Introductory Biochemistry

Introductory Biochemistry

CAROL HIGGINBOTHAM

Contents

Introductory Biochemistry is a text supporting a single-term biochemistry course as part of a one-year Introductory Chemistry course series often described as General, Organic and Biological Chemistry (GOB). The audience for this book includes students majoring in fields outside of the Sciences who need or want some knowledge of biochemistry. This includes students in pre-health fields, but also possibly students interested in Natural Resources, Agriculture or Business. The objective is to provide these students with an exposure to core ideas of the subject.

The text is a modified copy of Biochemistry Free for All, written and shared as an open educational resource by Dr. Kevin Ahern, Dr. Indira Rajagopal, and Taryln Tan at Oregon State University, Corvallis. The original audience for Ahern et al included students arriving to class with a more thorough background in both General and Organic Chemistry than GOB students. The original contains more information, more detail and remains openly available through Libretexts.

This book is likely to undergo additional modification in the future, including significant rewriting, and added support for concepts in biochemistry that draw from other subdisciplines of chemistry.

Additionally, while some interactive features have been removed in modifying the Ahern text, others have been made possible in this PressBooks format. Among those is the use of social annotation, which can be handled through hypothes.is, a free app that encourages reflective reading, questioning and social engagement around the text.

I have deep gratitude to Drs. Ahern and Rajagopal for production and dissemination of Biochemistry Free for All, and to the students who also contributed to the contents of their book. Writing a text and providing useful graphics is an enormous amount of work. It is very generous for them to complete that ambitious project and then share their work so freely.

More information on the original text and information on authors can be found at:

https://biochem.oregonstate.edu/content/biochemistry-free-and-easy

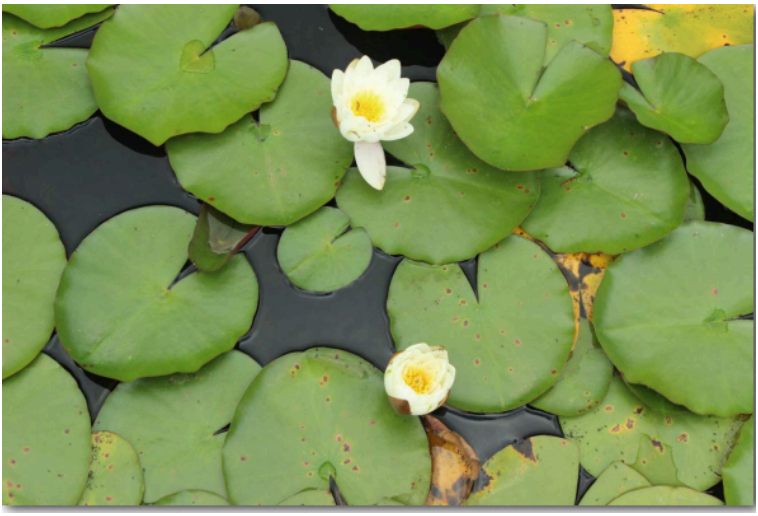
Dr. Kevin Ahern and Dr. Indira Rajagopal, Taralyn Tan

Contributors to chapters, students at Oregon State University at the time of contribution:

Martha Baker, Pehr Jacobson, Aleia Kim, Penelope Irving

1.1: Introduction - Basic Biology

The Bio of Biochemistry



Life is astoundingly diverse. It is not a straightforward exercise to count species, but carefully produced estimates do exist. One such estimate predicts approximately 8.7 million species exist globally, about 2.2 million of which are marine[1]. Huge numbers of species, which likely include the vast majority of all that exist, have yet to be described.

These organisms, representing the three great domains Eukarya, Bacteria and Archaea occupy every environmental niche imaginable, from the human gut to the frozen expanses of the Antarctic, and

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3160336/

from the rainforests of the Amazon basin to the acid waste washes of gold mines. Some organisms, like the tardigrades (Figure 1.1), or water bears, can withstand incredibly harsh conditions – from a few degrees above absolute zero to 300°F, the vacuum of outer space, and pressures greater than those at the greatest depths of the ocean. In the memorable words of Dr. Ian Malcolm in the movie, Jurassic Park, "Life finds a way".

On these
pages our aim's
to extract

Knowledge
out of a
mountain of
facts

Missing
excessive data

Will not really
matta

If our students
can learn and
relax

Despite the differing demands of existence in these widely varied environments, all living things share some common characteristics. The most noticeable of these is that from hummingbirds to humpback whales, from fungi to frogs, and from bacteria to birch trees, all living things are made up of cells. This fact was first discovered by Robert Hooke, in 1665 (Figure 1.2), when he used a microscope to look at a slice of cork and found that it seemed to be made up of tiny chambers that he named cells. Subsequent examination of other living things revealed that they too were, without exception, made up of cells. Today, we know that organisms in all three domains of life share this property – they are all made up of cells. For some, a single cell is the entire organism, while others are multicellular, like humans, wombats or weeping willows.

Figure 1.1 – Tardigrade Wikipedia

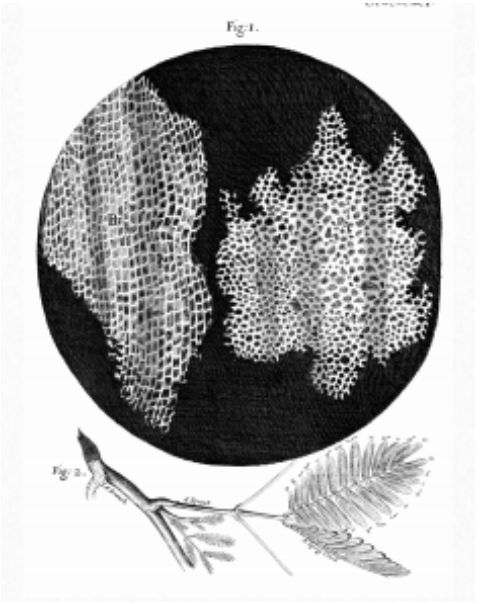


Figure 1.2 Slices of cork as seen by Hooke

The subject of this book is **biochemistry**, the science that explains life at the molecular level. The special characteristics of cells

influence the unique chemistry of life. It is, thus fitting to take a look at cells, the setting in which the molecular events of life take place.

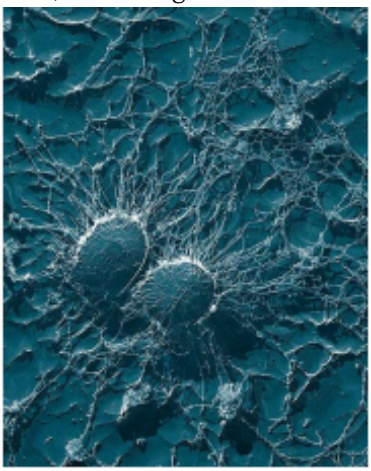


Figure 1.3 Bacterial cells Wikipedia Figure

"There are living systems; there is no "living matter". No substance, no single molecule, extracted and isolated from a living being possess, of its own, the aforementioned paradoxical properties. They are present in living systems, only; that is to say, nowhere below the level of the cell." – *Jacques Monod*

Cells

All cells, no matter what kind, have a plasma membrane that serves as a boundary for the cell, separating it from its surroundings. They also possess a genome made up of DNA that encodes the information for making the proteins required by the cell. To translate the information in the DNA and make the proteins it encodes, all cells have the machinery for protein synthesis, which include the structures named ribosomes and tRNAs. DNA is also the repository of information that gets copied and transmitted to the next generation, allowing living cells to reproduce.

Energy source	sunlight	photo-			-troph
	Preformed molecules	chemo-			
Electron donor	organic compound		organo-		
	inorganic compound		litho-		
Carbon source	organic compound			hetero-	
	inorganic compound			auto-	

Figure 1.4 – Organization of organisms by metabolic type Wikipedia

All cells obtain and use energy. The source of this energy is different in different organisms (Figure 1.4). Phototrophs are organisms that obtain metabolic energy from light, while chemotrophs get their energy from the oxidation of chemical fuels. Organisms that can capture energy from light or from chemical sources are termed autotrophs (auto=self, troph=nourishing). Others organisms are heterotrophs, which use the organic compounds made by other organisms as their energy source. Plants and other photosynthetic organisms are autotrophs. Animals such as ourselves are heterotrophs.

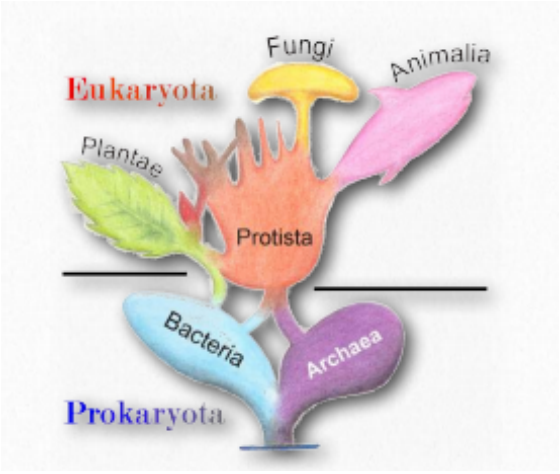


Figure 1.5 – Tree of life Wikipedia

Cells may be **aerobic** (i.e., use oxygen) or **anaerobic** (able to live

without oxygen). Some anaerobic cells require an environment free of oxygen, and are referred to as obligate anaerobes. Others are cells that can live with, or without, oxygen, and are called facultative anaerobes.

Prokaryotic and eukaryotic cells

Organisms may be divided into two major groups, the **prokaryotes** and the **eukaryotes**. The cells of the former lack a nucleus and other organelles, while those of the latter are characterized by numerous internal, membrane-bounded compartments, including a nucleus.

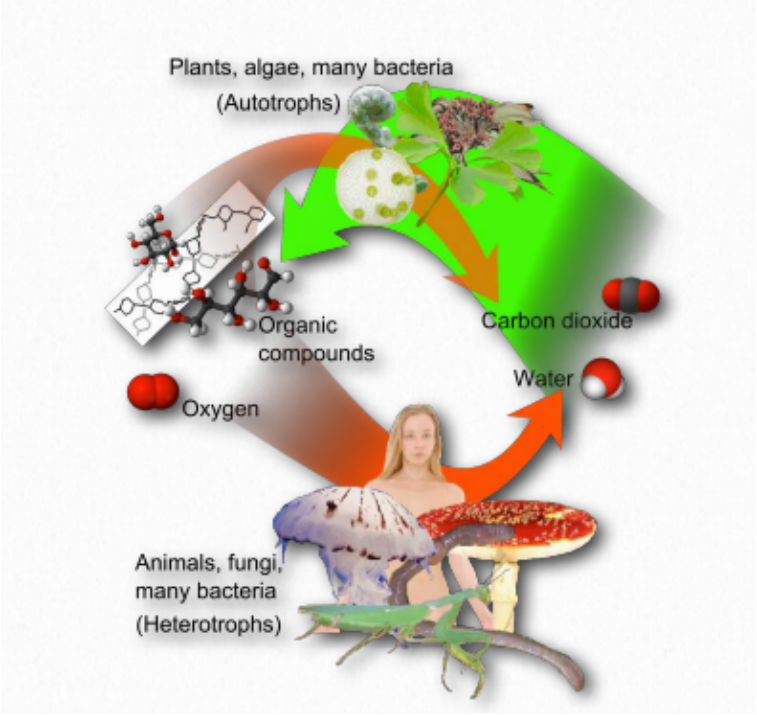


Figure 1.6 – Interplay between autotrophs and heterotrophs Wikipedia

Prokaryotes are unicellular and generally considerably smaller than their eukaryotic cousins, with sizes ranging from 0.5 to 5 μm in diameter. Prokaryotes typically have circular chromosomes, and may sometimes contain extra-chromosomal DNA elements (also

Image by usually circular) called plasmids. Although the DNA in prokaryotes is not wrapped around histones, as is the case for eukaryotes, prokaryotes have proteins associated with their DNA. The DNA-protein complexes in prokaryotes create a structure called a nucleoid, which is different from the eukaryotic nucleus in not being enclosed by a nuclear envelope (Figure 1.7).

Cellular Differences between Prokaryotes and Eukaryotes

Prokaryotes	Organelle	Eukaryotes
No definite nucleus	Nucleus	Present
Present	Cell Membrane	Present
None	Mitochondria	Present
None	Endoplasmic Reticulum	Present
None	Chloroplasts	Present in Plants

Aleia Kim Table 1.1

Prokaryotes may be divided into two broad categories, bacteria and archaea. These single-celled organisms are both ancient and widespread. Archaea were once thought to be a subgroup of bacteria, but have subsequently been shown to be a completely different group of organisms that are so distinct from both bacteria and eukaryotes that they now are classified in a domain of their own.

Bacteria

Like eukaryotic cells, bacterial cells have a plasma membrane surrounding them, but in addition, they also contain an exterior cell wall, comprised of an interlocked peptidoglycan network. On their exterior surfaces, bacteria have hair-like appendages called pili that allow them to adhere to other cells. Pili play an important role in bacterial conjugation, a process in which DNA is transferred between bacterial cells. In addition, bacterial cells may have flagella that enable them to move through their surroundings.

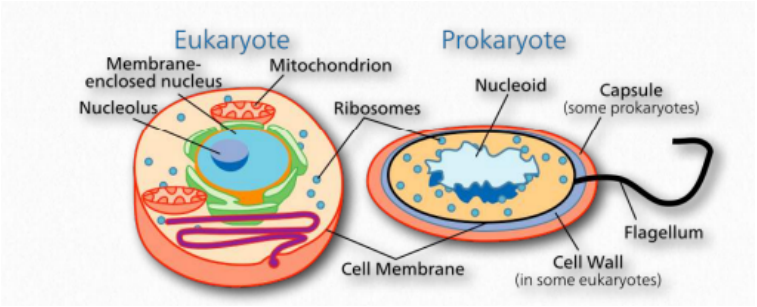


Figure 1.7 – Prokaryotic vs. eukaryotic cell structures (not to scale)

Excuse me for
feeling superia

To the life
forms we call
the bacteria

Students
know very well

There are no
organelles

To be found in
their tiny interia

Interestingly, bacteria can communicate using chemical signals, in a process called quorum sensing. These signaling mechanisms enable bacteria to assess conditions around them. Quorum sensing plays a role in the process of infection by bacterial pathogens as well as the formation of **biofilms**.

Archaea

The first archaeans to be studied were all found in harsh environments such as salt flats and hot springs. They were initially believed to live only in such extreme environments (Figure 1.8). We now know that archaeans can be found in every environment, moderate or extreme. Archaea have been found in the human gut, and in such huge numbers in marine plankton that it has been suggested that they may be the most abundant organisms on earth.

While they are unicellular and superficially resemble bacteria, archaea are in some respects more similar to eukaryotes. Their molecular machinery and ribosomes are much more like those of

eukaryotes than of prokaryotes. Archaea are unique among living organisms in their use of ether linkages to join the lipids used in their plasma membranes to glycerol. Not only are the ether linkages different from the ester linkages in all other forms of life, but the lipids themselves are different. In place of the fatty acids used in both bacterial and eukaryotic membranes, archaea use long isoprene-derived chains (Figure 1.9) This difference in membrane composition and structure makes archaeal membranes highly stable and may be advantageous in extreme conditions.

Figure 1.8 – Archaeans growing in acid mine waste

Like bacteria, archaea have a cell wall, but the composition is different. Some archaea have peptidoglycan-like molecules in their cell walls, while others build their cell walls entirely of glycoproteins and polysaccharides.

Eukaryotes

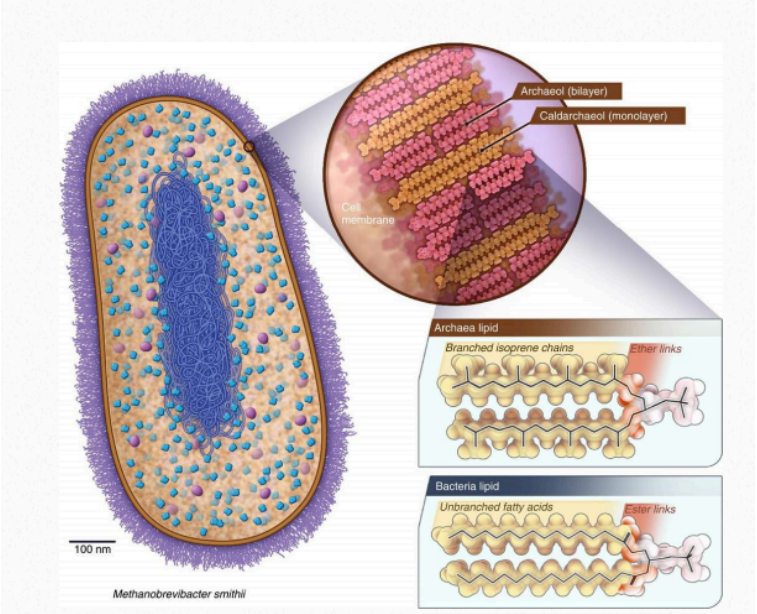


Figure 1.9 Archaeal membrane, top, showing unusual ether linkages and isoprene chains and bacterial membrane, below. Wikipedia

Eukaryotic cells are found in both unicellular and multicellular organizational schemes. Unicellular forms include yeast and many protists like Paramecium and Amoeba. Multicellular eukaryotes include plants, animals, and fungi. Eukaryotic cells are surrounded by a plasma membrane. Animal cells have no cell walls, whereas plant cells use cellulose, hemicellulose, and pectins to build cell walls outside their plasma membranes. Fungal cells have cell walls that contain the polymer chitin, which is also found in the exoskeletons of arthropods.

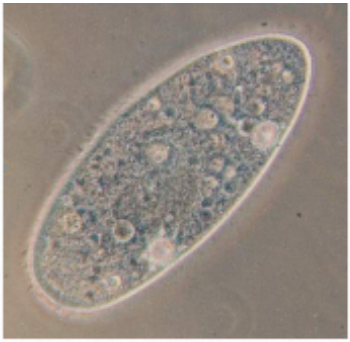


Figure 1.10 Paramecium Wikipedia

Eukaryotic cells are typically much larger (typically 10-100 μm) and contain considerably more DNA than prokaryotic cells. The most distinctive feature of eukaryotic cells, however, is the presence of a variety of internal membrane-bounded structures, called organelles.

Organelles

Eukaryotic cells are characterized by internal membrane-bounded compartments, or organelles. These compartments divide up the interior of the cell into discrete parts that have specialized functions.

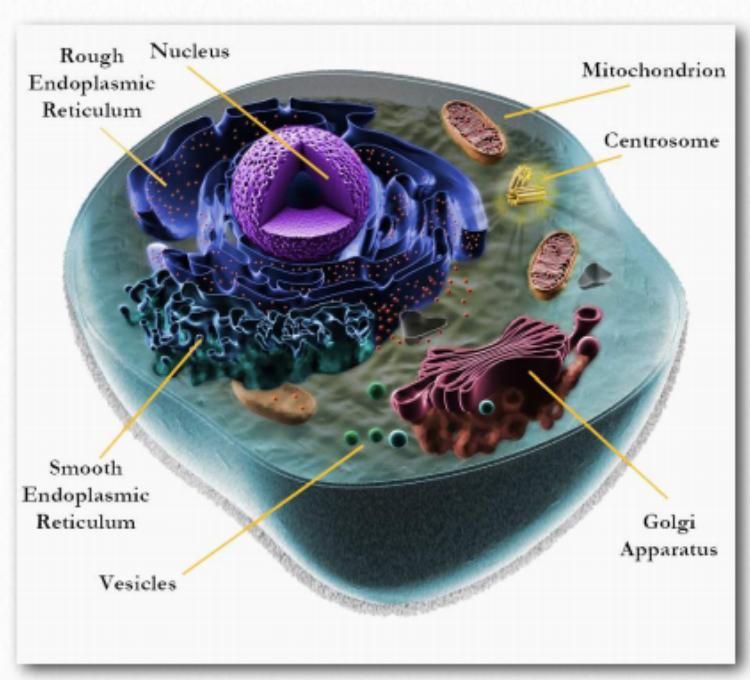


Figure 1.11 Animal cell structure Wikipedia

The presence of multiple compartments within the cell permits reactions requiring specific conditions to be carried out in isolation from the rest of the cell. For example, the formation of disulfide bonds in proteins is possible in the conditions within the endoplasmic reticulum, but would not readily occur in the different environment of the cytosol. The presence of membrane-bounded compartments also allows reactants to be more concentrated because of the smaller volume of the organelle.

Eukaryotic DNA in a cell is divided into linear bundles called chromosomes. Chromosomes contain the genomic DNA wrapped around cores of proteins called histones. The ends of linear eukaryotic chromosomes have telomeres, containing short repeated DNA sequences.

The chromosomes in eukaryotic cells are surrounded by the nuclear envelope, a double membrane structure that encloses the nucleus (Figure 1.12). Within the nucleus, there are enzymes

required for the replication and transcription of genetic information. The presence of the nuclear envelope also regulates which proteins can enter the nucleus at any given time.

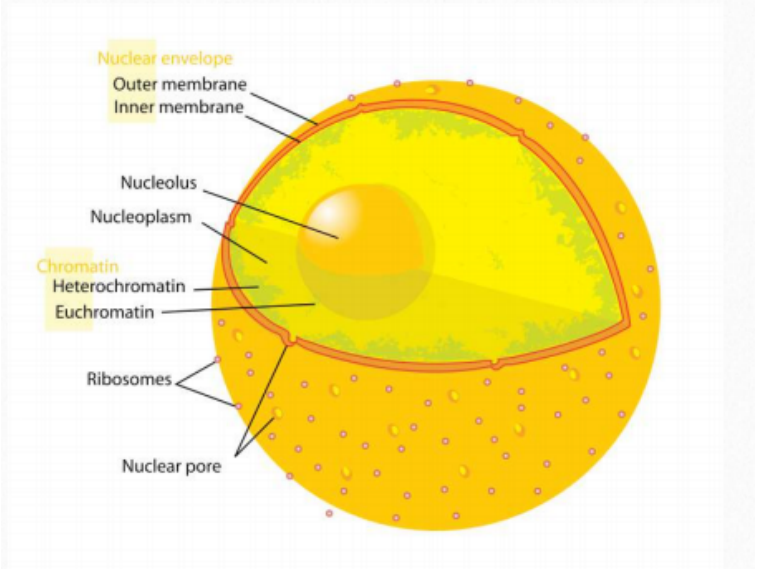


Figure 1.12 Cell nucleus

Mitochondria and chloroplasts have their own DNA, separate from and in addition to the nuclear DNA. This DNA is small and circular and resembles a prokaryotic chromosome. Mitochondria and chloroplasts also have their own ribosomes and tRNAs and can carry out their own protein synthesis. These organelles are likely derived from prokaryotes that once lived as **endosymbionts** within ancient eukaryotic cells and eventually became integrated into their host cells.

The cytoskeleton

Another interesting feature of eukaryotic cells is the presence of an internal skeleton-like structure called a cytoskeleton. The cytoskeleton is made up of a network of interlinking protein fibers Figure 1.13).

Although the word “skeleton” may suggest a rigid and fixed structure, the cytoskeleton is dynamic, with both microfilaments

and microtubules disassembling and rearranging themselves on an ongoing basis, as needed.

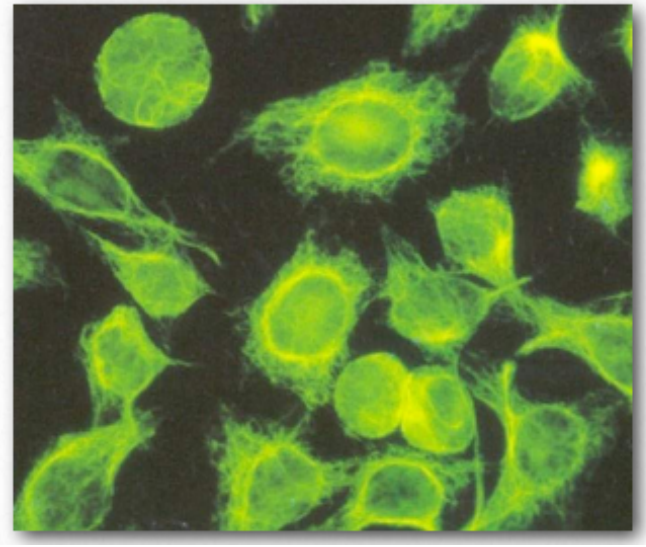


Figure 1.13 Cultured cells stained to show intermediate filaments Wikipedia

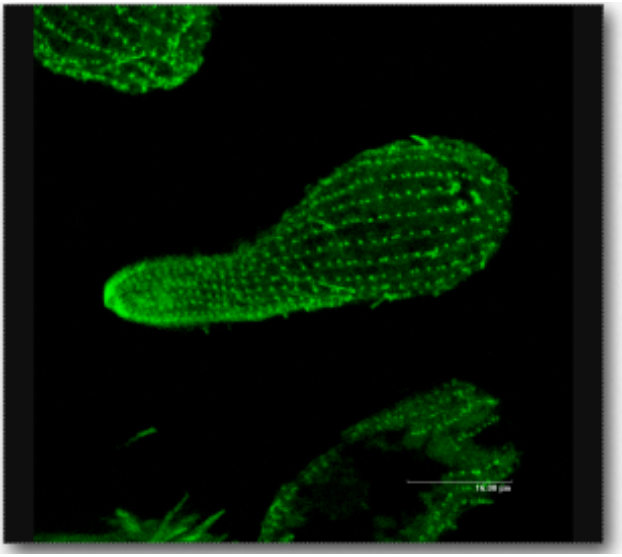


Figure 1.14 – β-tubulin in Tetrahymena Wikipedia

The cytoskeleton helps to organize the contents of the cell. Interactions among cytoskeletal proteins and components of the extracellular matrix are crucial in maintaining tissue structure.

Tissues

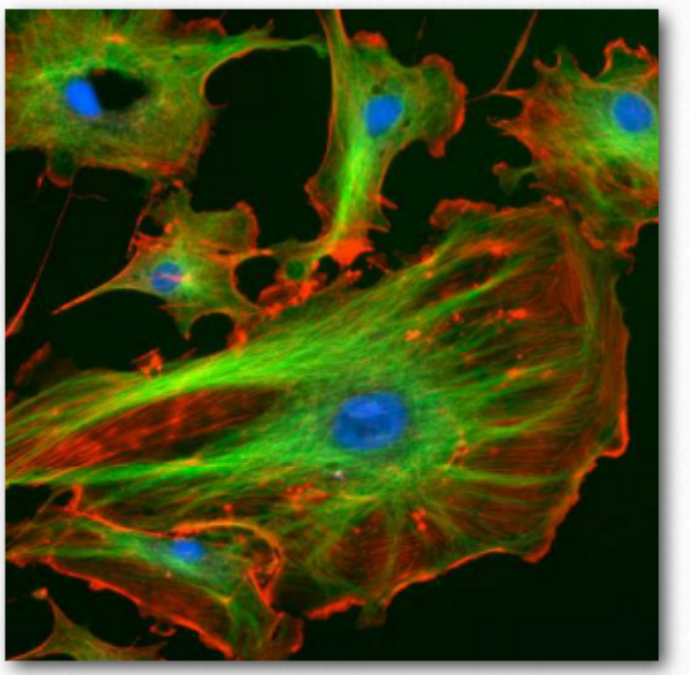


Figure 1.15 – The cytoskeleton. Actin filaments are in red, microtubules in green, nucleus in blue

Cells in multicellular organisms are organized into tissues that play specialized roles in the body. Animals have four types of tissues in their bodies – epithelium, connective tissue, nerve tissue, and muscle tissue.

Connective tissue

Of the four animal tissues, connective tissue is the one that serves as the "glue" to hold everything together. Connective tissue fills the gaps between all the other tissues of the body, including the nervous system.

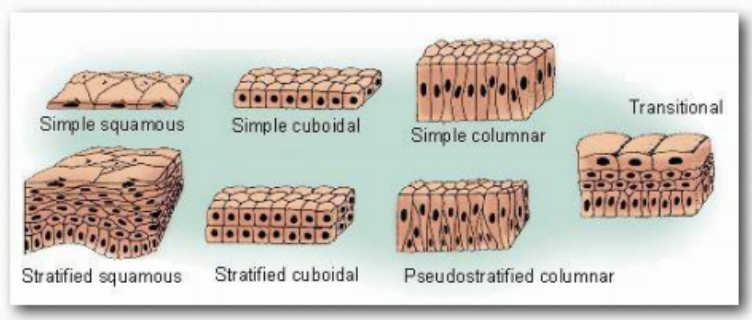


Figure 1.16 – Types of epithelial tissue

Nerve tissue

The nervous system of humans contains two main components. The brain and spinal cord comprise the central nervous system (CNS) and nerves branching from these make up the peripheral nervous system.

Nerve cells are also called neurons. Neurons can transmit and receive signals.

Nervous system functions include receiving input from senses, controlling muscles and glands, homeostasis, integration of information, and mental activity.

Muscle tissue

Mammals have three types of muscle tissue – 1) skeletal/striated muscle; 2) smooth, non-striated muscle; and 3) cardiac muscle. Cardiac muscle and smooth muscle are notable for contracting involuntarily. Both can be activated through nerve stimuli from the central nervous system or by innervation from the peripheral plexus or by endocrine/hormonal activation. Striated muscles, by contrast, only contract voluntarily by (mostly) conscious action influenced by the central nervous system.

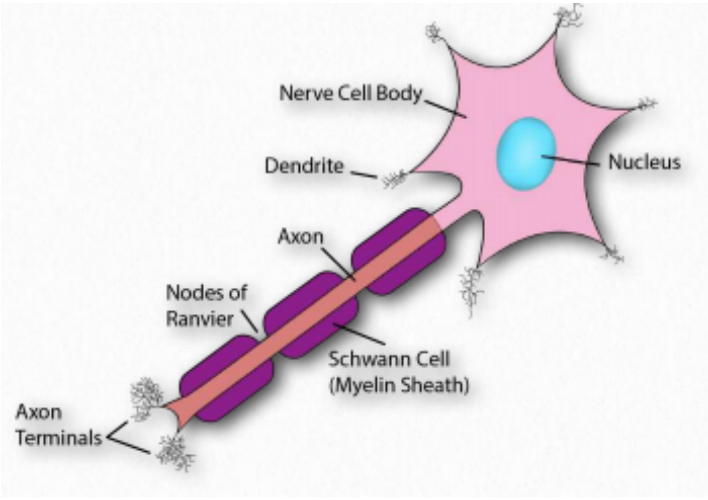


Figure 1.17 – Nerve cell anatomy Image by Pehr Jacobson

1.2: Introduction - Basic Chemistry

> "Organic chemistry is the chemistry of carbon compounds. Biochemistry is the chemistry of carbon compounds that crawl" -Michael Adams.

To understand biochemistry, one must possess at least a basic understanding of organic and general chemistry. In this brief section, we will provide a rapid review of some concepts necessary to understand cellular chemistry.

Chemistry is chemistry, whether in a cell or outside it, but biological chemistry is a particular subset of organic chemistry that often involves enormous macromolecules, and that happens in the aqueous environment of the cell.

Figure 1.18 shows organic functional groups common in biochemistry. You will encounter these functional groups as you study the reaction pathways that build and recycle the compounds of which cells are made.

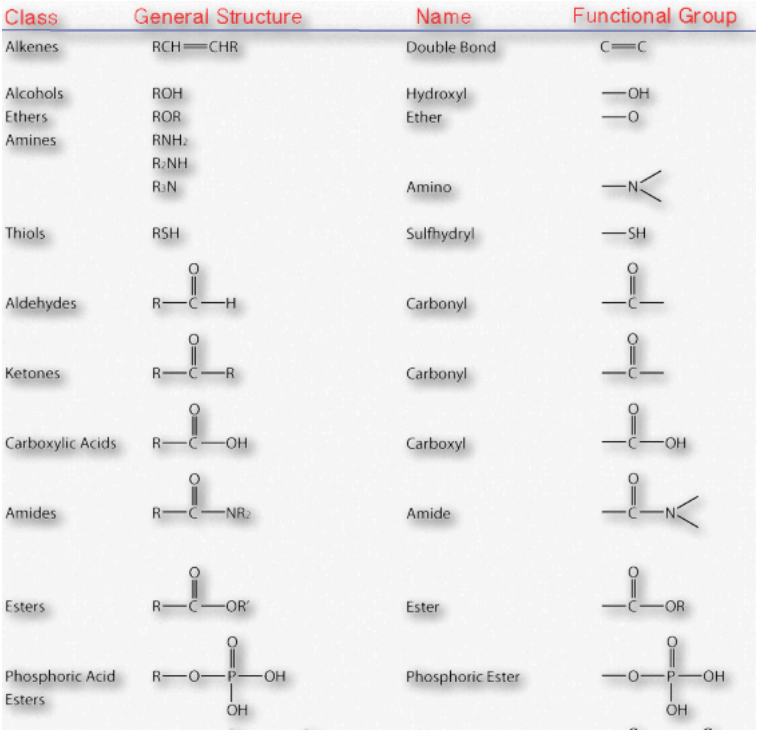


Figure 1.18 – *Important functional groups in biochemistry Image by Aleia Kim*

In addition to functional group theory, you will also make use of an understanding of covalent and ionic bonds as you learn biochemistry.

Covalent bonds, formed from the sharing of electrons between two atoms, hold together the structures of proteins, carbohydrates, lipids and nucleic acids. Ions and ionic bonds, formed when one atom donates an electron to another, appear in charged polyatomic ions and in dissolved minerals in cellular and body fluids.

Electronegativity

Electronegativity is a measure of the affinity a nucleus has for outer shell electrons (Table 1.2). High electronegativity corresponds to high affinity. Electrons in a covalent bond are held closer to the nucleus of an element with greater electronegativity compared to a nucleus with lower electronegativity.

Table 1.2

Electronegativities of Various Atoms

Atom	Electronegativity
Oxygen	3.5
Nitrogen	3.0
Sulfur	2.6
Carbon	2.5
Phosphorous	2.2
Hydrogen	2.1

Table 1.2 Image by Aleia Kim

For example, in a molecule of water, with hydrogen covalently bonded to oxygen, the electrons are "pulled" toward the oxygen, which is more electronegative. Because of this, there is a slightly greater negative charge near the oxygen atom on a water molecule, compared to the hydrogen. The unequal charge distribution in this bond makes it a polar covalent bond. The 3D arrangement of atoms in water, along with two polar covalent bonds, results in water exhibiting a molecular dipole with one side being somewhat negative and the other somewhat positive.

