solvents. Because they are grouped based on solubility properties, they are chemically more diverse than other groups of biomolecules. There are several distinct classes of lipids. Most lipids function as energy storage molecules or as structural components of membranes. Some are also homones, vitamins, and pigments.

II. Fatty Acids

A. Characteristics

1. They are carboxylic acids with hydrocarbon chains of 4 to 36 carbons with one acid group. The chain usually contains an even number of carbon atoms, with 16 or 18 being the most common number. **(PP 2)**

2. The chain is usually linear, but can be branched. A few fatty acids contain three-carbon rings or hydroxyl groups.

3. Some fatty acids are saturated (no double bonds). **(PP 3)** Some are monounsaturated (one double bond) or polyunsaturated (more than one double bond). For one double bond, the most common position is between C-9 and C-10. For several double bonds, the most common positions are C-9, C-12, and C-15. Double bonds are almost always unconjugated (-CH = CH – CH_2 – CH = CH -) rather than conjugated (- CH = CH – CH = CH -). The configuration is almost always cis rather than trans. Saturated molecules are extended; double bonds cause a kink. **(PP 4)**

4. Saturated fatty acids are waxy solids while unsaturated fatty acids are oily liquids.

B. Function

Free fatty acids are found in relatively low amounts in cells, but they are components and building blocks of many types of lipids.

C. Naming

Most fatty acids are given common names, and are described by a numbering system. **(PP 5)**

$CH_3(CH_2)_{14}COOH$	palmitic acid	16:O
$CH_3(CH_2)_{16}COOH$	stearic acid	18:O
$CH_3(CH_2)_7CH = CH(CH_2)_7COOH$	oleic acid	18:1 (Δ^9)
$CH_3(CH_2)_4CH = CHCH_2CH = CHCH_2CH = CHCH_2CH = CH(CH_2)_3COOH$		20:4 ($\Delta^{5,8,11,14}$)

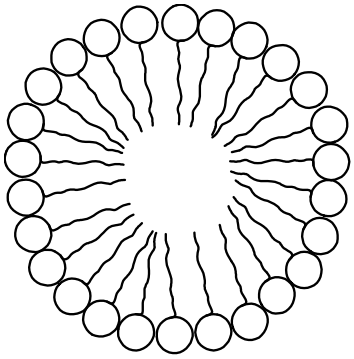
The first number is the number of carbon atoms, the second number is the number of double bonds, and the numbers with Δ describe the positions of the double bonds. **(PP 6-7)**

D. Reactions

Fatty acids placed in base (NaOH or KOH) form salts of fatty acids which are called soaps. **(PP 8)**

$$\underset{\text{fatty acid}}{CH_3(CH_2)_{16}COOH} + NaOH \rightarrow \underset{\text{soap}}{CH_3(CH_2)_{16}COO^- Na^+} + H_2O$$

Such salts are amphipathic since they possess a polar head (ionized carboxyl group) and a non-polar tail (hydrocarbon chain). While the original fatty acid is somewhat amphipathic, the fatty acid salt is strongly amphipathic. In water, soaps do not truly dissolve but disperse into micelles where the hydrophobic tails cluster to avoid water and the polar heads interact with water. **(PP 9)**

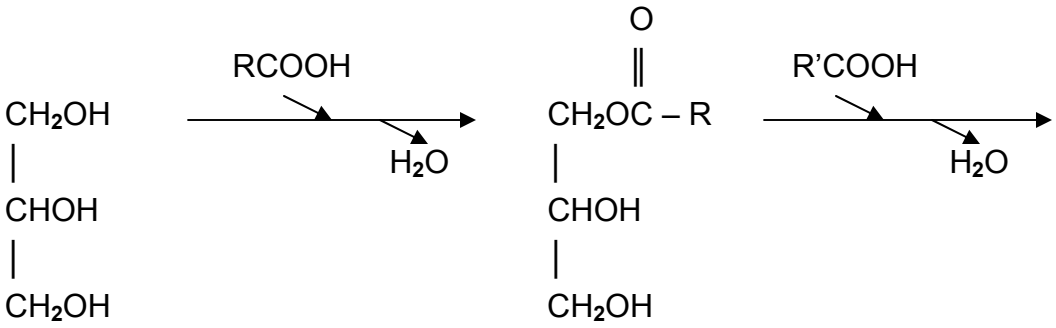


Fatty acid salts of Ca^{2+} and Mg^{2+} are very insoluble and precipitate out as white solids in hard water. Na^+ and K^+ salts can surround grease and disperse it, acting as soaps.

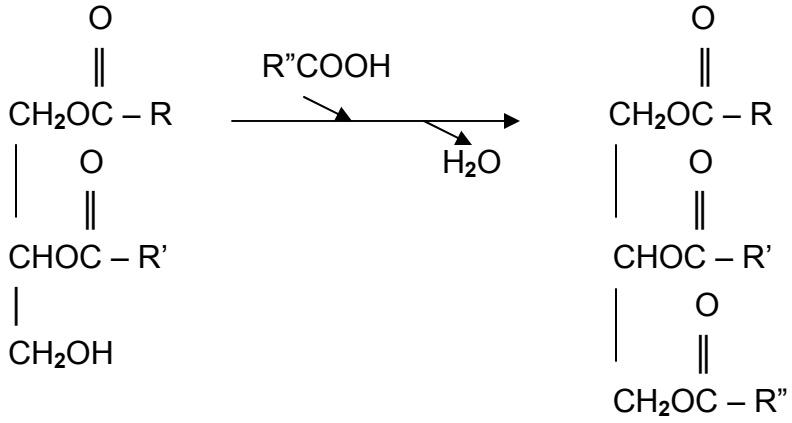
III. Triglycerides

A. Structure

Triglycerides (triacylglycerols) contain three fatty acids joined by ester bonds to a glycerol molecule. **(PP 10)** The three fatty acids may be the same or different. Without the -COOH group, they are even more non-polar than fatty acids. **(PP 11)**



monoglyceride **(PP 12)**



diglyceride

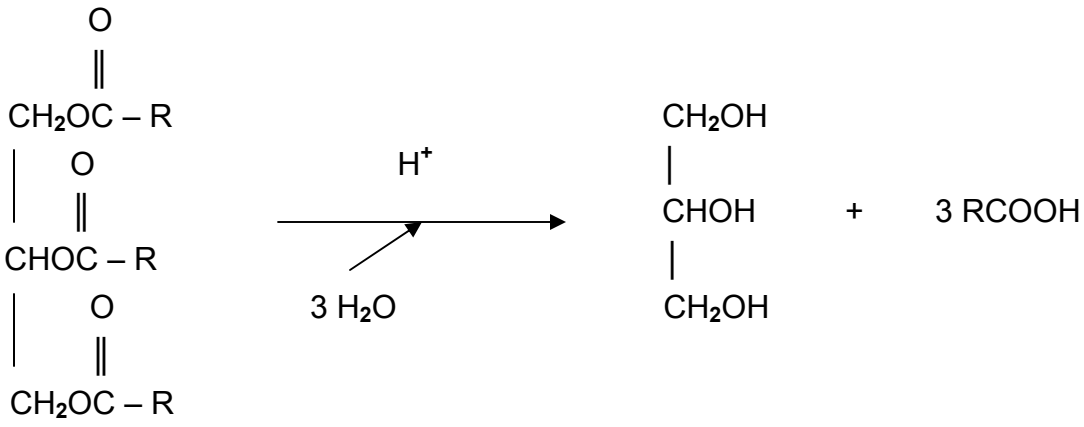
triglyceride **(PP 13)**

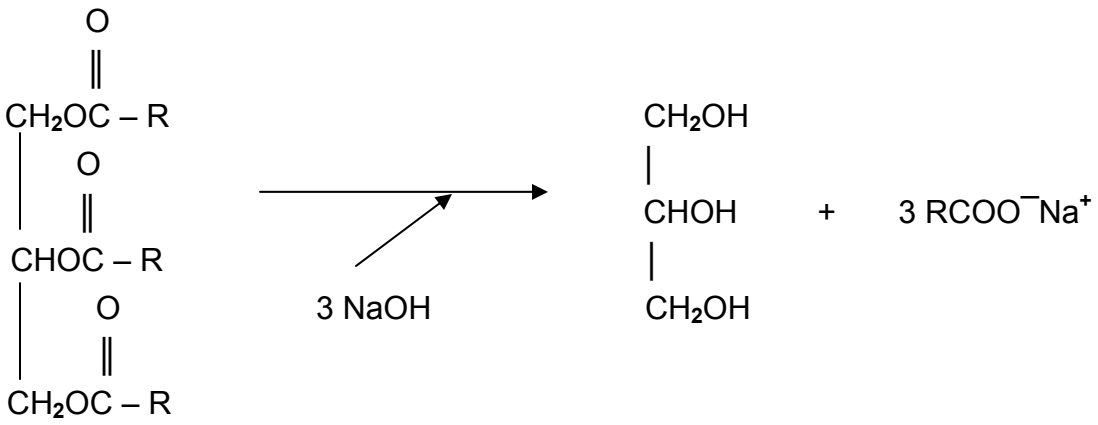
B. Function

Triglycerides store metabolic energy and provide insulation.

C. Reactions

Triglycerides can by hydrolyzed by acid or base. **(PP 14-15)**





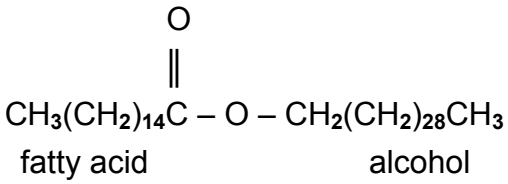
The second reaction is called saponification since it produces soaps. Lipids can be classified as saponifiable or non-saponifiable depending upon whether they contain fatty acids that can be released by base hydrolysis to form soaps. **(PP 16-17)**

Triglycerides can also be broken down by enzymes known as lipases.

IV. Waxes

A. Structure

Waxes are esters of long-chain fatty acids (14-36 carbons) with long-chain alcohols (16-30 carbons). **(PP 18)**



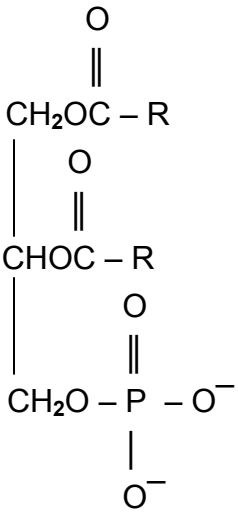
B. Function

Waxes function as metabolic fuels, and as protective coatings on hair, feathers, plants, etc.

V. Glycerophospholipids (phosphoglycerides)

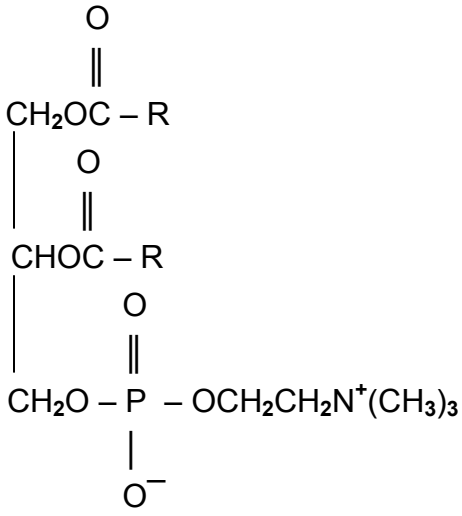
A. Structure

Glycerophospholipids contain glycerol, two fatty acids, and a phosphate group at C-3 with a polar group attached to it. **(PP 19-20)** They are derivatives of phosphatidic acid. **(PP 21)**



phosphatidic acid

Usually C-1 has a saturated fatty acid attached and C-2 has an unsaturated fatty acid. The groups that can attach to the phosphate include ethanolamine, choline, serine, and glycerol. **(PP 22-23)**



phosphatidylcholine

All glycerophospholipids always have a negative charge on the phosphate. The polar group may have additional charges. Thus they are amphipathic with a polar head (phosphate) and non-polar tail (fatty acids). **(PP 24-25)**

B. Function

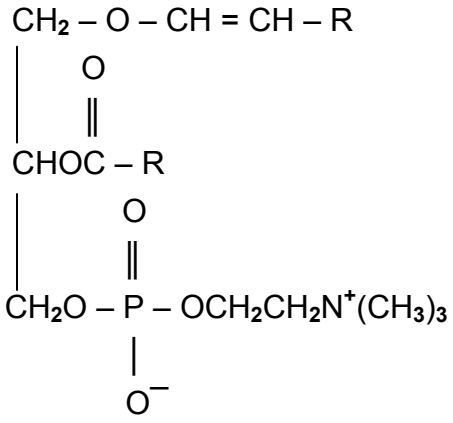
They are found in cell membranes.

C. Reactions

Glycerophospholipids can be hydrolyzed back to their components by acid, base, or phospholipases. Enzymatic breakdown can be involved in cell signals. In response to hormones, part of the lipid acts as an intracellular signal. **(PP 26)**

D. Ether-linked fatty acids

Some tissues contain ether lipids, where one of the ester bonds is instead an ether. While found in membranes, the significance of the bond is unknown. **(PP 27-28)**

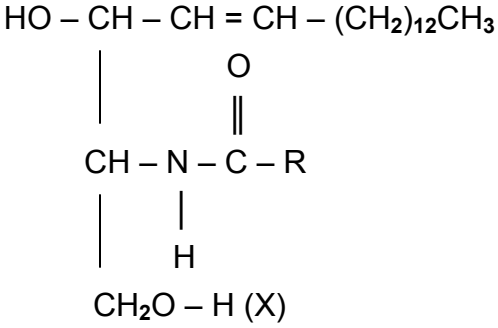


plasmalogen

VI. Sphingolipids

A. Structure

Sphingolipids contain one fatty acid, and one molecule of the long chain amino alcohol sphingosine, but no glycerol. These two molecules are joined to form a group of compounds called ceramides.**(PP 29)**



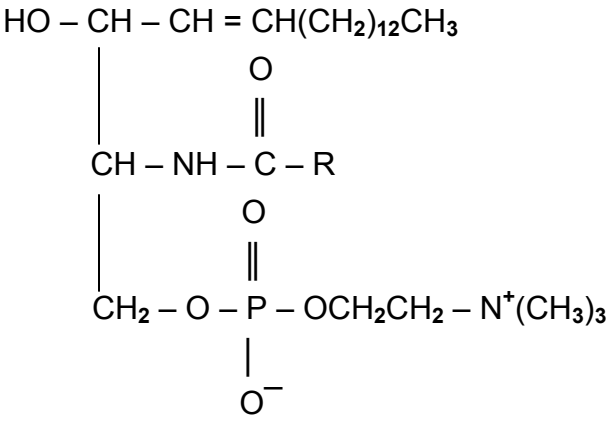
sphingosine

fatty acid

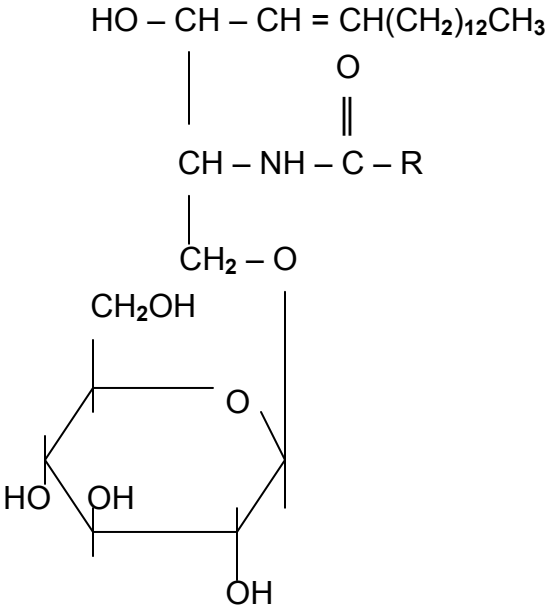
A polar group (X) can be added to form other types of sphingolipids. **(PP 30)**

B. Types

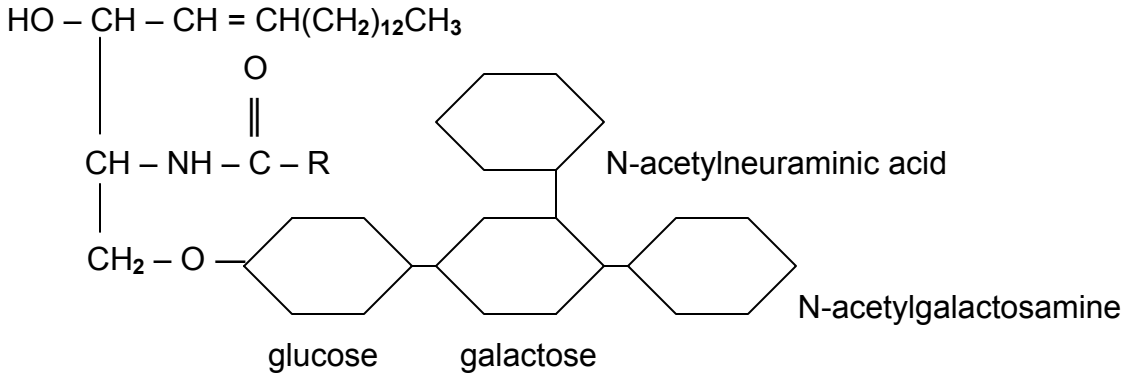
1. Sphingomyelins contain phosphocholine or phosphoethanolamine as the polar group. They are found in plasma membranes and in the myelin sheath. They have no net charge. **(PP 31)**



2. Cerebrosides contain a single sugar unit as the polar group. With galactose the lipid is found in neural cell plasma membranes; with glucose the lipid is found in nonneural membranes. They are uncharged overall. **(PP 32)**



3. Gangliosides contain a polar head made up of several (~ 4-6) sugar units. They are negatively charged due to the presence of sugar acids. They are found in various membranes including nerve cell membranes. **(PP 33)**



4. Globosides are similar in structure to gangliosides, but the oligosaccharide polar group contains 2-4 neutral sugar units, usually glucose, galactose, or N-acetylgalactosamine.

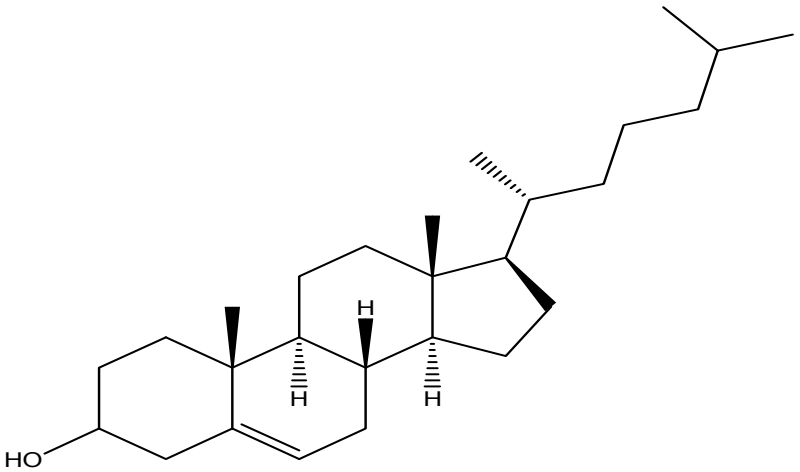
C. Function

Sphingolipids are more than just structural components of membranes. They are involved in recognition events on the cell surface, including such things as blood groups and toxin binding. The exact functions of many sphingolipids are unknown. **(PP 34-35)**

VII. Sterols

A. Structure

Sterols contain the steroid ring system of four fused rings, various side groups, and a hydroxyl group. Other steroids contain a carbonyl group. **(PP 36)**



cholesterol

They do not contain a fatty acid and so are non-saponifiable. They are amphipathic with the oxygen group making a polar head for the molecule.

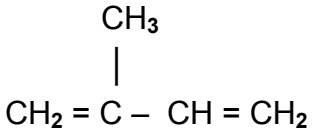
B. Function

Cholesterol is found in eukaryotic cell membranes. Sterols include bile acids, which help digest fats. The steroid hormones are mainly sex hormones that carry messages and so change metabolism. **(PP 37)** Vitamin D is also a steroid. It is a precursor of the hormone 1,25-dihydroxycholecalciferol which regulates calcium and phosphate metabolism (bones). **(PP 38)**

VIII. Isoprenoids (Terpenes)

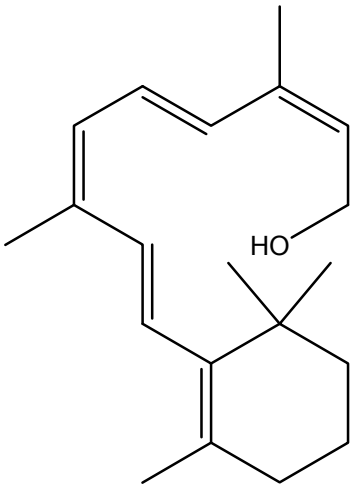
A. Structure

Isoprenoids are polymers of the 5-carbon unit isoprene **(PP 39)** and so are non-saponifiable.



isoprene

Vitamin A (retinol) is a terpene. **(PP 40)**



retinol

B. Function

Vitamin A functions in vision. Vitamin E prevents oxidative damage to membrane lipids. **(PP 41)** Vitamin K is required for blood clotting. **(PP 42)** Quinones function as electron carriers involved in energy production. Steroids are synthesized from 5-carbon compounds and so are technically terpenes.

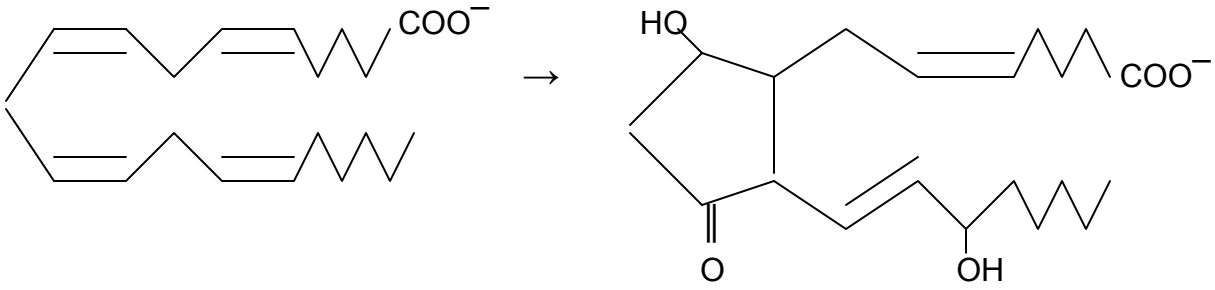
IX. Eicosanoids

A. Structure

Eicosanoids are fatty acid derivatives formed from arachidonic acid (20:4), a polyunsaturated fatty acid. **(PP 43)**

B. Types

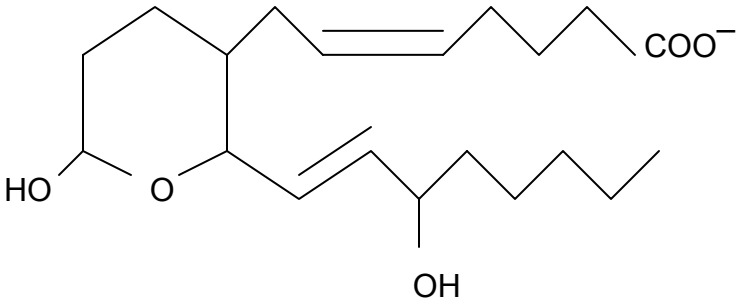
1. Prostaglandins contain a 5-membered ring.



Prostaglandin PGD_2

Prostaglandins are involved in smooth muscle contraction, regulation of blood flow, response to hormones, and regulating body temperature, inflammation, and pain. They are powerful biological signals.

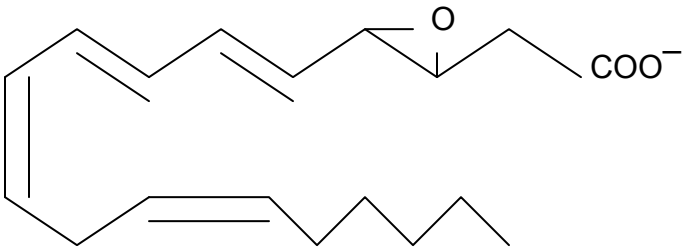
2. Thromboxanes have a 6-membered ring with an ether.



Thromboxane TXB_2

They function in formation of blood clots.

3. Leukotrienes contain three conjugated double bonds.



Leukotriene A

They are potent regulatory molecules, controlling such things as contraction of muscle linings in airways of lungs. **(PP 44-45)**

X. Membranes

Membranes contain the cell contents, and control the flow of substances in and out of the cell. They divide internal cell space into compartments and allow for cell-to-cell communication. They regulate pH and cell volume. **(PP 46)**

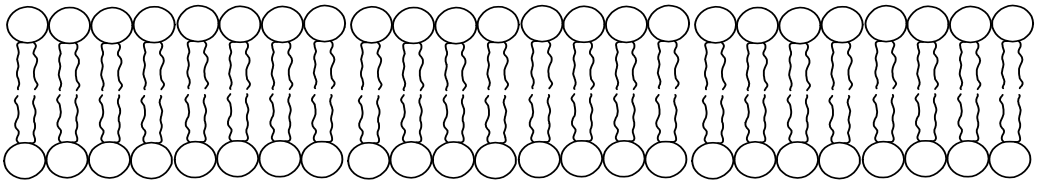
A. Composition

Membranes are composed of proteins and polar lipids. The proportions can vary enormously in different types of membranes, from 80% lipid/ 20% protein to 80% protein/ 20% lipid. The types of lipids are phospholipids, sphingolipids, and sterols, but again the proportions vary, with the sterols accounting for 0-50% of the lipid. **(PP 47)** The types of proteins vary even more depending upon the function of the cell. There can be 90% of one protein in specialized cells (rhodopsin in rod cells of retina), 20 major proteins, or hundreds of different proteins involved in transport, secretion, cell division, etc. Proteins can contain carbohydrate portions. Some may be anchored by covalent attachment to membrane lipids.

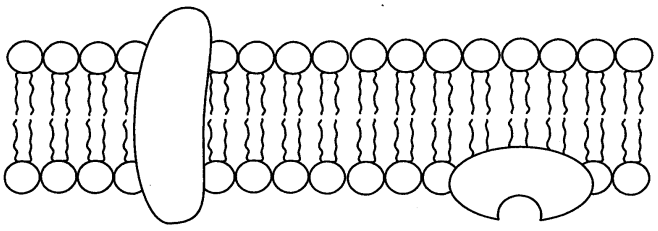
B. Membrane Structure

Membranes are 5-8 nm thick. Water and many non-polar molecules can move freely in and out of the membranes, but polar molecules and ions cannot (generally). They can move in and out only with help of transport proteins in the membrane.

The basic structure is that of a lipid bilayer. Two layers of lipid molecules arrange themselves with the polar portions facing outward toward the aqueous environment and the non-polar portions buried in the interior. **(PP 48)**



Proteins are embedded at irregular intervals. Some proteins, called integral membrane proteins, are firmly embedded in the membrane. They can either span the membrane or protrude only on one side. Other proteins are bound loosely to the membrane surface, termed peripheral membrane proteins. **(PP 49)**



Integral protein Peripheral protein

This is called the fluid mosaic model of membrane structure. Because the vast majority of interactions between lipid molecules and between lipid and protein molecules are non-covalent, the membrane structure is fluid and flexible, not rigid.

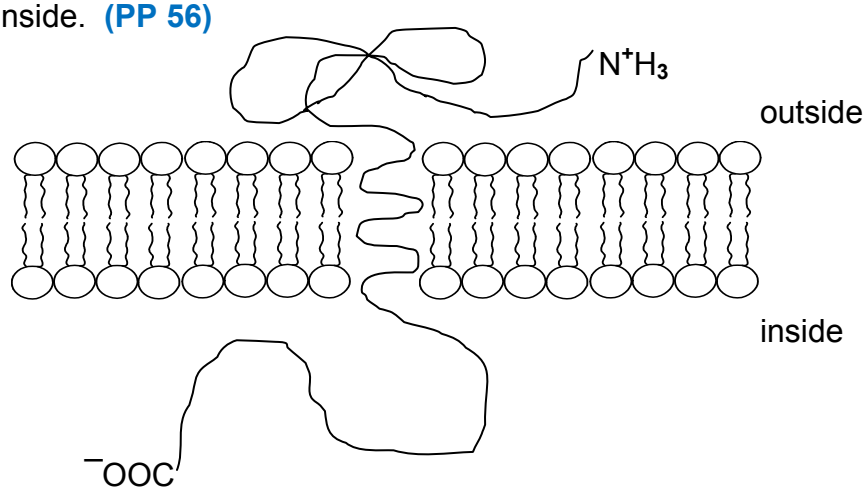
The membrane is asymmetric because different proteins face outward from the cell compared to the inside of the cell. Lipids are also asymmetrically distributed. Sphingomyelin and phosphatidylcholine are found preferentially on the outer surface of some membranes, while phosphatidylserine and phosphatidylethanolamine are found preferentially on the inner surface of the membrane. **(PP 50)**

While the lipid bilayer is stable, individual lipid molecules have great freedom of motion within the membrane. Lipid molecules can move laterally, and the hydrocarbon chains of the fatty acids are in motion due to rotation around the many carbon-carbon bonds. Lipid molecules can even flip from one side of the membrane to the other. **(PP 51)** Membranes can change the relative amounts of saturated and unsaturated fatty acids, depending upon conditions such as temperature. Saturated fatty acids fit well into an ordered array but the kinked unsaturated fatty acids disrupt order and make the membranes more fluid. **(PP 52)** Sterols, with a rigid ring, reduce the freedom of movement of fatty acid chains and so reduce fluidity. **(PP 53-54)**

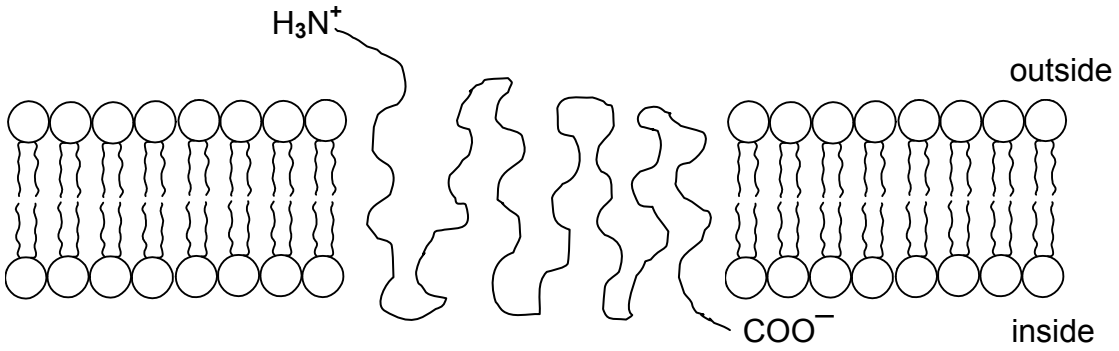
C. Membrane Proteins

Integral membrane proteins are firmly bound within the membrane while peripheral membrane proteins are loosely or reversibly bound to the membrane surface.

Integral proteins often span the membrane, but some are located on only one side. While some integral proteins are fully embedded in the membrane, others have large segments extending out into the surrounding solvent. **(PP 55)** A relatively simple integral membrane protein is glycophorin found in red blood cells. The N-terminal part of the protein glycophorin is on the outside and contains carbohydrate. The C-terminal part is on the inside. **(PP 56)**



Both terminal areas contain many polar amino acids and are hydrophilic, while the segment in the middle of the membrane is hydrophobic. The result is that the protein is asymmetrically oriented. Its orientation is specific and it does not flip-flop. Other integral membrane proteins, such as bacteriorhodopsin in bacteria, cross the membrane several times with hydrophobic α-helical segments. **(PP 57)**



Hydrophobic interactions between non-polar amino acids and the fatty acid chains anchor the protein firmly in the membrane. Some proteins are free to diffuse around and others are not. Certain proteins are covalently anchored to lipid molecules. Integral membrane proteins, because they contain extensive hydrophobic regions, are generally insoluble and so can be difficult to study. **(PP 58)**

Peripheral membrane proteins are held to the membrane by electrostatic interactions and H-bonds with the polar heads of membrane lipids and hydrophilic domains of integral membrane proteins. They are water-soluble. They may regulate membrane-bound enzymes, connect integral proteins to intracellular structures, or limit mobility of integral membrane proteins.

Certain proteins are covalently anchored to lipid molecules, which in turn anchor to the membrane through hydrophobic interactions. **(PP 59-61)**

XI. Lipoproteins

Another important function of lipids is as components of lipoproteins. Many lipoproteins are involved in transporting lipids around the body. Since lipids are water-insoluble, they cannot be transported freely. Instead they form complexes with specific proteins where the hydrophobic lipids aggregate at the core of the particle, while the hydrophilic amino acids of the protein are on the surface along with the polar groups of lipid molecules. **(PP 62)** There are four main classes of lipoproteins with different combinations of lipid and protein. **(PP 63-65)**

A. Chylomicrons contain 2% protein and 98% lipid. They transport triglycerides from the intestine to other tissues for use or storage.

B. Very low-density lipoproteins (VLDLs) contain 10% protein and 90% lipid. They transport excess triglycerides from the liver to adipose tissue for storage.

C. Low-density lipoproteins (LDLs) are 25% protein and 75% lipid, mostly cholesterol. They move cholesterol from the liver to other tissues.

D. High-density lipoproteins (HDLs) are 33% protein and 67% lipid. They collect cholesterol remaining in other lipoproteins, and recycle the cholesterol to the liver.

Chapter 9: Nucleotides

I. Introduction

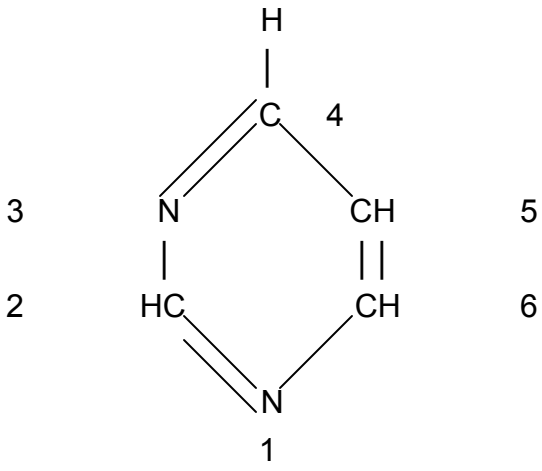
Nucleotides are the monomer units of nucleic acids. Just as amino acids are the building blocks of proteins, so nucleotides are the building blocks of nucleic acids. In addition to being nucleic acid components, nucleotides themselves function in energy storage and transfer during metabolic reactions. They also function as cofactors for certain enzymes, and as chemical signals in the response to hormones.

II. Nitrogenous Bases

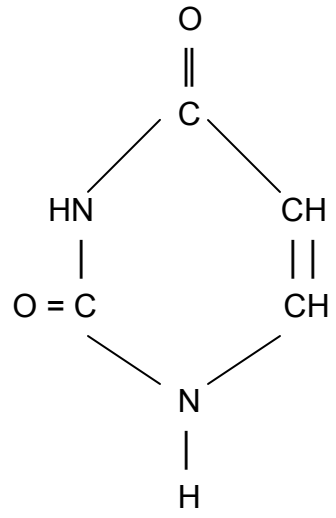
Nucleotides have three component parts, a nitrogenous base, a sugar, and a phosphate group. **(PP 2)** The most unusual of these is the nitrogenous base. These are heterocyclic compounds containing carbon and nitrogen in a 5 or 6-membered ring. The rings are unsaturated and have aromatic character. They are weak bases that are uncharged at pH = 7. There are two major types, pyrimidines and purines. **(PP 3)**

A. Pyrimidines

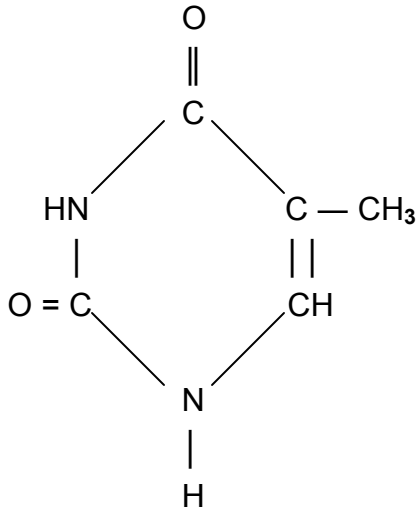
Pyrimidines are based on the parent compound pyrimidine, which contains one heterocyclic ring.



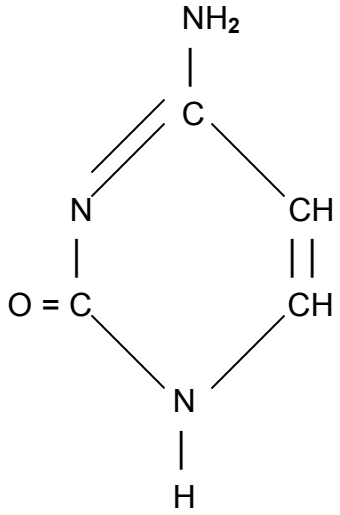
Numerous derivatives can be made, but three pyrimidines are predominant in nucleotides and nucleic acids. **(PP 4)**



uracil - U
2,4-dioxopyrimidine



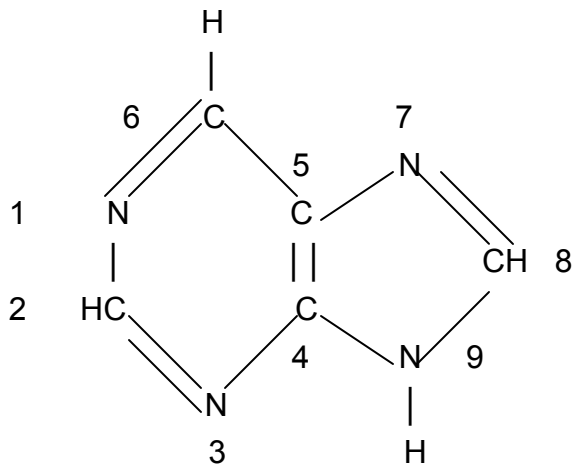
thymine - T
5-methyl -2,4-dioxopyrimidine



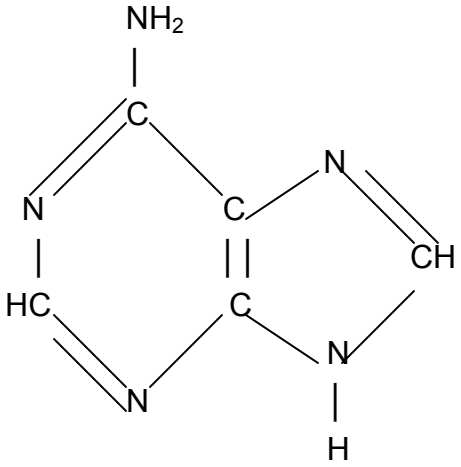
cytosine - C
4-amino -2-oxopyrimidine

B. Purines

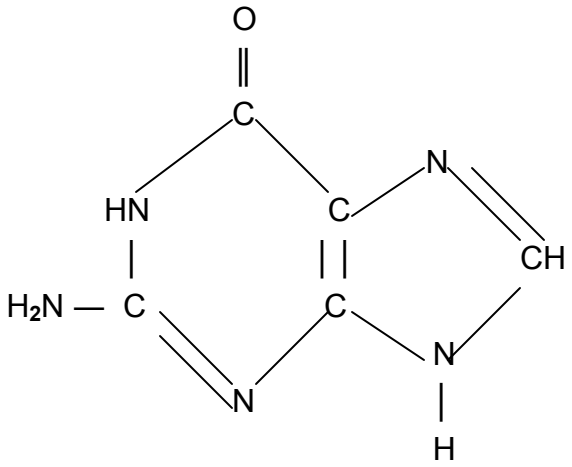
Purines are based on the parent compound purine, which is a derivative of pyrimidine. Purines consist of a pyrimidine ring fused with an imidazole ring to give two heterocyclic rings, one 5-membered and one 6-membered.



Two important purines are found in nucleic acids. **(PP 5)**



adenine - A
6-aminopurine



guanine - G
2-amino -6-oxopurine

C. General Features

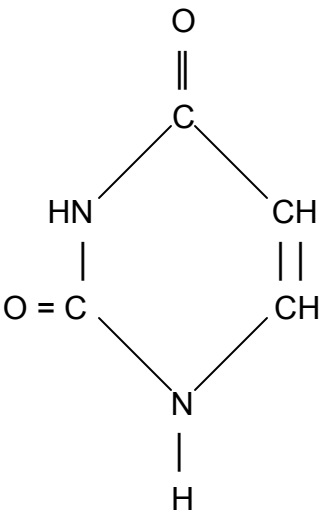
1. Both purines and pyrimidines have aromatic character and so are stable despite the presence of numerous double bonds. **(PP 6)**

2. Both types are weak bases but will be uncharged at pH = 7.0.

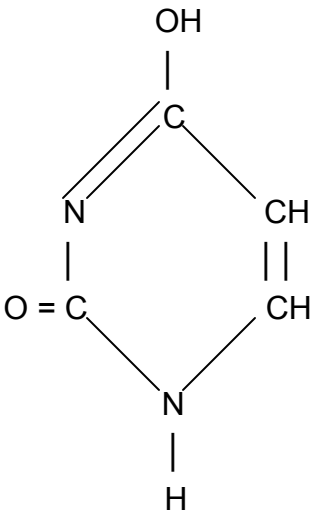
3. Pyrimidines are planar in shape. Purines are nearly planar with a slight pucker.

4. Both types have a great capacity for forming hydrogen bonds with the NH, NH_2, N, C = O groups.

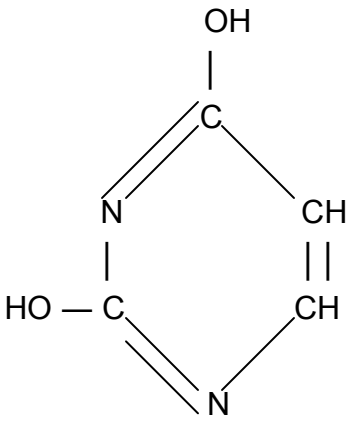
5. The forms shown above are the predominant forms at pH = 7.0. However, alternate tautomeric forms can exist, with the amounts of each form varying with pH. The tautomers of uracil are shown below. **(PP 7)**



lactam
(keto groups)



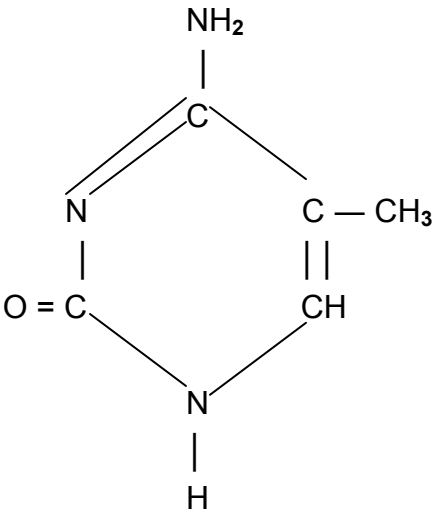
lactim



double lactim
(enol groups)

6. Rare bases

In addition to the five major bases, other unusual bases do exist and are sometimes found in nucleic acids. **(PP 8-9)** Often these rare bases have extra methyl groups. They still share the general characteristics of the more common bases. **(PP 10-11)**



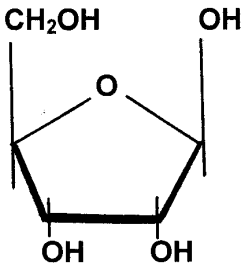
5-methylcytosine

III. Nucleosides

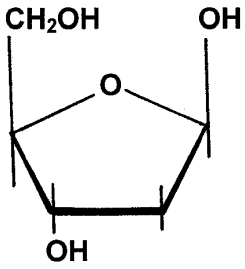
The second component of a nucleotide is a sugar. When a sugar is joined to a nitrogenous base, the resulting structure is called a nucleoside.

A. Sugars

Two types of sugars are found in nucleotides, both based on D-ribose. **(PP 12-13)**



β-D-ribofuranose

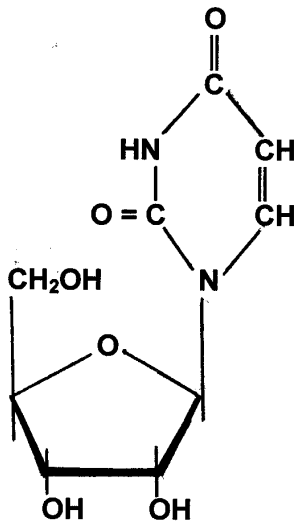


2-deoxy -β-D-ribofuranose
(lacks OH at 2 position)

B. Nucleoside Structure

When the sugar is linked to a nitrogenous base, a nucleoside is produced. The bond forms between the C-1 of the sugar and one of the nitrogens of the rings of the nitrogenous base. This is an N-glycosidic bond. It is relatively stable to alkali but is

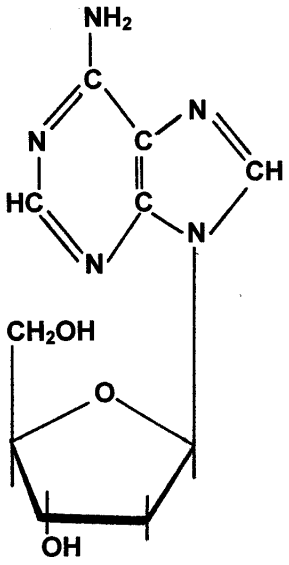
susceptible to acid hydrolysis (purines more than pyrimidines). The orientation of the bond is β. In pyrimidines the bond will be 1,1. **(PP 14)** In purines the bond is 1,9. **(PP 15)** To distinguish between the numbering of the sugar and the numbering of the nitrogenous base, primes are added to the numbers that refer to the sugar.



uridine

(1-β-D-ribofuranosyluracil)

The N-glycosidic bond links the 1 position of the pyrimidine to the 1' position of the ribose.



deoxyadenosine

(9-β-2'-deoxy-D-ribofuranosyladenine)

The N-glycosidic bond links the 9 position of the purine to the 1' position of the deoxyribose. **(PP 16)**

The sugar and the base are both nearly planar and lie at right angles to each other. Two conformations, syn and anti, are possible due to rotation around the N-glycosidic bond. Due to steric hindrance, the anti conformation is favored. **(PP 17-19)**

C. Nomenclature

To simplify the names, nucleosides are named by adding –dine or –sine to the name of the base. Deoxy as a prefix signifies the deoxyribose sugar. No prefix indicates the ribose sugar.

Base	Ribonucleoside	Deoxyribonucleoside
adenine	adenosine	deoxyadenosine
cytosine	cytidine	deoxycytidine
guanine	guanosine	deoxyguanosine
thymine	thymidine	deoxythymidine
uracil	uridine	deoxyuridine

IV. Nucleotides

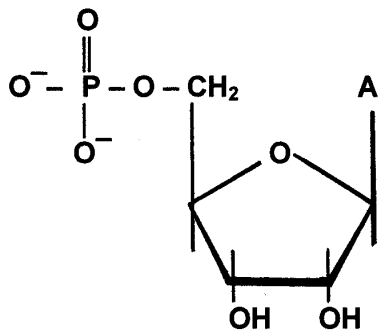
The third component of nucleotides is one or more phosphate groups. When a phosphate group is added to a nucleoside, the resulting compound is a nucleotide.

A. Types

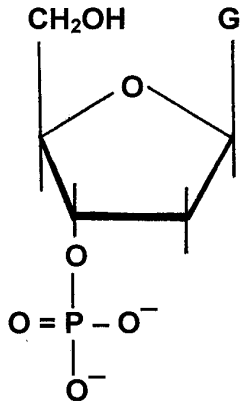
Nucleotides are distinguished by two features, where the phosphate group is attached and how many phosphate groups are attached.