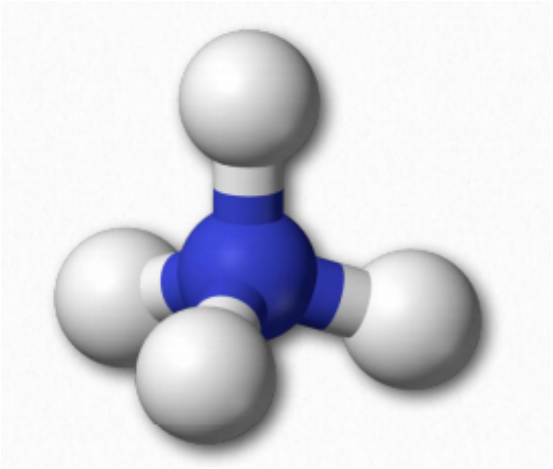


Figure 1.19 – *Tetrahedral structure Wikipedia*

Hydrogen bonds are intermolecular forces that occur between different molecules in close proximity to one another. In water, hydrogen bonding interactions are plentiful. They are the result of the attraction of the partial positive and partial negative ends of the molecular dipole (Figure 1.20).

Hydrogen bonds can also form between hydrogens with a partial positive charge and other strongly electronegative atoms, like nitrogen, with a partial negative charge.

Bonds between hydrogen and carbon do not form significant partial charges because the electronegativities of the two atoms are similar. Consequently, molecules containing many carbon-hydrogen bonds will not form hydrogen bonds and therefore, do not mix well with water. Such molecules are called hydrophobic. Other compounds with the ability to make hydrogen bonds are polar and can dissolve in water. They are called hydrophilic. Molecules possessing both characteristics are called amphiphilic.

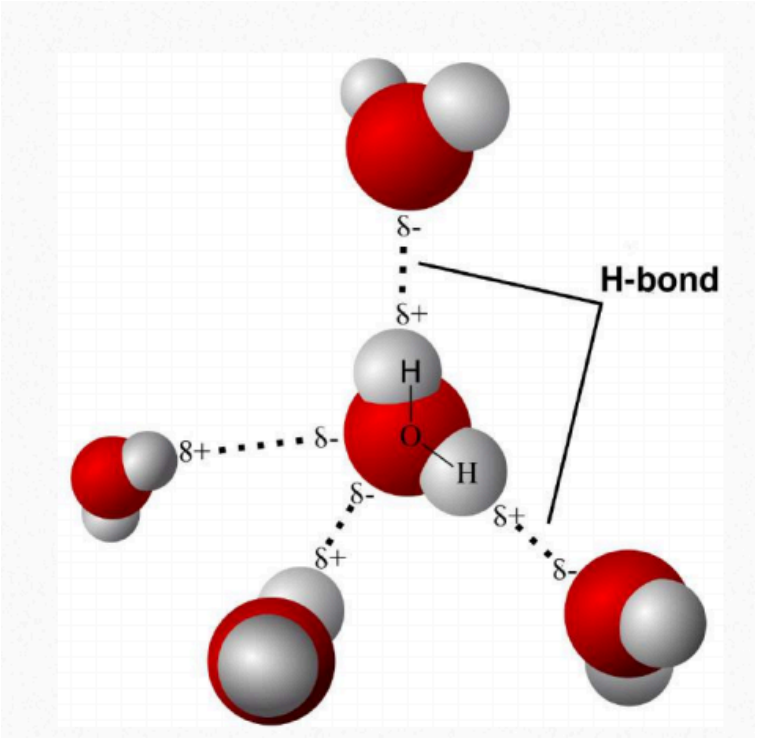


Figure 1.20 *Hydrogen bonds (dotted lines) between water molecules Wikipedia*

Weak interactions

Hydrogen bonds are one kind of intermolecular interaction between molecules. Other forms of intermolecular interactions that are important in biochemistry include weak interactions between a polar molecule and a transient dipole, or between two temporary dipoles. Although the individual hydrogen bonds, dipole-dipole or

other interactions are weak, they can result in quite strong interactions between molecules because of their large numbers.

Oxidation/reduction chemistry

Oxidation-reduction chemistry is common in biochemistry. Oxidation involves loss of electrons and reduction involves the gain of electrons. For every biological oxidation, there is a corresponding reduction – one molecule loses electrons to another molecule. Oxidation reactions tend to release energy and are a source of bioenergy for chemotrophic cells.

Ionization

Ionization of biomolecules, by contrast does not involve oxidation/reduction. In ionization, a hydrogen ion (H+) leaves behind its electron as it exits (leaving behind a negative charge) or joins a group (adding a positive charge). Biological ionizations typically involve carboxylic acid groups or amines, though phosphates or sulfates can also be ionized. A carboxylic acid group can have two ionization states – a zero charge form with the carboxylic acid holding its proton, and a -1 form where the hydrogen ion has been released. An amine also has two ionization states. A charge of zero on an amine corresponds to a nitrogen with three covalent bonds (usually in the form of C-NH2) and a charge of +1 corresponds to a nitrogen making four covalent bonds (usually X-NH3 +).

Stereochemistry

A carbon has the ability to make up to four single bonds (forming a tetrahedral structure). If it bonds to four different chemical groups, two different arrangements are possible, producing stereochemical "handedness"or chirality (Figure 1.21).

Enzymes have very specific 3-D structures, so the chemical reactions that involve enzyme catalysis are stereospecific: e.g. they produce just one stereoisomer. By contrast, the same molecules made without chiral reagents or processes produce equal amounts of both isomers.

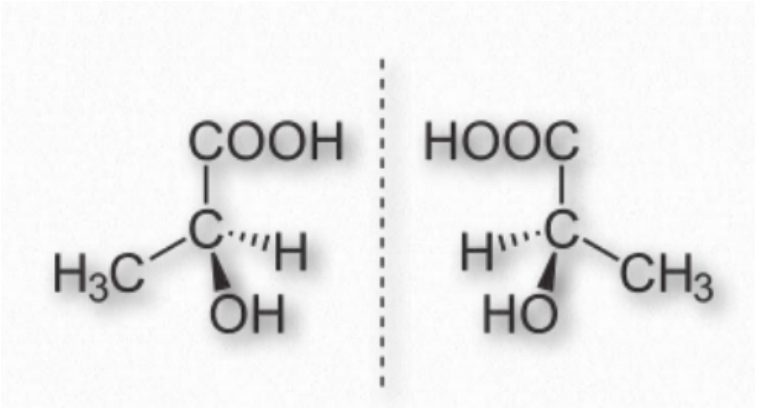


Figure 1.21 – *Mirror images of lactic acid*

Gibbs free energy

Energy calculations allow us to predict whether a reaction will be spontaneous, occurring without an input of energy. The factors that influence spontaneity include change in enthalpy (ΔH) and change in entropy (ΔS). Enthalpy is related to heat flow, and reactions that are favorable often give off heat. Entropy is related to order or

organization, and reactions that are favorable often involve a decrease in the order or organization of a system. These loose conceptual definitions are useful in casual considerations, but more structured definitions exist.

Enthalpy change and entropy change play off one another (with temperature another relevant factor) in this fashion:

$$G = H - TS$$

For a process, the change in the Gibbs free energy ΔG is given by \[ΔG = ΔH − TΔS\]

A negative ΔG corresponds to release of free energy. Reactions that release energy are exergonic and will occur spontaneously, whereas those that absorb energy are called endergonic and only occur with a continual input of energy from an external source.

In biochemistry, studies often have used a standard Gibbs free energy change ($\Delta G^{o'}$), measured with standard conditions of temperature, pressure, and concentration, and at pH = 7.

Catalysis

Catalysis is an increase in the rate of a reaction induced by a substance that is, itself, unchanged by the reaction. Because catalysts remain unchanged at the end of a reaction, a single catalyst molecule can be reused for many reaction cycles.

Proteins that catalyze reactions in cells are called enzymes, while ribozymes are RNA molecules that act as catalysts. Almost all biochemical reactions occur with the assistance of catalysts, which serve to encourage reactions but also control the chemistry that occurs in cells.

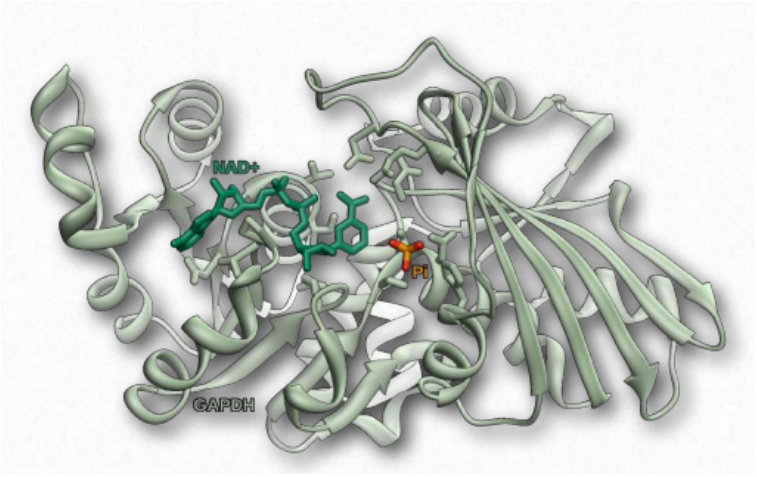


Figure 1.22 – Glyceraldehyde-3-phosphate dehydrogenase in the midst of catalysis Wikipedia

1.3: Introduction - Water and Buffers

When it comes to water, we're literally drowning in it, as water is by far the most abundant component of every cell. To understand life, we begin the discussion with the basics of water, because everything that happens in cells, even reactions buried deep inside enzymes, away from water, is influenced by water's chemistry.

The water molecule has wide 'V' shape (the HO-H angle is 104°) with uneven sharing of electrons between the oxygen and the hydrogen atoms (Figure 1.23). Oxygen, with its higher electronegativity, holds electrons closer to itself than the hydrogens do. The hydrogens, as a result, are described as having a partial positive charge (typically designated as δ+) and the oxygen has a partial negative charge (written as δ-). Thus, water is a polar molecule because charges are distributed around it unevenly, not symmetrically.

Water as a solvent

Water (Figure 1.23) is described as a solvent because of its ability to dissolve many, but not all, molecules. Molecules that are ionic or polar dissolve readily in water, but non-polar substances dissolve poorly in water, if at all. Oil, for example, which is non-polar, separates from water when mixed with it. On the other hand, sodium chloride, which ionizes, and ethanol, which is polar, are able to form hydrogen bonds, so both dissolve in water. Ethanol's solubility in water is crucial for brewers, winemakers, and distillers – but for this property, there would be no wine, beer or spirits.

As explained in an earlier section, we use the term hydrophilic to describe substances that interact well with water and dissolve in it and the term hydrophobic to refer to materials that are non-polar and do not dissolve in water. Table 1.3 illustrates some hydrophilic and hydrophobic substances. A third term, amphiphilic, refers to compounds that have both properties. Soaps, for example are amphiphilic, containing a long, non-polar aliphatic tail and a head that ionizes.

Hydrophilic vs Hydrophobic Compounds

Hydrophobic	Hydrophilic
Nonpolar hydrocarbons (hexane)	Ionic compounds (NaCl)
Lipids (fats and cholesterol)	Polar organic compounds (alcohols, ketones or carbonyls)
	Weak acids (phosphates, amino acids)
	Sugars/carbohydrates

Table 1.3 Image by Aleia Kim

Solubility

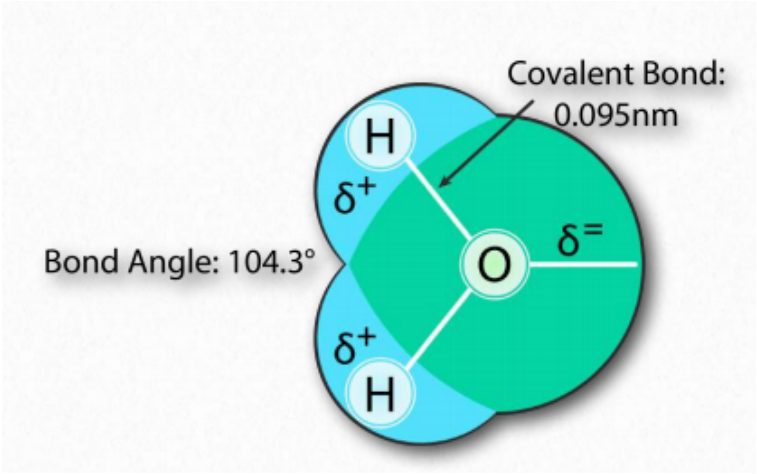


Figure 1.23 – Arrangement of atoms in water Image by Aleia Kim

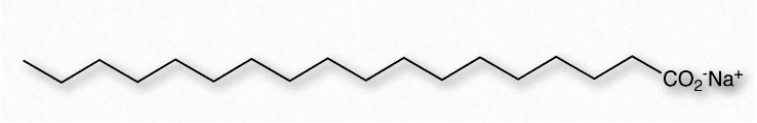


Figure 1.24 – Structure of a Soap

Let's consider why non-polar materials do not dissolve in water. We could imagine a situation where the process of dissolving involves the "surrounding" of each molecule of the nonpolar solute in water, just like each sodium and each chloride ion gets surrounded by water molecules as salt dissolves.

Water organization

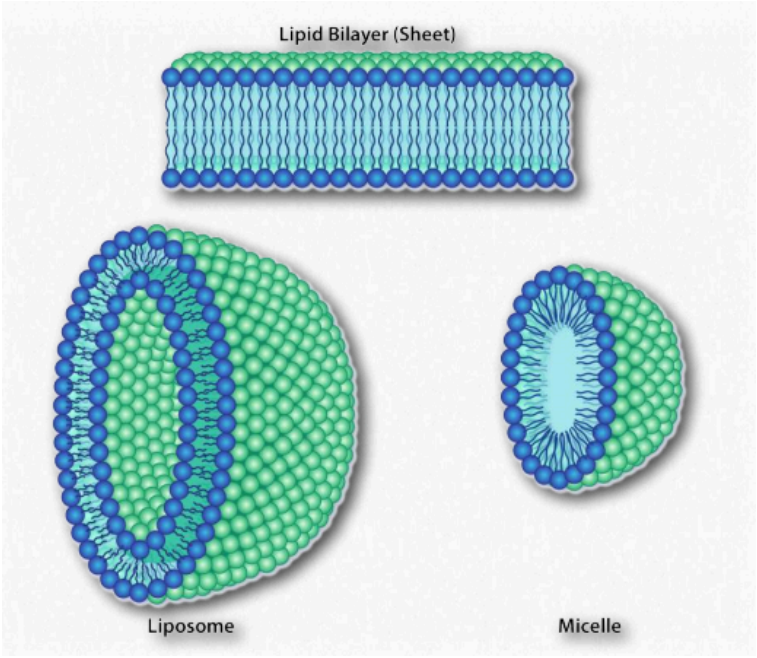


Figure 1.25 – Structures formed by amphiphilic substances in water. Image by Aleia Kim

There is a significant difference, though between surrounding a non-polar molecule with water molecules and surrounding ions (or polar compounds) with water molecules.

The difference is that since non-polar molecules don't really interact with water via intermolecular attractive interactions. The water behaves very differently than it does with ions or molecules that form hydrogen bonds. In fact, around each non-polar molecule, water gets very organized, aligning itself regularly. This increase in organization or order, described more technically as an increase in entropy, tends to be unfavorable in nature.

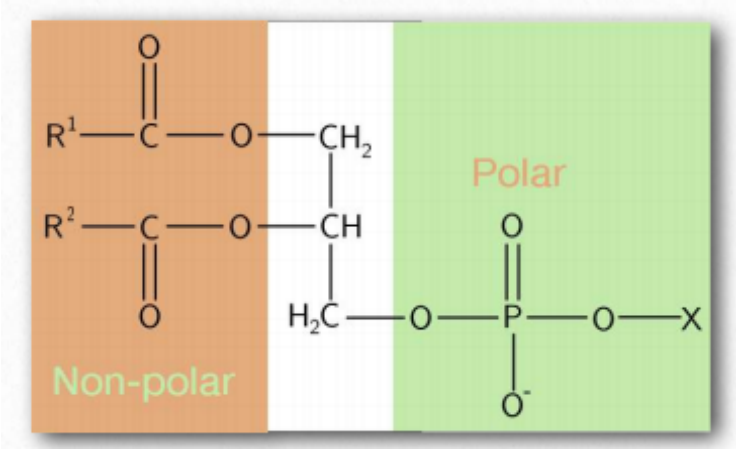


Figure 1.26 – A phospholipid – an amphiphilic substance

Further, when the non-polar material associates with itself and not water, then the water molecules are free to mix with one another, without being ordered. This type of increase in disorder is favored in nature. Therefore this phenomenon, described as an increase in entropy, drives the separation of non-polar substances from aqueous solutions.

Figure 1.27 – Vinegar (black) and oil (yellow) A mix of polar and nonpolar compounds Wikipedia

Amphiphilic substances

Next, we consider mixing of an amphiphilic substance, such as a soap, with water (Figure 1.24). The sodium ions attached to the fatty acids in soap readily come off in aqueous solution, leaving behind a negatively charged molecule at one end and a non-polar region at the other end. The ionization of the soap causes in an increase in entropy – two particles instead of one. The non-polar portion of the negatively charged soap ion is problematic – if exposed to water, it will cause water to organize and result in a decrease of entropy.

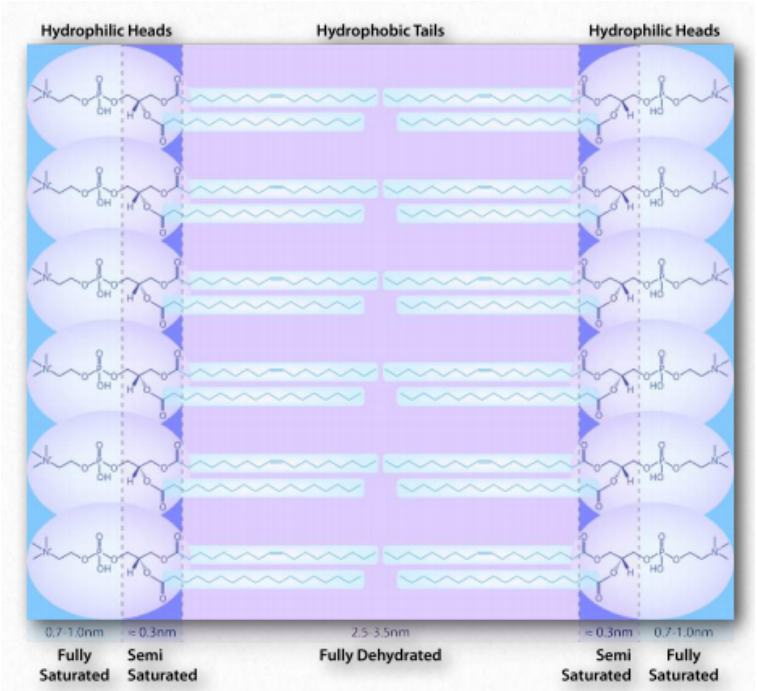


Figure 1.28 – Environment of a lipid bilayer. Water is concentrated away from the hydrophobic center, being saturated on the outside, semi-saturated near the head-tail junction and fully dehydrated in the middle. Image by Aleia Kim

Since we know fatty acids dissolve in water, there must be something else at play. There is. Just like the non-polar molecules in the first example associated with each other and not water, so too do the non-polar portions of the soap ions associate with each other and exclude water. The result is that the soap ions arrange themselves as micelles (Figure 1.25) with the non-polar portions on the interior of the structure away from water and the polar portions on the outside interacting with water.

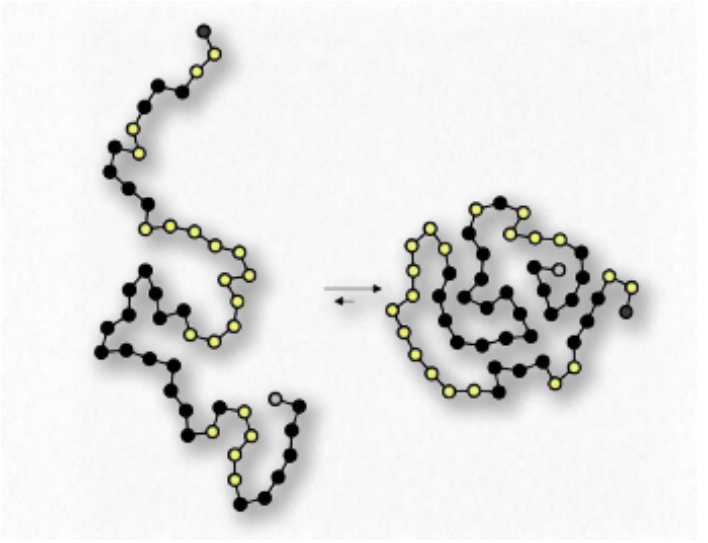


Figure 1.29 – Protein folding arranges hydrophobic amino acids (black dots) inside the protein

The interaction of the polar heads with water returns the water to its more disordered state. This favorable increase in disorder drives the formation of micelles.

A similar phenomenon based on the same forces drives the formation of bilayers in biological membranes. The amphiphilic molecules are different in that case. However the non-polar

portions of the molecules interact with each other to exclude water and the polar portions arrange themselves on the outsides of the bilayer (Figure 1.28).

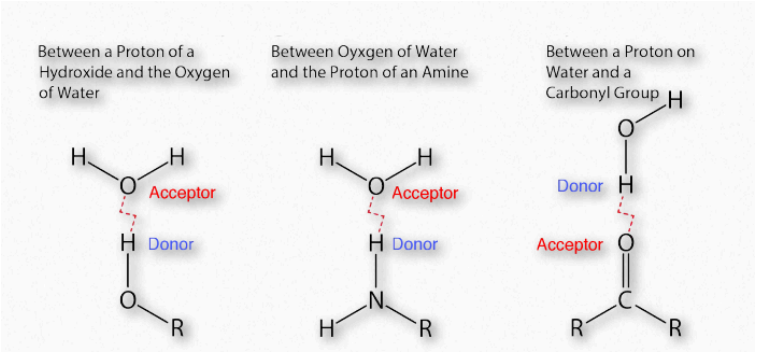


Figure 1.30 – Common hydrogen bonds in biochemistry Image by Aleia Kim

Yet another example is seen in the folding of globular proteins in the water interiors of cells. Nonpolar amino acids are found in the interior portions of proteins (water excluded). Interaction of the non-polar amino acids turns out to be a driving force for the folding of proteins as they are being made in a water-based solution.

Hydrogen bonds

The importance of hydrogen bonds in biochemistry (Figure 1.30) is hard to overstate. Linus Pauling himself said,

" *. . . . I believe that as the methods of structural chemistry are further applied to physiological problems it will be found that the significance of the hydrogen bond for physiology is greater than that of any other single structural feature.*"

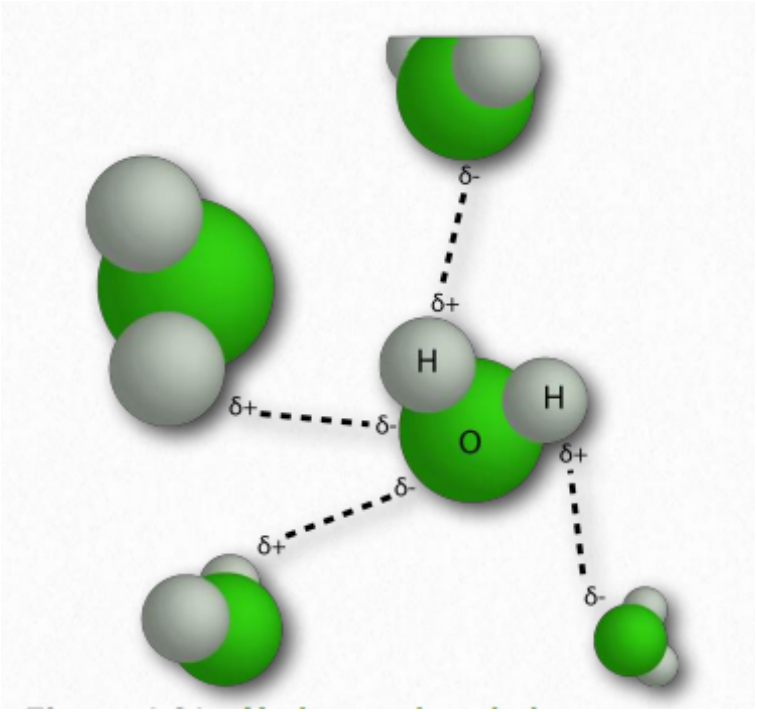


Figure 1.31 – Hydrogen bonds between water molecules Image by Pehr Jacobson

In 2011, an IUPAC task group gave an evidence-based definition of hydrogen bonding that states,

"*The hydrogen bond is an attractive interaction between a hydrogen atom from a molecule or a molecular fragment* X–H *in which* X *is more electronegative than* H*, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation.*"

As a reminder, hydrogen-bonding interactions are an intermolecular force that occurs between atoms on one molecule and those on another. These attractive forces are a consequence of highly polarized bonds: where one of the interacting molecules contains hydrogen attached to a very electronegative atom (F, O, or N) and the other contains lone pair, nonbonding electrons on an electronegative atom (F, O, N, or a halogen).

Partial Charges

Hydrogen bonding attractive interactions are due to partial charges on the interacting atoms. These tiny charges (δ+ and δ-) result when the partial positive charge of a hydrogen atom is attracted to the partial negative of another molecule.

In water, that means the hydrogen of one water molecule is attracted to the oxygen of another (Figure 1.31). Since water is an asymmetrical molecule, it means also that the charges are asymmetrical. Such an uneven distribution is what makes a molecular dipole. Polar molecules are important for interactions with other polar molecules and for dissolving ionic substances (Figure 1.32).

Hydrogen bonds are not exclusive to water. In fact, they are important forces holding together macromolecules that include proteins and nucleic acids. Hydrogen bonds occur within and between macromolecules.

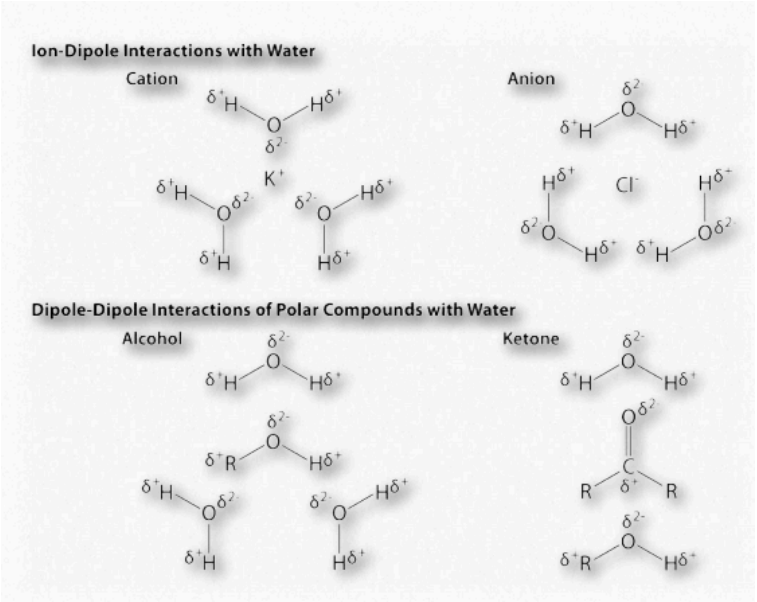


Figure 1.32 – Example dipole interactions in biochemistry Image by Aleia Kim

The complementary pairing that occurs between bases in opposite strands of DNA, for example, is based on hydrogen bonds. Each hydrogen bond is relatively weak (compared to a covalent bond, for example – Table 1.4), but collectively they can be quite strong.

Bond Energies

	Type of Bond	Bond Energy (kJ/mol)
Covalent Bonds	C—H	413
	O—H	460
Noncovalent Bonds	Hydrophobic Interaction	4-12
	Hydrogen Bond	20
	Ion-dipole Interaction	20

Table 1.4 Image by Aleia Kim

Benefits of weak interactions

Their weakness, however, is actually quite beneficial for cells, particularly as regards nucleic acids (Figure 1.33). The strands of DNA, for example, must be separated over short stretches in the processes of replication and the synthesis of RNA. Since only a few base pairs at a time need to be separated, the energy required to do this is small and the enzymes involved in the processes can readily take them apart, as needed. Hydrogen bonds also play roles in binding of substrates to enzymes, catalysis, and protein-protein interaction, as well as other kinds of binding, such as protein-DNA, or antibody-antigen.

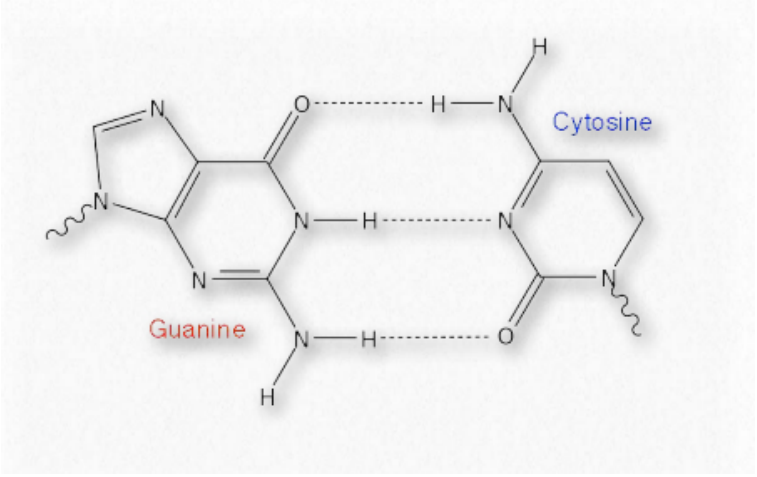


Figure 1.33 – Hydrogen bonds in a base pair of DNA Image by Aleia Kim

As noted, hydrogen bonds are weaker than covalent bonds (Table 1.4) and their strength varies form very weak (1-2 kJ/mol) to fairly strong (29 kJ/mol). Hydrogen bonds only occur over relatively short distances (2.2 to 4.0 Å). The farther apart the hydrogen bond distance is, the weaker the bond is.

The strength of the bond in kJ/mol represents the amount of heat that must be put into the system to break the bond – the larger

the number, the greater the strength of the bond. Hydrogen bonds are readily broken using heat. The boiling of water, for example, requires breaking of H-bonds. When a biological structure, such as a protein or a DNA molecule, is stabilized by hydrogen bonds, breaking those bonds destabilizes the structure and can result in denaturation of the substance – loss of structure. It is partly for this reason that most proteins and all DNAs lose their native, or folded, structures when heated to boiling.

Weak Acid pKa Values

Name	Chemical Structure of Acid	Chemical Structure of Salt	pKa
Acetic Acid	CH_3COOH	CH_3COO^-	4.76
Formic Acid	$HCOOH$	$HCOO^-$	3.75
Lactic Acid	$CH_3CHOHCOOH$	$CH_3CH—HCOO^-$	3.86
Pyruvic Acid	$CH_3COCOOH$	$CH_3C—COO^-$	2.50
Oxalic Acid (1)	$HOOC—COOH$	$HOOC—COO^-$	1.23
Oxalic Acid (2)	$HOOC—COO^-$	$OOC—COO^-$	4.19
Carbonic Acid (1)	H_2CO_3	HCO_3^-	6.37
Carbonic Acid (2)	HCO_3^-	CO_3^{2-}	10.20
Malic Acid (1)	$HOOC—CH_2—CHOH—COOH$	$HOOC—CH_2—CHOH—COO^-$	3.40
Malic Acid (2)	$HOOC—CH_2—CHOH—COO^-$	$^-OOC—CH_2—CHOH—COO^-$	5.26
Malonic Acid (1)	$HOOC—CH_2—COOH$	$HOOC—CH_2—COO^-$	2.83
Malonic Acid (2)	$HOOC—CH_2—COO^-$	$^-OOC—CH_2—COO^-$	5.69
Phosphoric Acid (1)	H_3PO_4	$H_2PO_4^-$	2.14
Phosphoric Acid (2)	$H_2PO_4^-$	HPO_4^{2-}	7.20
Phosphoric Acid (3)	HPO_4^{2-}	PO_4^{3-}	12.40
Succinic Acid (1)	$HOOC—CH_2—CH_2—COOH$	$HOOC—CH_2—CH_2—COO^-$	4.21
Succinic Acid (2)	$HOOC—CH_2—CH_2—COO^-$	$^-OOC—CH_2—CH_2—COO^-$	5.63

Image by Aleia Kim Table 1.5

For DNA molecules, denaturation results in complete separation of the strands from each other. For most proteins, this means loss of their characteristic three-dimensional structure and with it, loss of the function they performed. Though a few proteins can readily reassume their original structure when the solution they are in is cooled, most can't. This is one of the reasons that we cook our food. Proteins are essential for life, so denaturation of bacterial proteins results in death of any microorganisms contaminating the food.

The importance of buffers

Water can ionize to a slight extent (10^{-7} M) to form H^+ (proton) and OH^- (hydroxide). We measure the proton concentration of a solution with pH, which is the negative log of the proton concentration.

$pH = -Log[H^+]$

If the proton concentration, $[H^+]= 10^{-7}$ M, then the pH is 7. We could just as easily measure the hydroxide concentration with the pOH by the parallel equation,

$pOH = -Log[OH^-]$

In pure water, dissociation of a proton simultaneously creates a hydroxide, so the pOH of pure water is 7, as well. This also means that

$pH + pOH = 14$

Now, because protons and hydroxides can combine to form water, a large amount of one will cause there to be a small amount of the other. Why is this the case? In simple terms, if I dump 0.1 moles of H+ into a pure water solution, the high proton concentration will react with the relatively small amount of hydroxides to create water, thus reducing hydroxide concentration. Similarly, if I dump excess hydroxide (as NaOH, for example) into pure water, the proton concentration falls for the same reason.

Acids vs bases

Chemists use the term “acid” to refer to a substance which has protons that can dissociate (come off) when dissolved in water. They use the term “base” to refer to a substance that can absorb protons when dissolved in water. Both acids and bases come in strong and weak forms. (Examples of weak acids are shown in Table 1.5.) Strong

acids, such as HCl, dissociate completely in water. If we add 0.1 moles (6.02×10^{22} molecules) of HCl to a solution to make a liter, it will have 0.1 moles of H^+ and 0.1 moles of Cl^- or 6.02×10^{22} molecules of each . There will be no remaining HCl when this happens. A strong base like NaOH also dissociates completely into Na^+ and OH^- .