1. Position

The phosphate group will be attached to the ribose ring, but can vary as to which OH group it is joined to. The 5'-position is by far the most common, with the 3'-position also occurring.



adenosine -5’-monophosphate

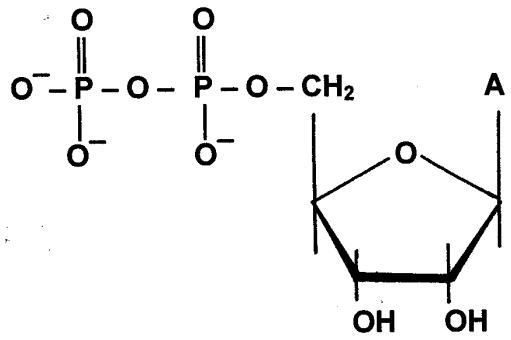


deoxyguanosine-3’-monophosphate

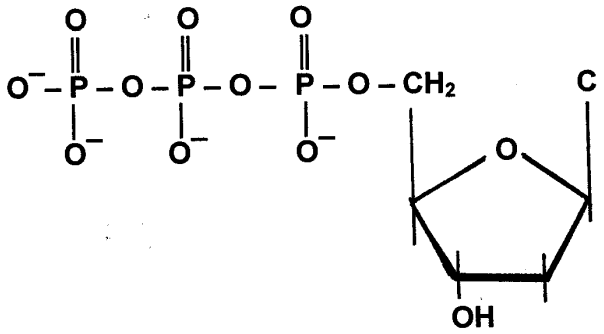
‘Mono’ is used to indicate the presence of one phosphate group. If no number is included in the name, such as adenosine monophosphate, then the assumption is that the phosphate group is at the 5’ position. At pH = 7.0, phosphate will be ionized and carry a negative charge of -2. **(PP 20-25)**

2. Number of phosphates

Many nucleotides have two or three phosphate groups, usually on the same carbon **(PP 26)**, although mixed 3',5' phosphates can occur.



adenosine-5’-diphosphate
charge = -3



deoxycytidine-5'-triphosphate
charge = -4

B. Nomenclature

To simplify the names, each compound is given an abbreviation. The first capital letter indicates the base, the second letter indicates the number of phosphates, and the third letter indicates phosphate. A small 'd' in front of the abbreviation signifies a deoxyribonucleotide. Without the 'd', it is assumed the compound is a ribonucleotide.

5' ribonucleotides (NMP = nucleoside monophosphate)

Base	NMP	NDP	NTP
A	AMP	ADP	ATP
G	GMP	GDP	GTP
C	CMP	CDP	CTP
U	UMP	UDP	UTP

5'-deoxyribonucleotides (dNMP = deoxyribonucleoside monophosphate)

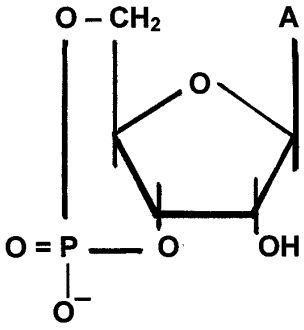
Base	dNMP	dNDP	dNTP
A	dAMP	dADP	dATP
G	dGMP	dGDP	dGTP
C	dCMP	dCDP	dCTP
T	dTMP	dTDP	dTTP

Because of the negative charges, nucleotides are often found in the cell complexed with divalent ions like Mg^{2+}. In the cell, phosphate groups are easily transferred by enzymes, so there is a constant interconversion among mono, di, and triphosphates. Free nucleosides and bases are found in cells only in very low levels. **(PP 27-30)**

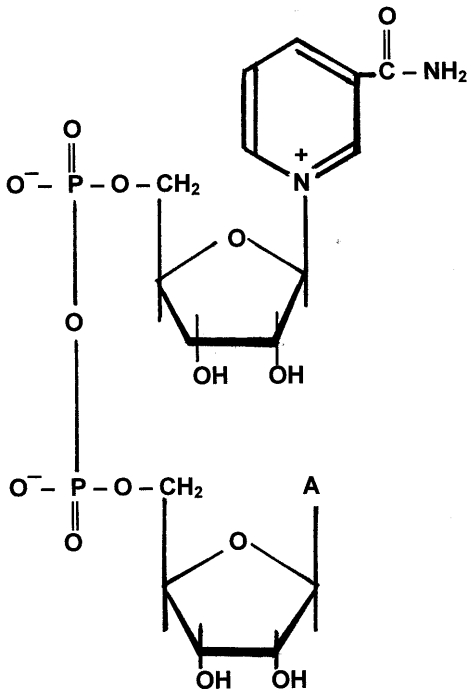
C. Function

ATP and other nucleotides are used for energy storage and transfer in cells as they lose or gain phosphate groups. They are also involved in enzyme regulation, participating in both allostery and covalent modification.

Unusual nucleotides, such as 3',5'-cyclic AMP (cAMP) mediate the action of hormones. **(PP 31)**



Some nucleotides function as enzyme cofactors, like nicotinamide adenine dinucleotide (NAD^+). **(PP 32)**



Chapter 10: Nucleic Acids

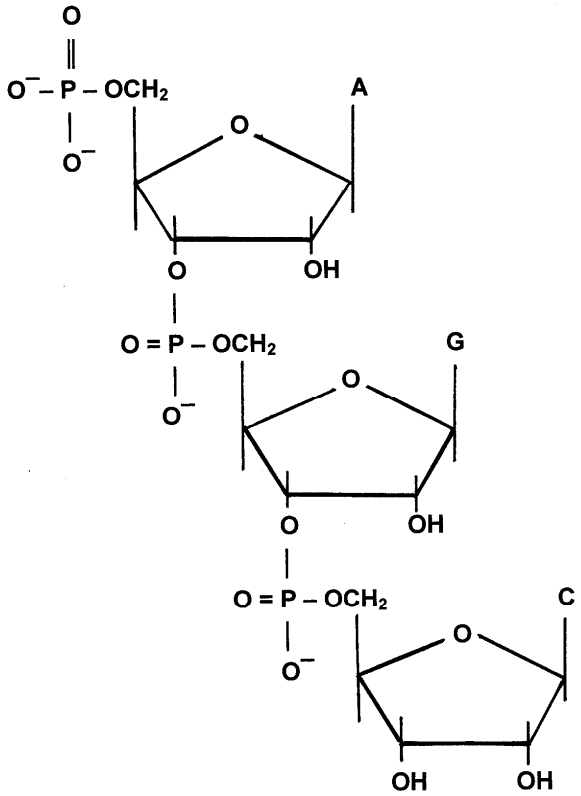
I. Introduction

Nucleic acids are long polymers of nucleotides. They function in the storage, transmission and usage of genetic information. There are two basic types of nucleic acids, DNA (deoxyribonucleic acid) and RNA (ribonucleic acid). DNA is the genetic material of the cell. It contains the cell's genes, which in turn determine all the characteristics of a cell. RNA (there are several types) functions in using the genetic information to make proteins which determine the nature of the cell.

II. Structure of Nucleic Acids

A. Phosphodiester Bonds - Primary Structure

Nucleotides are joined together by phosphodiester bonds into a nucleic acid. The sequence in which the different nucleotides are joined is the primary structure of a nucleic acid. **(PP 2)**



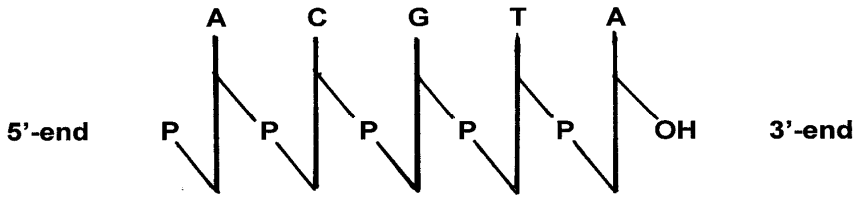
The bond is always a 3',5'-phosphodiester bond, linking the 5' carbon of one sugar ring to the 3' carbon of the next. RNA contains 2'-OH groups while DNA would lack them.

The covalent backbone thus consists of alternating phosphate and sugar groups. The bases are linked to the sugars but are not part of the phosphodiester bonds. The phosphate groups will be negatively charged at pH = 7.0.

The two ends of the chain are not identical. One end has a 5'-phosphate and the other has a 3'-OH. Thus the nucleic acid chain is said to have directionality or polarity.

The backbone is very hydrophilic while the bases are more hydrophobic. Nucleic acids range in size from less than 100 nucleotides to more than 10,000,000 nucleotides.

To consolidate the structure of nucleic acids, several notations can be used.



or pApCpGpTpA or pACGTA

The 5'-end by convention is on the left. **(PP 3)**

Nucleic acids are degraded back to nucleotides by acid (which also breaks N-glycosidic bonds). RNA degrades in base (due to the 2'-OH groups) but DNA does not. Nucleic acids can also be degraded by enzymes known as nucleases or phosphodiesterases. Nucleases are generally specific for either RNA (ribonucleases) or DNA (deoxyribonucleases). Some cleave at the end of a chain (exonucleases) while some cleave in the middle of a chain (endonucleases). Some produce 5'-nucleotides while some produce 3'-nucleotides.

B. RNA vs. DNA

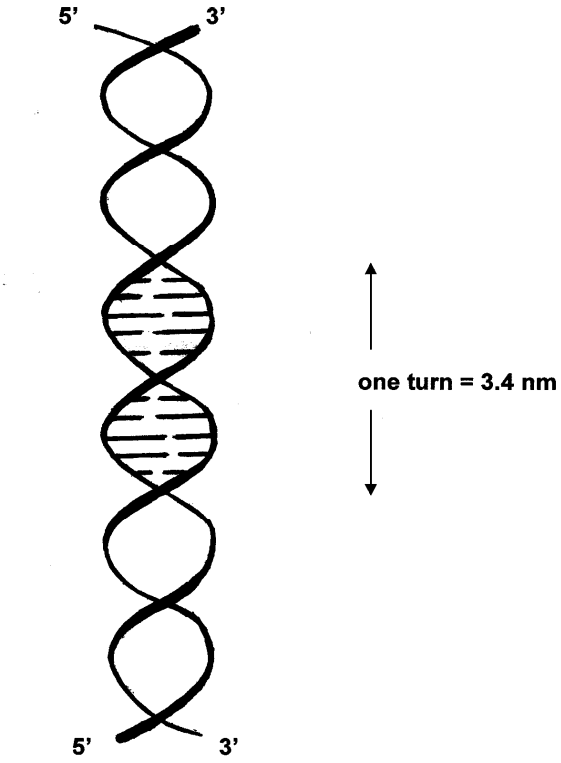
There are several differences between DNA and RNA. Most are generally though not universally true. **(PP 4-6)**

DNA	RNA
deoxyribose	ribose
A, C, G, T	A, C, G, U
larger, 10^3 - 10^8 nucleotides	smaller, 10^2 - 10^4 nucleotides
nucleus	cytoplasm
double-stranded	single-stranded
1-2 copies of one type	many copies of multiple types

C. Secondary Structure

Some nucleic acids, mainly double-stranded DNA, form more complex structures. The major form of secondary structure is the double-helix. The double-helix is a spiral shape, somewhat similar to the α-helix, except that two chains are involved. Two pieces of evidence helped in deducing the structure of the double-helix. First were the X-ray diffraction studies that showed a regular, repeating structure with definite dimensions (Franklin & Wilkins). **(PP 7)** Second were studies by Chargaff that showed a definite relationship between the amount of the bases, with A = T and G = C. In 1953, Watson & Crick put these results together to propose the structure of the double-helix, now known to be correct.

The double-helix, or duplex structure, has two chains in a spiral shape. **(PP 8)**

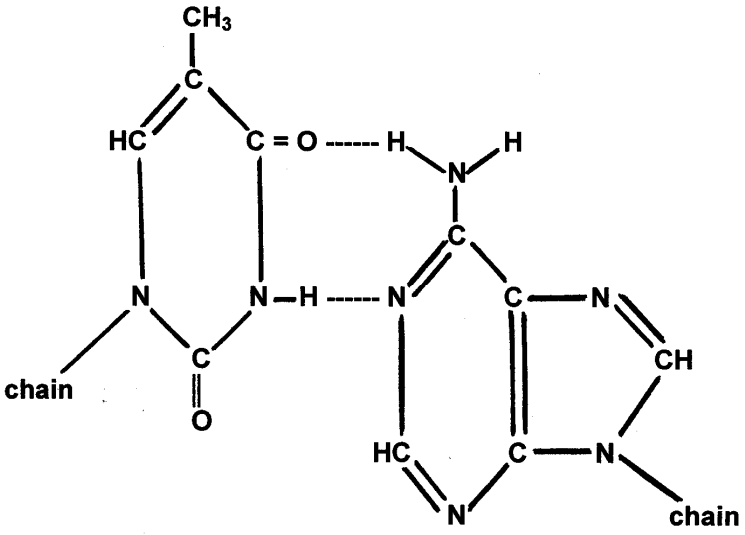


Each turn of the helix is 34 Å or 3.4 nm containing 10 base-pairs. (Each base-pair occupies 0.34 nm.) The width is 2.0 nm (20 Å). The helix is right-handed. The two chains have opposite directionality and so are antiparallel. The hydrophilic phosphates and sugars are on the outside where they can H-bond to water, and where repulsion between the two negatively-charged chains can be minimized. The more hydrophobic bases are protected in the interior. The bases lie perpendicular to the helix and parallel to each other. **(PP 9)** The helix contains two grooves on the exterior of unequal size, called the major groove and the minor groove which also spiral around the structure.

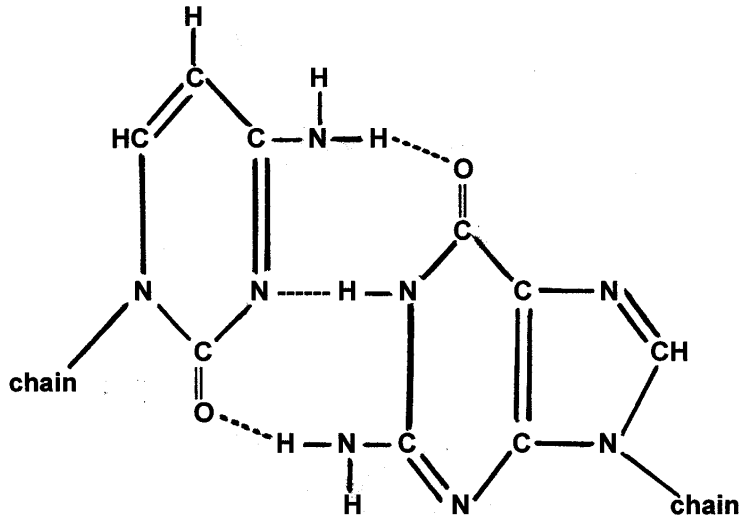
The two chains are held together by non-covalent forces. The first force is the relative hydrophilic vs. hydrophobic nature of the different parts of the nucleic acid. The second force is base stacking. The bases within a strand are parallel to each other, and there are van der Waals forces, dipole-dipole interactions, and hydrophobic interactions between them that stabilize the helix (and have effects on their electronic interactions and light absorbance).

The third force is hydrogen bonding. Only certain base combinations can hydrogen bond. **(PP 10)**

A – T base-pair (11.1Å wide with 2 H-bonds)



G – C base-pair (10.8Å wide with 3 H-bonds)



Other combinations of bases cannot fit within the helix or cannot properly hydrogen bond. This explains why the amount of A always equals the amount of T, and the amount of C equals the amount of G. Thus the base sequences of the strands are not identical, but rather are said to be complementary. **(PP 11)** This also explains the potential importance of tautomerization, since a base in a different tautomeric form will hydrogen bond differently. **(PP 12-15)**

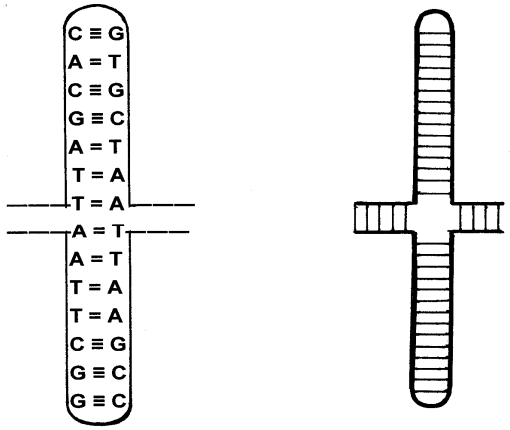
The double-helix is a somewhat rigid structure with fixed dimensions, but considerable flexibility also exists. Rotation around certain bonds can occur, as can bending and stretching of the strands. Local variations in structure occur depending upon the base sequence. In addition, the ends of a DNA molecule are not always base-paired, but vary with an average of 7 base-pairs frayed. The helix is not as stable at the end of a molecule because the last base-pair is not surrounded by other base-pairs. Even within the helix, each base-pair is not hydrogen-bonded at all times; momentary 'breathing' of DNA occurs.

Furthermore, the double-helix just described is not the only possible structure. This form, called B-DNA, is the major form of double-stranded DNA found in cells. Another form, the A-DNA form, occurs when DNA is crystallized and dehydrated. It is also a right-handed double-helix, but it has 11 base-pairs per turn and the bases are 20° from perpendicular. It does not appear to occur in cells. **(PP 16-18)** Another possible form is Z-DNA which is a left-handed helix with 12 base-pairs per turn and a zig-zag shape. DNA with alternating purines and pyrimidines can form Z-DNA, and short stretches may be found in cells. Their function is unknown but they may regulate gene expression.

Other more unusual secondary structures can also be found. When a segment of DNA contains a palindrome or inverted repeat, other structures are possible. **(PP 19)**

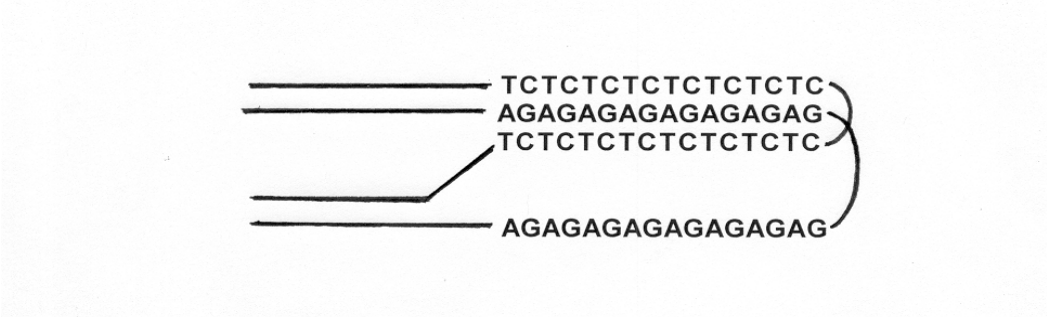
———TTAGCACGTGCTAA ———
———AATCGTGCACGATT ———

↓



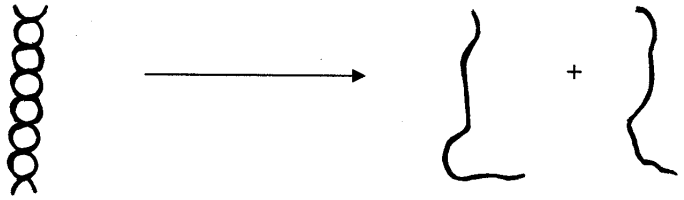
This double-stranded structure is called a cruciform. A single strand with this type of structure is called a hairpin. **(PP 20-21)** Such structures can have non-paired bases in the middle. These sequences tend to appear in regulatory regions of DNA. How many palindromes actually form cruciforms in cells is not known.

H-DNA contains 3 strands and occurs when one strand contains only purines or only pyrimidines in a long stretch. **(PP 22-26)**

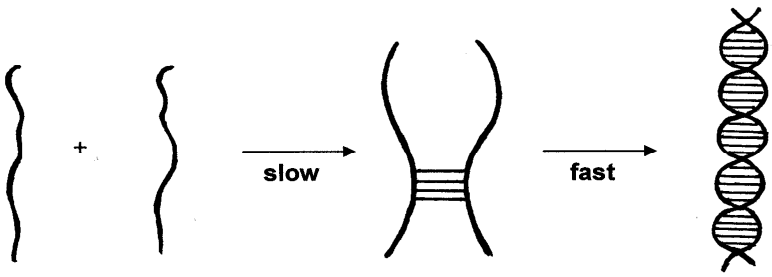


D. Denaturation

Duplex DNA can be denatured (or melted) if the forces between the bases are disrupted by heat, pH, or chemicals. No covalent bonds are broken. Distilled water also denatures DNA because negative charges on phosphate groups are not shielded by positive ions and repulsion occurs. **(PP 27)**



Denatured DNA can be renatured if the denaturing agent is removed slowly. Renaturation is a two-step process. The first slow step requires that two strands find an area of complementarity. Once these regions are H-bonded, the second fast step occurs where the rest of the molecule zips up because complementary bases are now aligned.



The ease with which a DNA molecule denatures depends upon the base content. Since G:C base-pairs have three H-bonds, they are more stable than A:T base-pairs with two H-bonds. Thus DNAs with high levels of G:C are more stable and more difficult to denature.

One DNA strand and one complementary RNA strand can also associate to form an RNA-DNA hybrid and the process is then called hybridization. **(PP 28)** RNA-DNA hybrids, as well as double-stranded RNA molecules, tend to assume an A-like conformation because the 2'-OH groups fit better in this structure. Even two nucleic acid strands that are not entirely complementary can partially renature or hybridize. The extent of hybridization depends on the extent of complementarity. This technique can be used to determine the relationship of different species, isolate a gene, or detect a certain sequence in a DNA molecule. **(PP 29-30)**

III. DNA as the Genetic Material

A. Evidence for DNA

In the early 1900's, proteins were better studied than nucleic acids, and it was thought that proteins would turn out to be genetic material, providing the needed diversity. Three different experiments showed rather that DNA was the genetic material.

1. Griffith - 1928, Avery - 1944

Griffith was working with the bacterium Streptococcus pneumoniae. A virulent type (S-type for smooth encapsulated colonies) would kill mice whereas a non-virulent strain (R-type for rough nonencapsulated colonies) would not. **(PP 31-32)** If a heat-killed extract of the virulent strain was mixed with live R-type bacteria and injected into a mouse, the mouse died and live S-type bacteria could be isolated. **(PP 33-34)** Thus something in the dead S-type extract transformed the non-virulent bacteria into the virulent strain (changed its genetics). Avery later showed that the transforming factor in the extract was not destroyed by proteases (so it was not protein) but was destroyed by deoxyribonucleases (so it was DNA). **(PP 35)**

2. Hershey and Chase - 1952

Using virus T-2 which infects E. coli bacteria, they radioactively labeled the viral protein coat with ^{35}S and the viral DNA with ^{32}P. **(PP 36)** When the virus infected the cell, the viral parts remaining outside the E. coli cell were sheared away in a blender (non-genetic parts) while the parts of the virus entering the cell (genetic material) were detected by the radioactive label. **(PP 37)** Only the ^{32}P-DNA entered the cell, not the ^{35}S-protein. **(PP 38)**

3. Chargaff - late 1940's

DNA has all the properties expected of the genetic material, according to Chargaff's work. **(PP 39-41)**

a. DNA from different tissues or different individuals of the same species has the same base composition.

b. DNA from different species has different base compositions. The closer the species, the more similar are the base compositions.

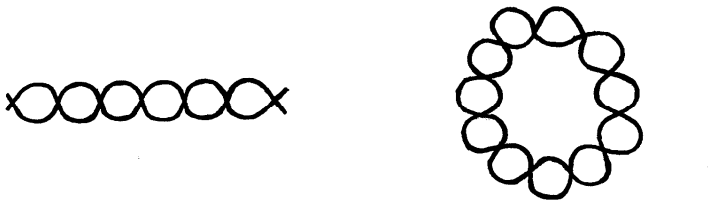
c. DNA does not vary with age, nutritional state or changing environment. Proteins do vary with these factors.

B. Genetic Material of Cells

Isolating intact DNA molecules can be very difficult due to their large size which makes them very fragile. However, when isolated intact, each species has a specific-sized DNA.

1. Viruses

Viruses vary enormously in their genetic material. **(PP 42)** Some have DNA (some single-stranded, some double-stranded). **(PP 43)** Others use RNA as the genetic material (single-stranded or double-stranded). The genetic material is usually contained in one nucleic acid molecule that can be linear or circular, depending upon the virus, that is packaged within the viral coat.



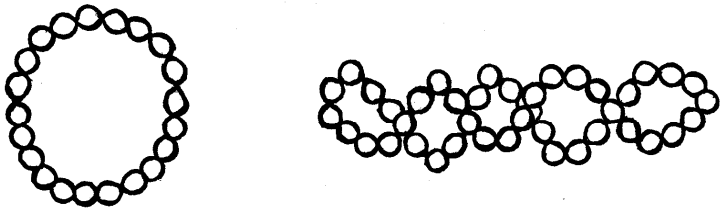
linear circular

Viruses have small chromosomes since they use many proteins of their host cell and so require a limited number of genes of their own. The chromosomes range in size from 5000-200,000 base-pairs with molecular weights of 3-100 million.

2. Bacteria - prokaryotes

E. coli DNA is typical of bacterial DNA. **(PP 44)** It has one circular double-stranded DNA molecule with 4.6 million base-pairs and a molecular weight of 2.9 billion. When extended, the DNA has a length of 1.6 mm, 850 times the length of the E. coli cell. **(PP 45)** The DNA must be compacted in order to fit in the cell. The DNA is held in the nuclear zone (bacteria have no nucleus) complexed with proteins and perhaps attached to the cell membrane at several spots. The proteins probably form a scaffold

which organizes the DNA into loops to help keep the DNA compact. Supercoiling of the DNA is also crucial to compacting. The double-helix is twisted to form a supertwisted or supercoiled molecule as opposed to a relaxed one. **(PP 46-48)**



relaxed supercoiled

Enzymes are needed to produce supercoiling since permanent supercoiling involves breaking and resealing the DNA strands. All the details of supercoiling and compacting are not understood. **(PP 49-50)**

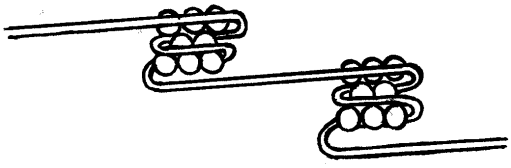
Bacteria also contain smaller, circular DNA molecules called plasmids that contain 10^3-10^5 base-pairs and can carry some genetic information. Plasmids can be picked up or lost by the cell.

3. Higher organisms-eukaryotes

Eukaryotic cells are more complex and contain more DNA than prokaryotes. Human cells have 600 times more DNA than E coli. The length of DNA in a single human cell is about two meters. This is subdivided into molecules (chromosomes), numbering 46 in humans. Each chromosome is a duplex, linear DNA molecule. **(PP 51)** The chromosomes are found inside a nucleus.

Chromosomes appear as well-defined bodies only during cell division. At other times, the chromosomes appear diffused. This is called chromatin, which consists of roughly equal amounts of protein and DNA, along with a small amount of RNA. The proteins come in two types, histone and non-histone. Histones are important because they function in compacting the DNA. There are five classes of histones. All are small (molecular weight = 11,000 to 21,000) and basic (rich in Arg and Lys), so they are positively charged to interact with negatively charged DNA. They are 80% α-helix, so they can easily interact with the double-helix. **(PP 52)**

Histones (8 molecules) form a core about which the DNA is wound, forming a nucleosome (the first level of DNA compacting). About 140 base-pairs are wound around the core. The bases between nucleosomes vary from 20-120 nucleotide pairs, depending upon the species (called linker DNA). The overall appearance is like beads on a string. **(PP 53)**



This achieves a degree of compaction of 6-7 fold, but this is not sufficient. Additional levels of compaction occur. Nucleosomes are arranged in a regular helical array called the 30 nm fiber (or solenoid) with six nucleosomes per turn, stabilized by one type of histone as well as non-histone proteins. **(PP 54)** This compacts the DNA a total of 40 fold. The 30 nm fiber appears to be arranged into loops with the help of a nuclear protein scaffold (overall compaction of 680 fold), and 18 loops then form a miniband (overall compaction of 12,000 fold). Many stacked minibands form a chromosome. However, the details of these complicated levels of structure are not completely understood. **(PP 55)**

Eukaryotic cells also contain DNA in mitochondria and chloroplasts. These molecules are relatively small, circular, and without histones. They may have originally been ancient bacteria that invaded other cells and evolved. **(PP 56-57)**

Chapter 11: DNA Replication

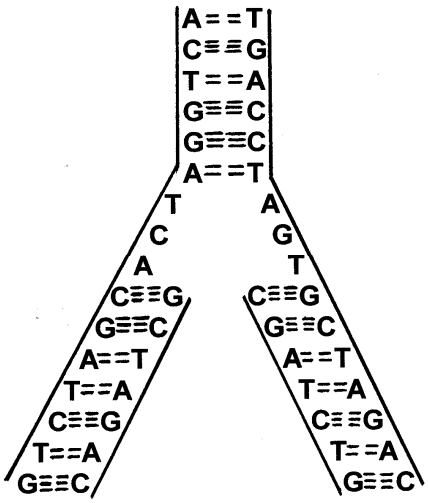
I. Introduction

Since DNA is the genetic material of a cell, when a cell divides it must provide each of the two daughter cells with a complete copy of the DNA. Since the parent cell has a complete copy, the DNA must somehow be duplicated prior to cell division. This is the process of DNA replication. Several things must be true of DNA replication if the two daughter cells are to get functional copies of the DNA. First, all the DNA must replicate to provide complete copies. Second, replication must be accurate: that is, the base sequence of the DNA must be faithfully duplicated. Third, the complex structure of DNA (double-helix, supercoiling, nucleosomes) must be maintained or at least restored after replication.

II. Basic Features

A. Base-pairing and templates

With the discovery of the double-helix, Watson and Crick proposed a logical mechanism for replication. Since base-pairing is dictated by the double-helix, each strand contains the information for making the other strand. Thus if the two strands of a double-helix were separated, each strand could act as a template for synthesis of a new, complementary strand. Since only certain base-pairs can form, one single-strand dictates the nucleotides that are used to fill in the second strand. Anywhere there is a T in a parental strand, an A will hydrogen bond. Anywhere there is a C, a G will bond. The new nucleotides are then joined together by phosphodiester bonds to form a new DNA strand. The result will be two DNA molecules with the same base sequences. This mechanism is not only logical, but provides a way for DNA replication to be accurate with regard to base sequence. **(PP 2)**

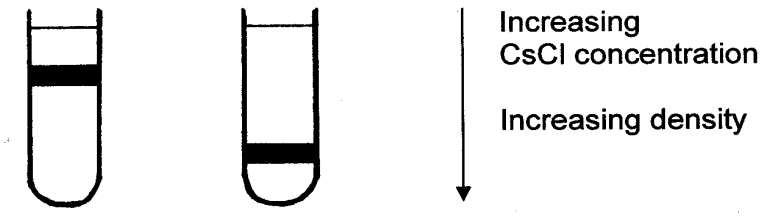


B. Semi-conservative replication

If Watson and Crick's idea was correct then replication would be what is termed semi-conservative. Each of the two resulting DNA molecules would consist of one old parental strand and one newly synthesized DNA strand. **(PP 3)** An alternative would be conservative replication, where the two old strands ended up together in one molecule and two new strands formed the other DNA molecule. A third possibility would be dispersive replication, in which both resulting molecules were composed of strands where old DNA and new DNA were mixed within each strand. **(PP 4)**

To distinguish among these possibilities, Meselson and Stahl in 1957 grew E. coli for several generations in medium where the nitrogen source was labeled with ^{15}N (heavy isotope). DNA labeled with ^{15}N has a density 1% greater than that of normal DNA. The E. coli cells were then switched to medium containing normal ^{14}N (light isotope). Thus all old DNA would be heavy, while newly made DNA would be light. After one generation (one DNA doubling) in light medium and after two generations, the DNA was isolated and analyzed.

The density of DNA can be measured using a technique known as CsCl density gradient centrifugation. A CsCl solution is placed in a centrifuge tube. When centrifuged at high speed, the CsCl forms a concentration gradient within the tube, with more concentrated CsCl near the bottom. Since the density of the solution varies with concentration, the solution is less dense at the top and gradually gets more dense toward the bottom. If DNA is included in the solution, it will migrate until it reaches its density. Thus heavy DNA will form a band in the centrifuge tube at a lower position than light DNA. **(PP 5)**

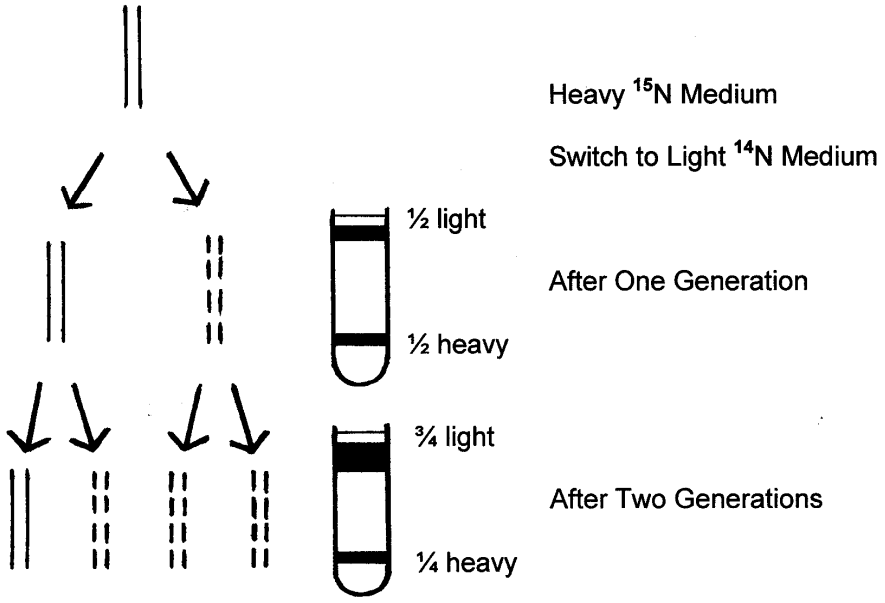


Light ^{14}N DNA Heavy ^{15}N DNA

Different results are expected for semi-conservative replication as compared to either conservative replication or dispersive replication. **(PP 6-7)**

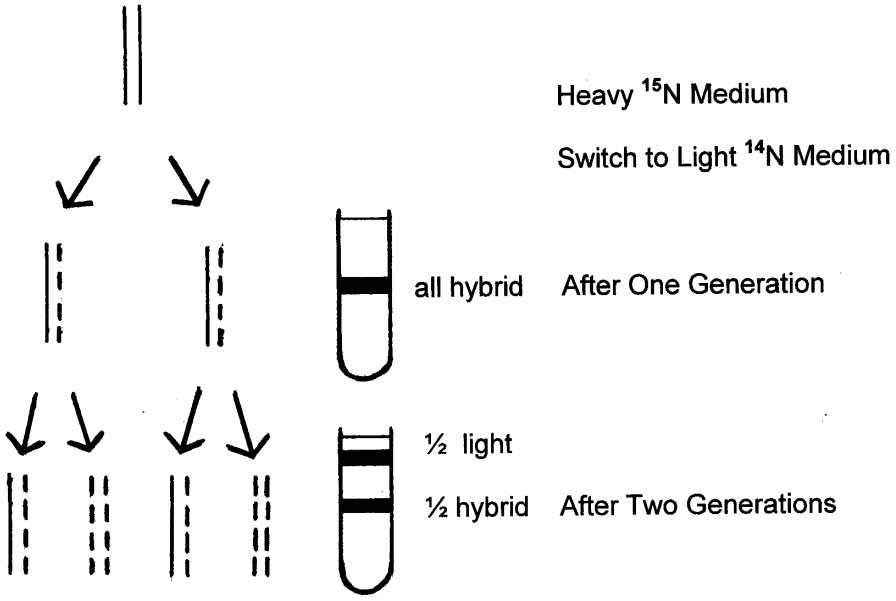
With conservative replication, analysis of the double-stranded DNA after one generation would show that half of it is heavy and half is light, and two corresponding bands would be seen in CsCl gradients. After two generations the DNA would be ¼ heavy and ¾ light, again producing two bands.

Conservative



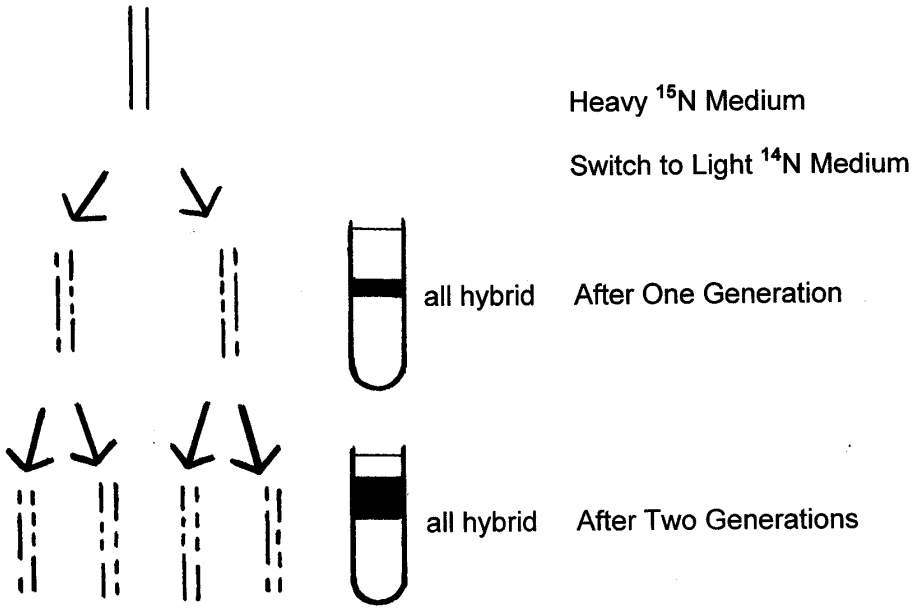
With semi-conservative replication, the DNA after one generation would be all hybrid consisting of one heavy strand and one light strand, and one band with intermediate density would be seen in CsCl gradients. After two generations, the DNA would be ½ hybrid and ½ light, resulting in two bands.

Semi-Conservative



With dispersive replication, the mixture of light and heavy DNA within all the strands would produce all hybrid DNA after one generation, and one band would be seen in CsCl gradients. After two generations, there would still be a mixture of light and heavy DNA within the strands, and one broad band with intermediate density would still be seen.

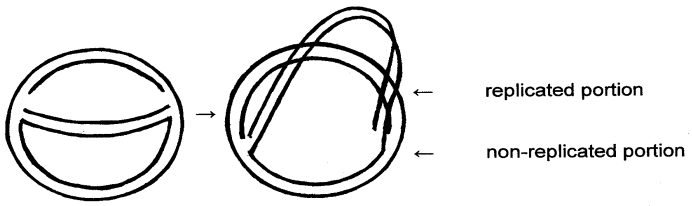
Dispersive



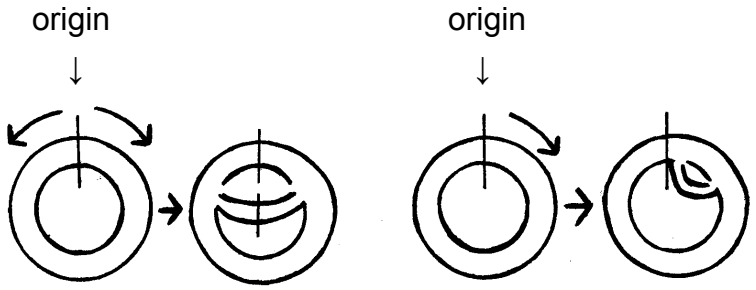
Thus different results are predicted for the different possible mechanisms of DNA replication. The actual experimental analysis showed results that indicate semi-conservative replication. **(PP 8-9)** This is true in all organisms and confirmed the Watson-Crick hypothesis. **(PP 10-13)**

C. Replicating DNA molecules

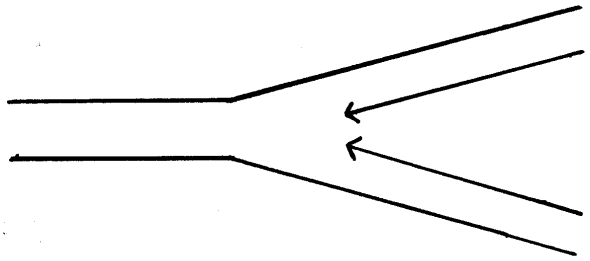
E. coli with its single circular chromosome was extensively studied with regard to DNA replication. The replicating molecule was examined at different stages of replication. This can be done by radioactively labeling the DNA (allow bacteria to replicate in medium containing radioactive precursors of DNA), isolating DNA, and spreading it on a photographic plate. This then forms an image of the DNA. This technique is called autoradiography. The results showed replicating molecules known as θ-structures. **(PP 14)**



Similar studies revealed two additional facts. First, replication always begins at the same place on the E coli chromosome. This is a specific site, designated oriC, with a specific sequence of 245 base-pairs which is always recognized as the starting place for replication. Second, DNA replication is bidirectional, meaning that replication proceeds in both directions around the circular chromosome until the new strands meet 180° from the origin. The alternative is unidirectional replication where replication proceeds in only one direction, moving 360° around the circle and back to the origin. **(PP 15)**



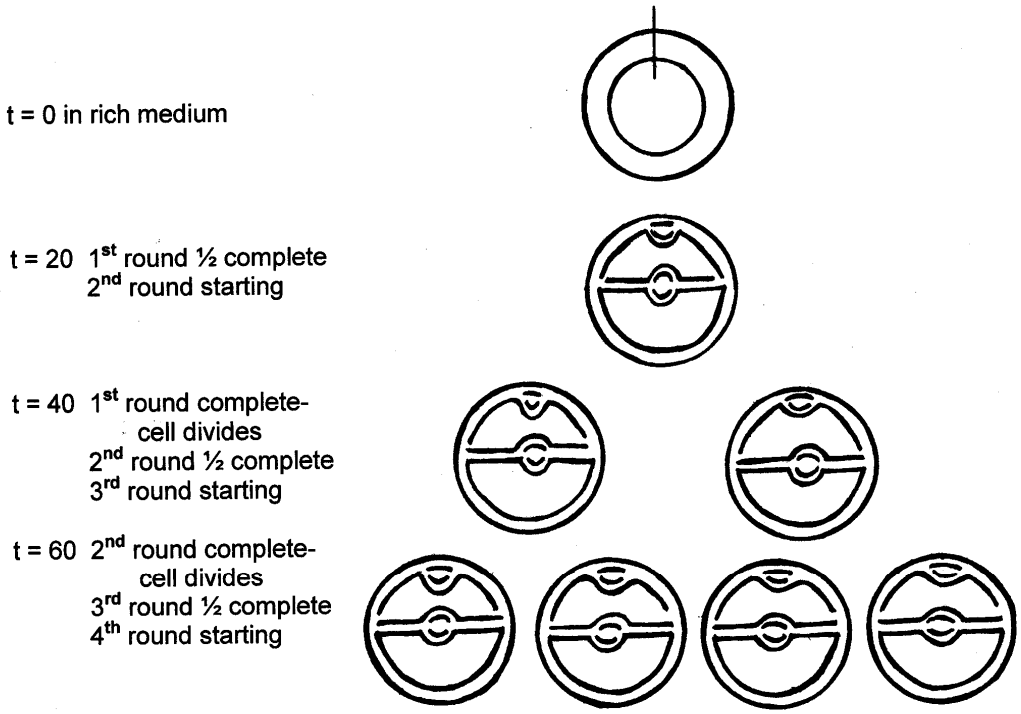
All replication, except in some plasmids and viruses, is bidirectional rather than unidirectional. Each intersection where replication is actively occurring is a replication fork. **(PP 16)** With bidirectional replication, there are two such forks in a θ-structure.



← replication/ fork movement

The rate of DNA replication is very fast in E. coli, about 45,000 nucleotides added per fork per minute at 37°C. This rate will vary with temperature, but at 37° is constant. However, E. coli at 37°C will have different generation times depending upon the nutritional state of the medium. In rich medium bacteria divide every 20-30 minutes, while in minimal medium the generation time is 2-3 hours. However, DNA replication always takes 40 minutes at 37°C. Bacteria cannot change the rate of replication, but they can control how often they initiate DNA replication. However, once replication is started, it continues at a fixed rate and goes to completion. Thus in minimal medium,

the bacteria spend 40 minutes replicating DNA and the rest of the 2-3 hours growing and accumulating enough energy. In rich medium, a generation time of 20 minutes is maintained by having two or more rounds of replication going on at the same time, each of which takes 40 minutes to complete. **(PP 17-18)**



A round of replication is completed every 20 minutes, allowing the cell to divide, and the two daughter DNA molecules are already undergoing subsequent rounds of DNA replication. Controlling replication through the frequency of initiation (rather than through varying the rate of polymerization) means that once the cell starts replication it is committed to finishing replication. This is efficient since replication must be completed to be beneficial to the cell. **(PP 19-20)**

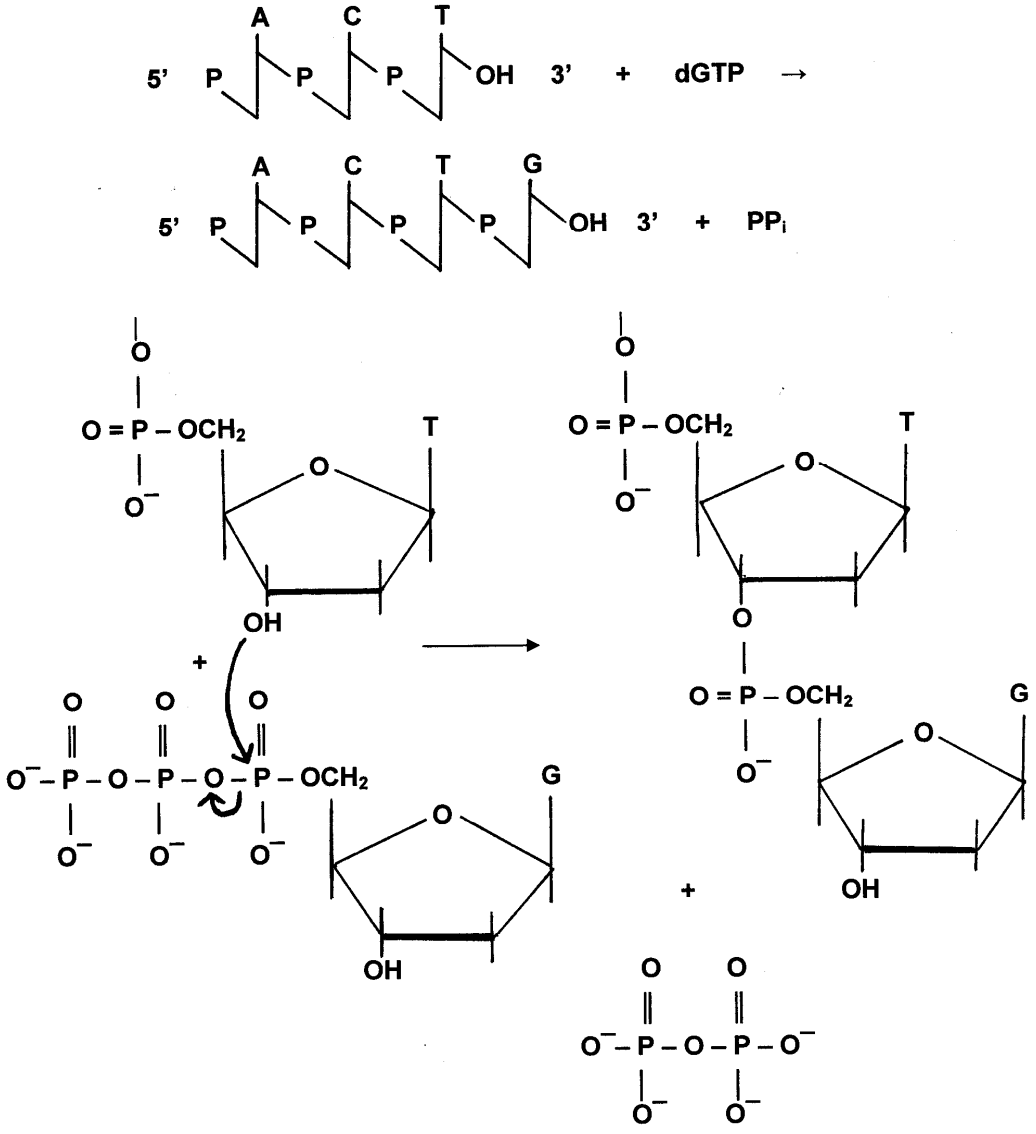
D. DNA Polymerization

The first enzyme found in E. coli to polymerize DNA is called DNA polymerase I. It is a single chain with a molecular weight of 103,000. The enzyme uses dNTPs in the following reaction.

$$\underset{\text{DNA}}{(\text{dNMP})_n} + \text{dNTP} \rightarrow \underset{\substack{\text{DNA} \\ \text{(longer by one} \\ \text{nucleotide)}}}{(\text{dNMP})_{n+1}} + \underset{\downarrow \\ 2\,P_i}{PP_i}$$

The breaking of the pyrophosphate bonds of the dNTP and PP_i provides energy for making the phosphodiester bonds which explains why the enzyme will not use dNDPs or dNMPs. It is also specific for DNA; ribonucleotides are not used.