Clearing Confusion - Students are often puzzled and expect that

$$[H^+] = [A^-]$$

because the dissociation equation shows one of each from HA. This is, in fact, true ONLY when HA is allowed to dissociate in pure water. Usually the HA is placed into solution that has protons and hydroxides to affect things. Those protons and /or hydroxides change the H^+ and A^- concentration unequally, since A^- can absorb some of the protons and/or HA can release H^+ when influenced by the OH^- in the solution. Therefore, one must calculate the proton concentration from the pH using the Henderson Hasselbalch equation.

$$pH = pK_a + \log([Ac^-]/[HAc])$$

Weak Acids

Weak acids and bases differ from their strong counterparts. When you put one mole of acetic acid (HAc) into pure water, only about 5% of the HAc molecules dissociate into H^+ and Ac^- . Clearly, weak acids are very different from strong acids. Weak bases behave similarly, except that they accept protons, rather than donate them. Since we can view everything as a form of a weak acid, we will not use the term weak base here.

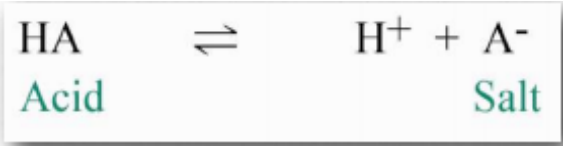
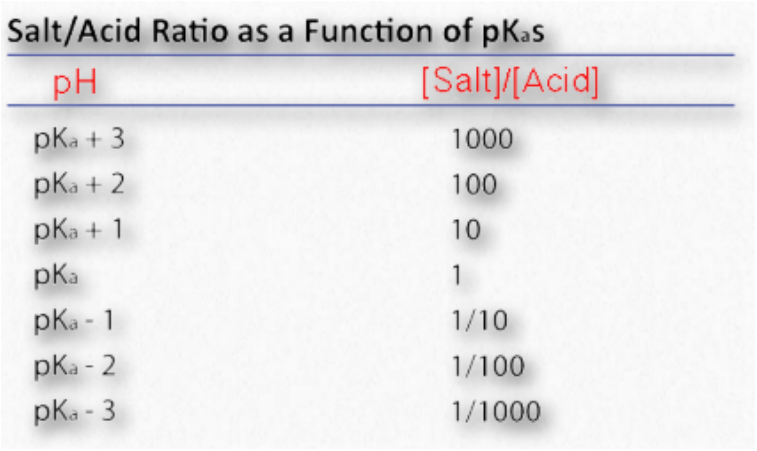


Figure 1.34 – Dissociation of a weak acid Image by Aleia Kim

Students are often puzzled and expect that $[H^+] = [A^-]$ because the dissociation equation shows one of each from HA. This is, in fact, true ONLY when HA is allowed to dissociate in pure water. Usually the HA is placed into solution that has protons and hydroxides to affect things. Those protons and /or hydroxides change the H^+ and A^- concentration unequally, since A^- can absorb some of the protons and/or HA can release H^+ when influenced by the OH^- in the solution. Therefore, one must calculate the proton concentration from the pH using the Henderson Hasselbalch equation.

$$pH = pKa + \log([Ac^-]/[HAc])$$



Salt/Acid Ratio as a Function of pK_as

pH	[Salt]/[Acid]
$pK_a + 3$	1000
$pK_a + 2$	100
$pK_a + 1$	10
pK_a	1
$pK_a - 1$	1/10
$pK_a - 2$	1/100
$pK_a - 3$	1/1000

Image by Aleia Kim Table 1.6

Solutions that contain a weak acid and also its related ('conjugate') base act as buffers. These solutions resist changes to pH even when additional acid or base is added to the solution. Biological systems almost always contain these buffers, with physiological pH values 'set' by them at levels that are optimal for the functioning of that cell or cell compartment.

How buffers work: analogy to a UPS

Weak acids are critical for life because their affinity for protons causes them to behave like a UPS. We're not referring to the UPS that is the United Parcel Service, but instead, to the encased battery backup systems for computers called Uninterruptible Power Supplies that kick on to keep a computer running during a power

failure. The battery in a laptop computer is a UPS, for example.

We can think of weak acids as Uninterruptible Proton Suppliers within certain pH ranges, providing (or absorbing) protons as needed. Weak acids thus help to keep the H+ concentration (and thus the pH) of the solution they are in relatively constant.

Consider the bicarbonate/carbonic acid system. Figure 1.35 shows what happens when H_2CO_3dissociates. Adding hydroxide ions (by adding a strong base like NaOH) to the solution causes the H^+ ions to react with OH^- ions to make water. Consequently, the concentration of H^+ ions would go down and the pH would go up.

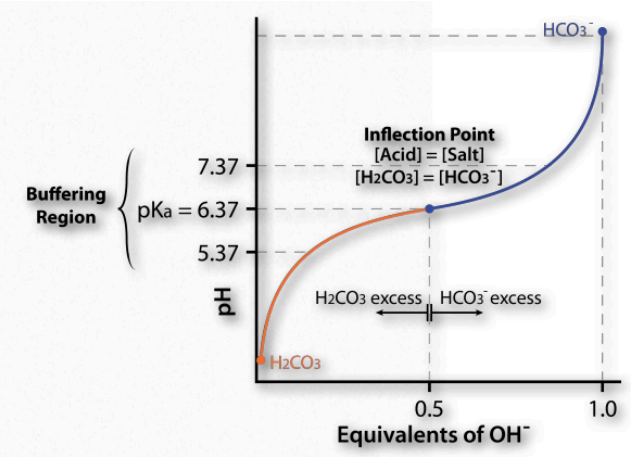


Figure 1.35 – Titration curve for carbonic acid Image by Aleia Kim

However, in contrast to the situation with a solution

of pure water, there is a backup source of H^+ available in the form of H_2CO_3. Here is where the UPS function kicks in. As protons are taken away by the added hydroxyl ions (making water), they are partly replaced by protons from the H_2CO_3. This is why a weak acid is a buffer. It resists changes in pH by releasing protons to compensate for those "used up" in reacting with the hydroxyl ions.

Why do we care about pH? Because biological molecules can, in some cases, be exquisitely sensitive to changes in it. As the pH of a solution changes, the charges of molecules in the solution can change, as you will see. Changing charges on biological molecules, especially proteins, can drastically affect how they work and even whether they work at all.

Henderson-Hasselbalch

It is useful to be able to predict the response of the H_2CO_3 system to changes in H^+ concentration. The Henderson-Hasselbalch equation

defines the relationship between pH and the ratio of HCO_3^- and H_2CO_3. It is

$pH = pKa + \log([HCO_3^-]/[H_2CO_3])$

This simple equation defines the relationship between the pH of a solution and the ratio of HCO_3^- and H_2CO_3 in it. The new term, called the pKa, is defined as

$pKa = -\text{Log } K_a$,

just as

$pH = -\text{Log } [H^+]$.

The Ka is the acid dissociation constant and is a measure of the strength of an acid. For a general acid, HA, which dissociates as

$HA \rightleftarrows H^+ + A^-$, $K_a = [H^+][A^-]/[HA]$

Thus, the stronger the acid, the more protons that will dissociate from it when added to water and the larger the value its Ka will have. Large values of Ka translate to lower values of pKa. As a result, the lower the pKa value is for a given acid, the stronger the weak acid is.

Constant pKa

Please note that pKa is a constant for a given acid. The pKa for carbonic acid is 6.37. By comparison, the pKa for formic acid is 3.75. Formic acid is therefore a stronger acid than acetic acid. A stronger acid will have more protons dissociated at a given pH than a weaker acid.

Now, how does this translate into stabilizing pH? Figure 1.35 shows a titration curve. In this curve, the titration begins with the conditions at the lower left (very low pH). At this pH, the H_2CO_3 form predominates, but as more and more OH- is added (moving to the 45 Why do we care about pH? Because biological molecules can, in some cases, be exquisitely sensitive to changes in it. As the pH of a solution changes, the charges of molecules in the solution can

change, as you will see. Changing charges on biological molecules, especially proteins, can drastically affect how they work and even whether they work at all right), the pH goes up, the amount of HCO3- goes up and (correspondingly), the amount of H_2CO_3 goes down. Notice that the curve "flattens" near the pKa (6.37).

Buffering region

Flattening of the curve tells us is that the pH is not changing much (not going up as fast) as it did earlier when the same amount of hydroxide was added. The system is resisting a change in pH (not stopping the change, but slowing it) in the region of about one pH unit above and one pH unit below the pKa. Thus, the buffering region of the carbonic acid/ bicarbonate buffer is from about 5.37 to 7.37. It is maximally strong at a pH of 6.37.

Now it starts to become apparent how the buffer works. HA can donate protons when extras are needed (such as when OH- is added to the solution by the addition of NaOH). Similarly, A- can accept protons when extra H+ are added to the solution (adding HCl, for example). The maximum ability to donate or accept protons comes when

$[A^-] = [HA]$

This is consistent with the Henderson Hasselbalch equation and the titration curve. When $[A^-] = [HA]$, pH = 6.37 + Log(1). Since Log(1) = 0, pH = 6.37 = pKa for carbonic acid. Thus for any buffer, the buffer will have maximum strength and display flattening of its titration curve when $[A^-] = [HA]$ and when pH = pKa. If a buffer has more than one pKa (Figure 1.36), then each pKa region will display the behavior.

Buffered vs non-buffered

To understand how well a buffer protects against changes in pH,

consider the effect of adding .01 moles of HCl to 1.0 liter of pure water (no volume change) at pH 7, compared to adding it to 1.0 liter of a 1M acetate buffer at pH 4.76. Since HCl completely dissociates, in 0.01M (10-2 M) HCl you will have 0.01M H+. For the pure water, the pH drops from 7.0 down to 2.0 (pH = -log(0.01M)).

By contrast, the acetate buffer's pH after adding the same amount of HCl is 4.74. Thus, the pure water solution sees its pH fall from 7 to 2 (5 pH units), whereas the buffered solution saw its pH drop from 4.76 to 4.74 (0.02 pH units). Clearly, the buffer minimizes the impact of the added protons compared to the pure water.

Buffer capacity

It is important to note that buffers have capacities limited by their concentration. Let's imagine that in the previous paragraph, we had added the 0.01 moles HCl to an acetate buffer that had a concentration of 0.01M and equal amounts of Ac- and HAc. When we try to do the math in parallel to the previous calculation, we see that there are 0.01M protons, but only 0.005M A- to absorb them. We could imagine that 0.005M of the protons would be absorbed, but that would still leave 0.005M of protons unbuffered. Thus, the pH of this solution would be approximately

pH = -log(0.005M) = 2.30

Exceeding buffer capacity dropped the pH significantly compared to adding the same amount of protons to a 1M acetate buffer. Consequently, when considering buffers, it is important to recognize that their concentration sets their limits. Another limit is the pH range in which one hopes to control proton concentration.

Multiple ionizable groups

Now, what happens if a molecule has two (or more) ionizable groups? It turns out, not surprisingly, that each group will have its own pKa and, as a consequence, will have multiple regions of buffering.

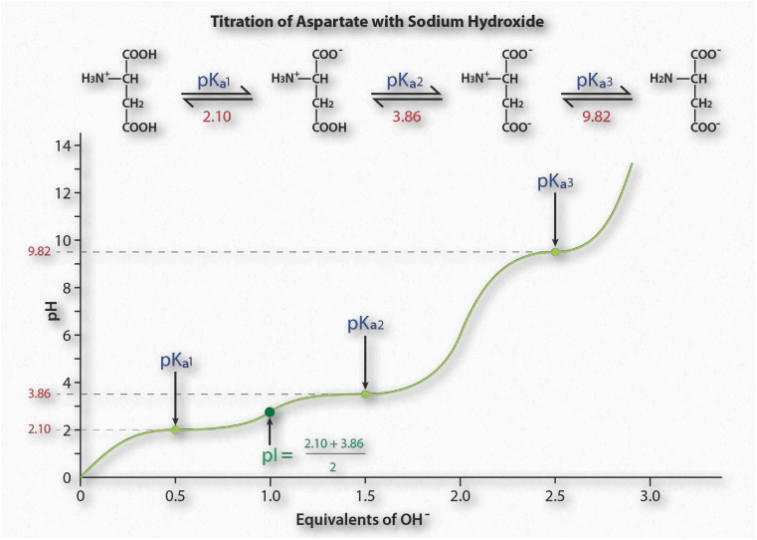


Figure 1.36 – Titration of an acidic amino acid Image by Aleia Kim

Figure 1.36 shows the titration curve for the amino acid aspartic acid. Note that in- stead of a single flattening of the curve, as was seen for acetic acid, aspartic acid's titration curve displays three such regions. These are individual buffering regions, each centered on the respective pKa values for the carboxyl group and the amine group.

Aspartic acid has four possible charges: +1 (α-carboxyl group, α-amino group, and Rgroup carboxyl each has a proton), 0 (α- carboxyl group missing proton, α- amino group has a proton, R-group carboxyl has a proton), -1 (α-carboxyl group and R-group carboxyl each lack a proton, α-amino group retains a proton), -2 (α-carboxyl, R-group carboxyl, and α-amino groups all lack extra proton).

Prediction

How does one predict the charge for an amino acid at a given pH? A good rule of thumb for estimating charge is that if the pH is more than one unit below the pKa for a group (carboxyl or amino), the proton is on. If the pH is more than one unit above the pKa for the group, the proton is off. If the pH is NOT more than one or less than one pH unit from the pKa, this simple assumption will not work.

Further, it is important to recognize that these rules of thumb are estimates only. The pI (pH at which the charge of a molecule is zero) is an exact value calculated as the average of the two pKa values on either side of the zero region. It is calculated at the average of the two pKa values around the point where the charge of the molecule is zero. For aspartic acid, this corresponds to $pK_{a1}$and pK_{a2}.

References

1. http://www.lpi.usra.edu/lunar/missions/apollo/ apollo_12/ experiments/surveyor/
2. Arunan, Elangannan; Desiraju, Gautam R.; Klein, Roger A.; Sadlej, Joanna; Scheiner, Steve; Alkorta, Ibon; Clary, David C.; Crabtree, Robert H.; Dannenberg, Joseph J.; Hobza, Pavel; Kjaergaard, Henrik G.; Legon, Anthony C.; Mennucci, Benedetta; Nesbitt, David J. (2011). "Definition of the hydrogen bond". Pure Appl. Chem. 83 (8): 1637–1641. doi:10.1351/PAC-REC-10-01-02

2.1: Prelude to Structure and Function

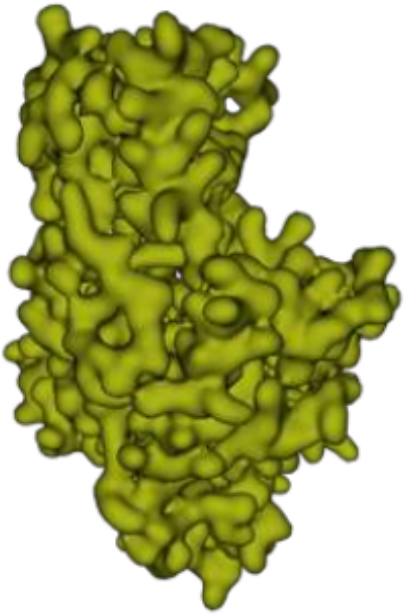
"*The man who does not read good books has no advantage over the man who cannot read them.*" Mark Twain

In this chapter, we will examine the structures of the major classes of biomolecules, with an eye to understanding how these structures relate to function.

As noted earlier, water is the most abundant molecule in cells, and provides the aqueous environment in which cellular chemistry happens. Dissolved in this water are ions including sodium, potassium and calcium. But the distinctiveness of biochemistry derives from the vast numbers of complex, large, carbon compounds, that are made by living cells.

You have probably learned that the major classes of biological molecules are proteins, nucleic acids, carbohydrates and lipids. The

first three of these major groups are **macromolecules** that are built as long polymers made up of smaller subunits or monomers, like strings of beads. The lipids, while not chains of monomers, also have smaller subunits that are assembled in various ways to make the lipid components of cells, including membranes. The chemical properties and three dimensional conformations of these molecules determine all the molecular interactions upon which life depends. Whether building structures within cells, transferring information, or catalyzing reactions, the activities of biomolecules are governed by their structures. The properties and shapes of macromolecules, in turn, depend on the subunits of which they are built.



***Interactive* 2.1:** *The enzyme Hexokinase: as for all enzymes, the activity of hexokinase depends on its structure. Protein Database (PDB)*

We will begin our more detailed examinations of biochemicals by considering proteins. Later we will return to consider the structures of the other biomolecules mentioned here.

2.2: Structure and Function - Amino Acids

All of the proteins on earth are made up of the same 20 amino acids. Linked together in long chains called polypeptides, amino acids are the building blocks for the vast assortment of proteins found in all living cells.

"*It is one of the more striking generalizations of biochemistry …that the twenty amino acids and the four bases, are, with minor reservations, the same throughout Nature.*" – Francis Crick

All amino acids have the same basic structure, shown in Figure 2.1. At the center of each amino acid is a carbon called the **α carbon** and attached to it are four groups – a hydrogen, a carboxylic acid group, an amine group, and an R-group, sometimes referred to as a variable group or side chain. The α carbon, carboxylic acid, and amino groups are common to all amino acids, so the R-group

is the only variable feature. With the exception of glycine, which has an R-group consisting of a hydrogen atom, all of the amino acids in proteins have four different groups attached to them and consequently can exist in two mirror isomeric forms.

The designations used in organic chemistry are not generally applied to amino acid nomenclature, but a similar system uses L and D to describe these enantiomers. Nature has not distributed the stereoisomers of amino acids equally. Instead, with only very minor exceptions, every amino acid found in cells and in proteins is in the L configuration.

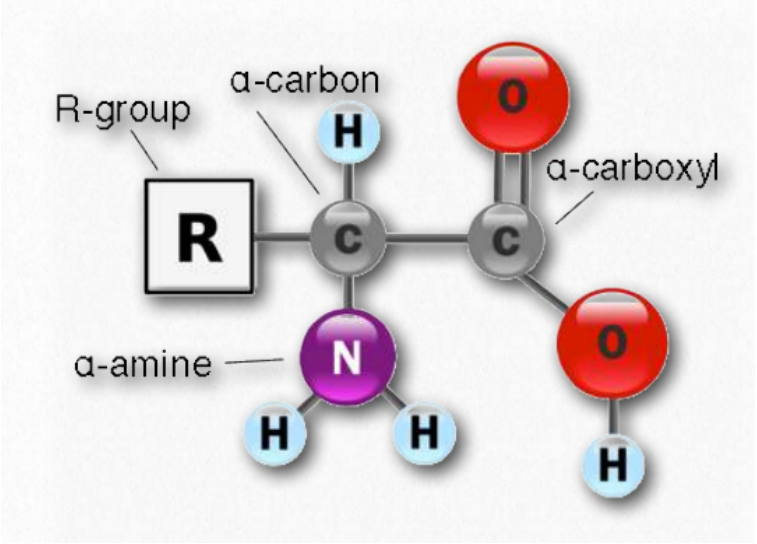


Figure 2.1 – General amino acid structure

There are 22 amino acids that are found in proteins and of these, only 20 are specified by the universal genetic code. The others, selenocysteine and pyrrolysine use tRNAs that are able to base pair with stop codons in the mRNA during translation. When this happens, these unusual amino acids can be incorporated into proteins. Enzymes containing selenocysteine, for example, include glutathione peroxidases, tetraiodothyronine 5′ deiodinases, thioredoxin reductases, formate dehydrogenases, glycine reductases, and selenophosphate synthetase. Pyrrolysine-

containing proteins are much rarer and are mostly confined to archaea.

Essential and non-essential

Nutritionists divide amino acids into two groups – essential amino acids and non-essential amino acids. Essential amino acids must be included in our diet because our cells can't synthesize them. What is essential varies considerably from one organism to another and even differ in humans, depending on whether they are adults or children. Table 2.1 shows essential and non-essential amino acids in humans.

Some amino acids that are normally nonessential, may need to be obtained from the diet in certain cases. Individuals who do not synthesize sufficient amounts of arginine, cysteine, glutamine, proline, selenocysteine, serine, and tyrosine, due to illness, for example, may need dietary supplements containing these amino acids.

Essential	Non-Essential
Histidine	Alanine
Isoleucine	Arginine
Leucine	Asparagine
Lysine	Aspartic acid
Methionine	Cysteine
Phenylalanine	Glutamic acid
Threonine	Glutamine
Tryptophan	Glycine
Valine	Proline
	Selenocysteine
	Serine
	Tyrosine

Table 2.1 - Essential and non-essential amino acids

Table 2.1 – Essential and non-essential amino acids

Non-protein amino acids

There are also amino acids found in cells that are not incorporated into proteins. Common examples include ornithine and citrulline. Both of these compounds are intermediates in the urea cycle, an important metabolic pathway.

R-group chemistry

Non-Polar	Carboxyl	Amine	Aromatic	Hydroxyl	Other
Alanine	Aspartic Acid	Arginine	Phenylalanine	Serine	Asparagine
Glycine	Glutamic Acid	Histidine	Tryptophan	Threonine	Cysteine
Isoleucine		Lysine	Tyrosine	Tyrosine	Glutamine
Leucine					Selenocysteine
Methionine					Pyrrolysine
Proline					
Valine					

Table 2.2 – Amino acid categories (based on R-group properties)

Amino acids can be classified based on the chemistry of their R-groups. It is useful to classify amino acids in this way because it is these side chains that give each amino acid its characteristic properties. Thus, amino acids with (chemically) similar side groups can be expected to function in similar ways, for example, during protein folding. The specific category divisions may vary, but all systems are attempts to organize and understand the relationship between an amino acid's structure and its properties or behavior as part of a larger system.

Non-polar amino acids

Amino acids in this group include:

- Alanine (Ala/A)
- Glycine (Gly/G)
- Isoleucine (Ile/I)
- Leucine (Leu/L)
- Methionine (Met/M)
- Valine (Val/V)

The amino acids in this group have nonpolar, hydrophobic R groups. When incorporated into globular proteins they tend to pack inward among other hydrophobic groups. In proteins that embed

themselves into or through membranes, these amino acids can orient themselves toward hydrophobic portions of the inside of the membrane.

The small R groups here are more readily packed into tight formations. Proline is exceptional in that it has an R group that folds back and covalently bonds to the backbone of the amino acid, creating a more rigid element in a protein chain that reduces free movement of the polypeptide chain. Additionally, proline can undergo hydroxylation reactions, stabilizing the protein structure. This occurs in collagen with the aid of ascorbic acid (Vitamin C). One symptom of the vitamin C deficiency syndrome 'scurvy' is the reduced quality of collagen in tissues, including the skin and gums. This can lead to the deterioration and loss of teeth.

Amino acid	Short	Abbrev.	Side chain	Hydro-phobic	pKa	Polar	pH	Small	Tiny	Aromatic or Aliphatic	van der Waals volume
Alanine	A	Ala	$-CH_3$	X	-	-	-	X	X	-	67
Cysteine	C	Cys	$-CH_2SH$	-	8.18	-	acidic	X	X	-	86
Aspartic acid	D	Asp	$-CH_2COOH$	-	3.90	X	acidic	X	-	-	91
Glutamic acid	E	Glu	$-CH_2CH_2COOH$	-	4.07	X	acidic	-	-	-	109
Phenylalanine	F	Phe	$-CH_2C_6H_5$	X	-	-	-	-	-	Aromatic	135
Glycine	G	Gly	-H	X	-	-	-	X	X	-	48
Histidine	H	His	$-CH_2-C_3H_3N_2$	-	6.04	X	weak basic	-	-	Aromatic	118
Isoleucine	I	Ile	$-CH(CH_3)CH_2CH_3$	X	-	-	-	-	-	Aliphatic	124
Lysine	K	Lys	$-(CH_2)_4NH_2$	-	10.54	X	basic	-	-	-	135
Leucine	L	Leu	$-CH_2CH(CH_3)_2$	X	-	-	-	-	-	Aliphatic	124
Methionine	M	Met	$-CH_2CH_2SCH_3$	X	-	-	-	-	-	-	124
Asparagine	N	Asn	$-CH_2CONH_2$	-	-	X	-	X	-	-	96
Pyrrolysine	O	Pyl	$-(CH_2)_4NHCOC_4H_5NCH_3$	-	-	X	weak basic	-	-	-	
Proline	P	Pro	$-CH_2CH_2CH_2-$	X	-	-	-	X	-	-	90
Glutamine	Q	Gln	$-CH_2CH_2CONH_2$	-	-	X	weak basic	-	-	-	114
Arginine	R	Arg	$-(CH_2)_3NH-C(NH)NH_2$	-	12.48	X	strongly basic	-	-	-	148
Serine	S	Ser	$-CH_2OH$	-	5.68	X	weak acidic	X	X	-	73
Threonine	T	Thr	$-CH(OH)CH_3$	-	5.53	X	weak acidic	X	-	-	93
Selenocysteine	U	Sec	$-CH_2SeH$	-	5.73	-	acidic	X	X	-	
Valine	V	Val	$-CH(CH_3)_2$	X	-	-	-	X	-	Aliphatic	105
Tryptophan	W	Trp	$-CH_2C_8H_6N$	-	5.885	X	weak basic	-	-	Aromatic	163
Tyrosine	Y	Tyr	$-CH_2-C_6H_4OH$	-	10.46	X	weak acidic	-	-	Aromatic	141

Figure 2.2 – Amino acid side chain properties Wikipedia

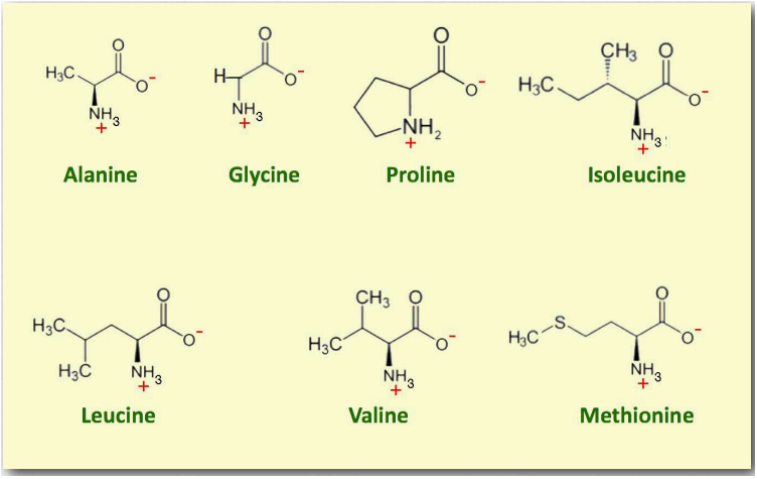


Figure 2.3 – Non-polar amino acids

Acidic Amino Acids (Carboxylic acid side chains)

Amino acids in this group include:

- Aspartic acid (Asp/D)
- Glutamic acid (Glu/E)

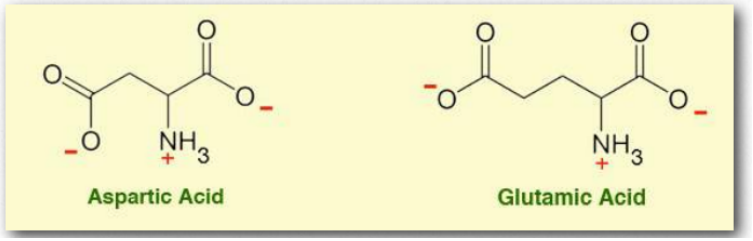


Figure 2.4 – Carboxyl amino acids

These amino acids each contain a carboxylic acid group as part of the variable group. At physiological pH, these groups exist primarily in their deprotonated state. It is easy to be confused if they are

drawn in this state, because their names include "acid" while the structure shows no ionizable proton and the charge on the R group is negative.

In addition to its role as a building block in proteins, glutamic acid (with the deprotonated form named "glutamate") is a neurotransmitter. It also is recognized by a receptor in our mouths, contributing to a taste sensation described as "umami." Many foods contain appreciable amounts of glutamate that are recognized by our taste receptors, and encourage us to eat these substances. Those foods frequently contain protein that has broken down to some degree: cooked meats, fermented sauces like Worcestershire or soy, tahini, broths, and yeast extracts.

Basic amino acids (Nitrogen-containing side chains)

Included in this group of amino acids are:

- Arginine (Arg/R)
- Histidine (His/H)
- Lysine (Lys/K)

Figure 2.5 – Amine amino acids

The variable group in each of these amino acids contains nitrogen, which imparts to the group the ability to exist in protonated and deprotonated states. They are frequently called basic, but also are

often drawn in their protonated state which is more prevalent at physiological pH.

Arginine (Arg/R) is interesting due to the fact it is an essential dietary amino acid for premature infants, who cannot synthesize it. In addition, surgical trauma, sepsis, and burns increase demand for arginine and proper healing can require dietary intake.

Histidine contains a nitrogen-containing imidazole functional group that has a pKa of 6. This means it can pick up or donate hydrogen ions in response to small changes in pH. In proteins, histidine frequently has important roles participating directly in reactions involving hydrogen ion transfer.

The R group on lysine is frequently chemically modified in order for it to make unusual linkages to other chemical groups or to take part in specific chemical reactions. Lysine is often added to animal feed because it is a limiting amino acid and is necessary for optimizing growth of animals raised for consumption.

Aromatic amino acids

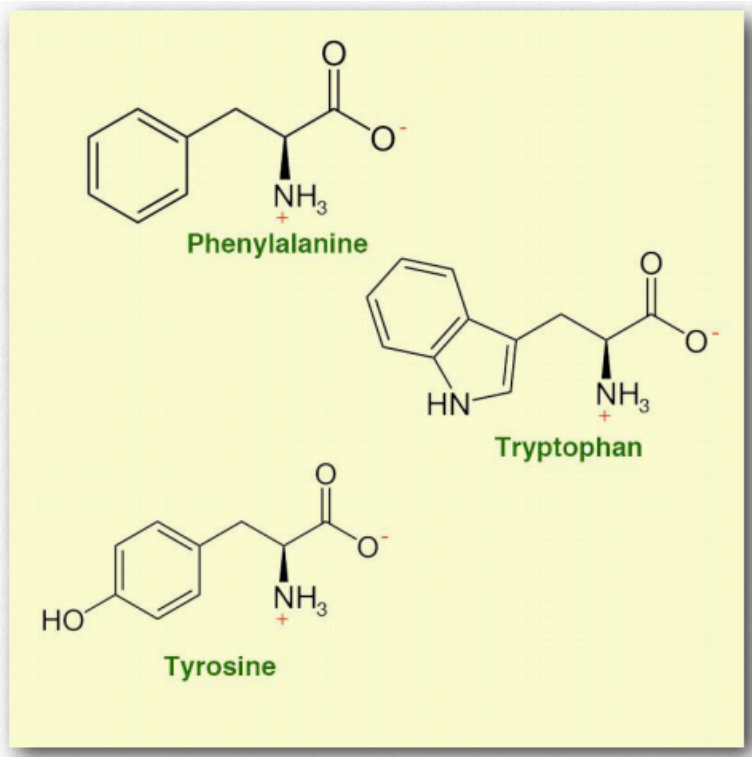


Figure 2.6 – Aromatic amino acids

Amino acids with aromatic side chains include:

- Phenylalanine (Phe/ F)
- Tryptophan (Trp/W)
- Tyrosine (Tyr/Y)

These amino acids are included in protein structures but also serve as precursors in some important biochemical pathways, leading to the production of hormones such as L-Dopa and serotonin.

Hydroxyl amino acids

This group includes

- Threonine (Thr/T)
- Serine (Ser/S)
- Tyrosine (already discussed as an aromatic amino acid)

The amino acids in this group contain alcohol groups, which can engage in hydrogen-bonding interactions. As part of protein molecules they are hydrophilic and can be oriented outward in watery environments. The alcohol group is subject to chemical reactions or modifications, for instance when carbohydrate groups are covalently linked to proteins.

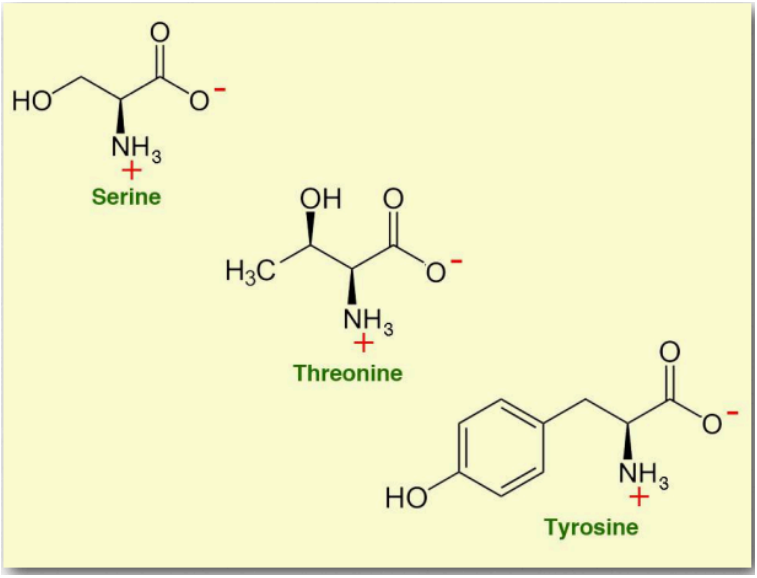


Figure 2.7 – Hydroxyl amino acids

Amino acid	Short	Abbrev.	Avg. mass (Da)	pI	pK_1 (α-COOH)	pK_2 (α-$^+NH_3$)
Alanine	A	Ala	89.09404	6.01	2.35	9.87
Cysteine	C	Cys	121.15404	5.05	1.92	10.70
Aspartic acid	D	Asp	133.10384	2.85	1.99	9.90
Glutamic acid	E	Glu	147.13074	3.15	2.10	9.47
Phenylalanine	F	Phe	165.19184	5.49	2.20	9.31
Glycine	G	Gly	75.06714	6.06	2.35	9.78
Histidine	H	His	155.15634	7.60	1.80	9.33
Isoleucine	I	Ile	131.17464	6.05	2.32	9.76
Lysine	K	Lys	146.18934	9.60	2.16	9.06
Leucine	L	Leu	131.17464	6.01	2.33	9.74
Methionine	M	Met	149.20784	5.74	2.13	9.28
Asparagine	N	Asn	132.11904	5.41	2.14	8.72
Pyrrolysine	O	Pyl	255.31			
Proline	P	Pro	115.13194	6.30	1.95	10.64
Glutamine	Q	Gln	146.14594	5.65	2.17	9.13
Arginine	R	Arg	174.20274	10.76	1.82	8.99
Serine	S	Ser	105.09344	5.68	2.19	9.21
Threonine	T	Thr	119.12034	5.60	2.09	9.10
Selenocysteine	U	Sec	168.053	5.47		
Valine	V	Val	117.14784	6.00	2.39	9.74
Tryptophan	W	Trp	204.22844	5.89	2.46	9.41
Tyrosine	Y	Tyr	181.19124	5.64	2.20	9.21

Figure 2.8 – Amino acid properties Wikipedia

Other amino acids

- Asparagine (Asn/N) is a polar amino acid. The amide on the functional group is not basic.
- Cysteine (Cys/C)
- Glutamine (Gln/Q)

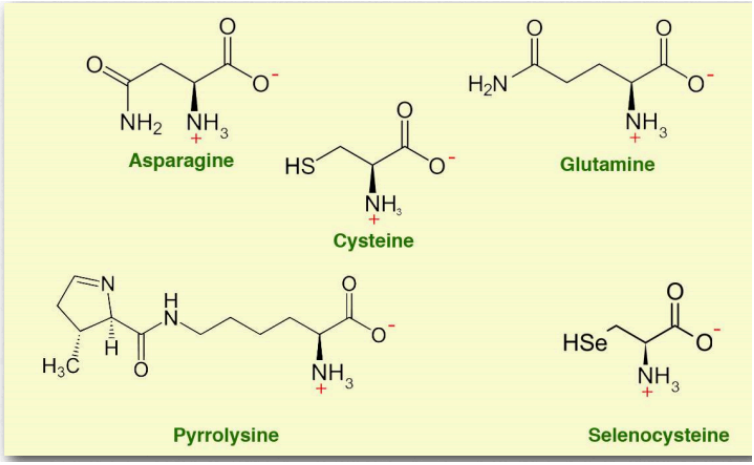


Figure 2.9 – Other amino acids

Cysteine, which contains a thiol. Thiols can react with one another via oxidation, forming disulfide links containing two covalently-linked sulfur atoms. Variable groups on methionine in protein chains can undergo such reactions, covalently tying the chains to one another with a short tether. Such disulfide links or bridges restrict the mobility of protein chains and contribute to more defined structures.