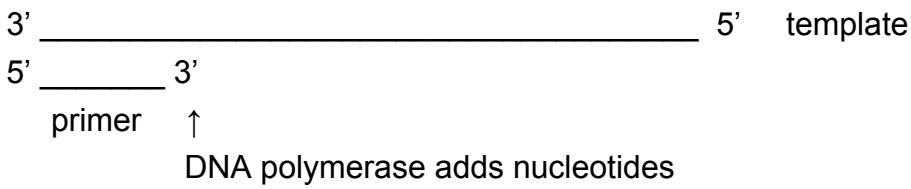
The enzyme specifically joins the new nucleotide to the 3'-end of the existing DNA molecule. **(PP 21-23)**



The DNA grows from the 5'-end toward the 3'-end (5'→3' polymerase).

Two other features of DNA polymerase are important. First, the enzyme requires a template, an exposed single-strand of DNA to which incoming nucleotides hydrogen-bond and so are correctly selected. Without a template, DNA polymerase I will not just join nucleotides randomly. This is consistent with the prediction that if DNA replication is to be accurate, base-pairing must guide the selection of each nucleotide to be

polymerized. Second, the enzyme requires a primer. The enzyme cannot start a new chain, but can only add nucleotides onto an existing chain. It can elongate, but it cannot initiate. **(PP 24-26)**



Some properties of DNA polymerase I (also called pol I) are not consistent with a replication enzyme. For instance, pol I adds nucleotides at a rate of 600 nucleotides/minute, which is too slow to replicate the chromosome in 40 minutes. Its processivity (the average number of nucleotides it polymerizes before it dissociates from the template) is 20, which is lower than expected for replication. In addition, mutants of polymerase I can replicate DNA. Polymerase I, it turns out, functions mainly in repair of DNA and to a limited extent in replication. Four other polymerases exist in E. coli. **(PP 27)** DNA polymerase II has multiple subunits (probably 7) and a molecular weight of about 90,000. It also has a relatively low polymerization rate and processivity, and functions in DNA repair. Polymerase III holoenzyme has at least ten subunits and a molecular weight of about 800,000. **(PP 28-29)** It polymerizes 30,000 - 50,000 nucleotides/minute, and is responsible for the vast majority of DNA replication. Three of its subunits (αε θ) form the core enzyme that has polymerase activity. The remaining subunits allow the enzyme to clamp onto the DNA template and result in a very high processivity of more than 500,000. **(PP 30-31)**

DNA polymerases I, II, and III all have the same basic properties, including using dNTPs, requiring a template and a primer, and polymerizing 5'→3' to accurately copy a DNA sequence. Polymerases IV and V function in DNA repair and are less accurate. **(PP 32-33)**

III. Replication Process

The replication of the chromosome is an enormously complicated process involving many proteins besides DNA polymerase III. **(PP 34)** The process can be divided into stages.

A. Initiation

The replication origin of E. coli (oriC) is 245 base-pairs. It contains three repeated sequences (13 bp) of one type and four or five repeated sequences (9 bp) of a second type. **(PP 35)** About 8 copies of a specific replication protein, called DnaA, bind to the

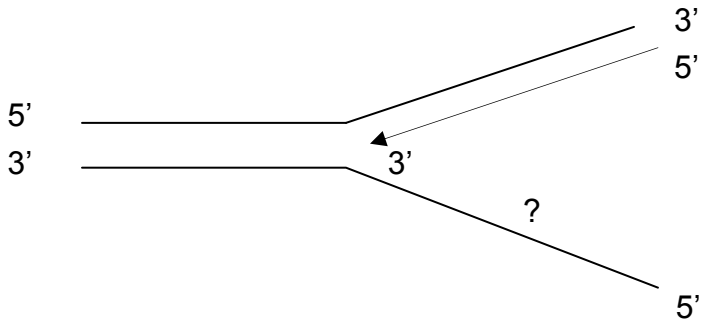
9-bp sequences and the DNA wraps around the protein, forming a nucleosome-like structure. This strains the A-T rich 13-bp sequences (called DUE-DNA unwinding element) and causes them to denature. **(PP 36)** Protein HU, FIS, and IHF are also needed, along with ATP. Then DnaC protein helps DnaB protein bind to the denatured region, and DnaB further unwinds the DNA in both directions. Primers are then synthesized so DNA can be replicated. **(PP 37)**

Additional proteins may also be involved. Control of the initiation process is not well-understood, but may involve interaction of the origin with the cell membrane and methylation of DNA within specific base sequences. **(PP 38-39)**

B. Elongation

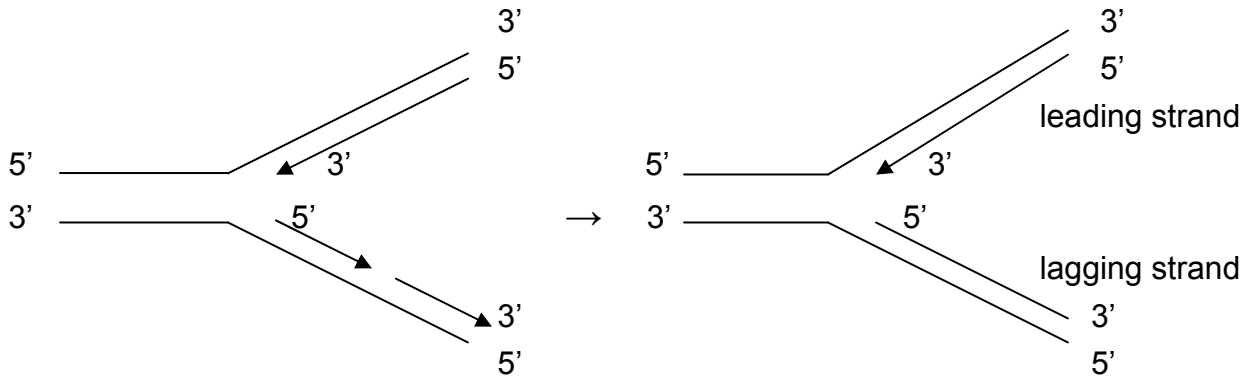
Elongation is more complicated than expected because of the properties of DNA polymerase III. First, DNA polymerase III cannot initiate a new DNA chain, so there is the question of how replication starts. DNA chains are initiated on short primers of RNA. Enzymes which polymerize RNA are similar to DNA polymerases, but RNA polymerases do not require a primer and so can initiate new chains.

An additional problem is caused by the directionality of DNA polymerase III in that it works only 5'⟶3'. However, in a replication fork, new DNA strands must be synthesized both 5'⟶3' and 3'⟶5'.



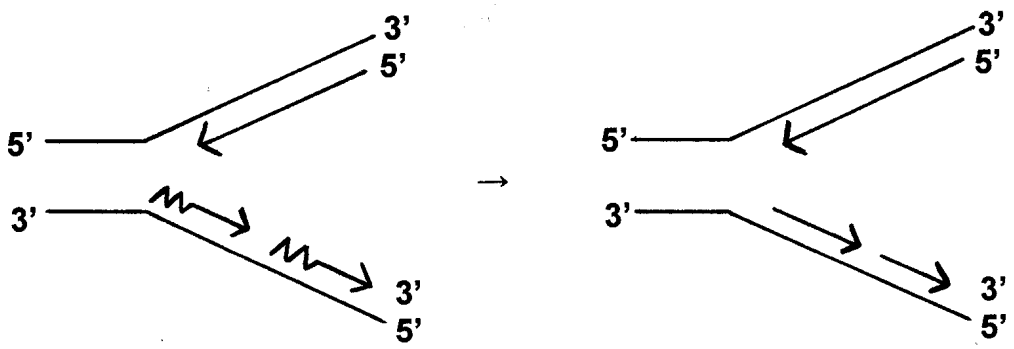
The two template DNA strands are antiparallel to each other, and each new DNA strand must also be antiparallel to its template. This means one new strand must be synthesized in the 5'⟶3' direction, which is what polymerases can do. The other strand, however, must be synthesized 3'⟶5', which no polymerases can do. The question is how is this new strand made. **(PP 40)**

Okazaki proposed that this 3'⟶5' strand was made discontinuously, as a series of short DNA pieces called Okazaki pieces. Each piece is synthesized 5'⟶3' and then joined together to give the appearance of a strand growing 3'⟶5'. **(PP 41)**



The strand that is made continuously is called the leading strand. The strand made discontinuously is called the lagging strand. Experiments demonstrated the presence of Okazaki pieces, not only in E. coli but in all cells.

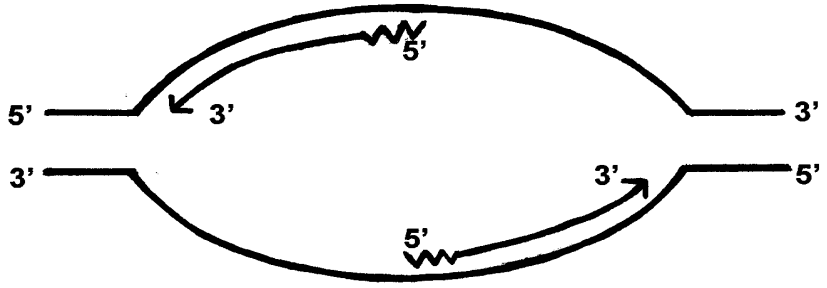
Okazaki's explanation solved one problem, but created another. Since DNA polymerases cannot initiate new chains, the problem now became to explain how the numerous Okazaki pieces are started. Okazaki proposed that each Okazaki piece started using an RNA primer (/\/\/\). **(PP 42)** Then, later during replication, the RNA primers are removed and filled in with DNA.



1. Leading strand synthesis

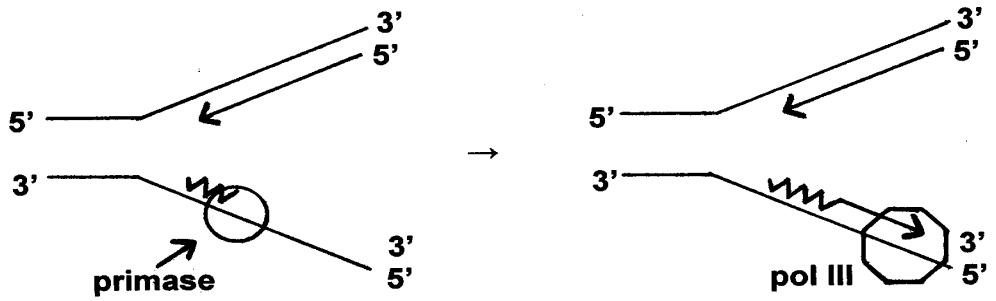
Putting all these features together, the following mechanism exists for leading strand synthesis. Once initiation has denatured the DNA at the origin, a short RNA primer (10-20 nucleotides) is synthesized at the origin by an RNA polymerase known as primase (DnaG protein). The cell's main RNA polymerase (also called RNA polymerase) may be involved as well. Primase binds to the DnaB protein to form a primosome complex, and other proteins may also be needed for the primosome. Once the primer is formed, DNA polymerase III extends this continuously by adding nucleotides, creating a long DNA chain that extends around the chromosome. There

are two leading strands that must be established at the origin, one in each of the two replication forks that create bi-directional replication. **(PP 43-44)**

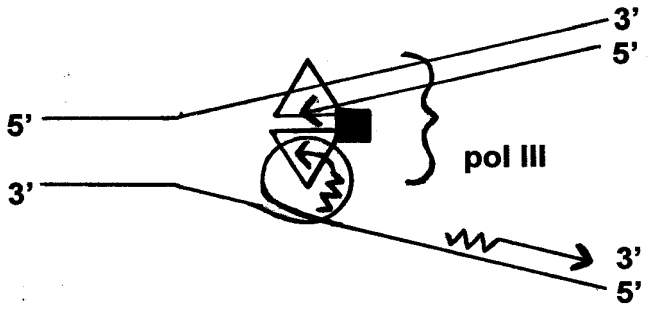


2. Lagging strand synthesis

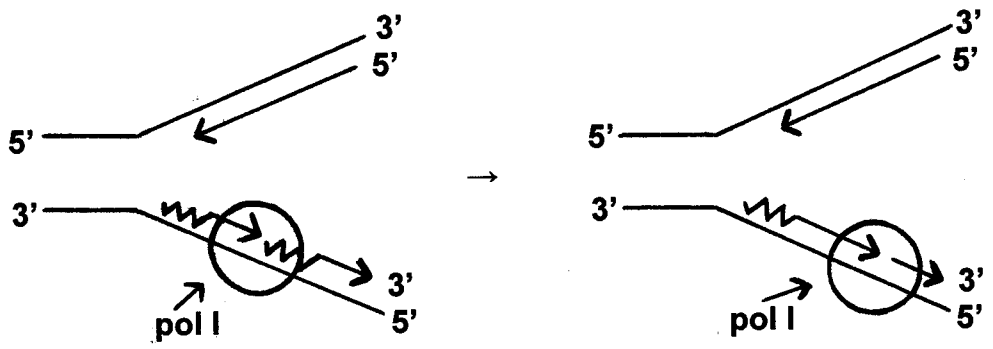
The primosome moves along the template with the replication fork. At various intervals, the primase enzyme makes a short RNA primer (10-20 nucleotides). DNA polymerase III then uses the primer to make an Okazaki piece (1000-2000 nucleotides). **(PP 45-47)**



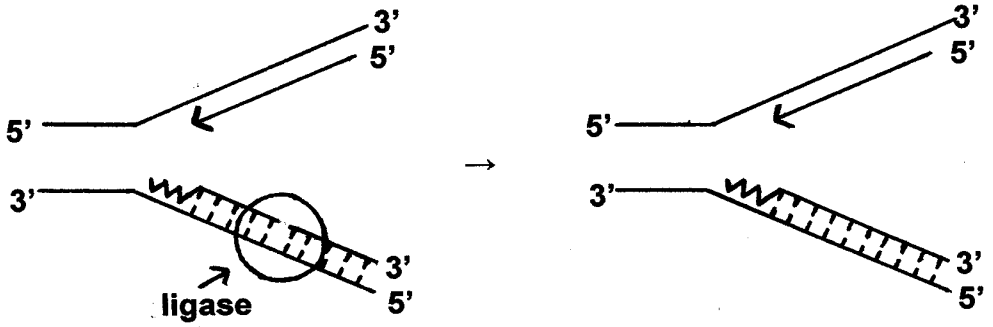
Leading and lagging strand synthesis in a replication fork are coordinated by DNA polymerase III holoenzyme. Two core enzymes (one for the leading strand and one for the lagging strand) are held clamped to the DNA strands by other holoenzyme subunits, and both are connected by additional subunits. Thus both DNA strands are replicated together. The lagging strand is probably looped so that the holoenzyme can work on both template strands at the same time. **(PP 48-51)** In addition, it is possible that pol III and other replication proteins remain attached to the cell membrane and the DNA moves through these replication factories, rather than the proteins moving along the DNA.



Removing the RNA primer is carried out by DNA polymerase I. In addition to its polymerase activity, polymerase I also has a 5'→3' exonuclease activity. Polymerase I attaches to the 3'-end of an Okazaki fragment and begins to add deoxyribonucleotides. When it bumps into the RNA primer of the next Okazaki piece, it removes the RNA primer while filling in with DNA. Thus the RNA primer is replaced with DNA by pol I using both its polymerase and exonuclease activities simultaneously. Polymerase III does not have this exonuclease activity. **(PP 52)**



After the primer is removed, there remains a nick in the DNA backbone, even though all the deoxyribonucleotides are in place and all the bases are paired. The DNA polymerases cannot make the last phosphodiester bond because the 5'-side of the nick has only one phosphate group as a result of removing the RNA primer. A different enzyme, called DNA ligase, seals the nick and makes the last phosphodiester bond. The reaction requires the input of energy from NAD^+ or ATP, depending upon the organism. **(PP 53)**



Following the action of DNA ligase, the Okazaki pieces are completely joined together to create the new lagging strand.

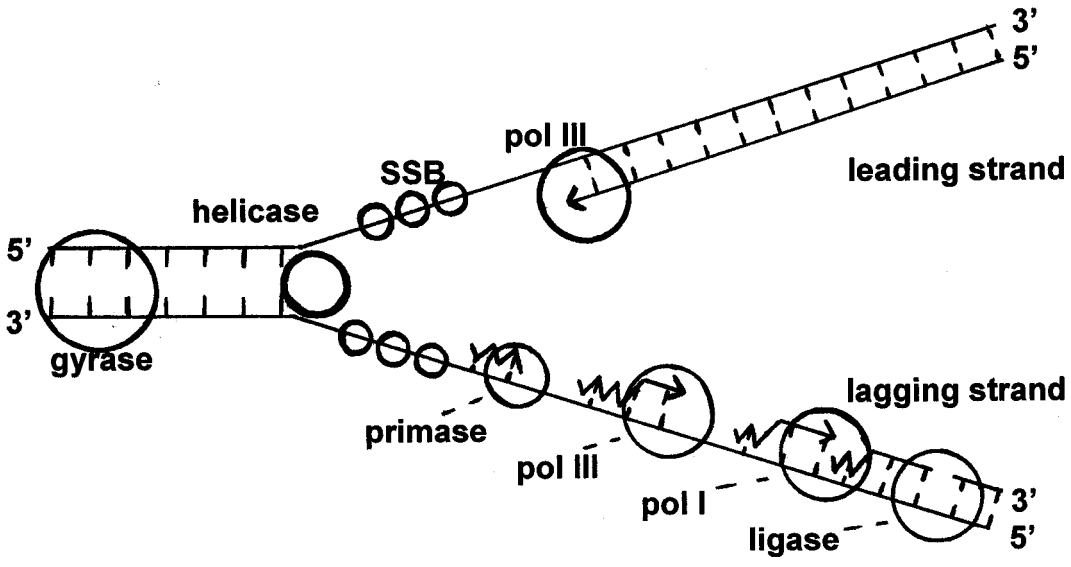
3. Geometry problems

Several other things must happen for DNA to replicate. First, the double-helix must be separated in two single strands in the replicating area and this area of denaturation must move with the replication fork. Enzymes known as helicases, mainly DnaB protein, unwind the DNA just ahead of replication, using the energy of ATP to break the hydrogen bonds and separate the two parental strands. **(PP 54)** Second, the DNA strands, once separated, have to be kept apart. The single-stranded regions are prevented from renaturing by single-stranded DNA binding protein (SSB) that binds cooperatively to single-stranded DNA. **(PP 55)** Third, as the double-helix unwinds in one area of the circular DNA, supercoils build up further along the molecule. The DNA can only twist so much before the tension must be relieved. **(PP 56-57)** Enzymes called topoisomerases, specifically DNA gyrase, can nick the DNA and adjust the supercoiling.

4. Other proteins

A number of other proteins are probably involved in DNA replication but their function is not yet defined. **(PP 58-59)**

5. Summary of elongation **(PP 60-66)**



C. Termination

The final separation of the two completed DNA molecules must be accomplished, but details of the process are not well-understood. The termination region of the DNA (Ter) contains a number of short repeats recognized by the Tus protein, which prevents unwinding of the double-helix and stops the replication forks. **(PP 67)** Topoisomerases then untangle the two completed DNA molecules. **(PP 68)**

IV. Accuracy

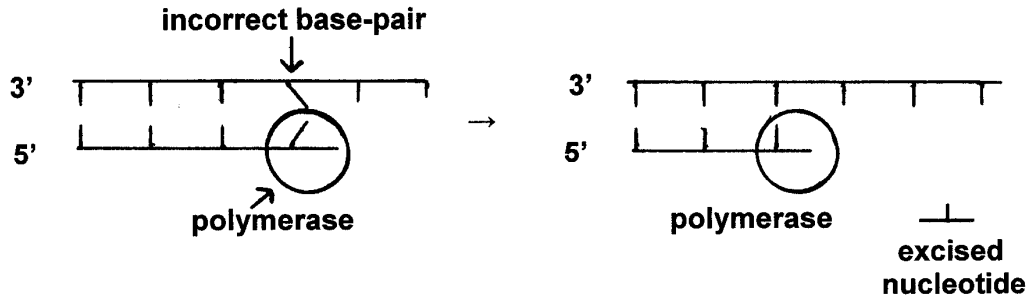
Replication must be very accurate since inserting a wrong nucleotide will change the genetic information (create a mutation) and this is usually harmful to the organism.

Base-pairing (correct H-bonding) is one major way of ensuring accuracy, but other factors also contribute to accuracy.

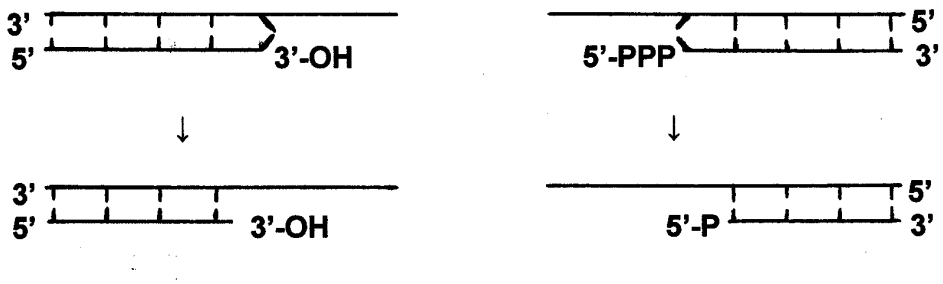
One factor is the requirement of DNA polymerases for a primer. Base-pairing is not very accurate until a stable double-helix already exists, which requires about one turn of the helix. Therefore, if DNA polymerases started a new chain, the first nucleotides laid down would contain many errors in base-pairing. By having a primer that is later removed and filled in accurately, mistakes with those first nucleotides do not become a permanent part of the DNA.

However, even with a primer, base-pairing is calculated to produce one wrong base in every 10^4-10^5 nucleotides due to tautomerization, mismatched base-pairings, etc. This seems like a small error rate, but with size of the E. coli chromosome (4×10^6 bp) this would mean 40-400 errors in every round of replication, which is unacceptable. The actual observed rate of error is one in 10^9-10^{10} nucleotides, which means only one error in every 1000-10,000 replications, which is acceptable.

This increase in accuracy is due to the fact that DNA polymerases have what is known as proof-reading activity. In addition to the 5'→3' polymerase activity, the enzymes also have a 3'→5' exonuclease activity. **(PP 69)** The DNA polymerases not only check for the proper nucleotide before they form the phosphodiester bond, but also check the base-pair again after the phosphodiester bond is formed. **(PP 70-71)** If the hydrogen-bonding is not correct, the polymerase 'backspaces' and removes the incorrect nucleotide and tries again. **(PP 72-74)** This double check system increases the accuracy of DNA polymerases by 10^2-10^3 fold. Mutants whose DNA polymerase III lacks proof-reading ability have much higher rates of mutation than normal cells.