- Selenocysteine (Sec/U) is a component of selenoproteins found in all kingdoms of life. Twenty five human proteins contain selenocysteine. It is a component in several enzymes, including glutathione peroxidases and thioredoxin reductases. It is not coded for by the standard genetic code.
- Pyrrolysine (Pyl/O) is a twenty second amino acid, but is rarely found in proteins. Like selenocysteine, it is not coded for in the genetic code and must be incorporated by unusual means.

Ionizing groups

Some, but not all amino acids have R-groups that can ionize. The

charge of a protein then arises from the charges of the amine group, the carboxyl group, and the sum of the charges of the ionized R-groups. Titration/ionization of aspartic acid is depicted in Figure 2.10. Ionization (or deionization) within a protein's structure can have significant effect on the overall conformation of the protein and, since structure is related to function, a major impact on the activity of a protein.

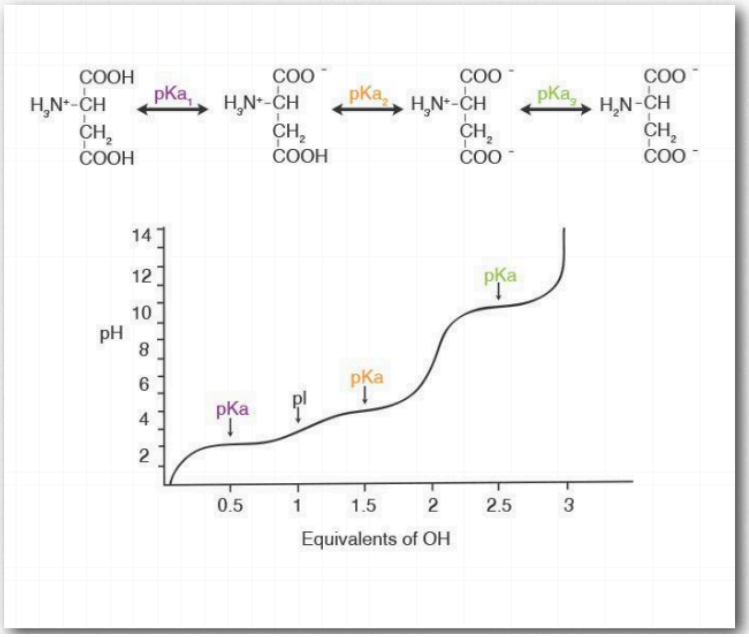


Figure 2.10 – Titration curve for aspartic acid Image by Penelope Irving

Building Polypeptides

Although amino acids serve other functions in cells, their most

important role is as constituents of proteins. Proteins, as we noted earlier, are polymers of amino acids.

Amino acids are linked to each other by peptide bonds, in which the carboxyl group of one amino acid is joined to the amino group of the next, with the loss of a molecule of water. Additional amino acids are added in the same way, by formation of peptide bonds between the free carboxyl on the end of the growing chain and the amino group of the next amino acid in the sequence. A chain made up of just a few amino acids linked together is called an oligopeptide (oligo=few) while a typical protein, which is made up of many amino acids is called a polypeptide (poly=many). The end of the peptide that has a free amino group is called the N-terminus (for NH2), while the end with the free carboxyl is termed the C-terminus (for carboxyl).

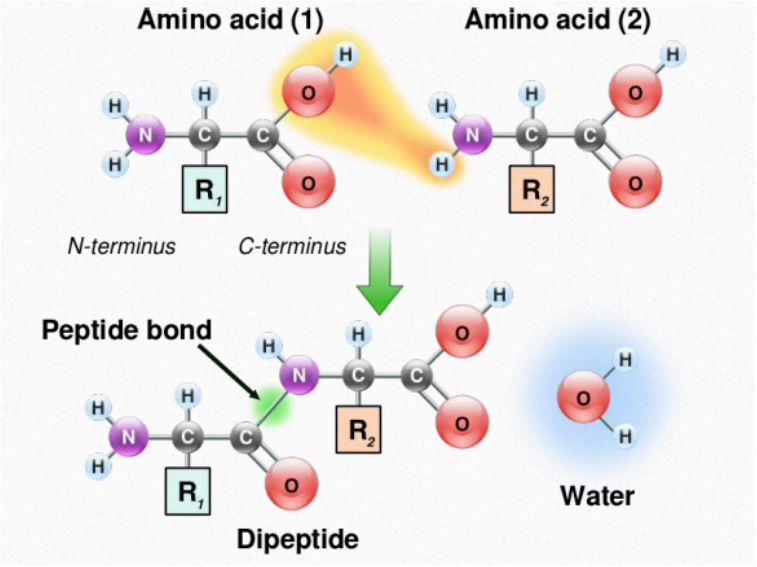


Figure 2.16 Formation of a peptide bond

As we've noted before, function is dependent on structure, and the string of amino acids must fold into a specific 3-D shape, or conformation, in order to make a functional protein. The folding of polypeptides into their functional forms is the topic of the next section.

2.3: Structure and Function- Proteins I

Proteins are the workhorses of the cell. Virtually everything that goes on inside of cells happens as a result of the actions of proteins. Among other things, protein enzymes catalyze the vast majority of cellular reactions, mediate signaling, give structure both to cells and to multicellular organisms, and exert control over the expression of genes. Life, as we know it, would not exist if there were no proteins. The versatility of proteins arises because of their varied structures.

Proteins are made by linking together amino acids, with each protein having a characteristic and unique amino acid sequence. To get a sense for the diversity of proteins that can be made using 20 different amino acids, consider that the number of different combinations possible with 20 amino acids is 20^n, where n=the number of amino acids in the chain. It becomes apparent that even a dipeptide made of just two amino acids joined together gives us 20^2 = 400 different combinations. If we do the calculation for a short peptide of 10 amino acids, we arrive at an enormous 10,240,000,000,000 combinations. Most proteins are much larger than this, making the possible number of proteins with unique amino acid sequences unimaginably huge.

Levels of Structure

The significance of the unique sequence, or order, of amino acids, known as the protein's primary structure, is that it dictates the 3-D conformation the folded protein will have. This conformation, in

turn, will determine the function of the protein. We shall examine protein structure at four distinct levels (Figure 2.17) – 1) how sequence of the amino acids in a protein (primary structure) gives identity and characteristics to a protein (Figure 2.18); 2) how local interactions between one part of the polypeptide backbone and another affect protein shape (secondary structure); 3) how the polypeptide chain of a protein can fold to allow amino acids to interact with each other that are not close in primary structure (tertiary structure); and 4) how different polypeptide chains interact with each other within a multi-subunit protein (quaternary structure).

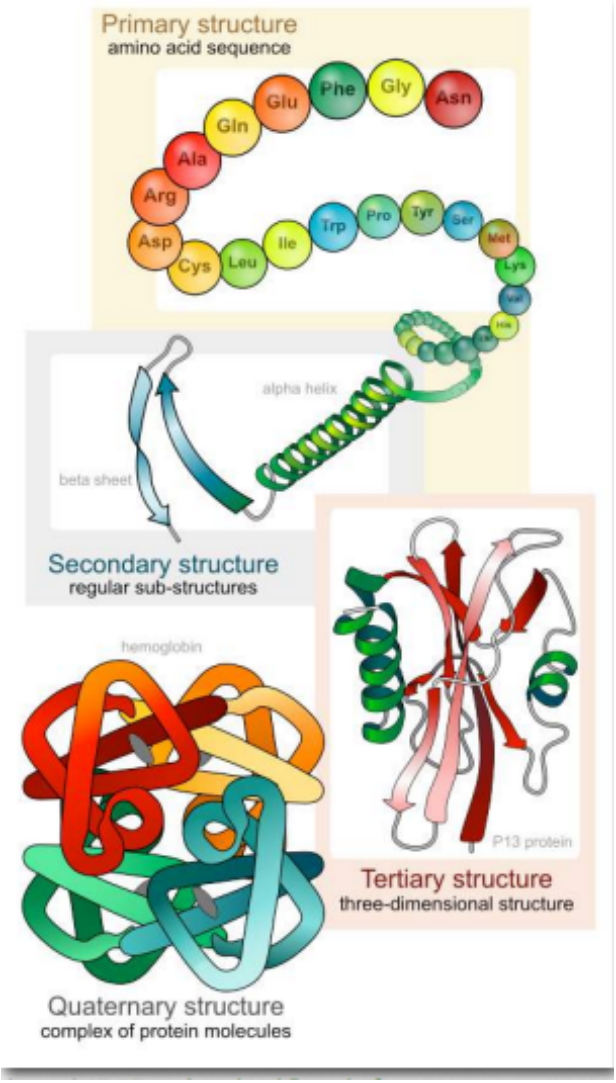


Figure 2.17 – Four Levels of Protein Structure

At this point, we should provide a couple of definitions. We use the term polypeptide to refer to a single polymer of amino acids. It may or may not have folded into its final, functional form. The term protein is sometimes used interchangeably with polypeptide, as in "protein synthesis". It is generally used, however, to refer to a folded,

functional molecule that may have one or more subunits (made up of individual polypeptides). Thus, when we use the term protein, we are usually referring to a functional, folded polypeptide or peptides. Structure is essential for function. If you alter the structure, you alter the function – usually, but not always, this means you lose all function. For many proteins, it is not difficult to alter the structure.

Proteins are flexible, not rigidly fixed in structure. As we shall see, it is the flexibility of proteins that allows them to be amazing catalysts and allows them to adapt to, respond to, and pass on signals upon binding of other molecules or proteins. However, proteins are not infinitely flexible. There are constraints on the conformations that proteins can adopt and these constraints govern the conformations that proteins display.

Subtle changes

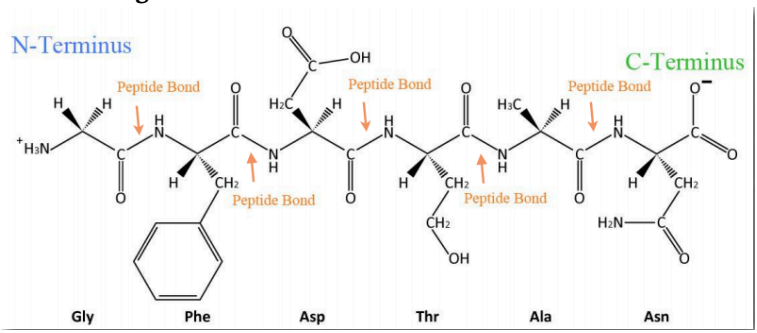


Figure 2.18 – Sequence of a simple polypeptide Wikipedia

Even very tiny, subtle changes in protein structure can give rise to big changes in the behavior of proteins. Hemoglobin, for example, undergoes an incredibly small structural change upon binding of one oxygen molecule, and that simple change causes the remainder of the protein to gain a considerably greater affinity for oxygen that the protein didn't have before the structural change.

Sequence, structure and function

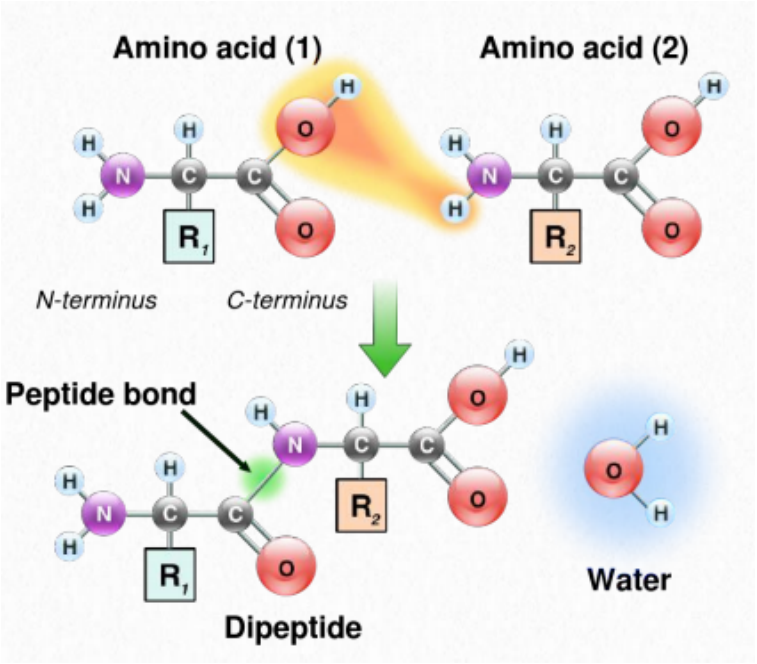


Figure 2.19 Linking of amino acids through peptide bond formation

As discussed earlier, the number of different amino acid sequences possible, even for short peptides, is very large. No two proteins with different amino acid sequences (primary structure) have identical overall structure. The unique amino acid sequence of a protein is reflected in its unique folded structure. This structure, in turn, determines the protein's function. This is why mutations that alter amino acid sequence can affect the function of a protein.

Protein Synthesis

Synthesis of proteins occurs in the ribosomes and proceeds by

joining the carboxyl terminus of the first amino acid to the amino terminus of the next one (Figure 2.19). The end of the protein that has the free α-amino group is referred to as the amino terminus or N-terminus. The other end is called the carboxyl terminus or C-terminus , since it contains the only free α-carboxyl group. All of the other α-amino groups and α-carboxyl groups are tied up in forming peptide Figure 2.19 Linking of amino acids through peptide bond formation bonds that join adjacent amino acids together. Proteins are synthesized starting with the amino terminus and ending at the carboxyl terminus.

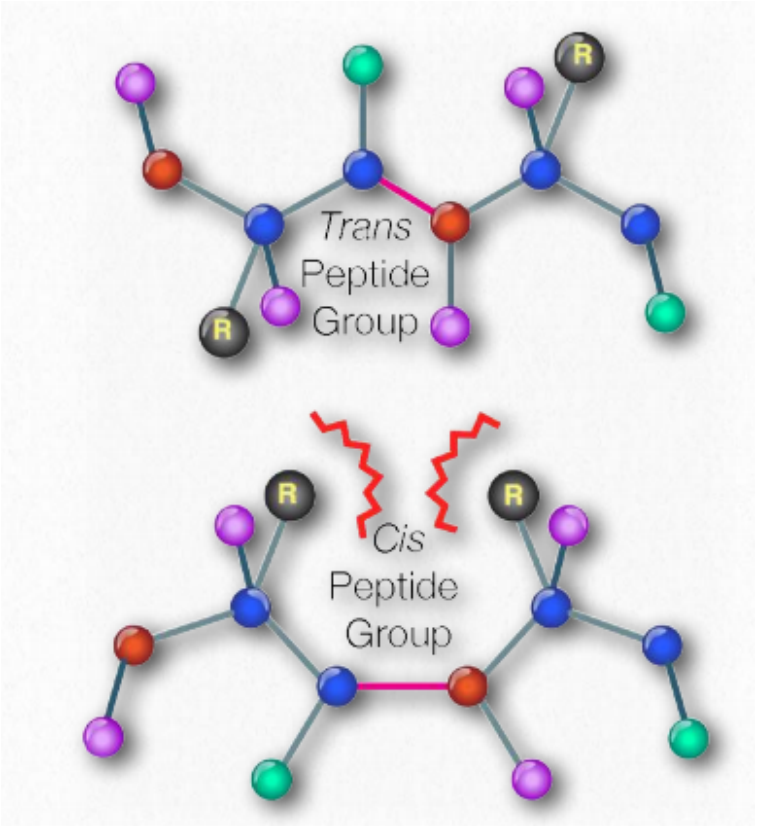


Figure 2.20 – Cis vs trans orientation of R-groups around peptide bond Image by Aleia Kim

Schematically, in Figure 2.18, we can see how sequential R-groups

of a protein are arranged in an alternating orientation on either side of the polypeptide chain. Organization of R-groups in this fashion is not random. Steric hindrance can occur when consecutive R-groups are oriented on the same side of a peptide backbone (Figure 2.20)

Primary Structure

Primary structure is the ultimate determinant of the overall conformation of a protein. The primary structure of any protein arrived at its current state as a result of mutation and selection over evolutionary time. Primary structure of proteins is mandated by the sequence of DNA coding for it in the genome. Regions of DNA specifying proteins are known as coding regions (or genes).

The base sequences of these regions directly specify the sequence of amino acids in proteins, with a one-to-one correspondence between the codons (groups of three consecutive bases) in the DNA and the amino acids in the encoded protein. The sequence of codons in DNA, copied into messenger RNA, specifies a sequence of amino acids in a protein. (Figure 2.21).

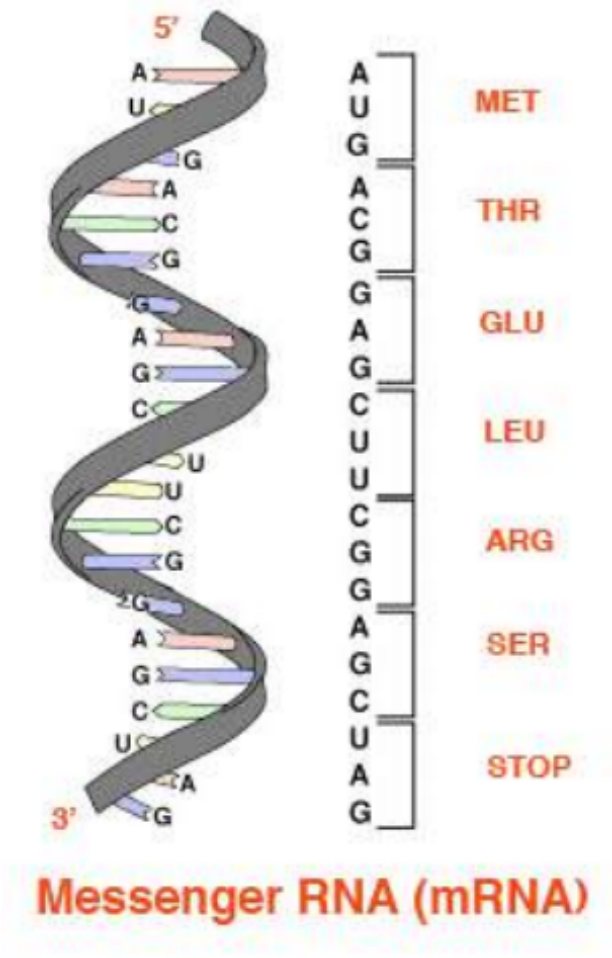


Figure 2.21 – From RNA to amino acids – the genetic code Wikipedia

The order in which the amino acids are joined together in protein synthesis starts defining a set of interactions between amino acids even as the synthesis is occurring. That is, a polypeptide can fold even as it is being made. The order of the R-group structures and resulting interactions are very important because early interactions

affect later interactions. This is because interactions start establishing structures – secondary and tertiary. If a helical structure (secondary structure), for example, starts to form, the possibilities for interaction of a particular amino acid Rgroup may be different than if the helix had not formed (Figure 2.22). R-group interactions can also cause bends in a polypeptide sequence (tertiary structure) and these bends can create (in some cases) opportunities for interactions that wouldn't have been possible without the bend or prevent (in other cases) similar interaction possibilities.

Secondary Structure

As protein synthesis progresses, interactions between amino acids close to each other begin to occur, giving rise to local patterns called secondary structure. These secondary structures include the well known α- helix and β-strands. Both were predicted by Linus Pauling, Robert Corey, and Herman Branson in 1951. Each structure has unique features.

α-helix

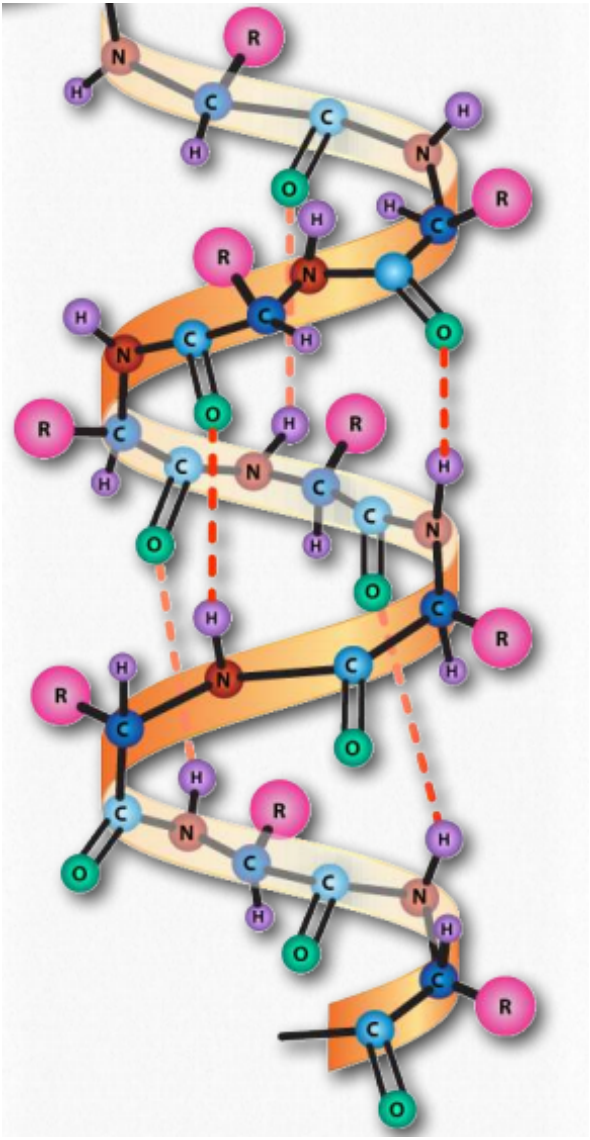


Figure 2.22 – The α-helix. Hydrogen bonds (dotted lines) between the carbonyl oxygen and the amine hydrogen stabilize the structure. Image by Aleia Kim

The α-helix has a coiled structure, with 3.6 amino acids per turn

of the helix (5 helical turns = 18 amino acids). Helices are predominantly right handed – only in rare cases, such as in sequences with many glycines can left handed α- helices form. In the α-helix, hydrogen bonds form between C=O groups and N-H groups in the polypeptide backbone that are four amino acids distant. These hydrogen bonds are the primary forces stabilizing the α-helix.

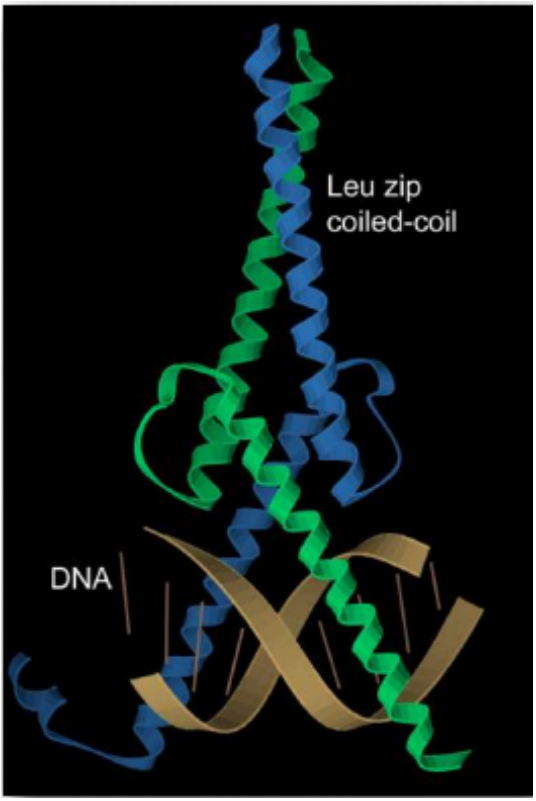


Figure 2.23 – α-helices in a protein with a leucine zipper structural domain. The α-helices are shown in blue and green and are bound to a DNA double helix in brown.

We use the terms rise, repeat, and pitch to describe the parameters of any helix. The repeat is the number of residues in a helix before it begins to repeat itself. For an α-helix, the repeat is 3.6

amino acids per turn of the helix. The rise is the distance the helix elevates with addition of each residue. For an α-helix, this is 0.15 nm per amino acid. The pitch is the distance between complete turns of the helix. For an α-helix, this is 0.54 nm. The stability of an α-helix is enhanced by the presence of the amino acid aspartate.

Figure 2.24 – α-helix sculpture outside Linus Pauling's boyhood home Wikipedia

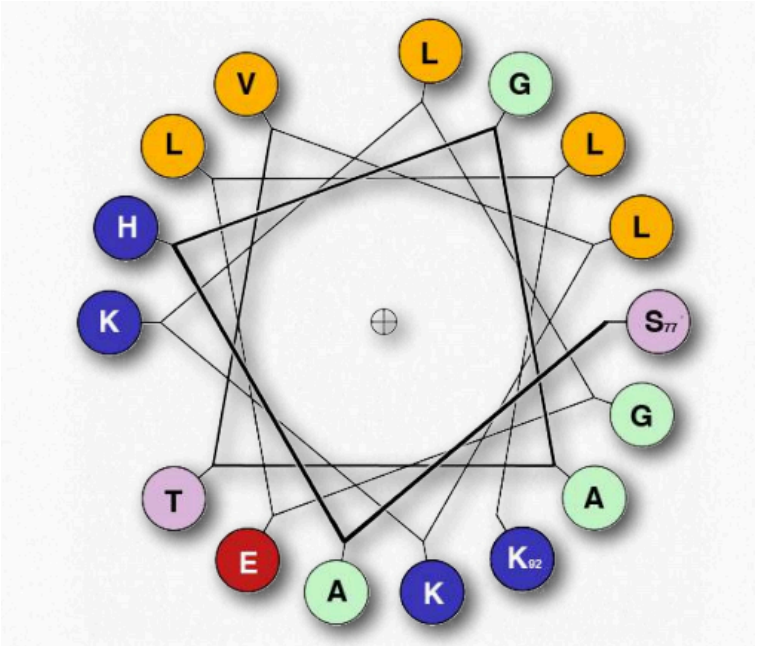


Figure 2.25 – Helical Wheel Representation of an α-Helix. The one letter genetic code is used. The helix starts at Serine #77 at the right and ends at lysine #92 in the lower right. Hydrophobic amino acids are shown in yellow and ionizing amino acids are shown in blue. Hydrophobic amino acids tend to interact with each other and not with ionizing amino acids. Wikipedia

β strand/sheet

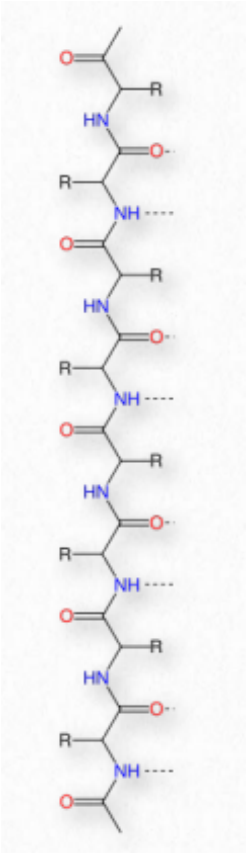


Figure 2.26 – β strand

A helix is, of course, a three-dimensional object. A flattened form of helix in two dimensions is a common description for a β- strand. Rather than coils, β-strands have bends and these are sometimes referred to as pleats, like the pleats in a curtain. β-strands can be organized to form elaborately organized structures, such as sheets, barrels, and other arrangements.

Higher order β-strand structures are sometimes called supersecondary structures), since they involve interactions between amino acids not close in primary sequence. These structures, too, are stabilized by hydrogen bonds between carbonyl oxygen atoms and hydrogens of amine groups in the polypeptide

backbone (Figure 2.28). In a higher order structure, strands can be arranged parallel (amino to carboxyl orientations the same) or anti-parallel (amino to carboxyl orientations opposite of each other (in Figure 2.27, the direction of the strand is shown by the arrowhead in the ribbon diagrams).

Turns

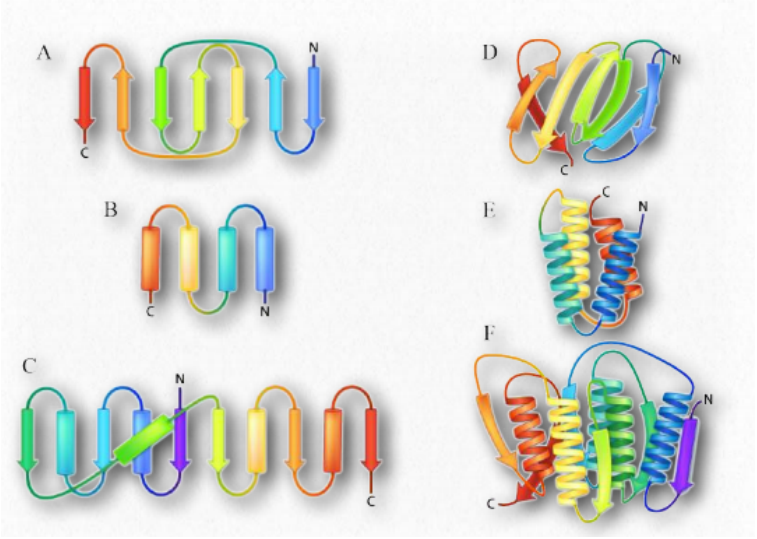


Figure 2.27 – Ribbon depictions of supersecondary β-sheets (A-D) and α-helix arrangements (E-F) Image by Aleia Kim

Turns (sometimes called reverse turns) are a type of secondary structure that, as the name suggests, causes a turn in the structure of a polypeptide chain. Turns give rise to tertiary structure ultimately, causing interruptions in the secondary structures (α-helices and β-strands) and often serve as connecting regions between two regions of secondary structure in a protein. Proline and glycine play common roles in turns, providing less flexibility (starting the turn) and greater flexibility (facilitating the turn), respectively.

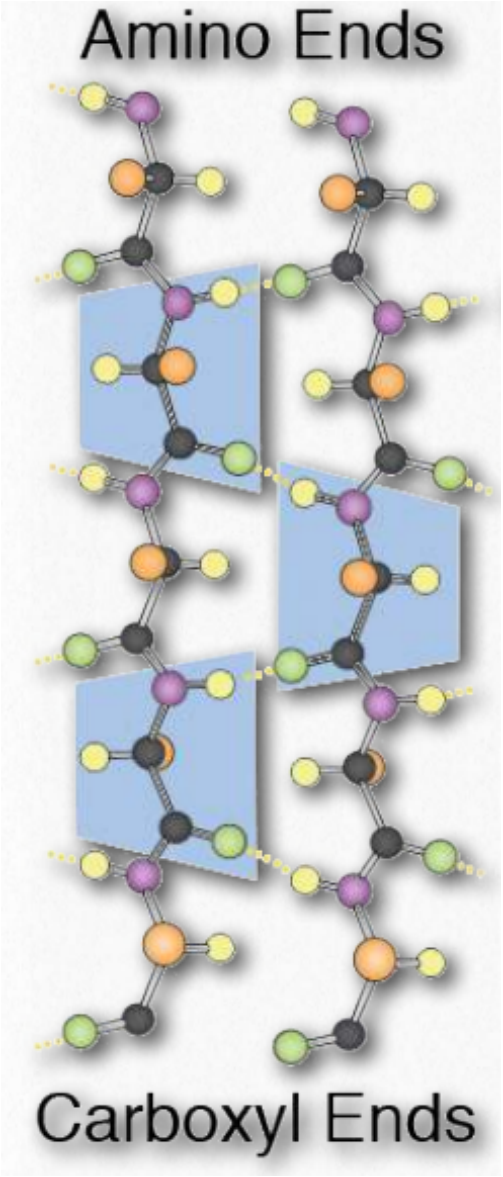


Figure 2.28 – Components of a β-sheet in a parallel arrangement. H-bonds in yellow. Image by Aleia Kim

There are at least five types of turns, with numerous variations of each giving rise to many different turns. The five types of turns are

- δ-turns – end amino acids are separated by one peptide bond
- γ-turns – separation by two peptide bonds
- β-turns – separation by three peptide bonds
- α-turns – separation by four peptide bonds
- π-turns – separation by five bonds

Of these, the β-turns are the most common form and the δ-turns are theoretical, but unlikely, due to steric limitations. Figure 2.29 depicts a β- turn.

3_{10} helices

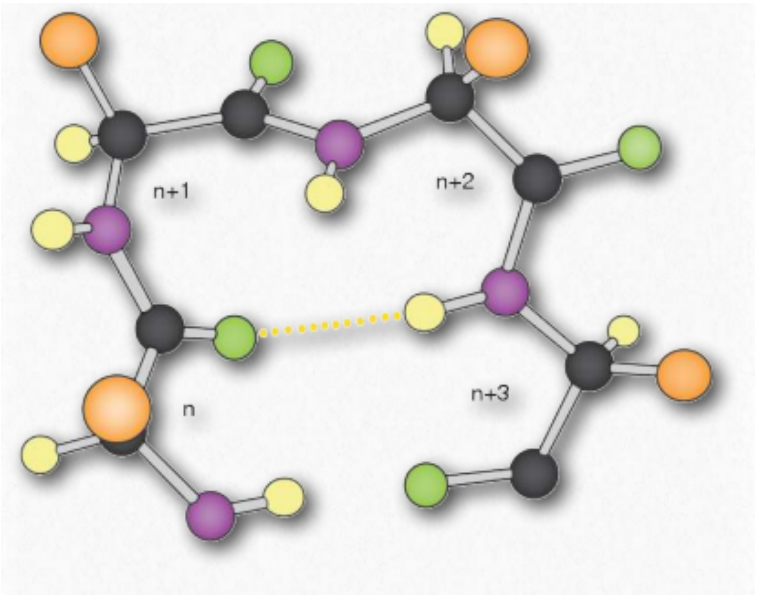


Figure 2.29 – β-turn. R-groups are shown in orange, hydrogens in yellow, carbons in charcoal, nitrogens in purple, and oxygens in green. A stabilizing hydrogen bond is indicated with the dotted line. Image by Aleia Kim

In addition to the α-helix, β-strands, and various turns, other regular, repeating structures are seen in proteins, but occur much less commonly. The 3_{10} helix is the fourth most abundant secondary structure in proteins, constituting about 10-15% of all helices. The

helix derives its name from the fact that it contains 10 amino acids in 3 turns. It is right-handed. Hydrogen bonds form between amino acids that are three residues apart. Most commonly, the 3_{10} helix appears at the amine or carboxyl end of an α-helix. Like the α-helix, the 3_{10} helix is stabilized by the presence of aspartate in its sequence.