Proof-reading further explains the need for a primer, since DNA polymerases are designed to check a previous base-pair before making the next phosphodiester bond. Thus they cannot put in the first nucleotide to start a new chain. In addition, proof-reading explains the 5'→3' directionality of polymerases. When the polymerase removes a nucleotide, it creates a 3'-OH, which is the same chemical group that is normally present at the reactive end of the chain and is ready for addition of the next incoming 5'-dNTP. If polymerases added onto the 5'-end, proof-reading would leave a 5'-phosphate. The monophosphate would have to be reactivated to a triphosphate before it could form a phosphodiester bond with the 3'-OH group of the incoming nucleotide. This would create an additional complication in the mechanism of replication. **(PP 75)**

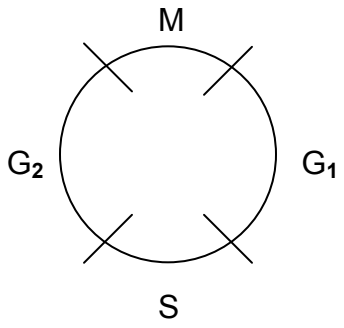


The last factor increasing fidelity is a separate enzyme system called the mismatch repair system. This system checks the new strands of DNA for mismatched bases immediately after replication and corrects any errors. The mismatch repair enzymes can distinguish the new DNA strands from the old parental strands because the old strands contain specific methyl groups on some of the nitrogenous bases, while the new strands have not yet been methylated. The mismatch repair system corrects mistakes in the new, non-methylated strands. **(PP 76)** This brings the overall error rate to 1 in 10^{9}-10^{10} nucleotides. **(PP 77-78)**

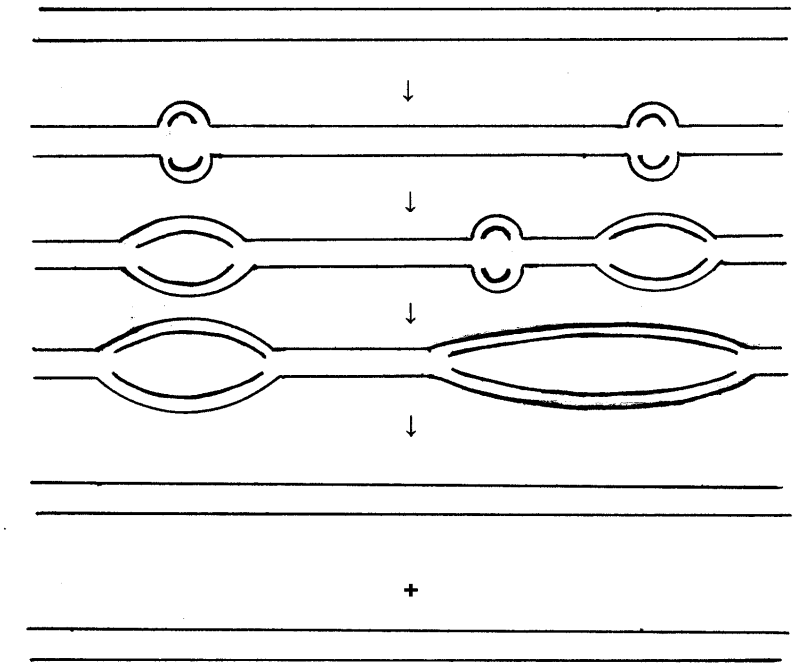
V. Replication in Eukaryotes

Eukaryotic replication has the same basic mechanism as replication in bacteria, but with several added complications. First, there is more DNA to replicate. Second, histones must be made for the new DNA. Third, nucleosomes and other compacting must be disassembled and reassembled as replication goes through a given area.

Unlike prokaryotes, eukaryotic replication takes place at a specific time. Eukaryotic cells have a cell cycle, with a specific series of events leading to mitosis (M). The entire cycle takes 12-72 hours and DNA replication (S phase) takes 1-3 hours, depending on the cell. **(PP 79)**



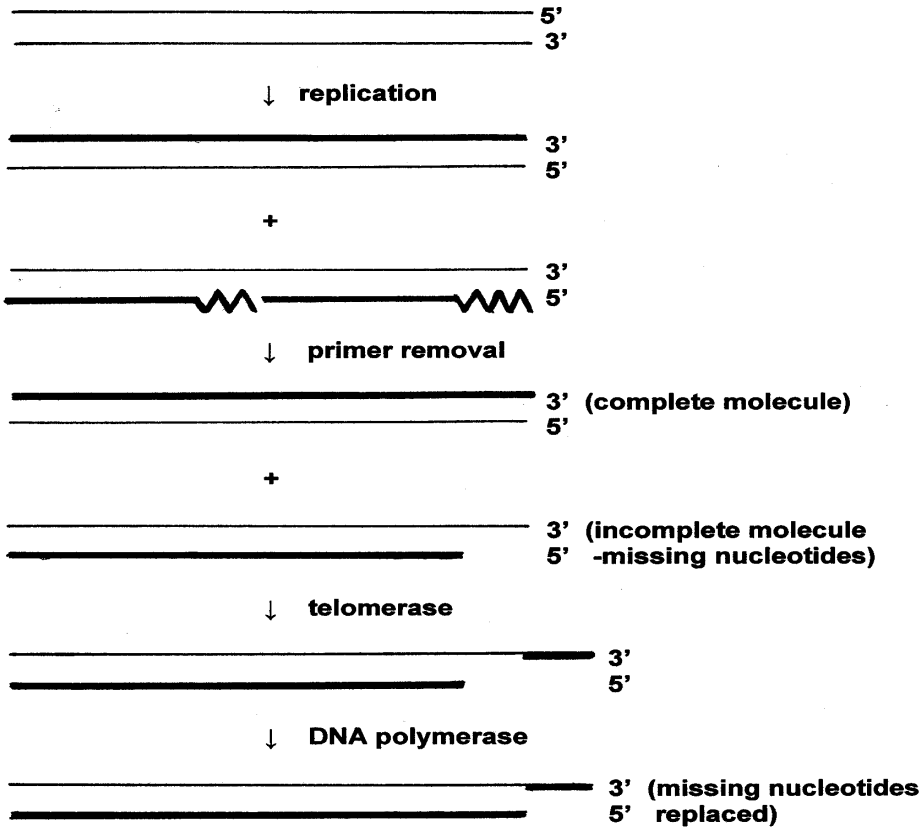
Eukaryotic DNA replication is slower than in prokaryotes, probably because of the nucleosomes. Eukaryotic polymerases polymerize at the rate of 10^3 bp/min per replication fork, considerably slower than in prokaryotes. If eukaryotic DNA had just one origin, it would take 30 days to replicate the DNA at this rate. To compensate, eukaryotes have not one origin but about 6000 origins per chromosome. The initiation sites are irregularly spaced along the chromosome, about 30,000 base-pairs apart. Initiations do not all occur at the same time, but occur throughout S phase. Initiation occurs on RNA primers and replication proceeds bidirectionally from each origin. This forms a series of replication bubbles (replicons) which eventually run into each other. **(PP 80)**



Within each bubble, the mechanism is basically the same as in bacteria. **(PP 81)** Leading strands are made continuously. Lagging strands are made discontinuously using RNA primers (8-12 nucleotides) and Okazaki pieces (50-300 nucleotides). Eukaryotes contain several DNA polymerases which have the same properties as those in E. coli. Other analogous replication proteins including primases and helicases have been found, but many replication proteins have yet to be isolated.

Nucleosomes are briefly unwound during replication, but not over large areas. Histones are made in large quantities during replication and new DNA is rapidly packaged into nucleosomes.

Eukaryotic chromosomes have another issue during replication because they are linear molecules. The ends of the chromosomes, called telomeres, cannot be completely replicated. **(PP 82)** The primer that is formed at the very end of the molecule cannot be replaced with DNA, resulting in the loss of some genetic material during each round of replication. This loss has been implicated in the process of aging. Some cells, including cancer cells, contain an enzyme called telomerase. This enzyme contains RNA which can act as a template to extend and complete the ends of the chromosomes. **(PP 83-86)**



Chapter 12: Transcription

I. Introduction

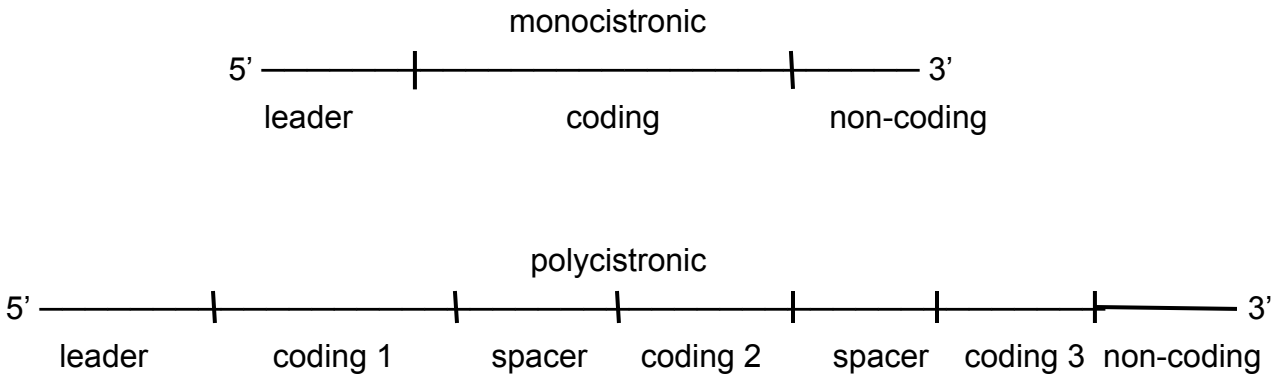
DNA structure and DNA replication explain the storage and transmission of genetic information. However, for a cell to utilize its own genetic information, not only the DNA but RNA is necessary. The overall scheme is as follows: DNA contains genes, each of which is a particular segment of the DNA with a specific base sequence. (There is nothing chemically distinctive about one gene as compared to another except the base sequence.) Every gene contains the information for making a particular protein (a polypeptide chain), and the information is held in the form of the base sequence of the gene. The cell's genes code for all the cell's proteins, both structural and enzymatic. The presence of these proteins determines what the cell looks like and what reactions can occur. The DNA itself, though, does not direct synthesis of the proteins. An intermediate molecule, called messenger RNA, is a go-between for the DNA and the protein. Thus the flow of information is from DNA to mRNA, and then to protein. This is called the central dogma. The process of making mRNA is called transcription. The process of making a protein from mRNA is called translation. **(PP 2-3)**

II. Messenger RNA

The existence of mRNA was first suspected because DNA stays in the nucleus of a cell while proteins are made in the cytoplasm. RNA was a logical candidate for carrying the information since the amount of RNA in a cell increases when protein synthesis increases. mRNAs are single-stranded molecules with the usual structure of a nucleic acid, containing ribonucleotides joined by phosphodiester bonds and the bases A, C, G, U. **(PP 4)** mRNA molecules vary in length from several hundred to over 10,000 nucleotides. A cell will normally have thousands of different mRNA types, each corresponding to a particular gene and used to make a particular protein.

mRNA comes in two basic types. If the mRNA molecule contains the information for making one polypeptide chain, it is called monocistronic. If it codes for two or more polypeptide chains, it is said to be polycistronic. **(PP 5)** Eukaryotic mRNAs are usually monocistronic while prokaryotic mRNAs are usually polycistronic. In addition, mRNA molecules contain regions that do not carry information for making the protein, called non-coding regions. The 5'-end of a mRNA molecule, called the leader sequence, is

non-coding. The 3'-end is also non-coding. In polycistronic mRNA, non-coding spacer regions are between the coding regions. **(PP 6-7)**



Not all RNA in a cell is mRNA. In fact, most RNA is of two other types, called transfer RNA (tRNA) and ribosomal RNA (rRNA) which function in protein synthesis.

III. RNA Polymerases

A. Reaction

In E. coli, mRNA is made by the enzyme RNA polymerase. It uses one strand of a DNA molecule as a template, picking out complementary nucleotides by base-pairing. Thus each mRNA molecule is complementary to a gene (complementary to one strand, same sequence as the other strand) and this is how the mRNA molecule contains the information for making a protein. Base-pairing is the same as in DNA, except U pairs with A. **(PP 8)**

In many respects, RNA polymerase is similar to the various DNA polymerases. It uses triphosphates as precursors (ribonucleotides) to supply energy for making phosphodiester bonds. It can use single-stranded or double-stranded DNA as a template, but not RNA. Base-pairing determines the nucleotide order and the enzyme makes the phosphodiester bonds. **(PP 9)** The enzyme polymerizes only 5'→3', creating an anti-parallel strand.

RNA polymerase also has some differences compared with DNA polymerases. It does not need a primer (it can initiate chains) and it does not proof-read (it has no 3'→5' exonuclease activity). The result is that transcription is not as accurate as replication, but it does not need to be. During transcription, the error rate is 1 in 10^4-10^5 nucleotides. Unlike DNA, mRNA molecules are not permanent cellular components.

mRNA molecules are routinely broken down after a few minutes (bacteria), hours, or days (eukaryotes). New mRNA molecules are then made. Thus an inaccurate mRNA may result in making some non-functional protein molecules, but the damage will probably not be serious or permanent. In contrast, mistakes in DNA persist and are passed to subsequent generations.

RNA polymerase unwinds the double-helix itself, keeping a small area of the DNA denatured (about 17 base-pairs) so the RNA can be made. **(PP 10)** As in replication, unwinding the helix causes supercoils to build up in other places which must be adjusted by topoisomerases.

Generally the RNA polymerase copies only one strand of the DNA in a given area of the chromosome. The template strand (also called the minus (-) strand, the non-coding strand, or the antisense strand), is copied. The other strand is called the non-template strand, sense strand, plus (+) strand, or coding strand. **(PP 11)** A strand may be (+) in one area of the chromosome and (-) in another. In some viruses, both strands in a given area contain information and are transcribed.

B. Structure

In E. coli, RNA polymerase holoenzyme has a molecular weight of 450,000 and contains six subunits in a $\alpha_2\beta\beta'\omega\sigma$ configuration. **(PP 12)** The σ subunit (of which there are several different types) binds loosely and is required only for starting transcription. It recognizes the beginning of genes where transcription should start. There is no point starting transcription in the middle of a gene since the mRNA will not have all the information for making a protein, so the σ subunit ensures that transcription begins only at the start of a gene. The core enzyme ($\alpha_2\beta\beta'\omega$) carries out the other functions of the enzyme. The β subunit binds NTPs. The β' subunit binds the enzyme to the DNA. Both β and β' subunits contribute to the active site. The α subunits help to assemble the enzyme and interact with some regulatory proteins. **(PP 13-14)**

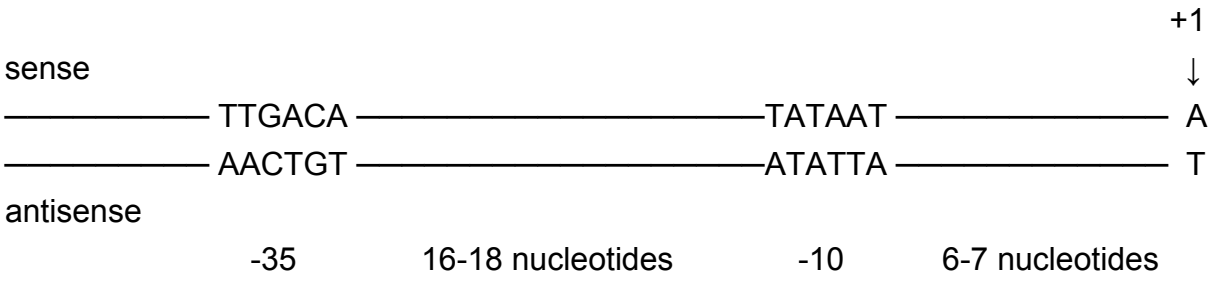
Whereas E. coli has only one RNA polymerase for making all mRNA, rRNA, and tRNA, eukaryotic cells have three kinds of RNA polymerase. RNA polymerase I makes rRNA, RNA polymerase II makes mRNA, and RNA polymerase III makes mainly tRNAs.

IV. Mechanism of Transcription

A. Initiation

Just prior to the gene in the DNA is an area of 50-100 nucleotides called the promoter. The promoter indicates the start of a gene but is not actually transcribed, and lies upstream relative to the start site of transcription. The start of transcription is designated as the +1 nucleotide, and the transcribed nucleotides lying downstream are

given increasing positive numbers. Nucleotides in the promoter are given negative numbers. It is here that RNA polymerase holoenzyme including the σ^{70} subunit (with a molecular weight of 70,000) recognizes specific sequences indicating the start of a gene, so RNA polymerase will bind and start transcribing in the correct position. Not all promoters have the same base sequence. However, two areas in most E. coli promoters are very similar. Centered at about 10 base-pairs before the start of the gene (-10 base-pairs) is a region called the Pribnow box. **(PP 15)** At about -35 nucleotides is another area called the -35 sequence. RNA polymerase (with the σ^{70} subunit bound) recognizes and binds to these areas. **(PP 16)**

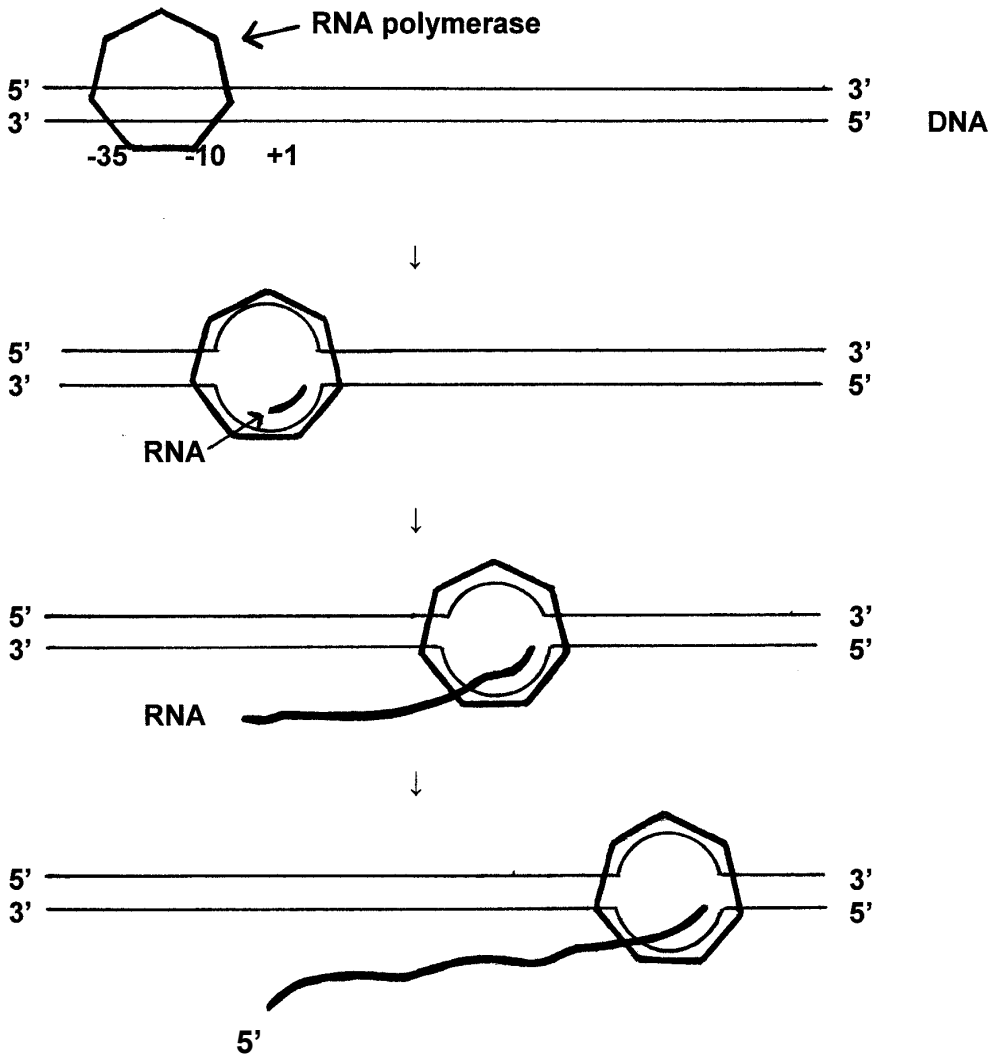


The sequences given are consensus sequences, but variations occur from promoter to promoter. Variations probably account for different rates of transcription for different genes. Some sequences interact better with RNA polymerase so transcription occurs frequently while other sequences bind RNA polymerase less strongly. The difference can be several orders of magnitude in how often transcription is initiated. This is one mechanism for controlling gene expression. Some specialized E. coli genes have different promoters recognized by other types of σ subunits. **(PP 17)**

It appears that RNA polymerase binds to the -35 sequence and the -10 sequence. The enzyme then unwinds 12-17 base-pairs of DNA around the start site of transcription. The A-T rich nature of this region and the σ subunit aid in unwinding. **(PP 18-20)**

B. Elongation

RNA polymerase then begins to make the RNA. The RNA transcript almost always begins with an A or a G. The RNA polymerase moves along the DNA, maintaining a small area of denatured bases. After a few nucleotides (6-10) are polymerized, the σ subunit drops off, and the rest of the enzyme (core enzyme) completes transcription. **(PP 21)** The rate of polymerization is 20-50 nucleotide/sec and varies with the G-C content of the DNA region. Once polymerization of a nucleotide has occurred, it rapidly dissociates from the DNA since the DNA double-helix is more stable and tends to reform. Only about 12 base-pairs of DNA-RNA hybrid exist at any time. **(PP 22)**

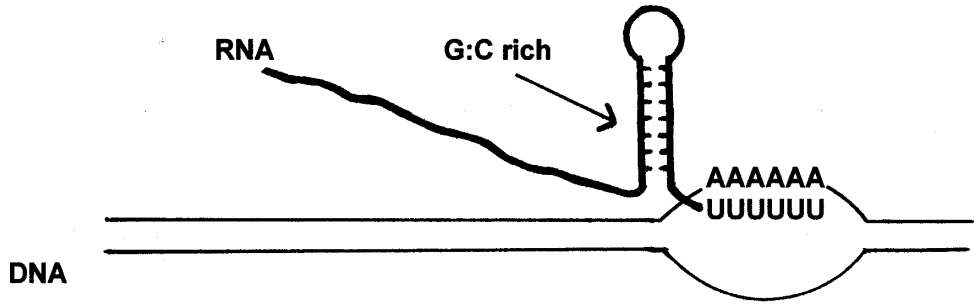


C. Termination

The end of a gene must also be signaled so that RNA polymerase does not go beyond that point. As with the start of the gene, the end is also marked by specific sequences.

In E. coli, termination falls into two main classes. One class depends upon a specific protein called rho (rho-dependent) while the other class does not (rho-independent).

Rho-independent sites have a potential hairpin structure at the termination site, which slows RNA polymerase and disrupts the RNA-DNA hybrid. **(PP 23)** Part of this hairpin is a G-C rich region which also slows down the RNA polymerase because it is difficult to denature. Following the hairpin, there is a stretch of 4-10 adenines in the template which are transcribed into uracils, forming a region of A-U base-pairs which are not very stable and so allow the RNA and enzyme to dissociate from the DNA. **(PP 24)**



Rho-dependent termination sites do not have the A-U area, but they usually do have a G-C rich hairpin region for slowing RNA polymerase. The rho protein has the ability to separate the RNA-DNA hybrid using the energy of ATP, and so acts to end transcription, but the details of this mechanism are not completely understood. **(PP 25-28)**

D. Eukaryotes

The mechanism of transcription is basically the same in eukaryotes, although the promoter sequences are different and vary depending upon which RNA polymerase is used. **(PP 29- 31)** Proteins known as transcription factors are required for initiation. **(PP 32)** Termination signals are not well-understood.

V. RNA Processing - Post-transcriptional Modification

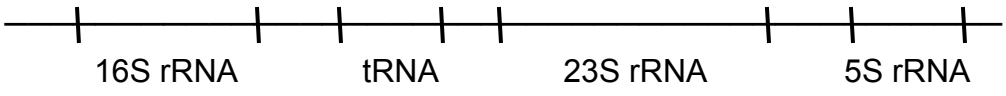
The RNA molecule made by RNA polymerase is not always functional and ready to use. Sometimes it must be modified before it is in its final, functional form. How much modification is needed depends upon the organism and the type of RNA. **(PP 33)**

A. Prokaryotes

1. mRNA does not need any modification.

2. tRNAs and rRNAs are made as longer primary transcripts and then processed into individual RNA molecules.

As an example, the three types of E. coli rRNA are made in one long RNA molecule, along with 1-2 tRNAs. (There are multiple copies of this set of genes.) **(PP 34)**



Certain bases are first methylated, and then specific ribonucleases cut this transcript into its component pieces. The individual pieces are further trimmed at both the 5'-end and the 3'-end. tRNAs undergo additional modifications, including the introduction of unusual bases. **(PP 35-38)**

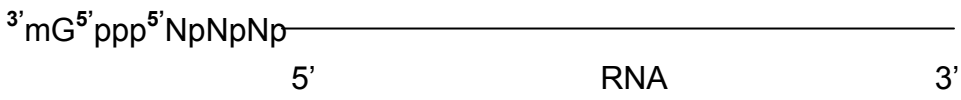
B. Eukaryotes

1. tRNAs and rRNAs are processed as they are in prokaryotes.