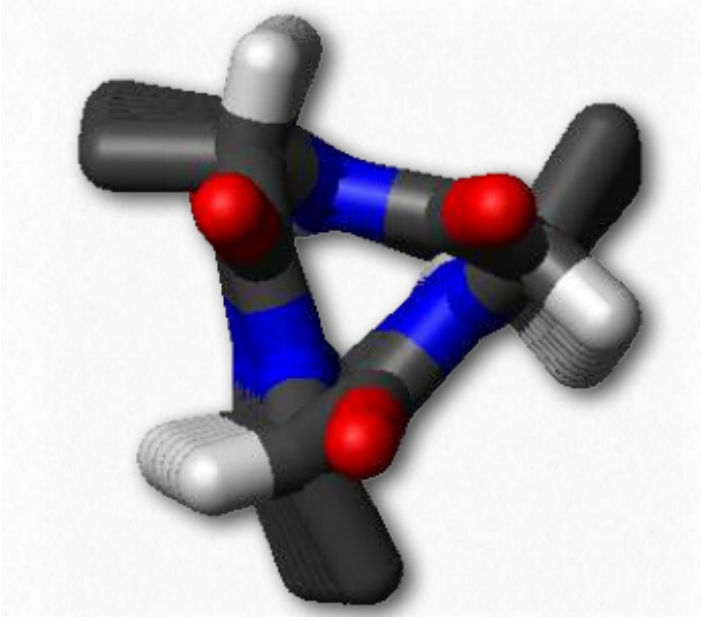


Figure 2.30 – Top view of a 310 Helix. Carbonyl groups are in red and pointed upwards. Note the almost perfect 3-fold symmetry Wikipedia

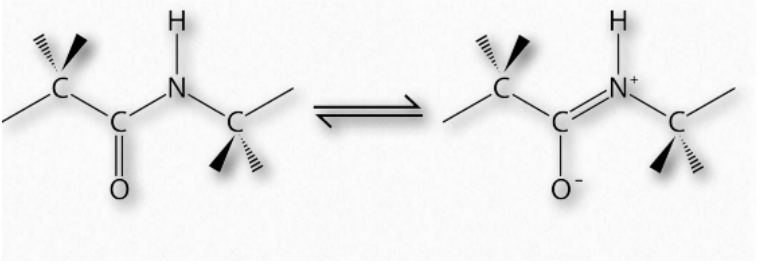


Figure 2.31 – Resonance of the peptide bond Wikipedia

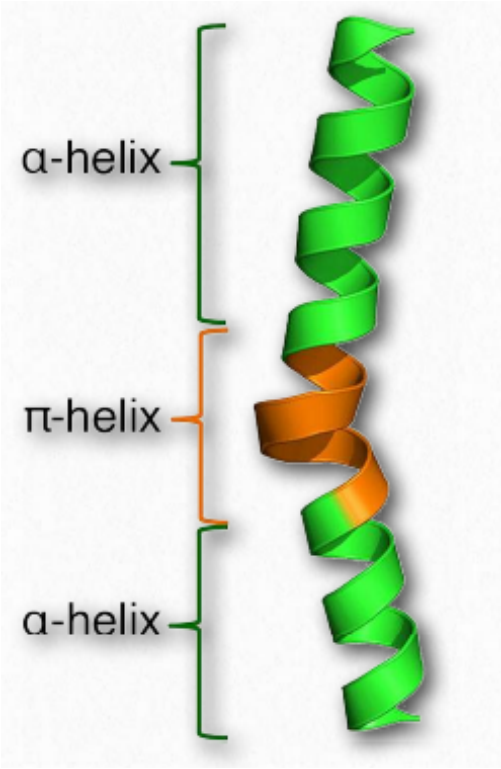


Figure 2.32 – π helix Wikipedia

π-helices

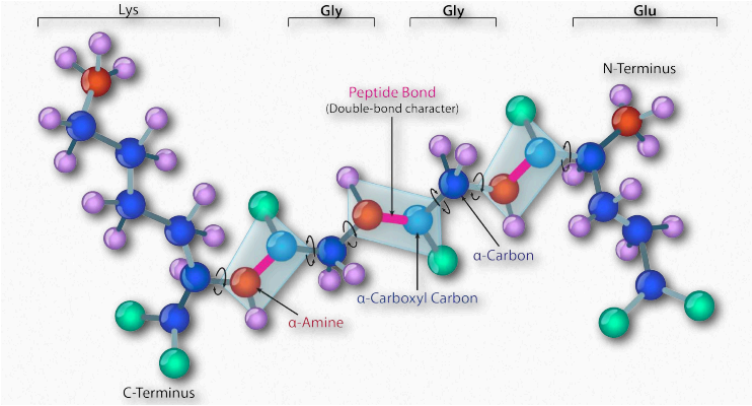


Figure 2.33 – Planes (light blue) defined by the double-bonded character of the peptide bond Image by Aleia Kim

A π-helix may be thought of as a special type of α- helix. Some sources describe it as an α-helix with an extra amino acid stuck in the middle of it (Figure 2.32). π-helices are not exactly rare, occurring at least once in as many as 15% of all proteins. Like the α- helix, the π-helix is right-handed, but where the α-helix has 18 amino acids in 5 turns, the π-helix has 22 amino acids in 5 turns. π-helices typically do not stretch for very long distances. Most are only about 7 amino acids long and the sequence almost always occurs in the middle of an α-helical region.

Ramachandran plots

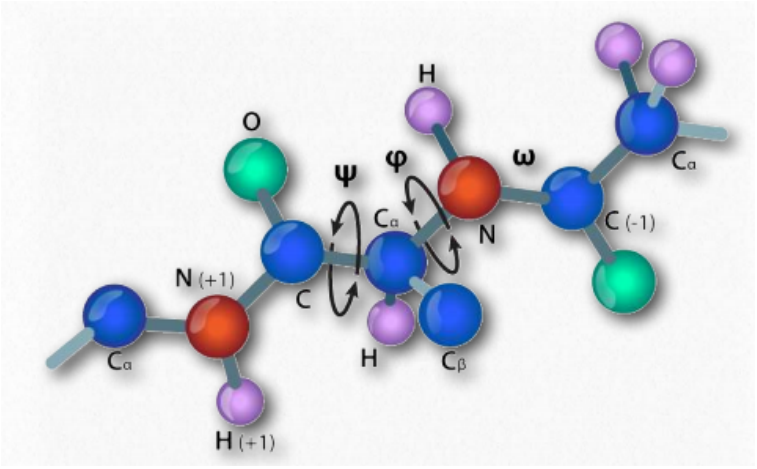


Figure 2.34 – ω, ψ, and φ rotational angles in a peptide Image by Aleia Kim

In 1963, G.N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan described a novel way to describe protein structure. If one considers the backbone of a polypeptide chain, it consists of a repeating set of three bonds. Sequentially (in the amino to carboxyl direction) they are 1) a rotatable bond (ψ) between α-carbon and α-carboxyl preceding the peptide bond (see HERE), 2) a non-rotatable peptide bond (ω) between the α-carboxyl and α-amine groups), and 3) a rotatable bond (φ) between the α-amine and α-carbon following

the peptide bond (see HERE). Note in Figures 2.33 and 2.34 that the amino to carboxyl direction is right to left.

The presence of the carbonyl oxygen on the α-carboxyl group allows the peptide bond to exist as a resonant structure, meaning that it behaves some of the time as a double bond. Double bonds cannot, of course, rotate, but the bonds on either side of it have some freedom of rotation. The φ and ψ angles are restricted to certain values, because some angles will result in steric hindrance. In addition, each type of secondary structure has a characteristic range of values for φ and ψ.

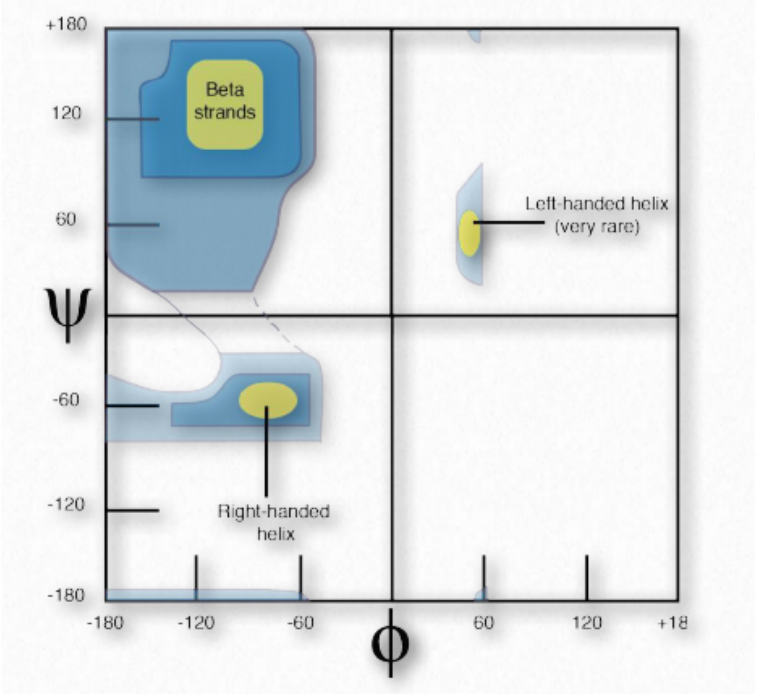


Figure 2.35 – Theoretical Ramachandran plot Image by Penelope Irving

Ramachandran and colleagues made theoretical calculations of the energetic stability of all possible angles from 0° to 360° for each of the φ and ψ angles and plotted the results on a Ramachandran

Plot (also called a φ-ψ plot), delineating regions of angles that were theoretically the most stable (Figure 2.35).

Three primary regions of stability were identified, corresponding to φ-ψ angles of β-strands (top left), right handed α- helices (bottom left), and lefthanded α-helices (upper right). The plots of predicted stability are remarkably accurate when compared to φ-ψ angles of actual proteins.

Secondary structure prediction

Amino Acid	α helix	Reverse turn	β sheet
Ala	1.41	0.82	0.72
Arg	1.21	0.90	0.84
Asn	0.76	1.34	0.48
Asp	0.99	1.24	0.39
Cys	0.66	0.54	1.40
Gln	1.27	0.84	0.98
Glu	1.59	1.01	0.52
Gly	0.43	1.77	0.58
His	1.05	0.81	0.80
Ile	1.09	0.47	1.67
Leu	1.34	0.57	1.22
Lys	1.23	1.07	0.69
Met	1.30	0.52	1.14
Phe	1.16	0.59	1.33
Pro	0.34	1.32	0.31
Ser	0.57	1.22	0.96
Thr	0.76	0.96	1.17
Trp	1.02	0.65	1.35
Tyr	0.74	0.76	1.45
Val	0.90	0.41	1.87

Table 2.3 – Relative tendencies of each amino acid to be in a

secondary structure. Higher values indicate greater tendency Image by Penelope Irving

By comparing primary structure (amino acid sequences) to known 3D protein structures, one can tally each time an amino acid is found in an α-helix, β-strand/sheet, or a turn. Computer analysis of thousands of these sequences allows one to assign a likelihood of any given amino acid appearing in each of these structures. Using these tendencies, one can, with up to 80% accuracy, predict regions of secondary structure in a protein based solely on amino acid sequence.

This is seen in Table 2.3. Occurrence in primary sequence of three consecutive amino acids with relative tendencies higher than one is an indicator that that region of the polypeptide is in the corresponding secondary structure. An online resource for predicting secondary structures called PSIPRED is available HERE.

Hydrophobicity

Amino Acid Hydropathy Scores

Amino Acid	One Letter Code	Hydropathy Score
Isoleucine	I	4.5
Valine	V	4.2
Leucine	L	3.8
Phenylalanine	F	2.8
Cysteine	C	2.5
Methionine	M	1.9
Alanine	A	1.8
Glycine	G	-0.4
Threonine	T	-0.7
Tryptophan	W	-0.9
Serine	S	-0.8
Tyrosine	Y	-1.3
Proline	P	-1.6
Histidine	H	-3.2
Glutamic acid	E	-3.5
Glutamine	Q	-3.5
Aspartic acid	D	-3.5
Asparagine	N	-3.5
Lysine	K	-3.9
Arginine	R	-4.5

Table 2.4 – Hydropathy Scores

The chemistry of amino acid Rgroups affects the structures they are most commonly found in. Subsets of their chemical properties can give clues to structure and, sometimes, cellular location. A prime example is the hydrophobicity (wateravoiding tendencies) of some Rgroups. Given the aqueous environment of the cell, such R-groups are not likely to be on the outside surface of a folded protein.

However, this rule does not hold for regions of protein that may be embedded within the lipid bilayers of cellular/ organelle

membranes. This is because the region of such proteins that form the transmembrane domains are are buried in the hydrophobic environment in the middle of the lipid bilayer.

Not surprisingly, scanning primary sequences for specifically sized/spaced stretches of hydrophobic amino acids can help to identify proteins found in membranes. Table 2.4 shows hydrophobicity values for R-groups of the amino acids. In this set, the scale runs from positive values (hydrophobic) to negative values (hydrophilic). A KyteDoolittle Hydropathy plot for the RET protooncogene membrane protein is shown in Figure 2.36. Two regions of the protein are very hydrophobic as can be seen from the peaks near amino acids 5-10 and 630-640. Such regions might be reasonably expected to be situated either within the interior of the folded protein or to be part of transmembrane domains.

Random coils

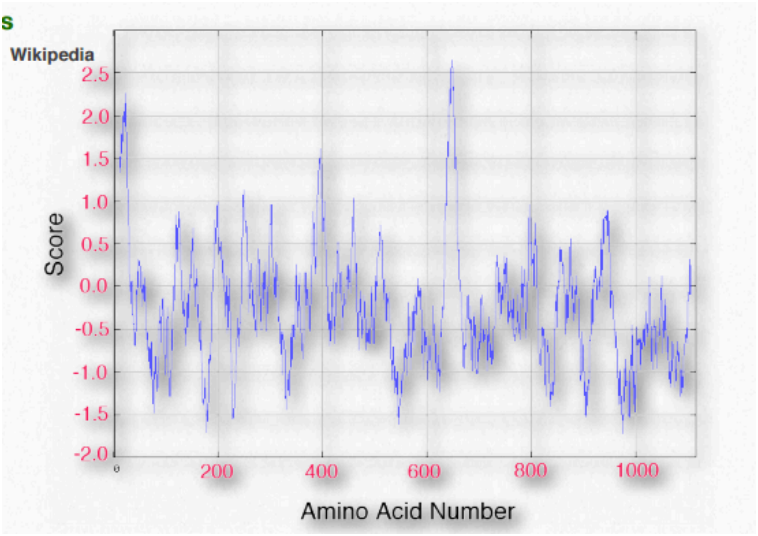


Figure 2.36 Kyte-Doolittle hydropathy plot for the RET protooncogene Wikipedia

Some sections of a protein assume no regular, discernible structure and are sometimes said to lack secondary structure, though they may have hydrogen bonds. Such segments are described as being in random coils and may have fluidity to their

structure that results in them having multiple stable forms. Random coils are identifiable with spectroscopic methods, such as circular dichroism Wikipedia and nuclear magnetic resonance (NMR) in which distinctive signals are observed. See also metamorphic proteins (HERE) and intrinsically disordered proteins (HERE).

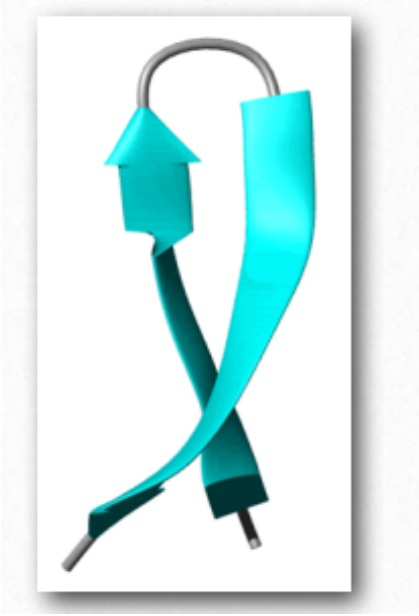


Figure 2.37 – Ribbon depiction of a β-hairpin. Shown are two β strands in turquoise interacting with each other.

Supersecondary structure

Another element of protein structure is harder to categorize because it incorporates elements of secondary and tertiary structure. Dubbed supersecondary structure (or structural motifs), these structures contain multiple nearby secondary structure components arranged in a specific way and that appear in multiple proteins. Since there are many ways of making secondary structures from different primary structures, so too can similar motifs arise from different primary sequences. An example of a structural motif is shown in Figure 2.37.

Tertiary structure

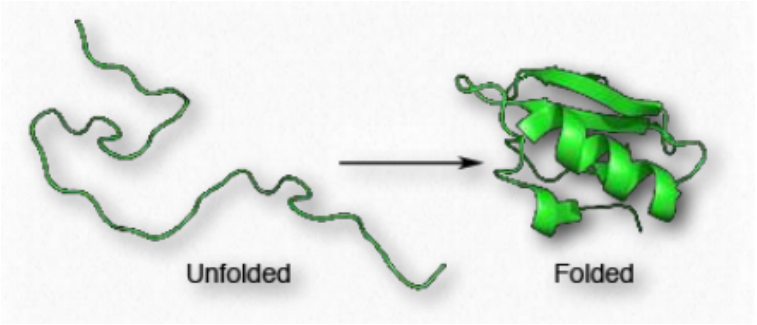


Figure 2.38 – Folding of a polypeptide chain

Proteins are distinguished from each other by the sequence of amino acids comprising them. The sequence of amino acids of a protein determines protein shape, since the chemical properties of each amino acid are forces that give rise to intermolecular interactions to begin to create secondary structures, such as α-helices and β-strands. The sequence also defines turns and random coils that play important roles in the process of protein folding.

Since shape is essential for protein function, the sequence of amino acids gives rise to all of the properties a protein has. As protein synthesis proceeds, individual components of secondary structure start to interact with each other, giving rise to folds that bring amino acids close together that are not near each other in primary structure (Figure 2.38). At the tertiary level of structure, interactions among the R-groups of the amino acids in the protein, as well as between the polypeptide backbone and amino acid side groups play a role in folding.

Globular proteins

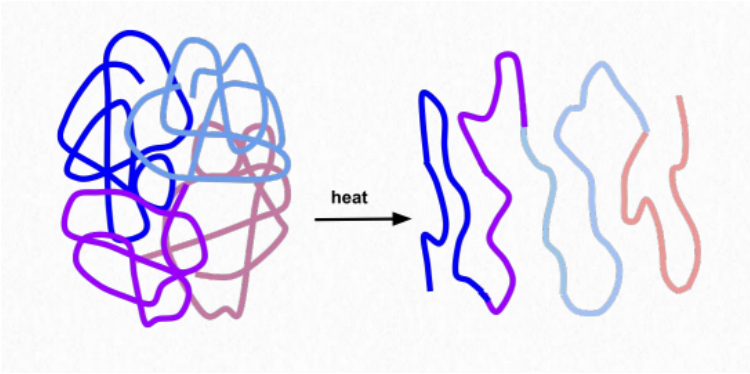


Figure 2.39 – Unfolding (denaturation) of a protein Wikipedia

Folding gives rise to distinct 3-D shapes in proteins that are non-fibrous. These proteins are called globular. A globular protein is stabilized by the same forces that drive its formation. These include ionic interactions, hydrogen bonding, hydrophobic forces, ionic bonds, disulfide bonds and metallic bonds. Treatments such as heat, pH changes, detergents, urea and mercaptoethanol overpower the stabilizing forces and cause a protein to unfold, losing its structure and (usually) its function (Figure 2.39). The ability of heat and detergents to denature proteins is why we cook our food and wash our hands before eating – such treatments denature the proteins in the microorganisms on our hands. Organisms that live in environments of high temperature (over 50°C) have proteins with changes in stabilizing forces – additional hydrogen bonds, additional salt bridges (ionic interactions), and compactness may all play roles in keeping these proteins from unfolding.

Protein stabilizing forces

Before considering the folding process, let us consider some of the forces that help to stabilize proteins.

Hydrogen bonds

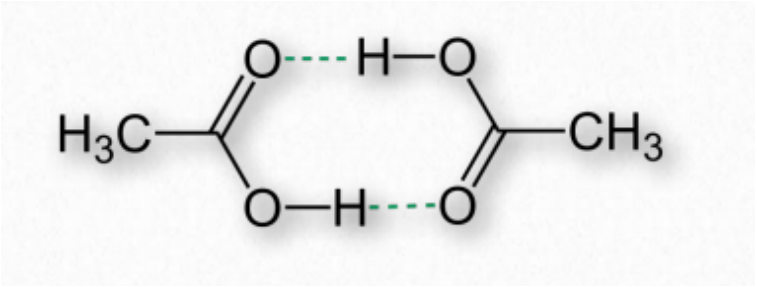


Figure 2.40 – Hydrogen bonds (dotted lines) between two molecules of acetic acid

Hydrogen bonds arise as a result of partially charged hydrogens found in covalent bonds. This occurs when the atom the hydrogen is bonded to has a greater electronegativity than hydrogen itself does, resulting in hydrogen having a partial positive charge because it is not able to hold electrons close to itself (Figure 2.40).

Hydrogen partially charged in this way is attracted to atoms, such as oxygen and nitrogen that have partial negative charges, due to having greater electronegativities and thus holding electrons closer to themselves. The partially positively charged hydrogens are called donors, whereas the partially negative atoms they are attracted to are called acceptors. (See Figure 1.30).

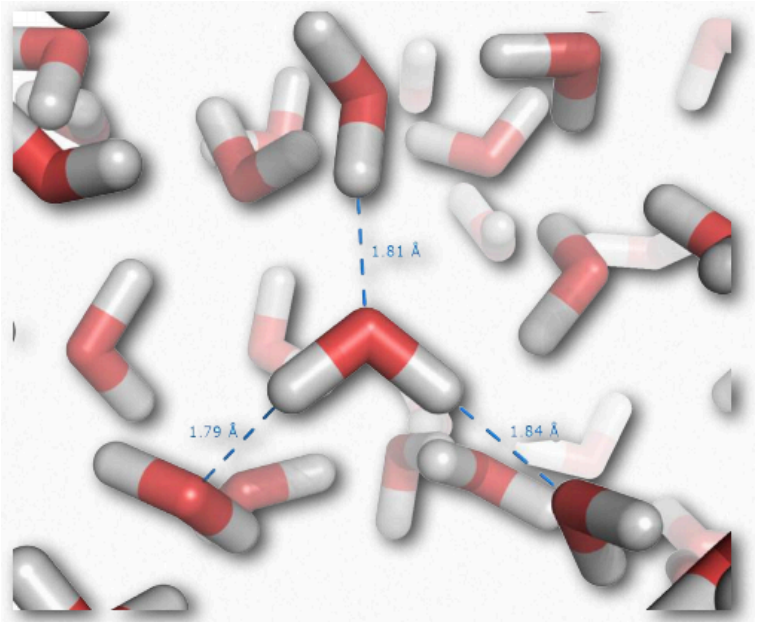


Figure 2.41 – Hydrogen bonding in liquid water Wikipedia

Individual hydrogen bonds are much weaker than a covalent bond, but collectively, they can exert strong forces. Consider liquid water, which contains enormous numbers of hydrogen bonds (Figure 2.41). These forces help water to remain liquid at room temperature. Other molecules lacking hydrogen bonds of equal or greater molecular weight than water, such as methane or carbon dioxide, are gases at the same temperature. Thus, the intermolecular interactions between water molecules help to "hold" water together and remain a liquid. Notably, only by raising the temperature of water to boiling are the forces of hydrogen bonding overcome, allowing water to become fully gaseous.

Hydrogen bonds are important forces in biopolymers that include DNA, proteins, and cellulose. All of these polymers lose their native structures upon boiling. Hydrogen bonds between amino acids that are close to each other in primary structure can give rise to regular repeating structures, such as helices or pleats, in proteins (secondary structure).

Ionic interactions

Ionic interactions are important forces stabilizing protein structure that arise from ionization of R-groups in the amino acids comprising a protein. These include the carboxyl amino acids (HERE), the amine amino acids as well as the sulfhydryl of cysteine and sometimes the hydroxyl of tyrosine.

Hydrophobic forces

Hydrophobic forces stabilize protein structure as a result of interactions that favor the exclusion of water. Non-polar amino acids (commonly found in the interior of proteins) favor associating with each other and this has the effect of excluding water. The excluded water has a higher entropy than water interacting with the hydrophobic side chains. This is because water aligns itself very regularly and in a distinct pattern when interacting with hydrophobic molecules.

When water is prevented from having these kinds of interactions, it is much more disordered that it would be if it could associate with the hydrophobic regions. It is partly for this reason that hydrophobic amino acids are found in protein interiors – so they can exclude water and increase entropy.

Disulfide bonds

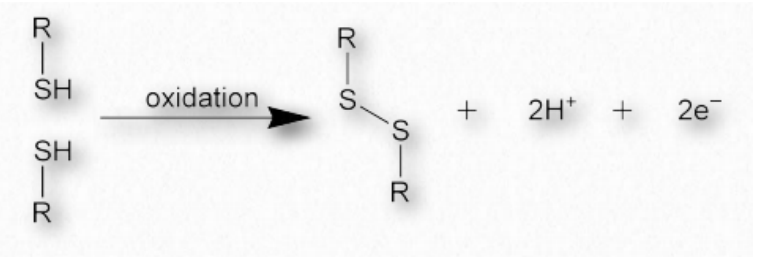


Figure 2.42 – Formation of a disulfide bond

Disulfide bonds, which are made when two sulfhydryl side-chains of cysteine are brought into close proximity, covalently join together different protein regions and can give great strength to the overall structure (Figures 2.42 & 2.43). An Ode to Protein Structure by Kevin Ahern The twenty wee amino A's Define a protein many ways Their order in a peptide chain Determines forms that proteins

gain And when they coil, it leaves me merry Cuz that makes structures secondary It's tertiary, I am told That happens when a protein folds But folded chains are downright scary When put together quaternary They're nature's wonders, that's for sure Creating problems, making cures A fool can fashion peptide poems But proteins come from ribosoems These joined residues of cysteine are sometimes referred to as cystine. Disulfide bonds are the strongest of the forces stabilizing protein structure.

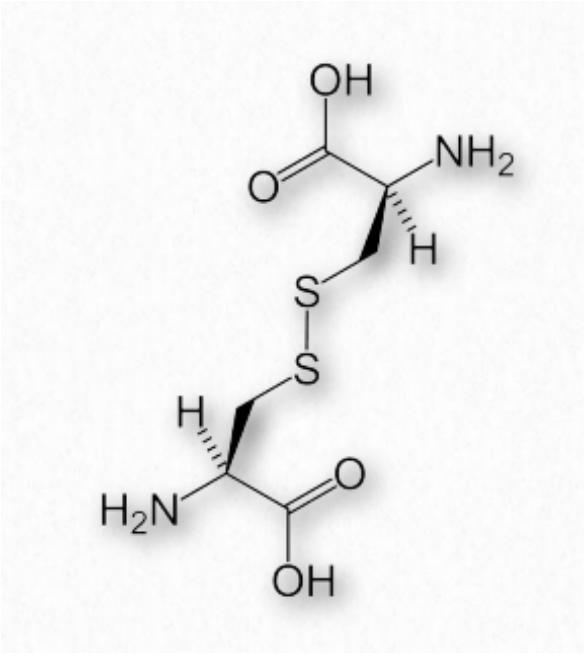


Figure 2.43 – Cystine – Two cysteines joined by a disulfide bond

van der Waals forces

van der Waals forces is a term used to describe various weak interactions, including those caused by attraction between a polar molecule and a transient dipole, or between two temporary dipoles. van der Waals forces are dynamic because of the fluctuating nature of the attraction, and are generally weak in comparison to covalent bonds, but can, over very short distances, be significant.

Post-translational modifications

Post-translational modifications can result in formation of covalent bonds stabilizing proteins as well. Hydroxylation of lysine and proline in strands of collagen can result in cross-linking of these groups and the resulting covalent bonds help to strengthen and stabilize the collagen.

Folding models

Two popular models of protein folding are currently under investigation. In the first (diffusion collision model), a nucleation event begins the process, followed by secondary structure formation. Collisions between the secondary structures (as in the β-hairpin in Figure 2.37) allow for folding to begin. By contrast, in the nucleation-condensation model, the secondary and tertiary structures form together.

Folding in proteins occurs fairly rapidly (0.1 to 1000 seconds) and can occur during synthesis – the amino terminus of a protein can start to fold before the carboxyl terminus is even made, though that is not always the case.

Folding process

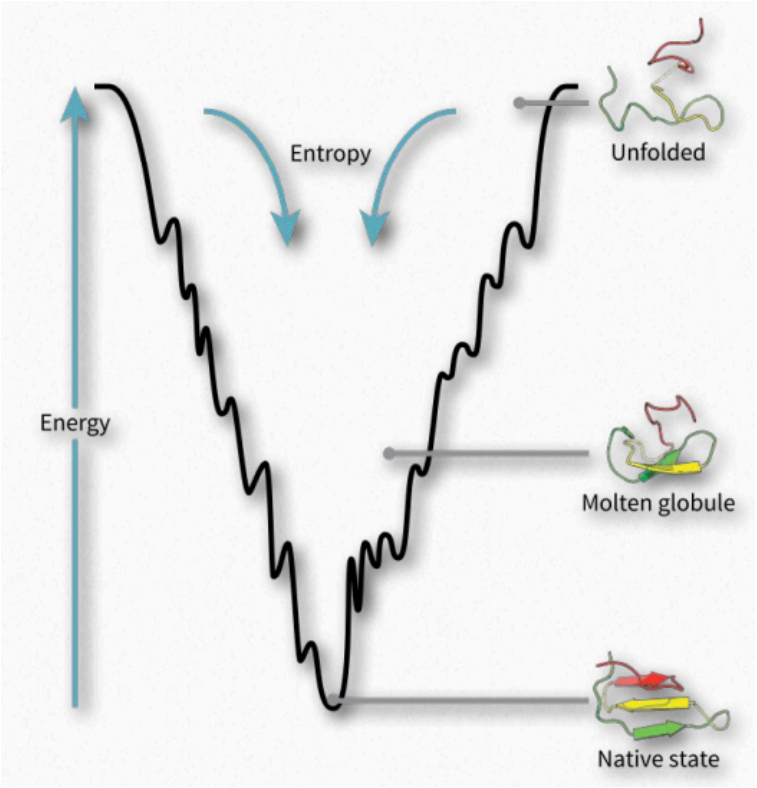


Figure 2.44 Folding funnel energy model of folding Wikipedia

Protein folding is hypothesized to occur in a "folding funnel" energy landscape in which a folded protein's native state corresponds to the minimal free energy possible in conditions of the medium (usually aqueous solvent) in which the protein is dissolved. As seen in the diagram (Figure 2.44), the energy funnel has numerous local minima (dips) in which a folding protein can become trapped as it moves down the energy plot. Other factors, such as temperature, electric/magnetic fields, and spacial considerations likely play roles.

If external forces affect local energy minima during folding, the process and end-product can be influenced. As the speed of a car going down a road will affect the safety of the journey, so too do energy considerations influence and guide the folding process,

resulting in fully functional, properly folded proteins in some cases and misfolded "mistakes" in others.

Getting stuck

As the folding process proceeds towards an energy minimum (bottom of the funnel in Figure 2.44), a protein can get "stuck" in any of the local minima and not reach the final folded state. Though the folded state is, in general, more organized and therefore has reduced entropy than the unfolded state, there are two forces that overcome the entropy decrease and drive the process forward.

The first is the magnitude of the decrease in energy as shown in the graph. Since $\Delta G = \Delta H - T\Delta S$, a decrease in ΔH can overcome a negative ΔS to make ΔG negative and push the folding process forward. Favorable (decreased) energy conditions arise with formation of ionic bonds, hydrogen bonds, disulfide bonds, and metallic bonds during the folding process. In addition, the hydrophobic effect increases entropy by allowing hydrophobic amino acids in the interior of a folded protein to exclude water, thus countering the impact of the ordering of the protein structure by making the ΔS less negative.

Structure prediction

Computer programs are very good at predicting secondary structure solely based on amino acid sequence, but struggle with determining tertiary structure using the same information. This is partly due to the fact that secondary structures have repeating points of stabilization based on geometry and any regular secondary structure (e.g., α-helix) varies very little from one to another. Folded structures, though, have an enormous number of possible structures as shown by Levinthal's Paradox.

Spectroscopy

Because of our inability to accurately predict tertiary structure based on amino acid sequence, proteins structures are actually determined using techniques of spectroscopy. In these approaches, proteins are subjected to varied forms of electromagnetic radiation and the ways they interact with the radiation allows researchers to determine atomic coordinates at Angstrom resolution from electron

densities (see X-ray crystallography) and how nuclei spins interact (see NMR).

Levinthal's paradox

In the late 1960s, Cyrus Levinthal outlined the magnitude of the complexity of the protein folding problem. He pointed out that for a protein with 100 amino acids, it would have 99 peptide bonds and 198 considerations for φ and ψ angles. If each of these had only three conformations, that would result in 3198 different possible foldings or 2.95×1094.

Even allowing a reasonable amount of time (one nanosecond) for each possible fold to occur, it would take longer than the age of the universe to sample all of them, meaning clearly that the process of folding is not occurring by a sequential random sampling and that attempts to determine protein structure by random sampling were doomed to fail. Levinthal, therefore, proposed that folding occurs by a sequential process that begins with a nucleation event that guides the process rapidly and is not unlike the funnel process depicted in Figure 2.44.