2. Unlike prokaryotes, mRNA in eukaryotes is heavily modified. The RNA as it is made in the nucleus is called heterogeneous nuclear RNA (hnRNA) or pre-mRNA. Three things happen to this before it becomes mature mRNA.

a. 5'-cap

The 5'-end of RNA is modified by a structure known as the 5'-cap. A 7-methyl guanosine is joined to the 5'-end via a 5'-5' triphosphate bond. This occurs shortly after synthesis of the RNA begins. **(PP 39)**



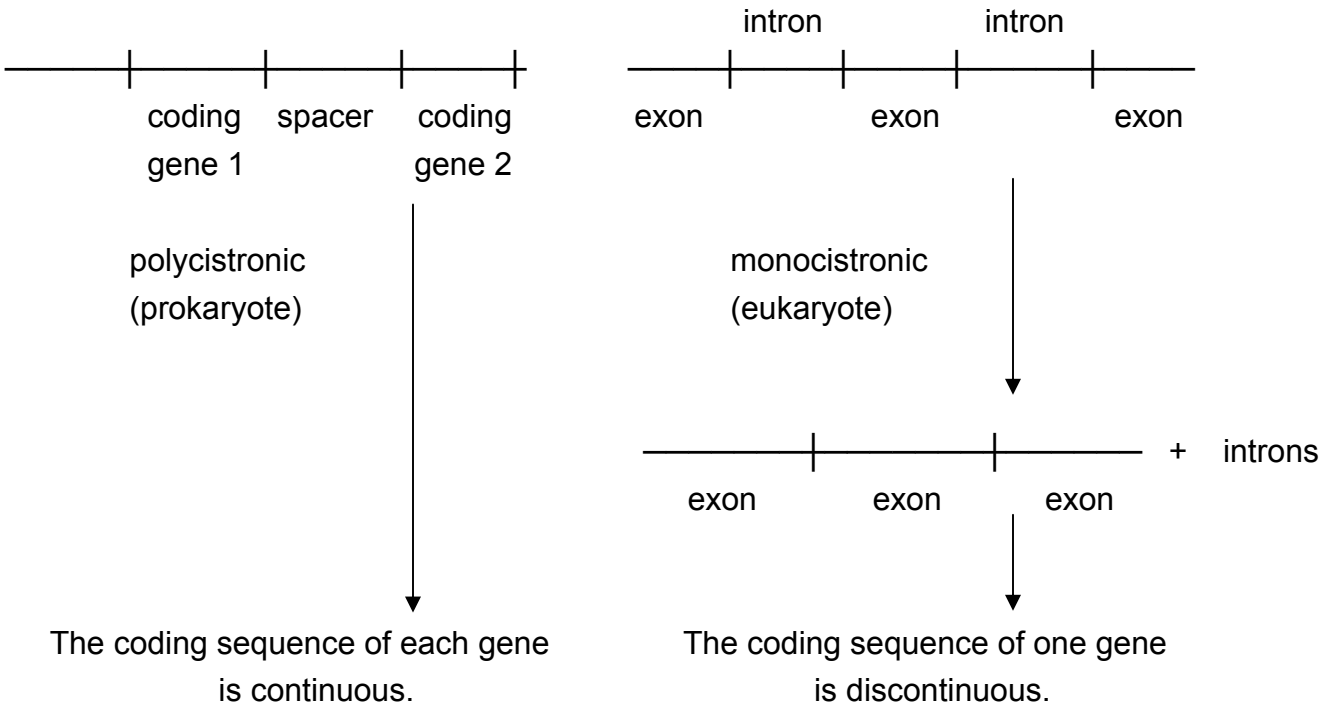
The guanosine is first added using GTP, then methylated, and sometimes the first and second nucleotides are also methylated. The cap binds to a specific protein and participates in binding mRNA to the ribosome (which is the site of protein synthesis). In addition, it may help to prevent premature degradation of the mRNA from the 5'-end.

b. 3'-poly(A) tail

Following transcription, a series of 80-250 adenines are added onto the 3'-end of the RNA. The adenines are added by the enzyme polyadenylate polymerase, also called poly(A) polymerase. First the 3'-end is cleaved close to a specific sequence (5'-AAUAAA-3'). The enzyme then adds the poly(A) tail using ATP but requiring no template. **(PP 40)** The poly(A) tail binds a specific protein which protects the mRNA from enzymatic degradation and also binds to the ribosome during protein synthesis. The poly(A) tail gets shorter as the mRNA gets older.

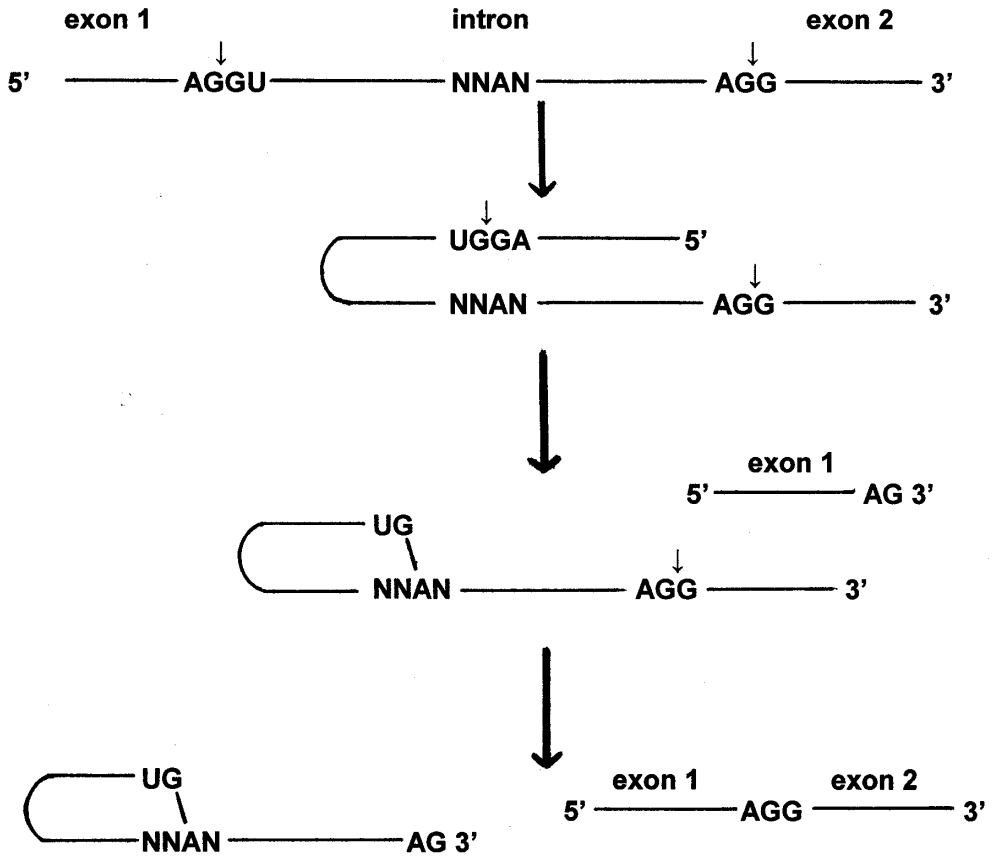
c. Removal of introns - RNA splicing

Eukaryotic pre-mRNA, though monocistronic, differs from prokaryotic mRNA in that it contains introns. Introns are non-coding regions that must be removed before the mRNA is functional. The coding regions of the gene (exons) are interrupted by introns. Introns occur within a gene, which distinguishes them from non-coding spacers between genes. **(PP 41)**



The introns must be removed and the exons joined back together to form one functional mRNA coding for one protein. **(PP 42)** This is the process of splicing. Introns occur in almost all eukaryotic genes (with a few exceptions) and in a few prokaryotic genes. Introns can number up to 200+ in a gene and range in size from 50 to more than 10,000 nucleotides. Exons are generally a few hundred nucleotides long. **(PP 43-44)**

Splicing must be extremely accurate. If a few bases of an exon are lost or a few bases of an intron stay in the mRNA, the genetic information will be changed. Specific sequences occur at intron-exon junctions of hnRNA which are recognized as splice sites. **(PP 45)** A group of specialized RNA molecules called small nuclear RNAs (snRNAs), 100-200 nucleotides in length, are required for splicing. Some of these are complementary to sequences within the intron or at the junctions. **(PP 46-47)** In conjunction with proteins, the snRNAs form small nuclear ribonucleoprotein complexes (snRNPs) which recognize splice sites and perform the splicing. The snRNPs associate into a very large complex called the spliceosome. **(PP 48)** To remove an intron, the 5'-splice site is first cleaved. The loose 5'-end of the intron is then joined to a specific branch site within the intron by a 2'-5' phosphodiester bond to create a lariat structure. **(PP 49)** Then the 3'-splice site is cleaved, the exons are joined together, and the intron (in lariat form) is eliminated and degraded.



There are several different mechanisms for removing introns, depending on the type and location of the RNA (mRNA, rRNA, tRNA, mitochondria, chloroplast, etc.) Some introns are self-splicing. No proteins are required and the RNA itself acts as an enzyme (ribozyme). **(PP 50-51)**

The function of introns is not understood. It may be that they allow for recombination of segments of proteins that results in new proteins and protein evolution. Alternatively, they may help in regulating the use of genes. Some eukaryotic primary transcripts can be processed in two or more different ways, allowing for production of two or more different proteins. **(PP 52)** Such transcripts can have two cleavage sites for poly(A) addition, or have alternate splicing patterns for intron removal. **(PP 53-54)**

d. Transport

All the modifications that convert pre-mRNA into mature mRNA occur in the nucleus. Once completed, mature mRNA must move out of the nucleus to the cytoplasm. Pre-mRNA must not leave the nucleus. **(PP 55)**

Chapter 13: Protein Synthesis

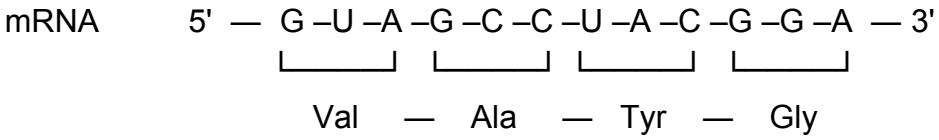
I. Introduction

As a result of transcription, an mRNA molecule has been synthesized and is now ready for protein synthesis (translation). The mRNA has a base sequence that contains the same information as a gene sequence in the DNA. This sequence contains the information for making a specific polypeptide chain. The process for making a protein requires not only the mRNA and amino acids, but many other components as well.

II. Genetic Code

A. Breaking the Code

Since the major difference between mRNA molecules is the base sequence, the sequence must somehow specify the order of the amino acids used to create the corresponding protein. Since there are four bases and twenty amino acids, one base cannot specify one amino acid. Likewise, two bases for each amino acid would not be sufficient since $4^2 = 16$ possible base combinations. The minimum number of bases needed to specify an amino acid is three, since $4^3 = 64$ possible combinations. This in fact is the case. Groups of three bases, called triplets or codons, specify a particular amino acid. The triplets are contiguous and do not overlap. **(PP 2-3)**

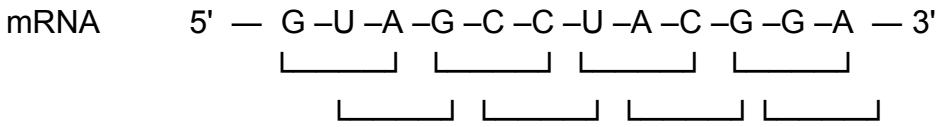


Which codon sequence specifies which amino acid (the genetic code) was determined in several ways. First, protein synthesis was performed using artificial mRNAs. For instance, poly(U) produced polyphenylalanine, showing UUU codes for Phe. Other mRNAs were made from different proportions of A and C. The expected frequency of each codon, AAA, AAC, ACC, etc., could be calculated and compared to the observed results. mRNAs with two alternating codons, like ACA**CAC**ACA**CAC,** produced a protein with two alternating amino acids. Second, trinucleotides could be used with the protein synthesis machinery to see which amino acid would also bind. Combining the results of all these experiments, the genetic code was broken. **(PP 4-5)**

B. Characteristics of the Code

Several important features of the code became apparent.

1. It is of crucial importance that 'reading' the mRNA during protein synthesis begins at the right point. If the reading frame of the triplets is displaced by even one base, the codons will be different and the protein will have totally different amino acids. **(PP 6)**



The starting point of translation, after the leader sequence, is an AUG codon, which is found by the protein synthesizing machinery at the beginning of translation.

2. Three of the 64 codons do not code for any amino acid. UAA, UAG, and UGA are called termination or stop codons and signal the end of a protein.

3. The genetic code is degenerate, meaning that with 20 amino acids and 61 codons, most amino acids have more than one codon. **(PP 7)** Leucine has six codons, glycine has four, phenylalanine has two, and only methionine and tryptophan have one each. For amino acids with multiple codons, it is usually the last base which varies while the first two are the same.

Glycine codons: 5' GGU 5' GGA 5' GGC 5' GGG

4. The genetic code is almost universal, with the same codons specifying the same amino acid in both prokaryotes and eukaryotes. A few exceptions have been found, most notably in mitochondria, where some codons vary. **(PP 8)**

5. It is apparent that a mutation is simply a change in the base sequence of the mRNA (and DNA). Silent mutations result in a fully active protein, either because the changed codon specifies the same original amino acid or because the substituted amino acid does not hurt the protein. More often, a base change results in a different amino acid and a non-functional or impaired protein (missense mutation). **(PP 9)** Mutations can also involve insertions and deletions of bases, which change the reading frame of the mRNA and so change the whole protein. **(PP 10)** If a codon is changed to a termination codon (nonsense mutation) the protein will end prematurely. The genetic code uses 61 of 64 codons (instead of only 20) to reduce the chance of a nonsense mutation. It is better to have three stop codons and 61 codons for amino acids compared to 44 stop codons and 20 codons for amino acids. A protein is more likely to be functional with one wrong amino acid than if it ends prematurely. **(PP 11-12)**

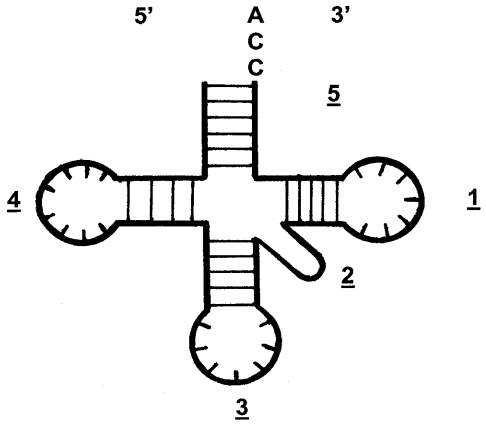
III. Components of Protein Synthesis

In addition to mRNA, two other types of RNA are needed, tRNA and rRNA. **(PP 13)** Other components are needed as well. **(PP 14)**

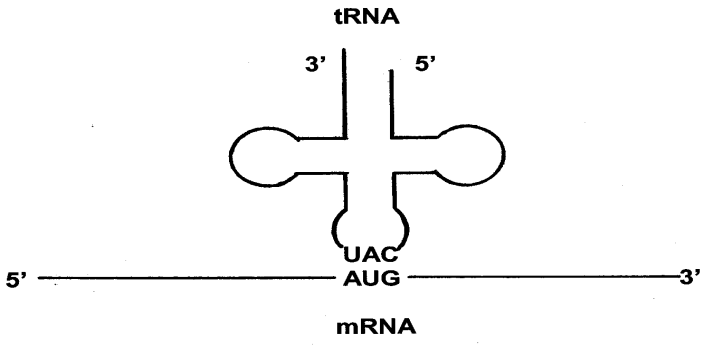
A. Transfer RNA (tRNA)

1. Structure

tRNA molecules bring the right amino acid in to pair with the right codon. Since the codon and amino acid cannot directly recognize each other, tRNA acts as an adapter. **(PP 15)** A tRNA molecule will recognize a specific codon at one end of its structure and bind to the right amino acid (for that codon) at the other end of its structure. **(PP 16)** Thus there must be numerous different tRNA molecules, at least one for each amino acid. tRNAs are small (73-94 nucleotides) with molecular weights of 23,000 - 30,000. All have C–C–A at the 3'-end and usually pG at the 5'-end. They can form a cloverleaf type secondary structure with intrastrand hydrogen bonds. Eight or more bases are modified into unusual bases. **(PP 17-18)**



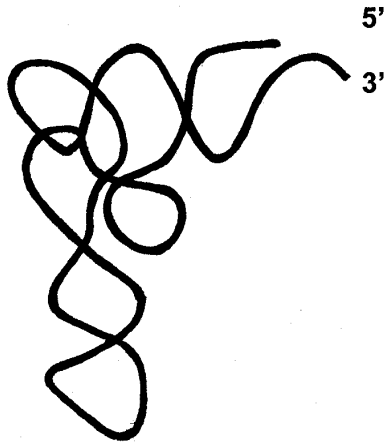
Area 1 is the TψC loop, containing ribothymidine and pseudouridine. Area 2 is the extra arm, of variable length in different tRNAs. Area 3 is the anticodon arm. The three middle bases are complementary to a codon. The anticodon is thus capable of base-pairing in an antiparallel manner to a specific codon. This is how the tRNA recognizes a codon and binds to mRNA.



Thus the anticodon allows the right tRNA with the right amino acid attached to align with a codon for that amino acid.

Area 4 is the DHU arm, containing dihydrouridine. This arm can vary somewhat in size in different tRNAs. Area 5 is the amino acid arm. This is where the amino acid becomes covalently attached to the 3'-end of the tRNA.

The actual structure is not a flat cloverleaf. Rather it is twisted into an L-shaped tertiary structure. **(PP 19)**



This structure is held together by base-stacking and some unusual H-bonds. Since these H-bonds are not part of a double helix, base-pairs such as G-U can occur. **(PP 20-21)**

2. Reaction of tRNA

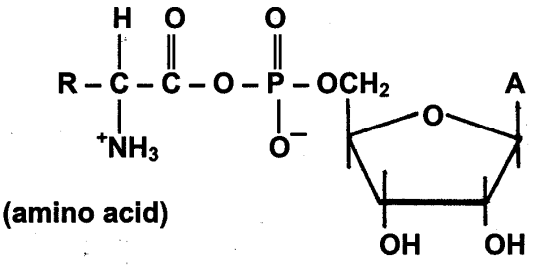
The tRNAs are joined to amino acids by a group of enzymes known as aminoacyl-tRNA synthetases. There are 20 different enzymes, one specific for each amino acid. The enzymes are very accurate, making sure to attach its amino acid to a tRNA with the right anticodon for that amino acid. The overall reaction is **(PP 22-23)**

$$\text{amino acid} + \text{tRNA} + \text{ATP} \rightarrow \text{aminoacyl} - \text{tRNA} + \text{AMP} + PP_i$$

The energy for joining the amino acid to the tRNA comes from breaking down ATP. The reaction occurs in two steps. First, ATP is broken to AMP + PP_i while the AMP and amino acid are joined and bound to the enzyme. **(PP 24)** Then the AMP is removed while the amino acid is moved to the tRNA. **(PP 25-26)**

First Step:

amino acid + ATP + E (enzyme) → E – aminoacyladenylate + PP_i



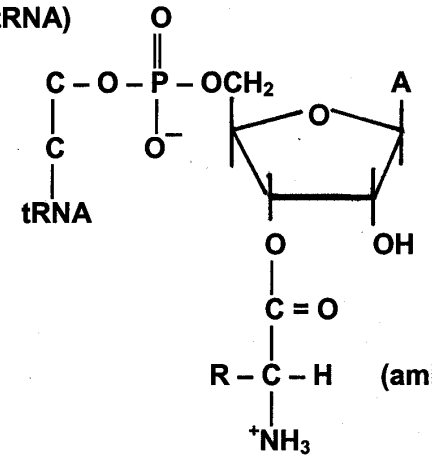
(amino acid)

structure of enzyme-bound aminoacyladenylate

(AMP)

Second Step:

E – aminoacyladenylate + tRNA → aminoacyl – tRNA + E + AMP



(amino acid)

structure of aminoacyl – tRNA

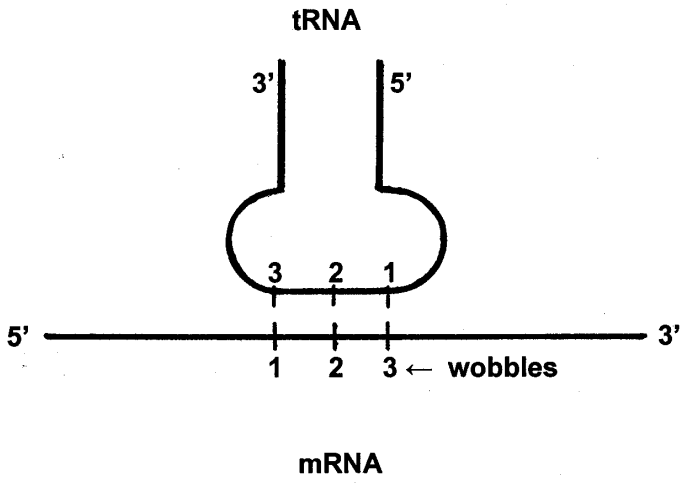
The amino acid is covalently attached to either the 2' or 3' OH group depending upon the specific enzyme.

Some aminoacyl tRNA synthetases must accurately distinguish between very similar amino acids, such as Val and Ile (Ile has one extra CH_2 group). **(PP 27)** The enzymes which must distinguish two amino acids increase accuracy by proofreading their reaction, which they do by having two active sites. **(PP 28)** In the first active site, the amino acid fits snugly and is converted into the aminoacyl-AMP. Thus the enzyme for Val cannot use Ile because Ile is too big. However, the enzyme for Ile will occasionally use Val since Val can bind (more loosely than Ile) to the first active site. After formation of the aminoacyl-AMP, the enzyme then attempts to put the aminoacyl-AMP into the second active site, which is smaller. The correct Ile-AMP will not fit into the second site, so the reaction proceeds with the tRNA. The incorrect Val-AMP will fit,

and when it enters the second site, it is hydrolyzed and the enzyme must start over again. Thus the enzyme checks the amino acid twice, once to eliminate amino acids which are too large and once to eliminate amino acids that are too small (double-sieve mechanism). In addition, some synthetases also check the final amino acid-tRNA and hydrolyze it if it is wrong. The result of these mechanisms is that the wrong amino acid is attached to a tRNA only one in 3000 times. The synthetases also check for the correct tRNA molecule, recognizing certain nucleotides (2-12) within the right tRNA molecule. **(PP 29)** The important nucleotides for recognition can occur in any part of the tRNA, depending upon the specific molecule. The specific conformation of the tRNA molecule is also important and is recognized. **(PP 30-31)**

3. Wobble Hypothesis

There are some numerical contradictions to be resolved, with 20 amino acids, 20 synthetase enzymes, and 61 codons. The first question is how many tRNAs are required, and it turns out that a minimum of 31 tRNAs are needed. However, most cells have more. Thus some amino acids have more than one tRNA. In such a case, the appropriate synthetase enzyme will recognize all the tRNAs for that amino acid, so only 20 synthetases are needed. A second problem is how 31 tRNAs with 31 anticodons can recognize 61 codons. It would seem that 61 tRNAs would be needed, but this is not the case. **(PP 32)** Some tRNAs can recognize more than one codon (all for the same amino acid) and the explanation is called the wobble hypothesis. This hypothesis assumes that the first two bases in a codon form strong, normal base-pairs with the anticodon, but the third base 'wobbles' and forms looser, more flexible H-bonds, sometimes with several different bases. **(PP 33)**



For instance, it has been found that if the first anticodon base is U, it can pair with either A or G. The following pairings are possible. **(PP 34-35)**

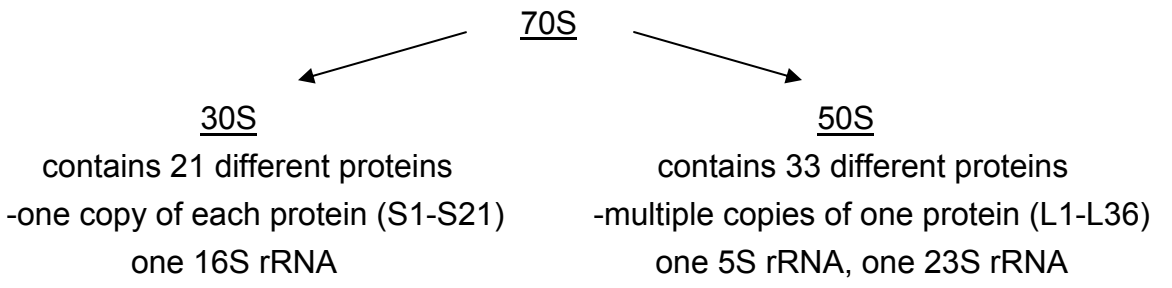
1st anticodon base	3rd codon base
U	A or G
G	C or U
I(inosine)	A, U, or C
C	G
A	U

With the wobble, 31 tRNAs can recognize 61 codons. This also explains why two codons for the same amino acid usually differ in the 3rd base, allowing one tRNA to recognize both codons. Codons differing in the first two bases require different tRNA molecules. The reason for this complexity involves the mechanism of protein synthesis. tRNAs must bind specifically to codons in mRNA, which is accomplished by two strong base-pairs and one weak one. The codon-anticodon interaction must be strong enough to be accurate, but it cannot be too strong either. Codons and anticodons must also dissociate during protein synthesis as well, and if this is too difficult, it will slow down protein synthesis. Three normal base-pairs would be too strong. Two normal base-pairs and one 'wobble' base-pair are strong enough for accuracy but weak enough to permit rapid dissociation as well. **(PP 36-37)**

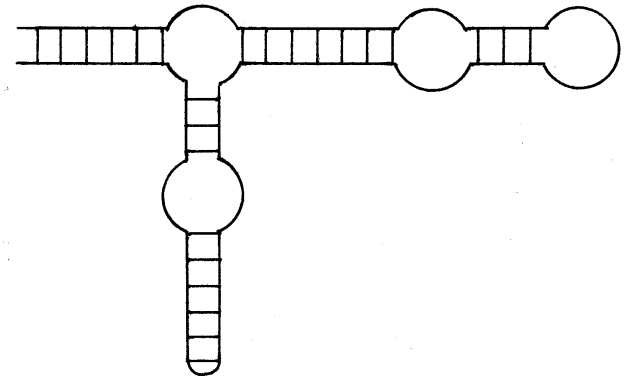
B. Ribosomes

Ribosomes are the sites of protein synthesis. An E. coli cell contains 15,000 ribosomes, each 65% RNA and 35% protein with a molecular weight = 2.5×10^6. These RNA-protein complexes align tRNAs with codons on the mRNA and make peptide bonds between amino acids. **(PP 38)**

The size of the ribosome is designated by its sedimentation coefficient, 70S. The ribosome consists of two unequal subunits, 50S and 30S. **(PP 39-40)**



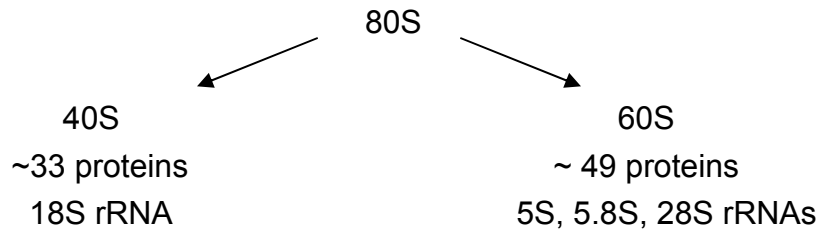
The proteins have molecular weights from 6000 to 75,000. **(PP 41)** Although purified, specific functions are largely unknown. The three rRNAs (5S = 120 nuc., 16S = 1542 nuc., 23S = 2904 nuc.) are all capable of extensive intrastrand hydrogen bonding and can assume shapes like complicated tRNA molecules. **(PP 42)**



5S rRNA

It is known that part of the 16S rRNA is complementary to the 5'-end of mRNA and binds the mRNA. The 23S rRNA has the catalytic ability to make peptide bonds.

Eukaryotic ribosomes are larger (80S) with more proteins and rRNA.



Ribosomes self-assemble. **(PP 43-44)**

IV. Mechanism of Protein Synthesis

Protein synthesis is a very complicated process, using up to 90% of the chemical energy for biosynthesis and requiring ~100 different proteins and ~40 different RNA molecules.

A. General Features

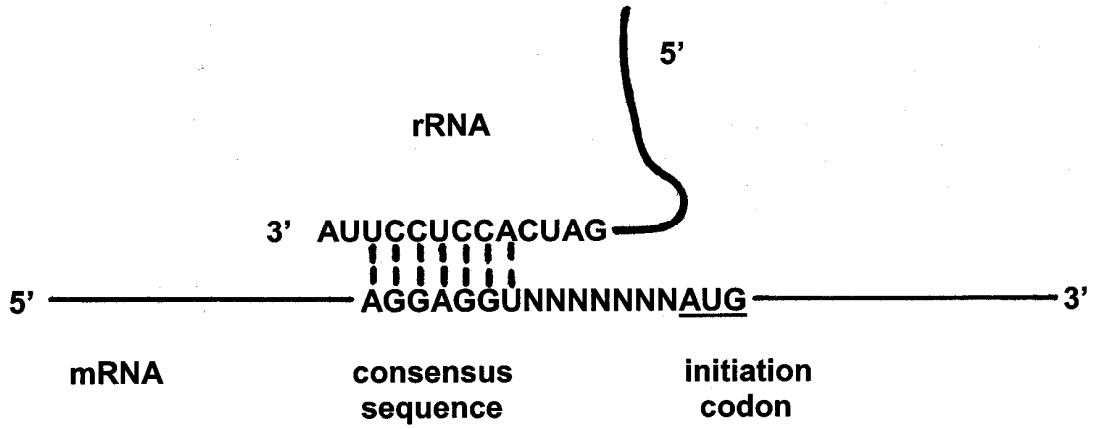
A typical protein is made in about ten seconds. The error rate is about 1 in 10^4 amino acids, sufficient because proteins are routinely broken down and replaced. The mRNA is read 5'→3'. The protein is made one amino acid at a time, starting at the N-terminal and finishing at the C-terminal. **(PP 45)**

B. Initiation

Protein synthesis begins with formation of an initiation complex. This requires the mRNA, the two ribosomal subunits, GTP, the first tRNA, and three additional proteins known as initiation factors. **(PP 46)**

GTP is split in several places during protein synthesis to provide energy, probably to bring about conformational changes. The three initiation factors (IF-1, IF-2, IF-3) are non-ribosomal proteins needed only during initiation. The first tRNA is always the same one, used only for initiation. The first amino acid is a methionine, for which there is only one codon, AUG. Although there is only one codon, there are two tRNAs for methionine. One is used only for initiation, the other is used for Met (AUGs) in the middle of a protein. The same synthetase attaches Met to both tRNAs. However, only the initiating tRNA is then further modified by adding a formyl group, created N-formylmethionine. **(PP 47-48)** The two tRNAs are therefore designated $tRNA^{Met}$ and $tRNA^{fMet}$.

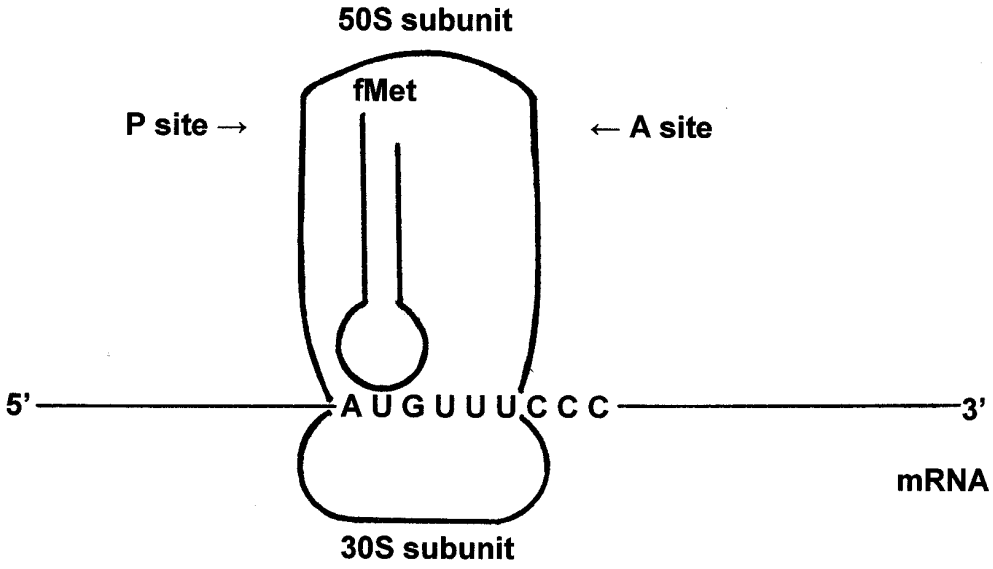
The first step of initiation involves the binding of IF-1 and IF-3 to the 30S subunit. IF-3 prevents the 50S subunit from binding prematurely and IF-1 assists in this function. The mRNA then binds to the 30S subunit and is correctly positioned such that the initiating AUG codon is in the right position to start protein synthesis. **(PP 49)** The positioning is accomplished because of an initiation signal in the leader sequence of the mRNA called the Shine-Dalgarno sequence. **(PP 50)** This sequence, found ~10 nucleotides before the AUG codon, is complementary to the 3'-end of 16S rRNA. **(PP 51)**



This mRNA-rRNA interaction results in correct positioning of the AUG codon.

Along with the mRNA, IF-2-GTP and fMet-$tRNA^{fMet}$ bind. **(PP 52)** The anticodon of the tRNA pairs with the AUG codon. Finally, the 50S subunit binds, GTP is hydrolyzed and GDP and P_i are released, and the three initiation factors are also released. **(PP 53)**

The resulting initiation complex has two important sites on the ribosome called the P site (peptidyl) and A site (aminoacyl) which can bind tRNAs. The initiator tRNA is in the P site, the only one it can bind to. All other tRNAs can bind only to the A site when they first interact with the ribosome. A third site for binding tRNAs, called the E site (exit site) also forms. **(PP 54-55)**



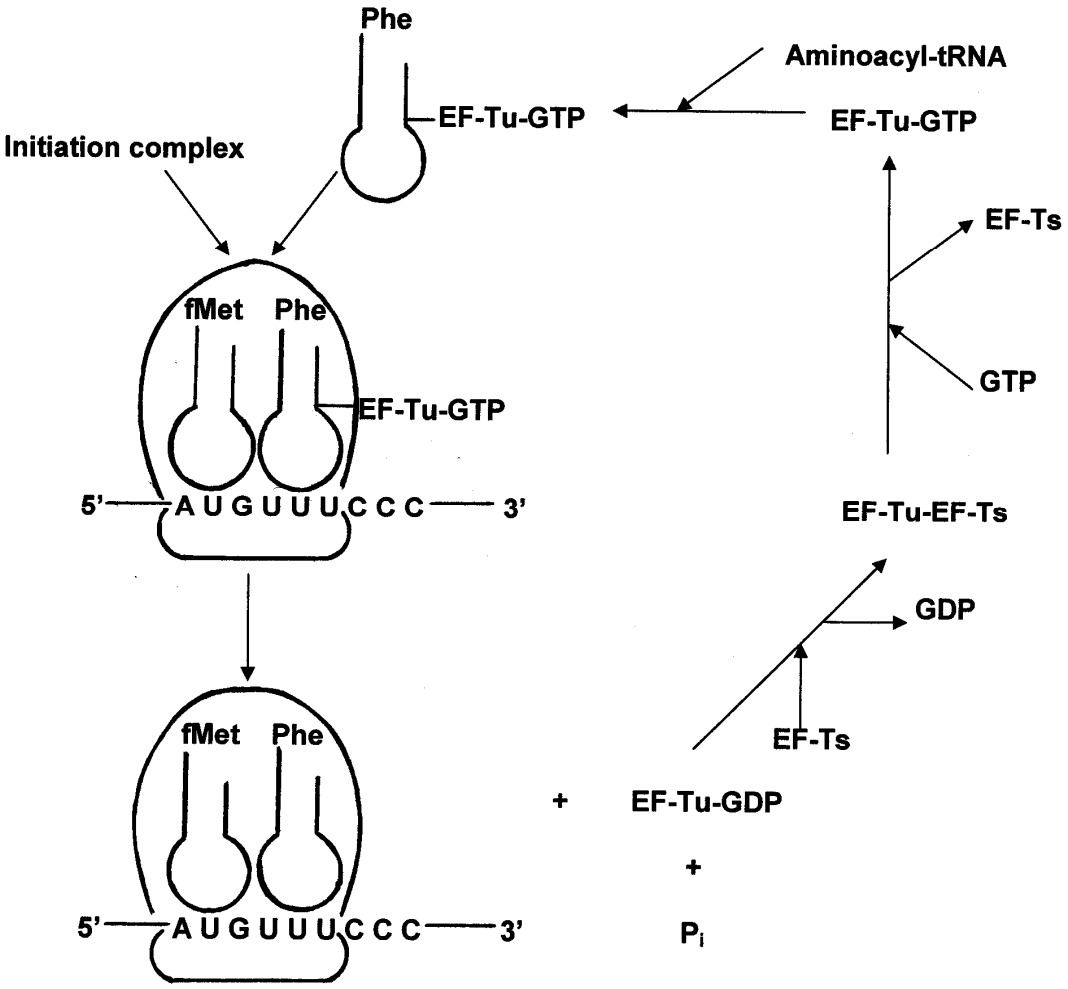
Initiation Complex

In eukaryotes, the process is basically the same. The initiating tRNA, though it is used only for initiation, binds Met, not fMet. **(PP 56)** At least a dozen initiation factors are known. **(PP 57)** There is no Shine-Dalgarno sequence, but the 5'-cap on the mRNA is essential.

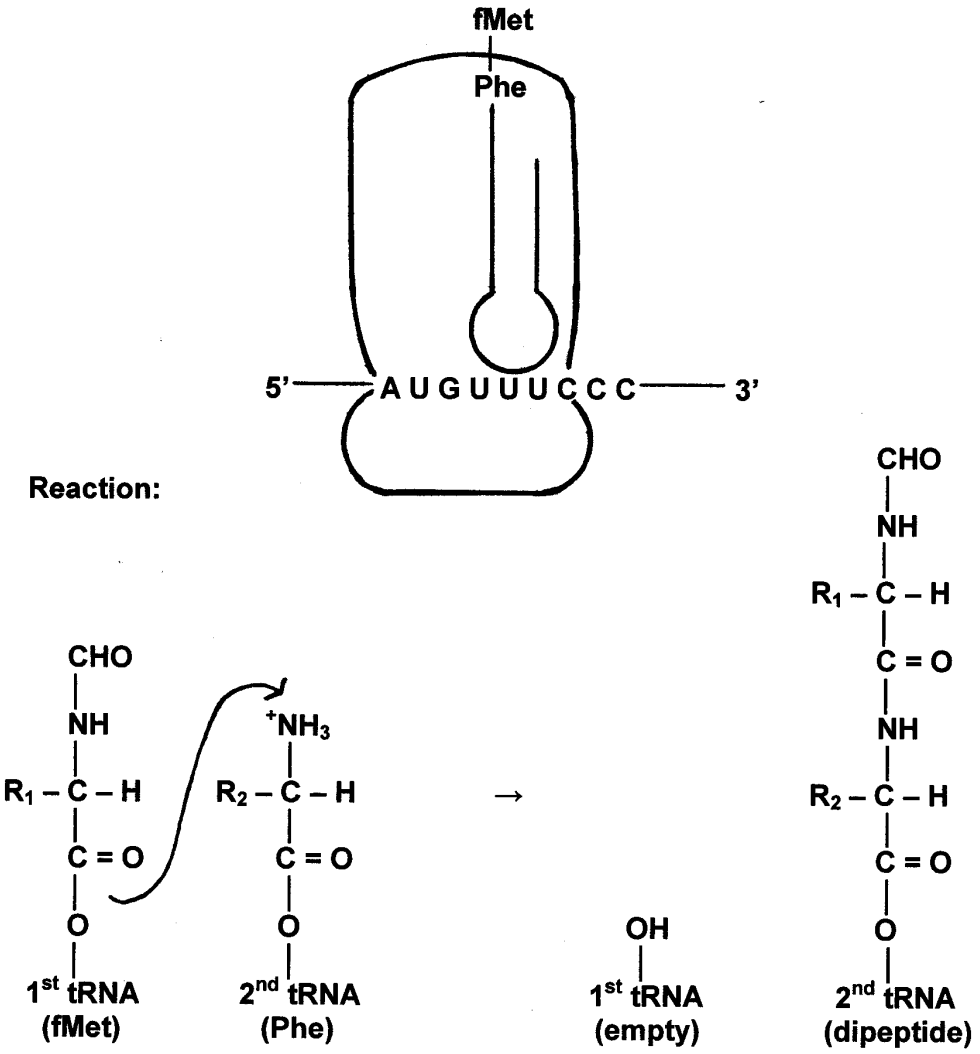
C. Elongation

Elongation is a series of repetitive steps resulting in adding an amino acid via a peptide bond. Besides the initiation complex, it requires the second tRNA, GTP and three non-ribosomal proteins called elongation factors, EF-Tu, EF-Ts, and EF-G. **(PP 58)**

The second tRNA binds to EF-Tu containing GTP. **(PP 59)** It then binds to the A site and the GTP is hydrolyzed, releasing P_i and EF-Tu-GDP. The EF-Tu-GDP complex is regenerated by EF-Ts, which allows exchange of the GDP for a GTP. Using EF-Tu and hydrolyzing GTP allows the tRNA complex to stay on the ribosome for a few milliseconds. If the tRNA is wrong, the codon-anticodon binding will dissociate before the EF-Tu reactions are complete, preventing a wrong amino acid from being joined. This is a form of proof-reading.

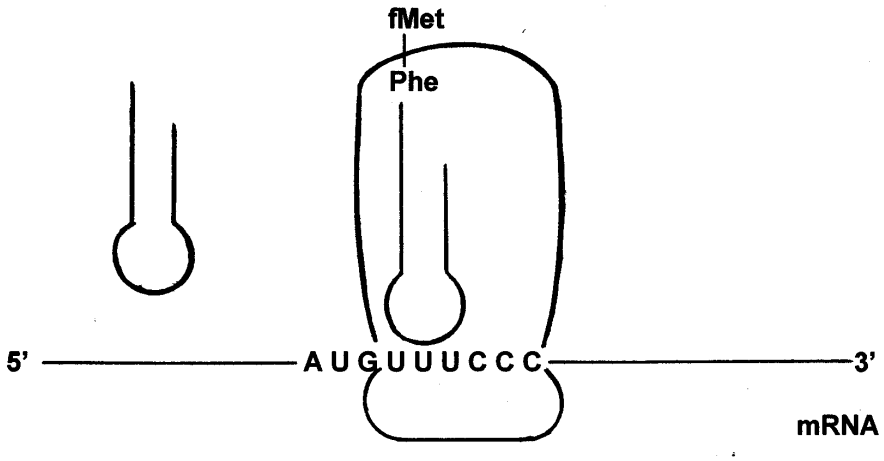


Now, a peptide bond is formed between the first and second amino acids. The dipeptide remains attached to the second tRNA while the first tRNA now has no amino acid. **(PP 60-61)**



The reaction is apparently catalyzed by the 23S rRNA, an example of an RNA enzyme. The activity is called peptidyl transferase.

The next stage is called translocation, where the ribosome moves along the mRNA by one codon. **(PP 62)** This movement puts the second tRNA (with attached dipeptide) into the P site and ejects the first tRNA from the P site. The first tRNA probably goes into the third ribosomal binding site, the E site (exit), before leaving the ribosome. The third mRNA codon is now in the A site, while the second codon is in the P site. This step requires EF-G. Energy for a change in conformation is supplied by GTP hydrolysis



The elongation cycle now repeats, with the third tRNA coming into the A site. For each amino acid added, two GTPs are used. The ribosome moves toward the 3'-end of the mRNA, one codon at a time.

In eukaryotes the process is basically the same. Three analogous EFs are known.

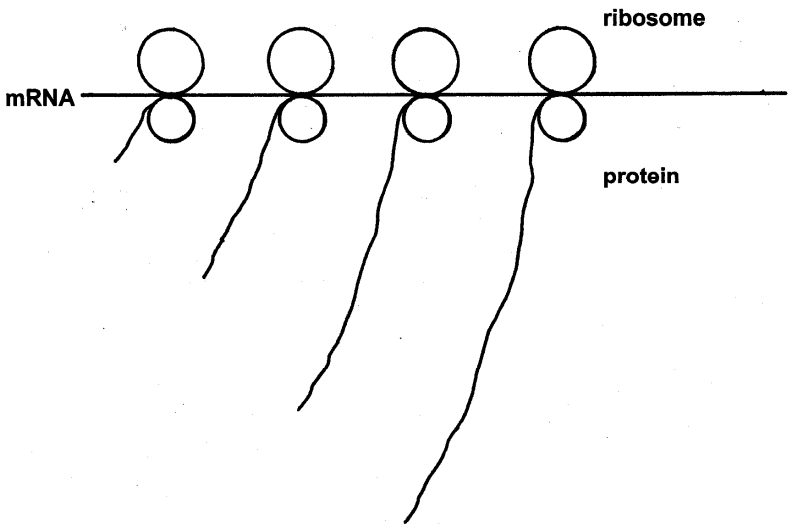
D. Termination

Elongation continues until the ribosome adds the last amino acid and encounters a termination codon (UAA, UAG, UGA) immediately after the last amino acid codon. Termination requires three termination or release factors. These proteins are called RF_1, RF_2, and RF_3. RF_1 recognizes UAG and UAA, while RF_2 recognizes UGA and UAA. These bind at a termination codon and induce the peptidyl transferase to hydrolyze the protein chain from the last tRNA. The function of RF_3 is to promote binding of RF_1 and RF_2, using GTP. Release of the protein causes the last tRNA to drop off, and the ribosome dissociates into subunits. **(PP 63-66)**

In eukaryotes, a single releasing factor recognizes all three termination codons.

E. Polysomes

In both prokaryotes and eukaryotes, multiple ribosomes can attach to the same mRNA, forming what are called polysomes. A ribosome covers about 80 nucleotides, so many can be interspersed along a mRNA. **(PP 67-68)**



In bacteria, transcription and translation are tightly coupled. Translation begins at the 5'-end of the mRNA before the 3'-end is even completed. **(PP 69-70)** This cannot happen in eukaryotes because RNA processing must first occur. In addition, eukaryotic mRNA is made in the nucleus while translation occurs in the cytoplasm.

V. Post-Translational Modification

The synthesized protein must now fold and undergo various modifications before it becomes functional. Folding, which can begin before the protein is fully made, often occurs spontaneously. Modifications involve several types. **(PP 71)**

A. The amino terminal is usually modified, removing the fMet and sometimes additional amino acids. Often the new N-terminal is acetylated.

B. The carboxyl terminal is often modified by removing some amino acids.

C. Modification of certain amino acids can occur. Ser, Thr, and Tyr can be phosphorylated. Lys and Glu can be methylated. **(PP 72)**

D. Carbohydrates can be added to make glycoproteins.

E. Other prosthetic groups are added.

F. Proteolytic cleavage is sometimes necessary for activity, including trimming a protein, cutting it into two chains, or eliminating some internal amino acids.

G. Disulfide bonds must be formed.

H. Protein Targeting

Proteins must reach their correct cellular location, and this process is especially complicated in eukaryotes. Proteins destined for the cytosol simply remain there. Proteins destined for secretion or the plasma membrane share part of their targeting pathway. An amino-terminal sequence marks them for movement to the endoplasmic reticulum. The signal can vary from 13 to 36 amino acids including ~12 hydrophobic amino acids in a stretch. Also found is a positively-charged amino acid near the N-terminal and a sequence of small amino acids near the cleavage site which eventually allows for removal of the signal sequence.

Protein synthesis starts on free ribosomes, generating the signal sequence. The sequence and the ribosome are then moved by a large complex (signal recognition particle) to the endoplasmic reticulum where protein synthesis is completed as the protein moves into the endoplasmic reticulum and the signal sequence is cleaved off. The proteins are then directed to various locations including the Golgi complex, the plasma membrane, lysosomes, or secretion from the cell. **(PP 73-74)**

Proteins destined for mitochondria or the nucleus have different signal sequences to move them to these locations. **(PP 75-76)**