Diseases of protein misfolding

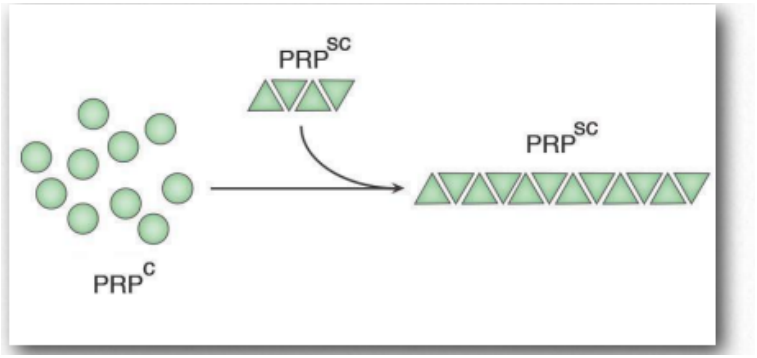


Figure 2.45 – Misfolding of the normal PRPc protein induced by PRPsc Image by Penelope Irving

The proper folding of proteins is essential to their function. It follows then that misfolding of proteins (also called proteopathy) might have consequences. In some cases, this might simply result in an inactive protein. Protein misfolding also plays a role in numerous

diseases, such as Mad Cow Disease, Alzheimers, Parkinson's Disease, and CreutzfeldJakob disease. Many, but not all, misfolding diseases affect brain tissue.

Figure 2.46 – Cows with Mad Cow Disease lose their ability to stand

Insoluble deposits

Misfolded proteins will commonly form aggregates called amyloids that are harmful to tissues containing them because they change from being soluble to insoluble in water and form deposits. The process by which misfolding (Figure 2.45) occurs is not completely clear, but in many cases, it has been demonstrated that a "seed" protein which is misfolded can induce the same misfolding in other copies of the same protein. These seed proteins are known as prions and they act as infectious agents, resulting in the spread of disease. The list of human diseases linked to protein misfolding is long and continues to grow. A Wikipedia link is HERE.

Prions

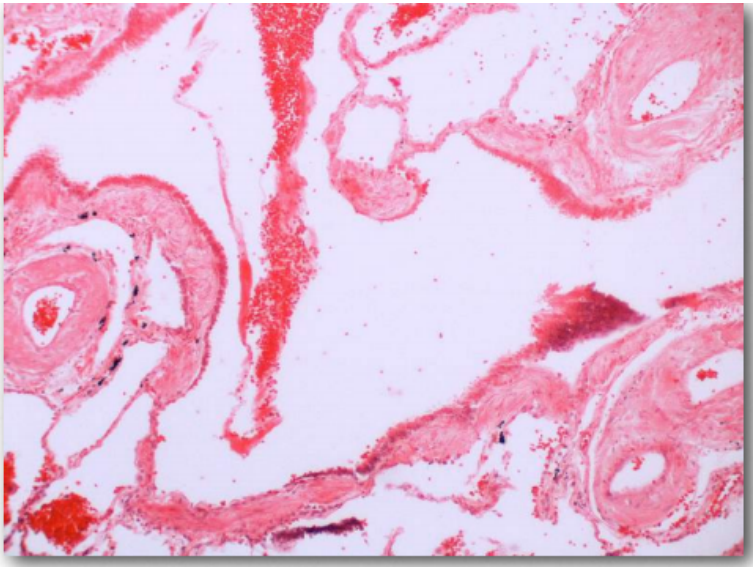


Figure 2.47 – Diffuse amyloidosis in a blood vessel (red dots) Wikipedia

Prions are infectious protein particles that cause transmissible spongiform encephalopathies (TSEs), the best known of which is Mad Cow disease. Other manifestations include the disease, scrapie, in sheep, and human diseases, such as CreutzfeldtJakob disease (CJD), Fatal Familial Insomnia, and kuru. The protein involved in these diseases is a membrane protein called PrP. PrP is encoded in the genome of many organisms and is found in most cells of the body. PrPc is the name given to the structure of PrP that is normal and not associated with disease. PrPSc is the name given to a misfolded form of the same protein, that is associated with the development of disease symptoms (Figure 2.45).

Misfolded

The misfolded PrPSc is associated with the TSE diseases and acts as an infectious particle. A third form of PrP, called PrPres can be found in TSEs, but is not infectious. The 'res' of PrPres indicates it is protease resistant. It is worth noting that all three forms of PrP have the same amino acid sequence and differ from each other only in the ways in which the polypeptide chains are folded. The most

dangerously misfolded form of PrP is PrPSc, because of its ability to act like an infectious agent – a seed protein that can induce misfolding of PrPc , thus converting it into PrPSc.

Function

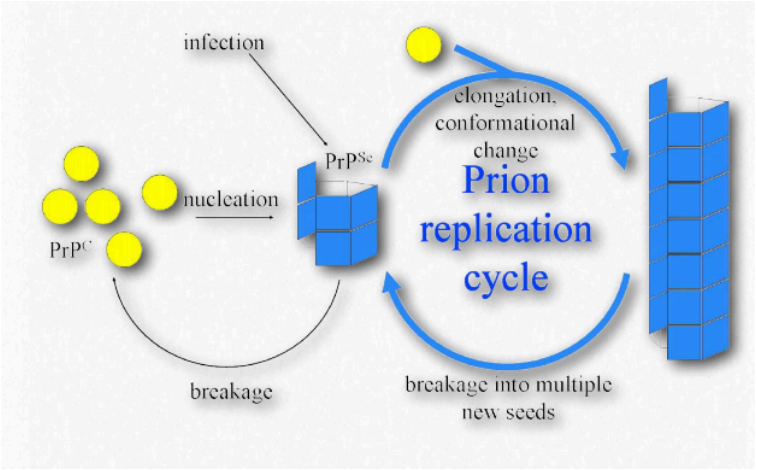


Figure 2.48 – One model of prion propagation Wikipedia

The function of PrPc is unknown. Mice lacking the PrP gene do not have major abnormalities. They do appear to exhibit problems with long term memory, suggesting a function for PrPc . Stanley Prusiner, who discovered prions and coined the term, received the Nobel Prize in Medicine in 1997 for his work. I think that if I chanced to be on A protein making up a prion I'd twist it and for goodness sakes Stop it from making fold mistakes

Amyloids

Amyloids are a collection of improperly folded protein aggregates that are found in the human body. As a consequence of their misfolding, they are insoluble and contribute to some twenty human diseases including important neurological ones involving prions. Diseases include (affected protein in parentheses) – Alzheimer's disease (Amyloid β), Parkinson's disease (α-synuclein), Huntington's disease (huntingtin), rheumatoid arthritis (serum amyloid A), fatal familial insomnia (PrPSc), and others.

Amino acid sequence plays a role in amyloidogenesis. Glutamine-

rich polypeptides are common in yeast and human prions. Trinucleotide repeats are important in Huntington's disease. Where sequence is not a factor, hydrophobic association between β-sheets can play a role.

Amyloid β

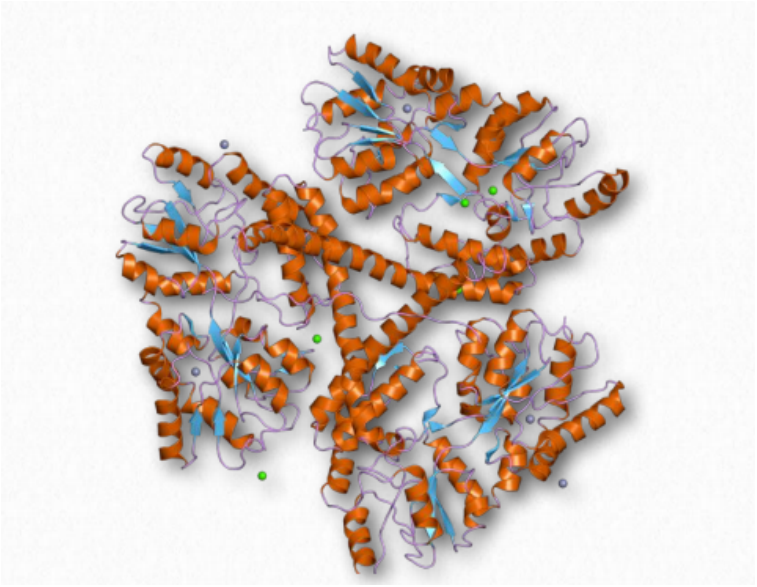


Figure 2.49 – Huntingtin

Amyloid β refers to collections of small proteins (36-43 amino acids) that appear to play a role in Alzheimer's disease. (Tau protein is the other factor.) They are, in fact, the main components of amyloid plaques found in the brains of patients suffering from the disease and arise from proteolytic cleavage of a larger amyloid precursor glycoprotein called Amyloid Precursor Protein, an integral membrane protein of nerve cells whose function is not known. Two proteases, β-secretase and γ- secretase perform this function. Amyloid β proteins are improperly folded and appear to induce other proteins to misfold and thus precipitate and form the amyloid characteristic of the disease. The plaques are toxic to nerve cells and give rise to the dementia characteristic of the disease.

It is thought that aggregation of amyloid β proteins during

misfolding leads to generation of reactive oxygen species and that this is the means by which neurons are damaged. It is not known what the actual function of amyloid β is. Autosomal dominant mutations in the protein lead to early onset of the disease, but this occurs in no more than 10% of the cases. Strategies for treating the disease include inhibition of the secretases that generate the peptide fragments from the amyloid precursor protein.

Huntingtin

Huntingtin is the central gene in Huntington's disease. The protein made from it is glutamine rich, with 6-35 such residues in its wild-type form. In Huntington's disease, this gene is mutated, increasing the number of glutamines in the mutant protein to between 36 and 250. The size of the protein varies with the number of glutamines in the mutant protein, but the wild-type protein has over 3100 amino acids and a molecular weight of about 350,000 Da. Its precise function is not known, but huntingtin is found in nerve cells, with the highest level in the brain. It is thought to possibly play roles in transport, signaling, and protection against apoptosis. Huntingtin is also required for early embryonic development. Within the cell, huntingtin is found localized primarily with microtubules and vesicles.

Trinucleotide repeat

The huntingtin gene contains many copies of the sequence CAG (called trinucleotide repeats), which code for the many glutamines in the protein. Huntington's disease arises when extra copies of the CAG sequence are generated when the DNA of the gene is being copied. Expansion of repeated sequences can occur due to slipping of the polymerase relative to the DNA template during replication. As a result, multiple additional copies of the trinucleotide repeat may be made, resulting in proteins with variable numbers of glutamine residues. Up to 35 repeats can be tolerated without problem. The number of repeats can expand over the course of a person's lifetime, however, by the same mechanism. Individuals with 36-40 repeats begin to show signs of the disease and if there are over 40, the disease will be present.

Molecular chaperones

The importance of the proper folding of proteins is highlighted by the diseases associated with misfolded proteins, so it is no surprise, then, that cells expend energy to facilitate the proper folding of proteins. Cells use two classes of proteins known as molecular chaperones, to facilitate such folding in cells. Molecular chaperones are of two kinds, the chaperones, and the chaperonins. An example of the first category is the Hsp70 class of proteins. Hsp stands for "heat shock protein", based on the fact that these proteins were first observed in large amounts in cells that had been briefly subjected to high temperatures. Hsps function to assist cells in stresses arising from heat shock and exposure to oxidizing conditions or toxic heavy metals, such as cadmium and mercury. However, they also play an important role in normal conditions, where they assist in the proper folding of polypeptides by preventing aberrant interactions that could lead to misfolding or aggregation. The Hsp70 proteins are found in almost all cells and use ATP hydrolysis to stimulate structural changes in the shape of the chaperone to accommodate binding of substrate proteins. The binding domain of Hsp70s contains a β-barrel structure which wraps around the polypeptide chain of the substrate and has affinity for hydrophobic side chains of amino acids. As shown in Figure 2.50, Hsp70 binds to polypeptides as they emerge from ribosomes during protein synthesis. Binding of substrate stimulates ATP hydrolysis and this is facilitated by another heat shock protein known as Hsp40. The hydrolysis of ATP causes the Hsp70 to taken on a closed conformation that helps shield exposed hydrophobic residues and prevent aggregation or local misfolding.

After protein synthesis is complete, ADP is released and replaced by ATP and this results in release of the substrate protein, which then allows the full length polypeptide to fold correctly.

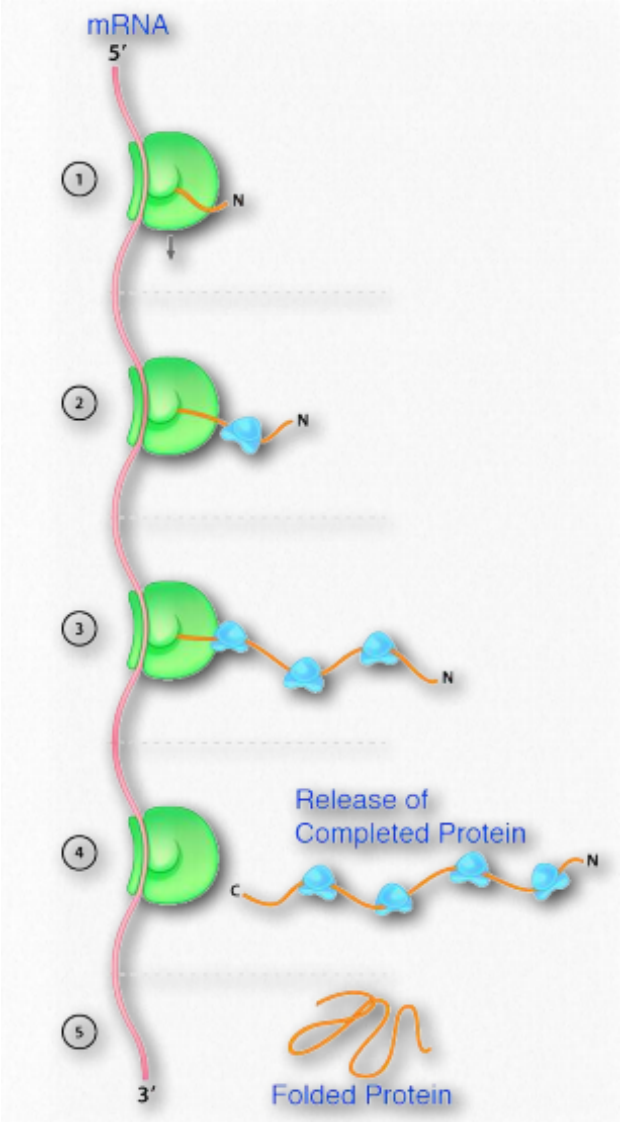


Figure 2.50 – Action of Hsp70 (blue) to facilitate proper folding of a protein (orange) Image by Aleia Kim

In heat shock

In times of heat shock or oxidative stress, Hsp70 proteins bind to unfolded hydrophobic regions of proteins to similarly prevent

them from aggregating and allowing them to properly refold. When proteins are damaged, Hsp70 recruits enzymes that ubiquitinate the damaged protein to target them for destruction in proteasomes. Thus, the Hsp70 proteins play an important role in ensuring not only that proteins are properly folded, but that damaged or nonfunctional proteins are removed by degradation in the proteasome.

Chaperonins

A second class of proteins involved in assisting other proteins to fold properly are known as chaperonins. There are two primary categories of chaperonins – Class I (found in bacteria, chloroplasts, and mitochondria) and Class II (found in the cytosol of eukaryotes and archaebacteria). The best studied chaperonins are the GroEL/ GroES complex proteins found in bacteria (Figure 2.51).

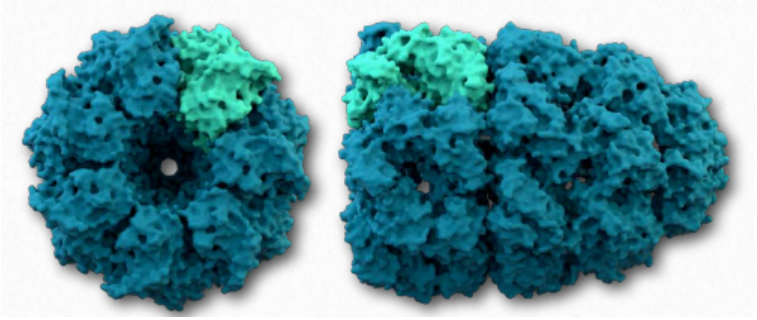


Figure 2.51 – View from bottom of GroEL (left) and GroEL/ GroES complex (right) Wikipedia

GroEL/GroES may not be able to undo aggregated proteins, but by facilitating proper folding, it provides competition for misfolding as a process and can reduce or eliminate problems arising from improper folding. GroEL is a double-ring 14mer with a hydrophobic region that can facilitate folding of substrates 15-60 kDa in size. GroES is a singlering heptamer that binds to GroEL in the presence of ATP and functions as a cover over GroEL. Hydrolysis of ATP by chaperonins induce large conformational changes that affect binding of substrate proteins and their folding. It is not known exactly how chaperonins fold proteins. Passive models postulate the

chaperonin complex functioning inertly by preventing unfavorable intermolecular interactions or placing restrictions on spaces available for folding to occur. Active models propose that structural changes in the chaperonin complex induce structural changes in the substrate protein.

Protein breakdown

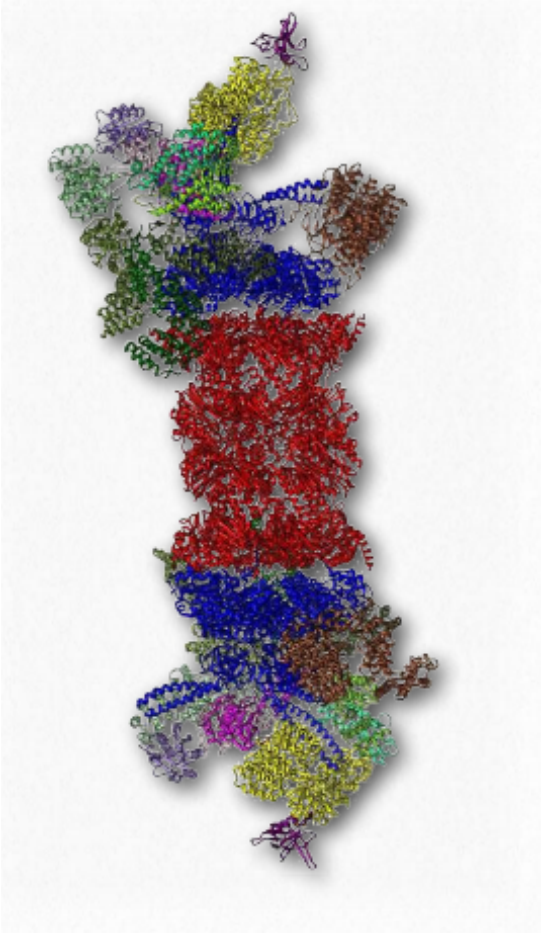


Figure 2.52 – 26S proteasome. Active site shown in red Wikipedia

Another protein complex that has an important function in the lifetime dynamics of proteins is the proteasome (Figure 2.52). Proteasomes, which are found in all eukaryotes and archaeans, as

well as some bacteria, function to break down unneeded or damaged proteins by proteolytic degradation. Proteasomes help to regulate the concentration of some proteins and degrade ones that are misfolded. The proteasomal degradation pathway plays an important role in cellular processes that include progression through the cell cycle, modulation of gene expression, and response to oxidative stresses.

Degradation in the proteasome yields short peptides seven to eight amino acids in length. Threonine proteases play important roles. Breakdown of these peptides yields individual amino acids, thus facilitating their recycling in cells. Proteins are targeted for degradation in eukaryotic proteasomes by attachment to multiple copies of a small protein called ubiquitin (8.5 kDa – 76 amino acids). The enzyme catalyzing the reaction is known as ubiquitin ligase. The resulting polyubiquitin chain is bound by the proteasome and degradation begins. Ubiquitin was named due to it ubiquitously being found in eukaryotic cells.

Ubiquitin

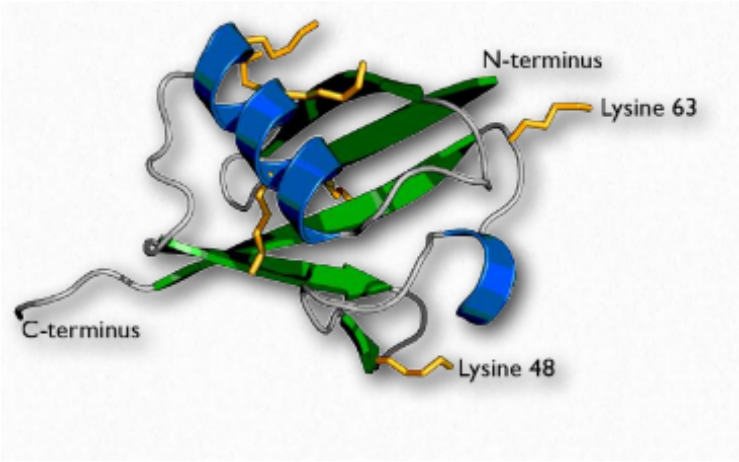


Figure 2.53 – Ubiquitin (lysine side chains shown in yellow) Wikipedia

Ubiquitin (Figure 2.53) is a small (8.5 kDa) multi-functional protein found in eukaryotic cells. It is commonly added to target proteins by action of ubiquitin ligase enzymes (E3 in Figure 2.54). One

(ubiquitination) or many (polyubiquitination) ubiquitin molecules may be added. Attachment of the ubiquitin is through the side chain of one of seven different lysine residues in ubiquitin.

The addition of ubiquitin to proteins has many effects, the best known of which is targeting the protein for degradation in the proteasome. Proteasomal targeting is seen when polyubiquitination occurs at lysines #29 and 48. Polyubiquitination or monoubiquitination at other lysines can result in altered cellular location and changed protein-protein interactions. The latter may alter affect inflammation, endocytic trafficking, translation and DNA repair.

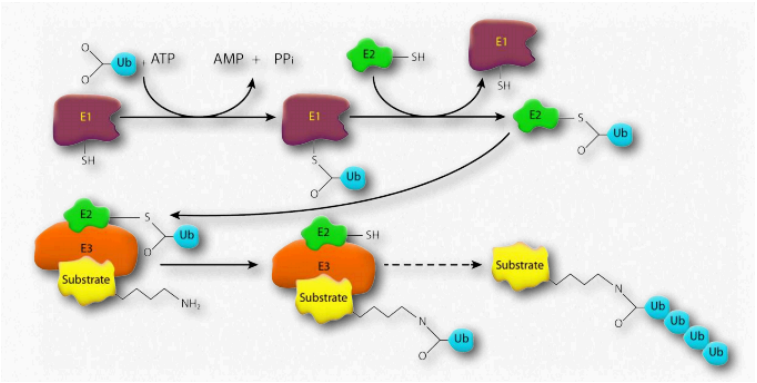
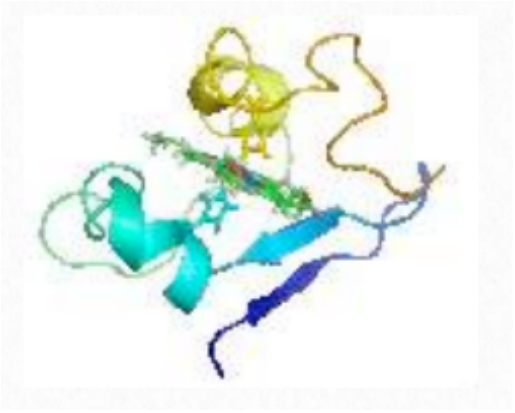


Figure 2.54 – Pathway for ubiquitination of a target substrate protein Image by Pehr Jacobson

Ubiquitin ligase malfunction

Parkin is a Parkinson's disease-related protein that, when mutated, is linked to an inherited form of the disease called autosomal recessive juvenile Parkinson's disease. The function of the protein is not known, but it is a component of the E3 ubiquitin ligase system responsible for transferring ubiquitin from the E2 protein to a lysine side chain on the target protein. It is thought that mutations in parkin lead to proteasomal dysfunction and a consequent inability to break down proteins harmful to dopaminergic neurons. This results in the death or malfunction of these neurons, resulting in Parkinson's disease.

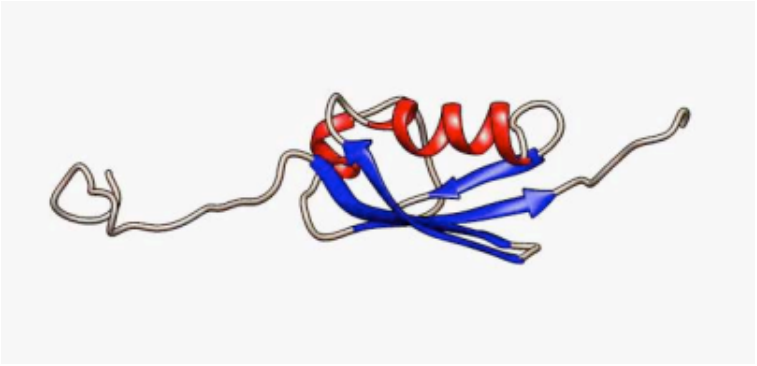
Intrinsically disordered proteins



Movie 2.1 – Dynamic movement of cytochrome C in solution Wikipedia

As is evident from the many examples described elsewhere in the book, the 3-D structure of proteins is important for their function. But, increasingly, it is becoming evident that not all proteins fold into a stable structure. Studies on the so-called intrinsically disordered proteins (IDPs) in the past cou- ple of decades has shown that many proteins are biologically active, even thought they fail to fold into stable structures. Yet other proteins exhibit regions that remain unfolded (IDP regions) even as the rest of the polypeptide folds into a structured form.

Intrinsically disordered proteins and disordered regions within proteins have, in fact, been known for many years, but were regarded as an anomaly. It is only recently, with the realization that IDPs and IDP regions are widespread among eukaryotic proteins, that it has been recognized that the observed disorder is a “feature, not a bug”.



Movie 2.2 SUMO-1, a protein with intrinsically disordered sections Wikipedia

Comparison of IDPs shows that they share sequence characteristics that appear to favor their disordered state. That is, just as some amino acid sequences may favor the folding of a polypeptide into a particular structure, the amino acid sequences of IDPs favor their remaining unfolded. IDP regions are seen to be low in hydrophobic residues and unusually rich in polar residues and proline. The presence of a large number of charged amino acids in the IDPs can inhibit folding through charge repulsion, while the lack of hydrophobic residues makes it difficult to form a stable hydrophobic core, and proline discourages the formation of helical structures. The observed differences between amino acid sequences in IDPs and structured proteins have been used to design algorithms to predict whether a given amino acid sequence will be disordered.

What is the significance of intrinsically disordered proteins or regions? The fact that this property is encoded in their amino acid sequences suggests that their disorder may be linked to their function. The flexible, mobile nature of some IDP regions may play a crucial role in their function, permitting a transition to a folded structure upon binding a protein partner or undergoing post-translational modification. Studies on several wellknown proteins with IDP regions suggest some answers. IDP regions may enhance the ability of proteins like the lac repressor to translocate along

the DNA to search for specific binding sites. The flexibility of IDPs can also be an asset in protein-protein interactions, especially for proteins that are known to interact with many different protein partners.

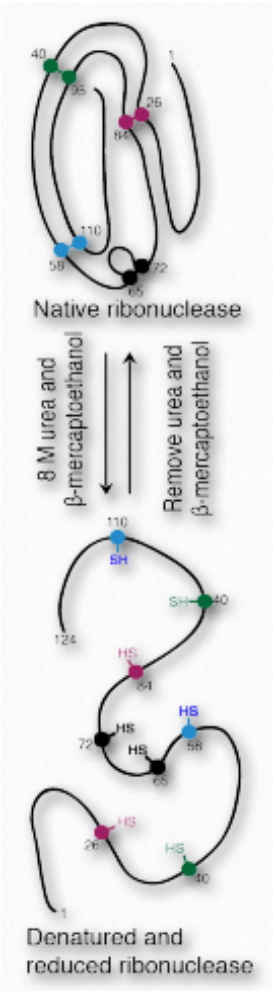


Figure 2.55 – Denaturation and renaturation of ribonuclease Wikipedia

For example, p53 has IDP regions that may allow the protein to interact with a variety of functional partners. Comparison of the

known functions of proteins with predictions of disorder in these proteins suggests that IDPs and IDP regions may disproportionately function in signaling and regulation, while more structured proteins skew towards roles in catalysis and transport. Interestingly, many of the proteins found in both ribosomes and spliceosomes are predicted to have IDP regions that may play a part in correct assembly of these complexes. Even though IDPs have not been studied intensively for very long, what little is known of them suggests that they play an important and underestimated role in cells.

Metamorphic proteins

Another group of proteins that have recently changed our thinking about protein structure and function are the so-called metamorphic proteins. These proteins are capable of forming more than one stable, folded state starting with a single amino acid sequence. Although it is true that multiple folded conformations are not ruled out by the laws of physics and chemistry, metamorphic proteins are a relatively new discovery. It was known, of course, that prion proteins were capable of folding into alternative structures, but metamorphic proteins appear to be able to toggle back and forth between two stable structures. While in some cases, the metamorphic protein undergoes this switch in response to binding another molecule, some proteins that can accomplish this transition on their own. An interesting example is the signaling molecule, lymphotactin. Lymphotactin has two biological functions that are carried out by its two conformers- a monomeric form that binds the lymphotactin receptor and a dimeric form that binds heparin. It is possible that this sort of switching is more widespread than has been thought.

Refolding denatured proteins

All information for protein folding is contained in the amino acid sequence of the protein. It may seem curious then that most proteins do not fold into their proper, fully active form after they have been+++ denatured and the denaturant is removed. A few do, in fact. One good example is bovine ribonuclease (Figure 2.55). Its

catalytic activity is very resistant to heat and urea and attempts to denature it don't work very well. However, if one treats the enzyme with β-mercaptoethanol (which breaks disulfide bonds) prior to urea treatment and/or heating, activity is lost, indicating that the covalent disulfide bonds help stabilize the overall enzyme structure and when they are broken, denaturation can readily occur. When the mixture cools back down to room temperature, over time some enzyme activity reappears, indicating that ribonuclease re-folded under the new conditions.

Interestingly, renaturation will occur maximally if a tiny amount of β-mercaptoethanol is left in the solution during the process. The reason for this is because β- mercaptoethanol permits reduction (and breaking) of accidental, incorrect disulfide bonds during the folding process. Without it, these disulfide bonds will prevent proper folds from forming.

Irreversible denaturation

Most enzymes, however, do not behave like bovine ribonuclease. Once denatured, their activity cannot be recovered to any significant There are not very many ways Inactivating RNase It's stable when it's hot or cold Because disulfides tightly hold If you desire to make it stall Use hot mercaptoethanol extent. This may seem to contradict the idea of folding information being inherent to the sequence of amino acids in the protein. It does not.

Most enzymes don't refold properly after denaturation for two reasons. First, normal folding may occur as proteins are being made. Interactions among amino acids early in the synthesis are not "confused" by interactions with amino acids later in the synthesis because those amino acids aren't present as the process starts.

Chaperonins' role

In other cases, the folding process of some proteins in the cell relied upon action of chaperonin proteins (see HERE). In the absence of chaperonins, interactions that might result in misfolding occur, thus preventing proper folding. Thus, early folding and the assistance of chaperonins eliminate some potential "wrong-folding"

interactions that can occur if the entire sequence was present when folding started.

Quaternary structure

A fourth level of protein structure is that of quaternary structure. It refers to structures that arise as a result of interactions between multiple polypeptides. The units can be identical multiple copies or can be different polypeptide chains. Adult hemoglobin is a good example of a protein with quaternary structure, being composed of two identical chains called α and two identical chains called β.

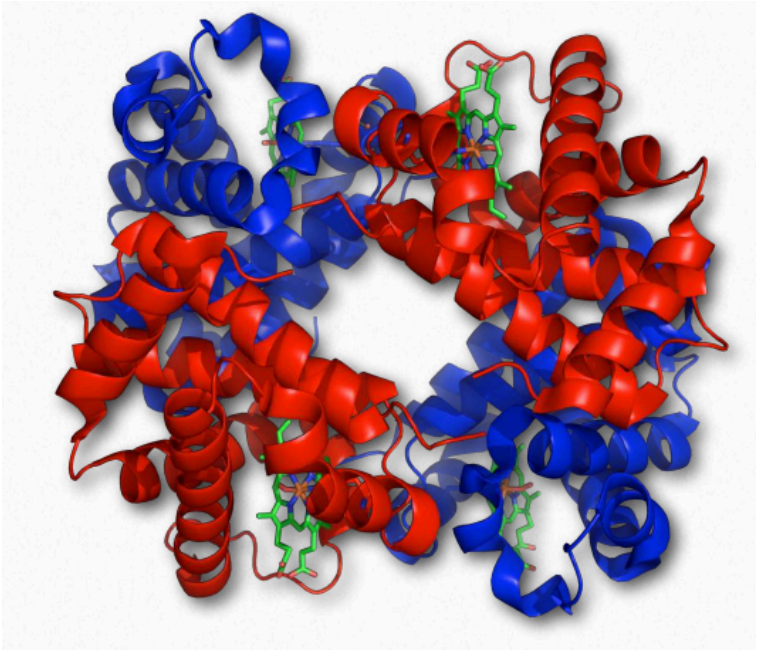
Though the α-chains are very similar to the β- chains, they are not identical. Both of the α- and the β-chains are also related to the single polypeptide chain in the related protein called myoglobin. Both myoglobin and hemoglobin have similarity in binding oxygen, but their behavior towards the molecule differ significantly. Notably, hemoglobin's multiple subunits (with quaternary structure) compared to myoglobin's single subunit (with no quaternary structure) give rise to these differences.

References

1. https://en.wikipedia.org/wiki/Van_der_W aals_force 105

2.4: Structure and Function-Proteins II

In this section, we hope to bring to life the connection between structure and function of proteins. So far, we have described notable features of the four elements (primary, secondary, tertiary, and quaternary) of protein structure and discussed example proteins/motifs exhibiting them. In this section, we will examine from a functional perspective a few proteins/domains whose function relies on secondary, tertiary, or quaternary structure. It is, of course a bit of a narrow focus to ascribe protein function to any one component of structure, but our hope is by presenting these examples, we can bring to life the way in which a protein's secondary, tertiary, and quarternary structure lead to the functions it has.



Hemoglobin Wikipedia

Fibrous proteins – secondary structure

Proteins whose cellular or extracellular roles have a strong structural component are composed primarily of primary and second structure, with little folding of the chains. Thus, they have very little tertiary structure and are fibrous in nature. Proteins exhibiting these traits are commonly insoluble in water and are referred to as fibrous proteins (also called scleroproteins). The examples described in this category are found exclusively in animals where they serve roles in flesh, connective tissues and hardened external structures, such as hair. They also contain the three common fibrous protein structures α -helices (keratins), β-strands/

sheets (fibroin & elastin) and triple helices (collagen). The fibrous proteins have some commonality of amino acid sequence. Each possesses an abundance of repeating sequences of amino acids with small, non-reactive side groups. Many contain short repeats of sequences, often with glycine.

Keratins

Figure 2.56 – The horns of an impala are composed of keratin Wikipedia

The keratins are a family of related animal proteins that take numerous forms. α-keratins are structural components of the outer layer of human skin and are integral to hair, nails, claws, feathers,

beaks, scales, and hooves. Keratins provide strength to tissues, such as the tongue, and over 50 different keratins are encoded in the human genome. At a cellular level, keratins comprise the intermediate filaments of the cytoskeleton. α- keratins primarily contain α-helices, but can also have β-strand/sheet structures. Individual α-helices are often intertwined to form coils of coiled structures and these strands can also be further joined together by disulfide bonds, increasing structural strength considerably. This is particularly relevant for α-keratin in hair, which contains about 14% cysteine. The odor of burned hair and that of the chemicals used to curl/uncurl hair (breaking/re-making disulfide bonds) arise from their sulfurous components. β-keratins are comprised of β-sheets, as their name implies.

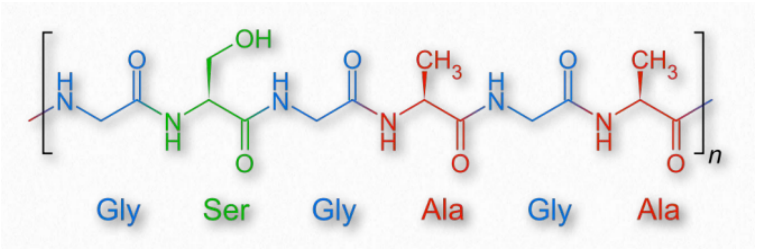


Figure 2.57 – The repeating amino acid sequence of fibroin

Fibroin

An insoluble fibrous protein that is a component of the silk of spiders and the larvae of moths and other insects, fibroin is comprised of antiparallel β-strands tightly packed together to form β- sheets. The primary structure of fibroin is a short repeating sequence with glycine at every other residue (Figure 2.57). The small R-groups of the glycine and alanine in the repeating sequence allows for the tight packing characteristic of the fibers of silk. Wikipedia link HERE Elastin As suggested by its name, elastin is a protein with elastic characteristics that functions in many tissues of the body to allow them to resume their shapes after expanding or contracting. The protein is rich in glycine and proline and can comprise over 50% of the weight of dry, defatted arteries.

Figure 2.58 – Weaving of a silk sari Wikipedia

Elastin

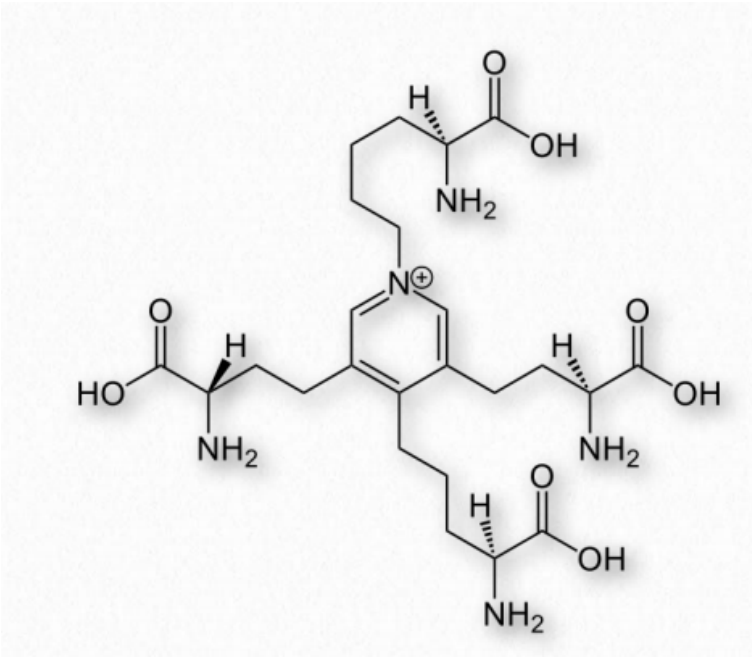


Figure 2.59 – Desmosine Wikipedia

is made by linking tropoelastin proteins together through lysine residues to make a durable complex crosslinked by desmosine. In arteries, elastin helps with pressure wave propagation for facilitating blood flow.

Collagen

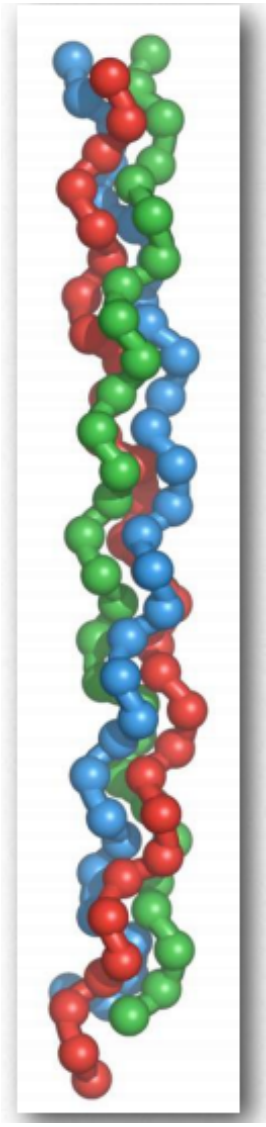


Figure 2.60 – Collagen's triple helix Wikipedia

Collagen is the most abundant protein in mammals, occupying up to a third of the total mass. There are at least 16 types of collagen. Its fibers are a major component of tendons and they are also found

abundantly in skin. Collagen is also prominent in cornea, cartilage, bone, blood vessels and the gut.

Collagen's structure is an example of a helix of helices, being composed of three lefthanded helical chains that each are coiled together in a right-handed fashion to make the collagen fiber (Figure 2.60). Each helix is stretched out more than an α-helix, giving it an extended appearance. On the inside of the triple helical structure, only residues of glycine are found, since the side chains of other amino acids are too bulky. Collagen chains have the repeating structure glycinem-n where m is often proline and n is often hydroxyproline (Figure 2.61).

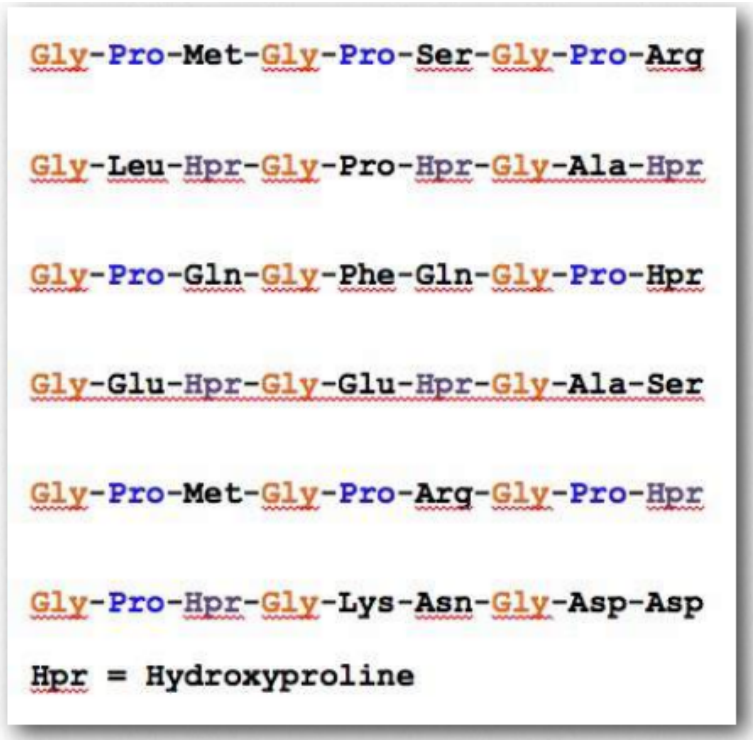


Figure 2.61 – Repeating sequences in collagen

Collagen is synthesized in a pre-procollagen form. Processing of the pre-procollagen in the endoplasmic reticulum results in glycosylation, removal of the 'pre' sequence, and hydroxylation of

lysine and proline residues (see below). The hydroxides can form covalent cross-links with each other, strengthening the collagen fibers. As pro-collagen is exported out of the cell, proteases trim it, resulting in a final form of collagen called tropocollagen.

Hydroxylation

Hydroxylation of proline and lysine side chains occurs post-translationally in a reaction catalyzed by prolyl-4-hydroxylase and lysyl-hydroxylase (lysyl oxidase), respectively. The reaction requires vitamin C. Since hydroxylation of these residues is essential for formation of stable triple helices at body temperature, vitamin C deficiency results in weak, unstable collagen and, consequently, weakened connective tissues. It is the cause of the disease known as scurvy. Hydrolyzed collagen is used to make gelatin, which is important in the food industry. collagens. Wikipedia link HERE

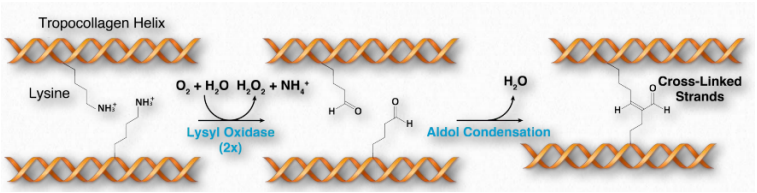


Figure 2.62 – Oxidation and cross-linking of lysine residues in tropocollagen. Only two strands of the triple helix are shown for simplicity Image by Aleia Kim

Lamins

Lamins are fibrous proteins that provide structure in the cell nucleus and play a role in transcription regulation. They are similar to proteins making up the intermediate filaments, but have extra amino acids in one coil of the protein. Lamins help to form the nuclear lamin in the interior of the nuclear envelope and play important roles in assembling and disassembling the latter in the process of mitosis. They also help to position nuclear pores. In the process of mitosis, disassembly of the nuclear envelope is promoted by phosphorylation of lamins by a protein called mitosis promoting factor and assembly is favored by reversing the reaction (dephosphorylation).

Structural domains – tertiary structure

Every globular protein relies on its tertiary structure to perform its function, so rather than trying to find representative proteins for tertiary structure (an almost impossible task!), we focus here on a few elements of tertiary structure that are common to many proteins. These are the structural domains and they differ from the structural motifs of supersecondary structure by being larger (25-500 amino acids), having a conserved amino acid sequence, and a history of evolving and functioning independently of the protein chains they are found in. Structural domains are fundamental units of tertiary structure and are found in more than one protein. A structural domain is selfstabilizing and often folds independently of the rest of the protein chain.

Leucine zipper

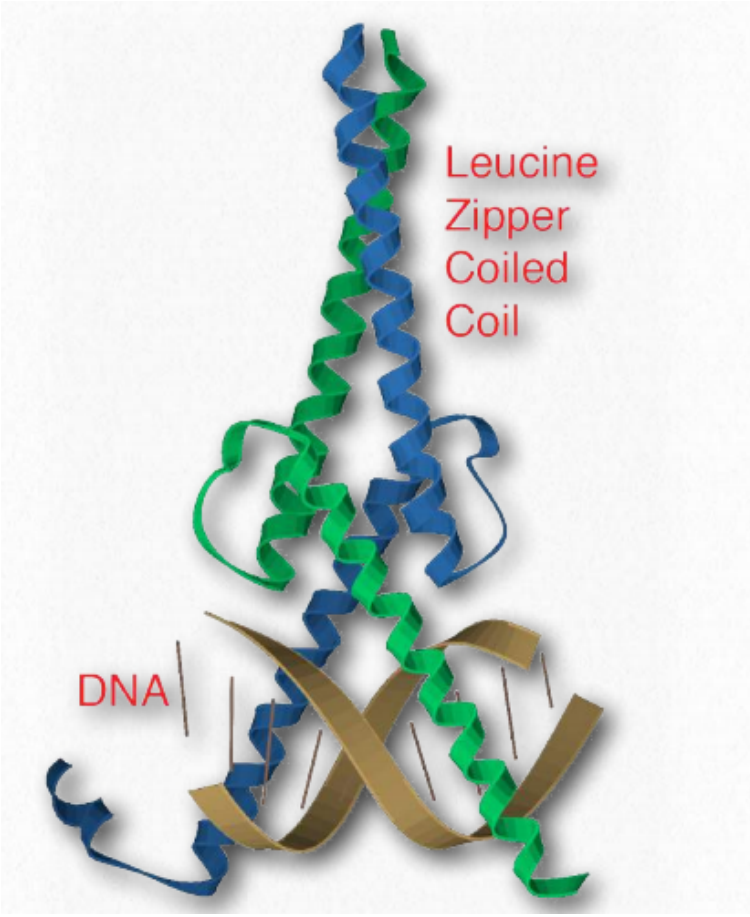


Figure 2.63 – Leucine zipper bound to DNA Wikipedia

A common feature of many eukaryotic DNA binding proteins, leucine zippers are characterized by a repeating set of leucine residues in a protein that interact like a zipper to favor dimerization. Another part of the domain has amino acids (commonly arginine and lysine) that allow it to interact with the DNA double helix (Figure 2.63). Transcription factors that contain leucine zippers include Jun-B, CREB, and AP-1 fos/ jun.

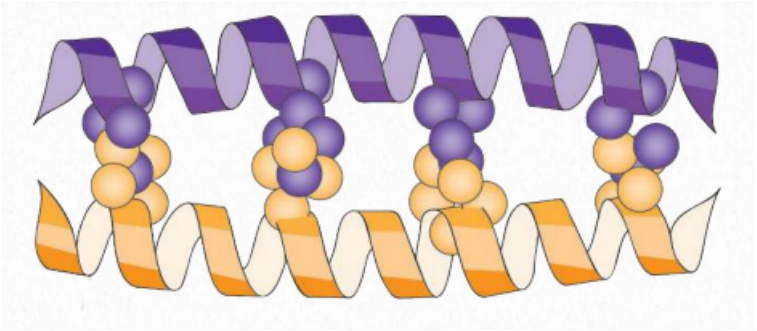


Figure 2.64 – Leucine zipper structure. Leucines are indicated by orange and purple balls. Image by Penelope Irving

Zinc fingers

The shortest structural domains are the zinc fingers, which get their name from the fact that one or more coordinated zinc ions stabilize their finger-like structure. Despite their name, some zinc fingers do not bind zinc. There are many structural domains classified as zinc fingers and these are grouped into different families. Zinc fingers were first identified as components of DNA binding transcription factors, but others are now known to bind RNA, protein, and even lipid structures. Cysteine and histidine side chains commonly play roles in coordinating the zinc.

Src SH_2 domain

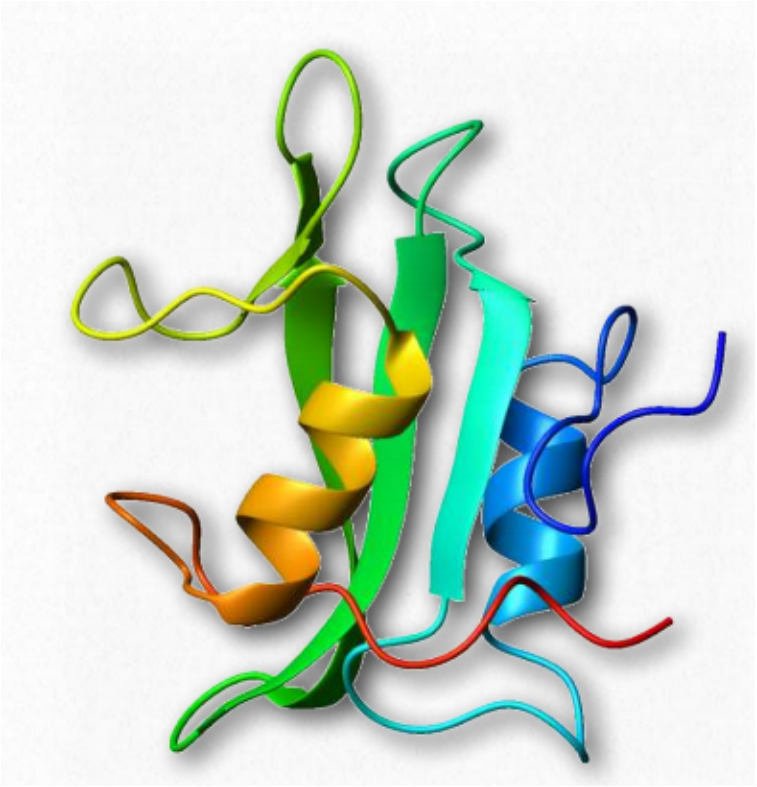


Figure 2.65 – SH2 Domain Wikipedia

The Src oncoprotein contains a conserved SH_2structural domain that recognizes and binds phosphorylated tyrosine side chains in other proteins (Figure 2.65). Phosphorylation is a fundamental activity in signaling and phosphorylation of tyrosine and interaction between proteins carrying signals is critically needed for cellular communication. The SH_2domain is found in over 100 human proteins.

Helix-turn-helix domain

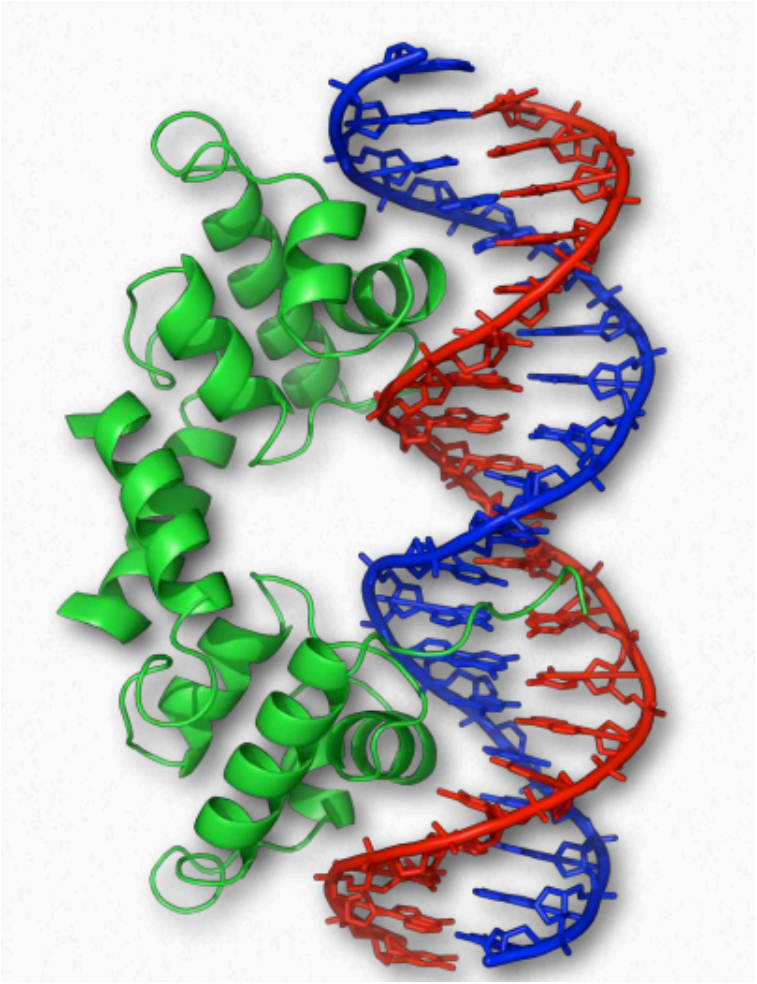


Figure 2.66 – Helix-Turn-Helix Domain of a Protein Bound to DNA Wikipedia

Helix-turn-helix is a common domain found in DNA binding proteins, consisting of two α-helices separated by a small number of amino acids. As seen in Figure 2.66, the helix parts of the structural domain interact with the bases in the major groove of DNA. Individual α-helices in a protein are part of a helix-turn-helix structure, where the turn separates the individual helices.

Pleckstrin homology domain

Pleckstrin Homology (PH) domains are protein domains with important functions in the process of signaling. This arises partly from the affinity for binding phosphorylated inositides, such as PIP2 and PIP3, found in Figure 2.66 – Helix-Turn-Helix Domain of a Protein Bound to DNA Wikipedia Figure 2.65 – SH2 Domain Wikipedia biological membranes. PH domains can also bind to G-proteins and protein kinase C. The domain spans about amino acids and is found in numerous signaling proteins. These include Akt/Rac Serine/ Threonine Protein Kinases, Btk/ltk/Tec tyrosine protein kinases, insulin receptor substrate (IRS-1), Phosphatidylinositolspecific phospholipase C, and several yeast proteins involved in cell cycle regulation.

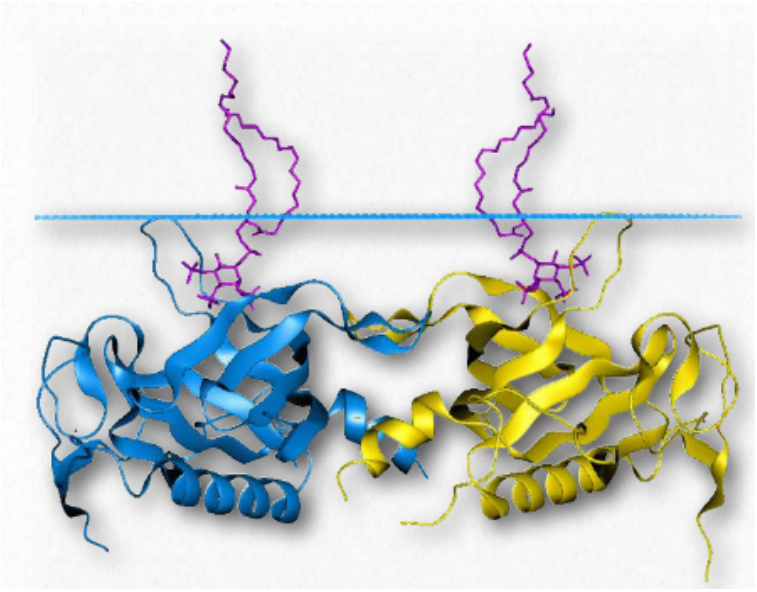


Figure 2.67 – Pleckstrin homology domain of Btk tyrosine protein kinase. The protein is embedded in a membrane (above blue line) Wikipedia

Structural globular proteins

Enzymes catalyze reactions and proteins such as hemoglobin perform important specialized functions. Evolutionary selection has reduced and eliminated waste so that we can be sure every protein

in a cell has a function, even though in some cases we may not know what it is. Sometimes the structure of the proFigure 2.68 – Relationship of basement membrane to epithelium, endothelium, and connective tissue tein is its primary function because the structure provides stability, organization, connections other important properties. It is with this in mind that we present the following proteins.

Basement membrane

The basement membrane is a layered extracellular matrix of tissue comprised of protein fibers (type IV collagen) and glycosaminoglycans that separates the epithelium from other tissues (Figure 2.68). More importantly, the basement membrane acts like a glue to hold tissues together. The skin, for example, is anchored to the rest of the body by the basement membrane. Basement membranes provide an interface of interaction between cells and the environment around them, thus facilitating signaling processes. They play roles in differentiation during embryogenesis and also in maintenance of function in adult organisms.

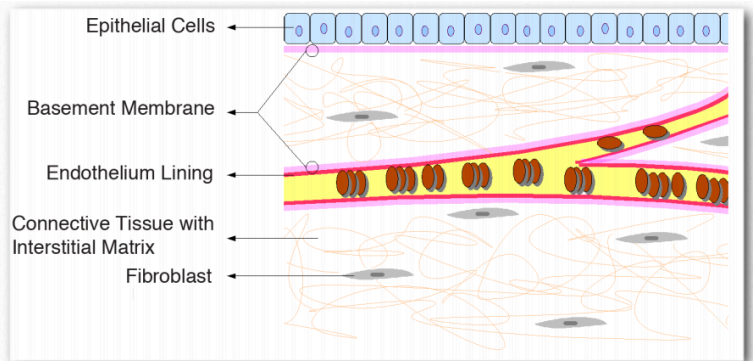


Figure 2.68 – Relationship of basement membrane to epithelium, endothelium, and connective tissue

Actin

Actin is the most abundant globular protein found in most types of eukaryotic cells, comprising as much as 20% of the weight of muscle cells. Similar proteins have been identified in bacteria (MreB) and archaeons (Ta0583). Actin is a monomeric subunit able to

polymerize readily into two different types of filaments. Microfilaments are major component of the cytoskeleton and are acted on by myosin in the contraction of muscle cells (See HERE). Actin will be discussed in more detail in the next section HERE.

Intermediate Filaments

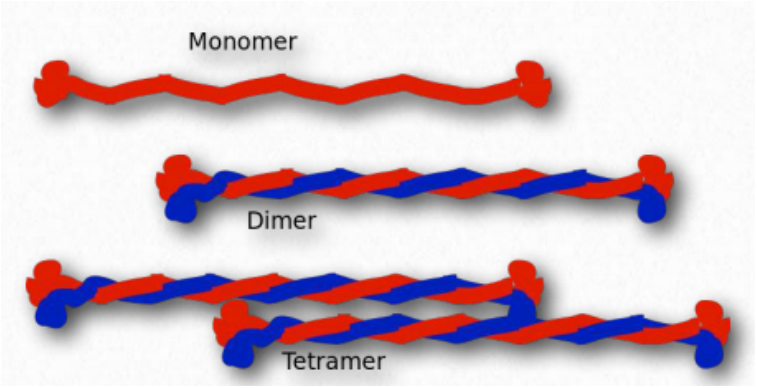


Figure 2.69 – Assembly of intermediate filaments

Intermediate filaments are a part of the cytoskeleton in many animal cells and are comprised of over 70 different proteins. They are called intermediate because their size (average diameter = 10 nm) is between that of the microfilaments (7 nm) and the microtubules (25 nm).

The intermediate filament components include fibrous proteins, such as the keratins and the lamins, which are nuclear, as well as cytoplasmic forms. Intermediate filaments give flexibility to cells because of their own physical properties. They can, for example, be stretched to several times of their original length.

Six types

There are six different types of intermediate filaments. Type I and II are acidic or basic and attract each other to make larger filaments. They include epithelial keratins and trichocytic keratins (hair components). Type III proteins include four structural proteins – desmin, GFAP (glial fibrillary acidic protein), peripherin, and vimentin. Type IV also is a grouping of three proteins and one multiprotein structure (neurofilaments). The three proteins are α-internexin, synemin, and syncoilin. Type V intermediate filaments

encompass the lamins, which give structure to the nucleus. Phosphorylation of lamins leads to their disassembly and this is important in the process of mitosis. The Type VI category includes only a single protein known as nestin.

Tubulin

A third type of filament found in cells is that of the microbutules. Comprised of a polymer of two units of a globular protein called tubulin, microtubules provide "rails" for motor proteins to move organelles and other "cargo" from one part of a cell to another. Microtubules and tubulin are discussed in more detail HERE.

Vimentin

Vimentin (Figure 2.70) is the most widely distributed protein of the intermediate filaments. It is expressed in fibroblasts, leukocytes, and blood vessel endothelial cells. The protein has a significant role maintaining the position of organelles in the cytoplasm, with attachments to the nucleus, mitochondria, and endoplasmic reticulum (Figure 2.70). Vimentin provides elasticity to cells and resilience that does not arise from the microtubules or microfilaments. Wounded mice that lack the vimentin gene survive, but take longer to heal wounds than wild type mice. Vimentin also controls the movement of cholesterol from lysosomes to the site of esterification. The result is a reduction in the amount of cholesterol stored inside of cells and has implications for adrenal cells, which must have esters of cholesterol.

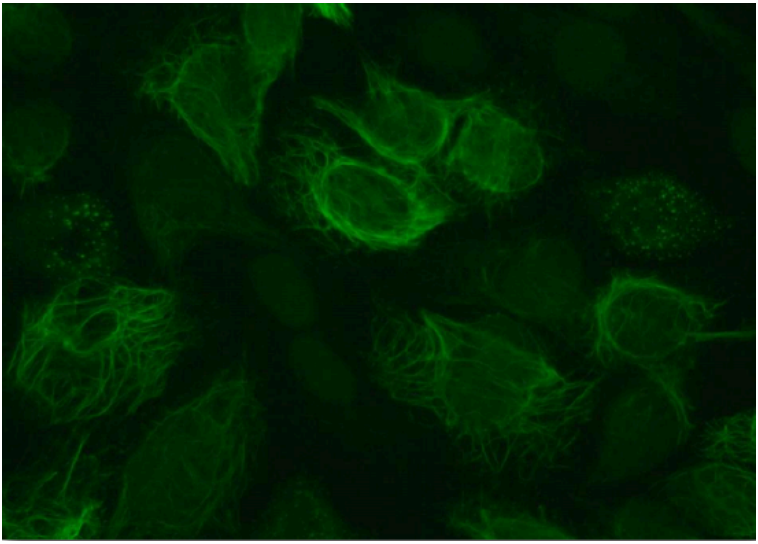


Figure 2.70 – Vimentin in cells Wikipedia

Mucin

Mucins are a group of proteins found in animal epithelial tissue that have many glycosyl residues on them and typically are of high molecular weight (1 to 10 million Da). They are gel-like in their character and are often used for lubrication. Mucus is comprised of mucins. In addition to lubrication, mucins also help to control mineralization, such as bone formation in vertebrate organisms and calcification in echinoderms. They also play roles in the immune system by helping to bind pathogens. Mucins are commonly secreted onto mucosal surfaces (nostrils, eyes, mouth, ears, stomach, genitals, anus) or into fluids, such as saliva. Because of their extensive mucosylation, mucins hold a considerable amount of water (giving them the "slimy" feel) and are resistant to proteolysis.

Vinculin

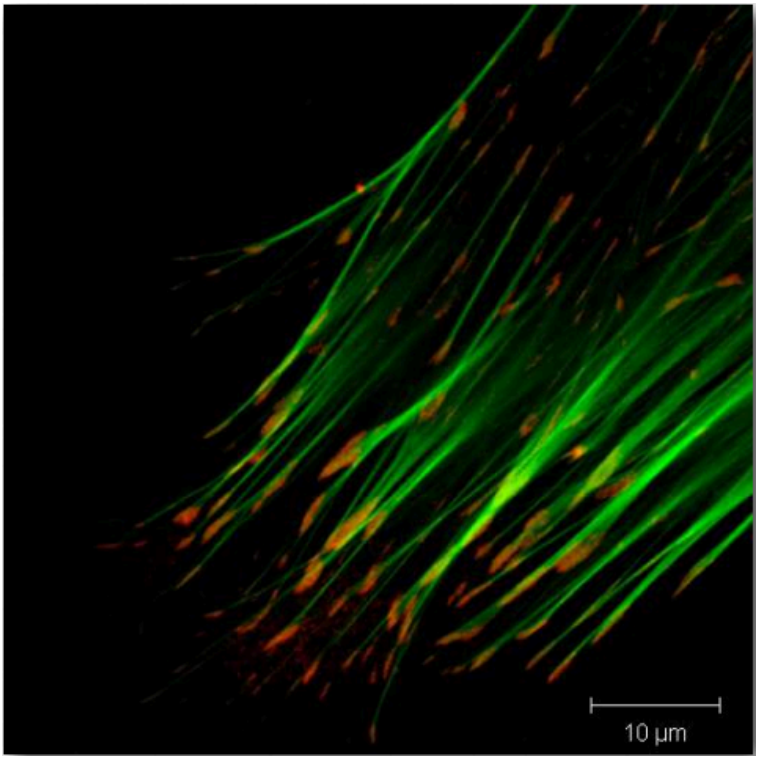


Figure 2.71 – Actin filaments (green) attached to vinculin in focal adhesion (red) Wikipedia

Vinculin (Figure 2.72) is a membrane cytoskeletal protein found in the focal adhesion structures of mammalian cells. It is found at cell-cell and cell-matrix junctions and interacts with integrins, talin, paxillins and F-actin. Vinculin is thought to assist (along with other proteins) in anchoring actin microfilaments to the membrane (Figure 2.71). Binding of vinculin to actin and to talin is regulating by polyphosphoinositides and can be inhibited by acidic phospholipids.

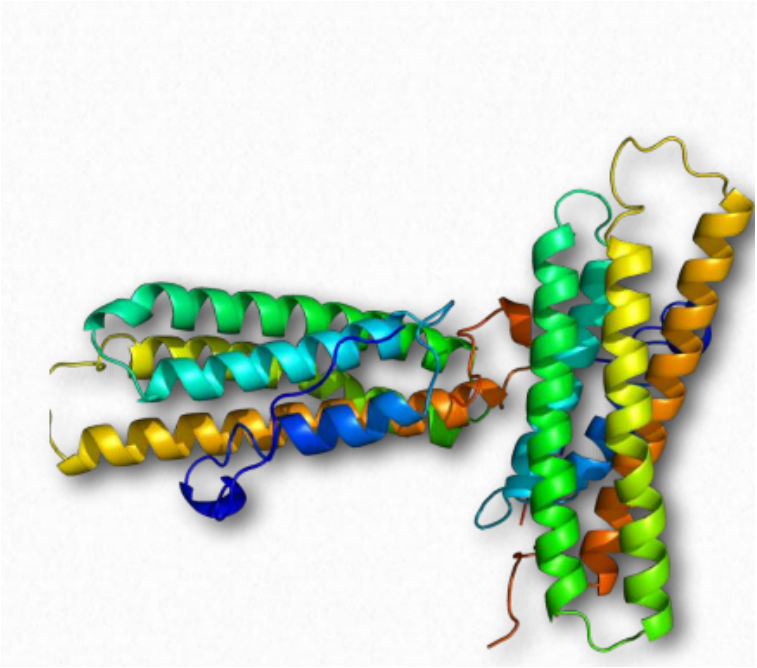


Figure 2.72 – Vinculin Wikipedia

Syndecans

Syndecans are transmembrane proteins that make a single pass with a long amino acid chain (24-25 residues) through plasma membranes and facilitate G proteincoupled receptors' interaction with Figure 2.71 – Actin filaments (green) attached to vinculin in focal adhesion (red) Wikipedia ligands, such as growth factors, fibronectin, collagens (I, III, and IV) and antithrombin-1. Syndecans typically have 3-5 heparan sulfate and chondroitin sulfate chains attached to them.

Heparan sulfate can be cleaved at the site of a wound and stimulate action of fibroblast growth factor in the healing process. The role of syndecans in cell-cell adhesion is shown in mutant cells lacking syndecan I that do not adhere well to each other. Syndecan 4 is also known to adhere to integrin. Syndecans can also inhibit the spread of tumors by the ability of the syndecan 1 ectodomain to

suppress growth of tumor cells without affecting normal epithelial cells.

Defensin

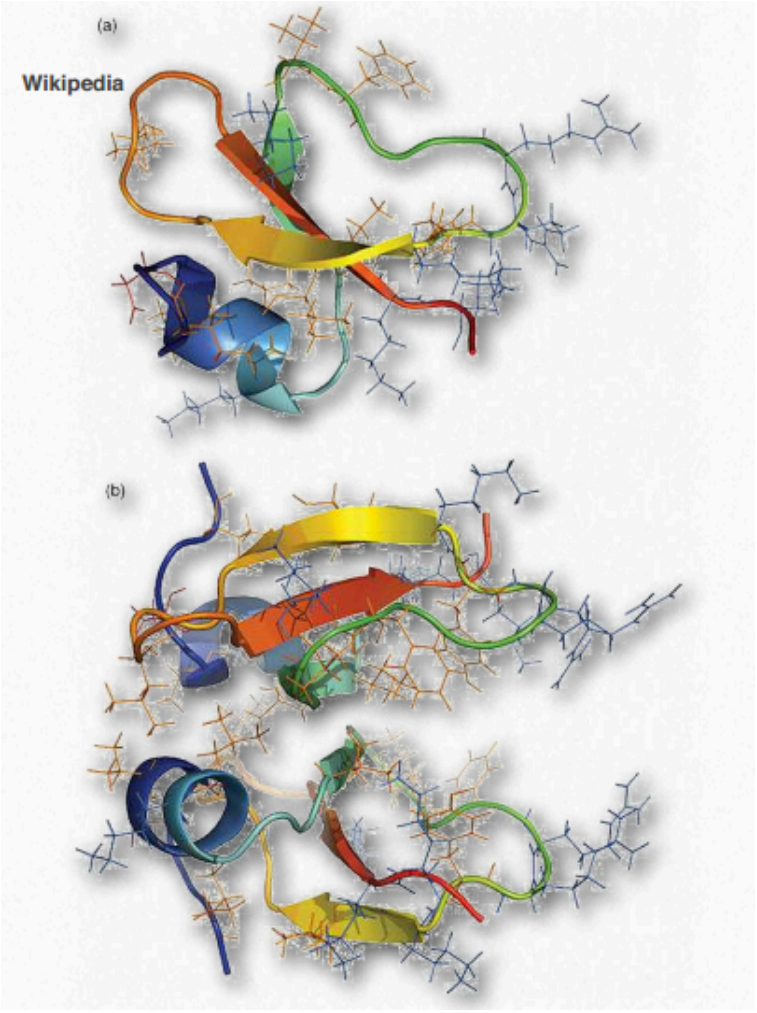


Figure 2.73 – Defensin monomer (top) and dimer (bottom) – Cationic residues in blue, hydrophobic residues in orange, and anionic residues in red Wikipedia

Defensins (Figure 2.73) are a group of small cationic proteins (rich in cysteine residues) that serve as host defense peptides in

vertebrate and invertebrate organisms. They protect against infection by various bacteria, fungi, and viruses. Defensins contain between 18 and 45 amino acids with (typically) about 6- 8 cysteine residues. In the immune system, defensins help to kill bacteria engulfed by phagocytosis by epithelial cells and neutrophils. They kill 120 Figure 2.72 – Vinculin Wikipedia bacteria by acting like ionophores – binding the membrane and opening pore-like structures to release ions and nutrients from the cells.

Focal adhesions

In the cell, focal adhesions are structures containing multiple proteins that mechanically link cytoskeletal structures (actin bundles) with the extracellular matrix. They are dynamic, with proteins bringing and leaving with signals regarding the cell cycle, cell motility, and more almost constantly. Focal adhesions serve as anchors and as a signaling hub at cellular locations where integrins bind molecules and where membrane clustering events occur. Over 100 different proteins are found in focal adhesions.

Focal adhesions communicate important messages to cells, acting as sensors to update information about the status of the extracellular matrix, which, in turn, adjusts/ affects their actions. In sedentary cells, they are stabler than in cells in motion because when cells move, focal adhesion contacts are established at the "front" and removed at the rear as motion progresses. This can be very important in white blood cells' ability to find tissue damage.

Ankyrin

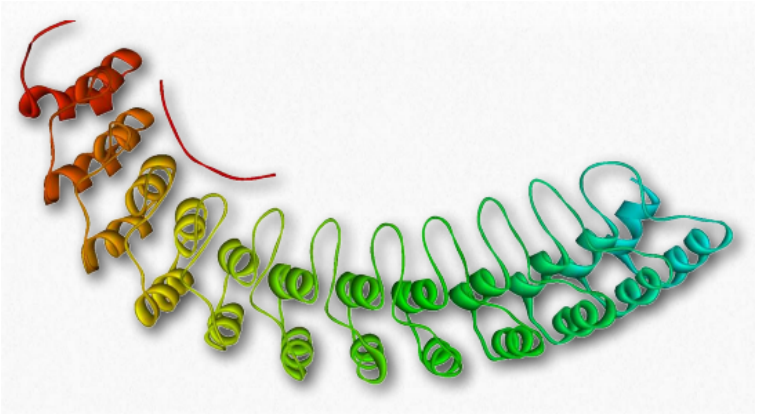


Figure 2.74 – Ankyrin's membrane-binding domain

Ankyrins (Figure 2.74) are a family of membrane adaptor proteins serving as "anchors" to interconnect integral membrane proteins to the spectrin-actin membrane cytoskeleton. Ankyrins are anchored to the plasma membrane by covalently linked palmitoyl-CoA. They bind to the β subunit of spectrin and at least a dozen groups of integral membrane proteins. The ankyrin proteins contain four functional domains: an N-terminal region with 24 tandem ankyrin repeats, a central spectrin-binding domain, a "death domain" interacting with apoptotic proteins, and a C-terminal regulatory domain that is highly varied significantly among different ankyrins.

Spectrin

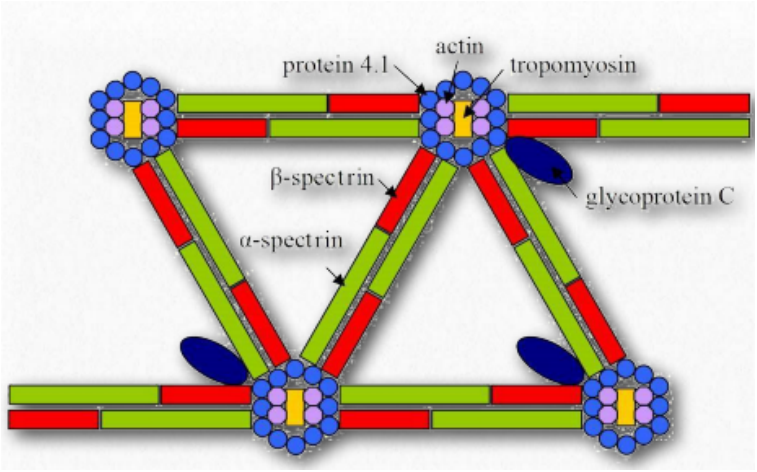


Figure 2.75 – Spectrin and other proteins in the cytoskeleton Wikipedia

Spectrin (Figures 2.75 & 2.76) is a protein of the cellular cytoskeleton that plays an important role in maintaining its structure and the integrity of the plasma membrane. In animals, spectrin gives red blood cells their shape. Spectrin is located inside the inner layer of the eukaryotic plasma membrane where it forms a network of pentagonal or hexagonal arrangements.

Spectrin fibers collect together at junctional complexes of actin and is also attached to ankyrin for stability, as well as numerous integral membrane proteins, such as glycophorin.

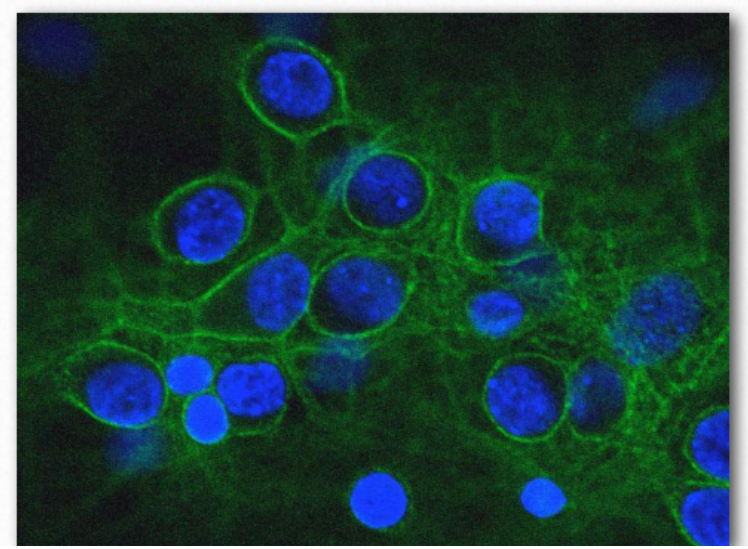


Figure 2.76 – Spectrin (green) and nuclei (Blue) Wikipedia

Integrins

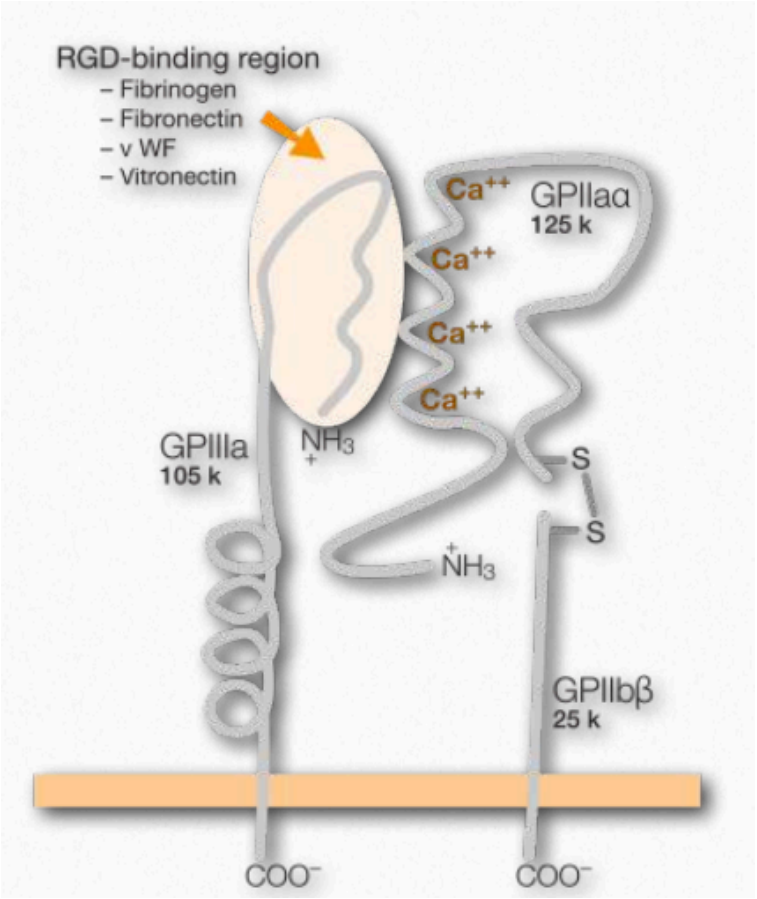


Figure 2.77 – Integrin and its binding site (on top left) Wikipedia

In multicellular organisms, cells need connections, both to each other and to the extracellular matrix. Facilitating these attachments at the cellular end are transmembrane proteins known as integrins (Figure 2.77). Integrins are found in all metazoan cells. Ligands for the integrins include collagen, fibronectin, laminin, and vitronectin. Integrins function not only in attachment, but also in communication, cell migration, virus linkages (adenovirus, for example), and blood clotting. Integrins are able to sense chemical and mechanical signals about the extracellular matrix and move that information to intracellular domains as part of the process of signal

transduction. Inside the cells, responses to the signals affect cell shape, regulation of the cell cycle, movement, or changes in other cell receptors in the membrane. The process is dynamic and allows for rapid responses as may be necessary, for example in the process of blood clotting, where the integrin known as GPIbIIIa (on the surface of blood platelets) attaches to fibrin in a clot as it develops.

Integrins work along with other receptors, including immunoglobulins, other cell adhesion molecules, cadherins, selectins, and syndecans. In mammals the proteins have a large number of subunits – 18 α- and 8 β-chains. They are a bridge between its links outside the cell to the extracellular matrix (ECM) and its links inside the cell to the cytoskeleton. Integrins play central role in formation and stability of focal adhesions. These are large molecular complexes arising from clustering of integrin-ECM connections. In the process of cellular movement, integrins at the "front" of the cell (in the direction of the movement), make new attachments to substrate and release connections to substrate in the back of the cell. These latter integrins are then endocytosed and reused.

Integrins also help to modulate signal transduction through tyrosine kinase receptors in the cell membrane by regulating movement of adapters to the plasma membrane. β1c integrin, for example, recruits the Shp2 phosphatase to the insulin growth factor receptor to cause it to become dephosphorylated, thus turning off the signal it communicates. Integrins can also help to recruit signaling molecules inside of the cell to activated tyrosine kinases to help them to communicate their signals.

Cadherins

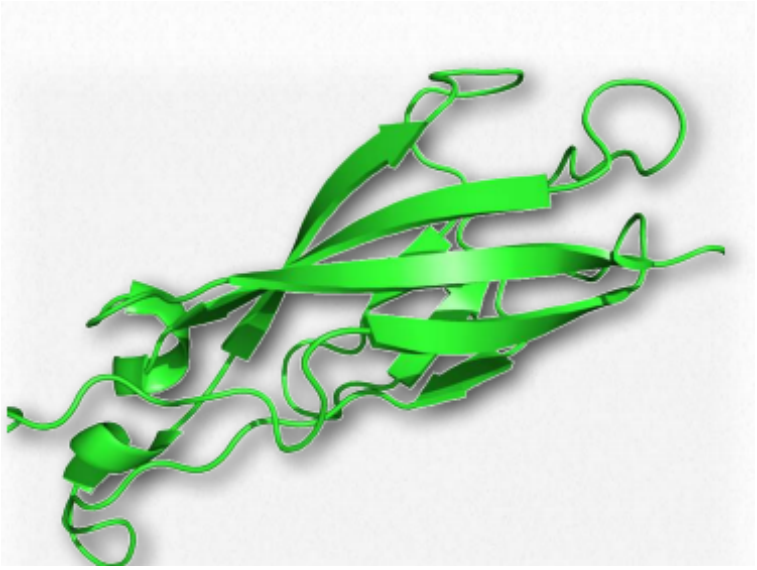


Figure 2.78 – Extracellular ectodomain of a cadherin

Cadherins (Figure 2.78) constitute a type-1 class of transmembrane proteins playing important roles in cell adhesion. They require calcium ions to function, forming adherens junctions that hold tissues together (See Figure 2.69). Cells of a specific cadherin type will preferentially cluster with each other in preference to associating with cells containing a different cadherin type. Caderins are both receptors and places for ligands to attach. They assist in the proper positioning of cells in development, separation of different tissue layers, and cell migration.

Selectins

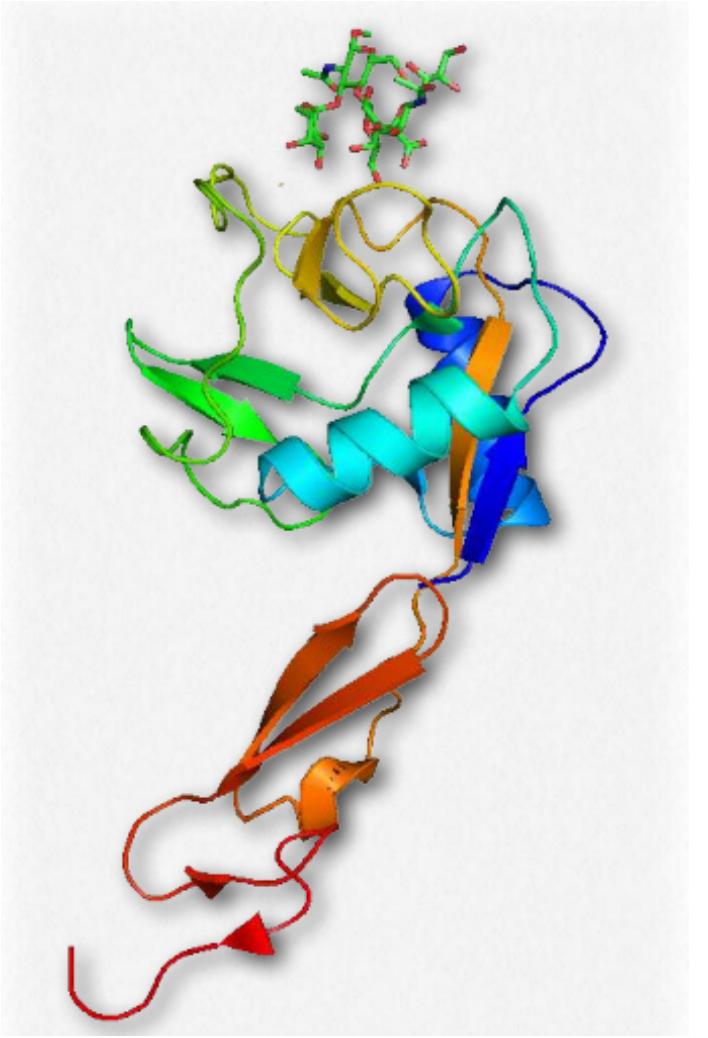


Figure 2.79 – Selectin bound to a sugar Wikipedia

Selectins (Figure 2.79) are cell adhesion glycoproteins that bind to sugar molecules. As such, they are a type of lectin – proteins that bind sugar polymers (see HERE also). All selectins have an N-terminal calcium-dependent lectin domain, a single transmembrane domain, and an intracellular cytoplasmic tail.

There are three different types of selectins, 1) E-selectin

(endothelial); 2) L (lymphocytic; and 3) P (platelets and endothelial cells. Selectins function in lymphocyte homing (adhesion of blood lymphocytes to cells in lymphoid organs), in inflammation processes, and in cancer metastasis. Near the site of inflammation, P-selectin on the surface of blood capillary cells interacts with glycoproteins on leukocyte cell surfaces. This has the effect of slowing the movement of the leukocyte. At the target site of inflammation, E- selectin on the endothelial cells of the blood vessel and L-selectin on the surface of the leukocyte bind to their respective carbohydrates, stopping the leukocyte movement. The leukocyte then crosses the wall of the capillary and begins the immune response. Selectins are involved in the inflammatory processes of asthma, psoriasis, multiple scleroris, and rheumatoid arthritis.

Laminins

Laminins are extracellular matrix glycoproteins that a major components of the basal lamina and affect cell differentiation, migration, and adhesion. They are secreted into the extracellular matrix where they are incorporated and are essential for tissue maintenance and survival. When laminins are defective, muscles may not form properly and give rise to muscular dystrophy.

Laminins are associated with fibronectin, entactin, and perlecan proteins in type IV collagen networks and bind to integrin receptors in the plasma membrane. As a consequence, laminins contribute to cellular attachment, differentiation, shape, and movement. The proteins are trimeric in structure, having one α-chain, a β-chain, and a γ-chain. Fifteen combinations of different chains are known.

Vitronectin

Vitronectin is a glycoprotein (75kDa) found in blood serum (platelets), the extracellular matrix, and in bone. It promotes the process of cell adhesion and spreading and binds to several protease inhibitors (serpins). It is secreted from cells and is believed to play roles in blood clotting and the malignancy of tumors. One domain of vitronectin binds to plasminogen activator inhibitor and acts to stabilize it. Another domain of the protein binds to cellular integrin

proteins, such as the vitronectin receptor that anchors cells to the extracellular matrix.

Catenins

Catenins are a family of proteins interacting with cadherin proteins in cell adhesion (Figure 2.69). Four main types of catenins are known, α-, β-, γ-, and δ-catenin. Catenins play roles in cellular organization before development occurs and help to regulate cellular growth. α-catenin and β-catenin are found at adherens junctions with cadherin and help cells to maintain epithelial layers. Cadherins are connected to actin filaments of the cytoskeleton and catenins play the critical role. Catenins are important for the process whereby cellular division is inhibited when cells come in contact with each other (contact inhibition).

When catenin genes are mutated, cadherin cell adhesions can disappear and tumorigenesis may result. Catenins have been found to be associated with colorectal and numerous other forms of cancer.

Glycophorins

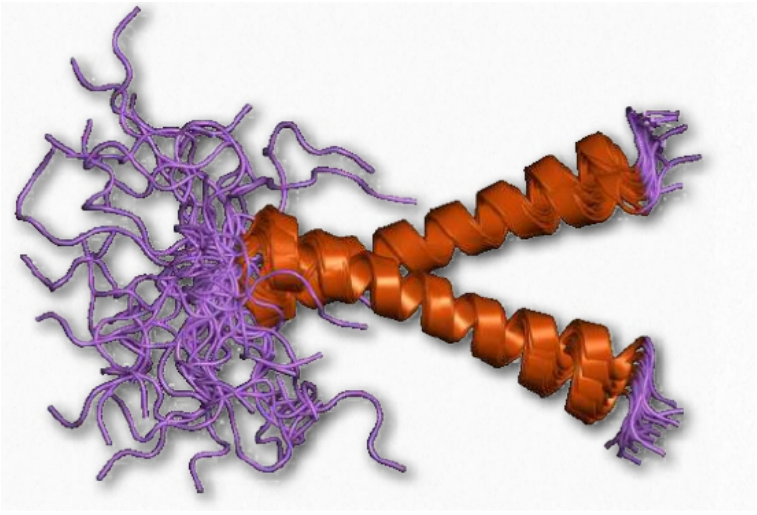


Figure 2.80 – Glycophorin a

All of the membrane proteins described so far are notable for the connections they make to other proteins and cellular structures.

Some membrane proteins, though, are designed to reduce cellular connections to proteins of other cells. This is particularly important for blood cells where “stickiness” is undesirable except where clotting is concerned.

Glycophorins (Figure 2.80) are membrane-spanning sialoglycoproteins of red blood cells. They are heavily glycosylated (60%).and rich in sialic acid, giving the cells a very hydrophilic (and negatively charged) coat, which enables them to circulate in the bloodstream without adhering to other cells or the vessel walls.

Five glycophorins have been identified – four (A,B,C,and D) from isolated membranes and a fifth form (E) from coding in the human genome. The proteins are abundant, forming about 2% of the total membrane proteins in these cells. Glycophorins have important roles in regulating RBC membrane mechanical properties and shape. Because some glycophorins can be expressed in various nonerythroid tissues (particularly Glycophorin C), the importance of their interactions with the membrane skeleton may have a considerable biological significance.

Cooperativity and allosterism – quaternary structure

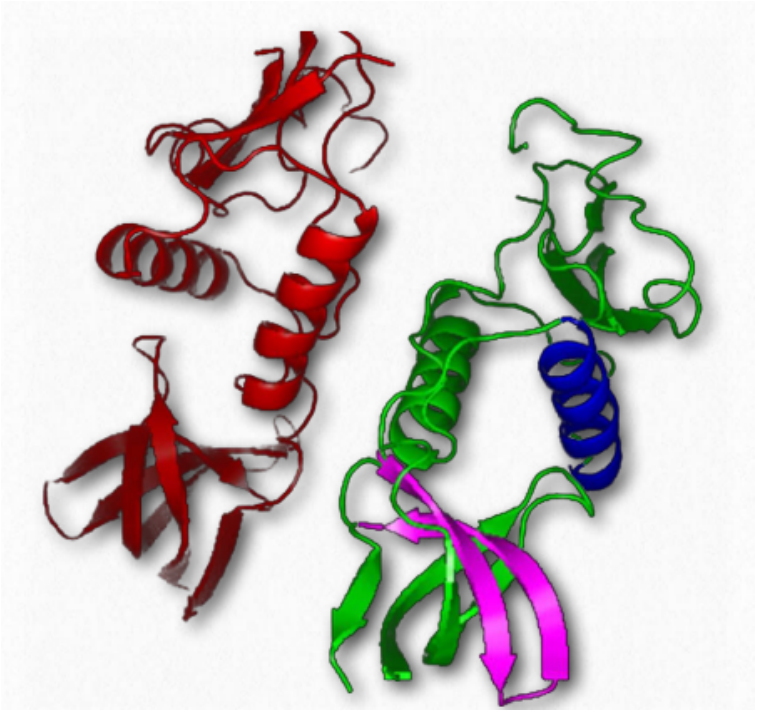


Figure 2.81 – Two polypeptide units of a protein interact in quaternary structure Wikipedia

Quaternary structure, of course describes the interactions of individual subunits of a multi-subunit protein (Figure 2.81). The result of these interactions can give rise to important biological phenomena, such as cooperative binding of substrates to a protein and allosteric effects on the action of an enzyme.

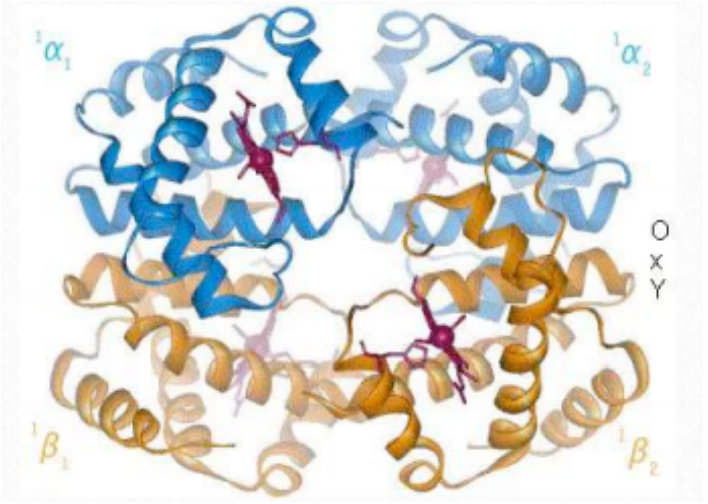
Allosteric effects can occur by a series of mechanisms, but a common feature is that binding of an effector to an enzyme subunit causes (or locks) the enzyme in either a Tstate (less activity) or an R-state (more activity). Effectors can be enzyme substrates (homotropic effectors) or non-substrates (heterotropic effectors). Allosterism will be covered in more depth in the Catalysis chapter HERE.

We begin our consideration of quaternary structure with a discussion of cooperativity, how it arises in the multi-subunit

protein hemoglobin and how its properties contrast with those of the related, single subunit protein myoglobin.

Cooperativity

Cooperativity is defined as the phenomenon where binding of one ligand molecule by a protein favors the binding of additional molecules of the same type. Hemoglobin, for example, exhibits cooperativity when the binding of an oxygen molecule by the iron of the heme group in one of the four subunits causes a slight conformation change in the subunit. This happens because the heme iron is attached to a histidine side chain and binding of oxygen 'lifts' the iron along with the histidine ring (also known as the imidazole ring).



Movie 2.3 – Hemoglobin's structural changes on binding oxygen Wikipedia

Since each hemoglobin subunit interacts with and influences the other subunits, they too are induced to change shape slightly when the first subunit binds to oxygen (a transition described as going from the T-state to the R-state). These shape changes favor each of the remaining subunits binding oxygen, as well. This is very important in the lungs where oxygen is picked up by hemoglobin, because the binding of the first oxygen molecule facilitates the

rapid uptake of more oxygen molecules. In the tissues, where the oxygen concentration is lower, the oxygen leaves hemoglobin and the proteins flips from the R-state back to the Tstate.

CO_2 transport

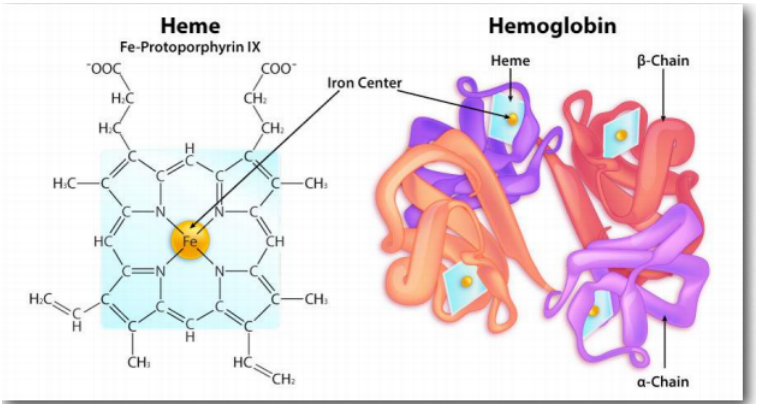


Figure 2.82 – Heme structure within hemoglobin Image by Aleia Kim

Cooperativity is only one of many fascinating structural aspects of hemoglobin that help the body to receive oxygen where it is needed and pick it up where it is abundant. Hemoglobin also assists in the transport of the product of cellular respiration (carbon dioxide) from the tissues producing it to the lungs where it is exhaled. Like the binding of oxygen to hemoglobin, binding of other molecules to hemoglobin affects its affinity for oxygen. The effect is particularly pronounced when comparing the oxygen binding characteristics of hemoglobin's four subunits with the oxygen binding of the related protein myoglobin's single subunit (Figure 2.83).

Different oxygen binding

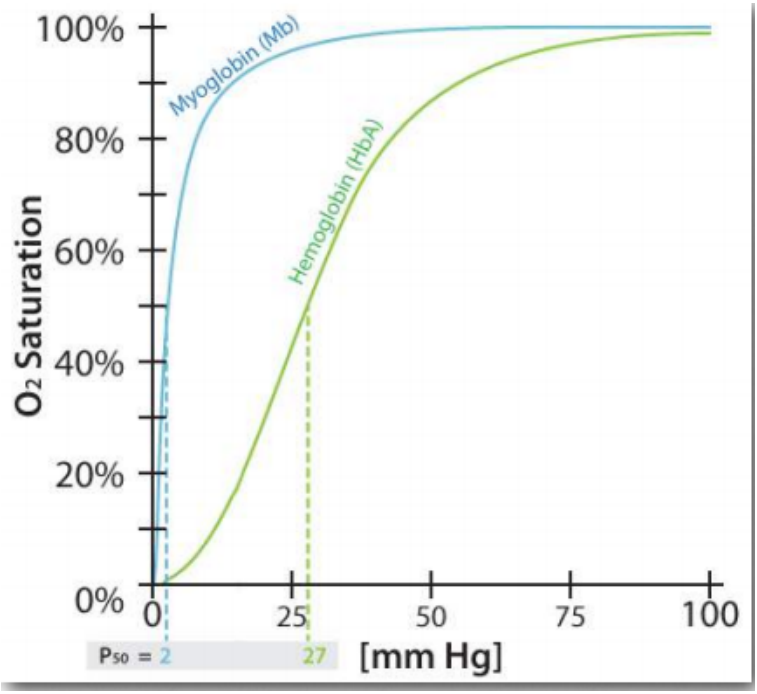


Figure 2.83 – Oxygen binding affinities for myoglobin and hemoglobin Figure by Aleia Kim

Like hemoglobin, myoglobin contains an iron in a heme group that binds to oxygen. The structure of the globin protein in myoglobin is very similar to the structure of the globins in hemoglobin and hemoglobin is thought to have evolved from myoglobin in evolutionary history. As seen in Figure 2.83, the binding curve of hemoglobin for oxygen is S-shaped (sigmoidal), whereas the binding curve for myoglobin is hyperbolic. What this tells us is that hemoglobin's affinity for oxygen is low at a low concentration oxygen, but increases as the oxygen concentration increases. Since myoglobin very quickly saturates with oxygen, even under low oxygen concentrations, it says that its affinity for oxygen is high and doesn't change.

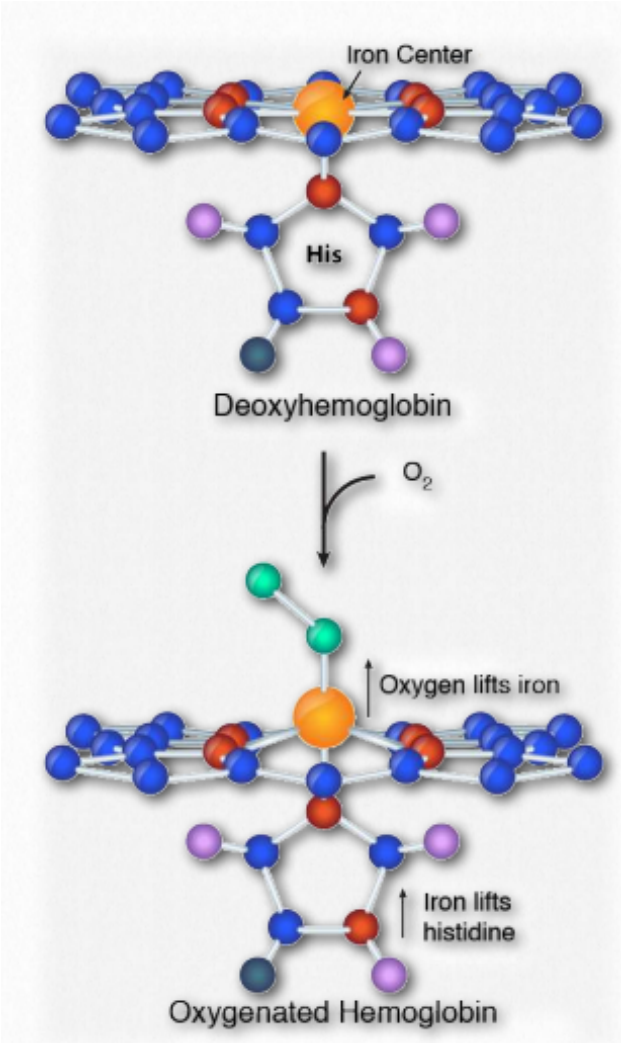
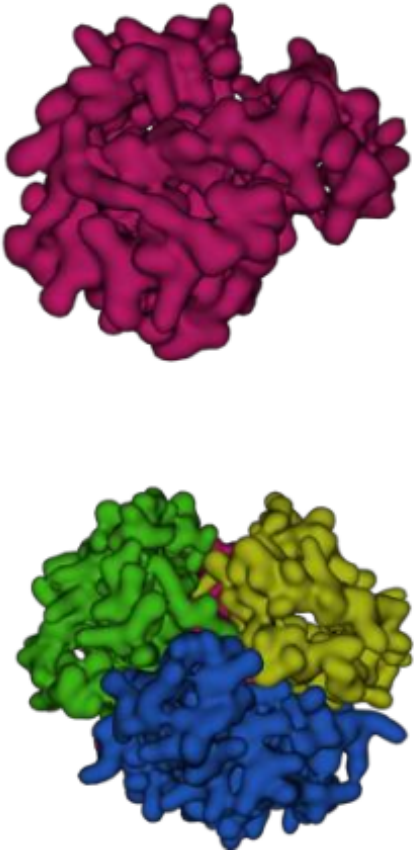


Figure 2.84 – Binding of oxygen at the heme center of hemoglobin
Image by Aleia Kim

Because myoglobin has only a single subunit, binding of oxygen by that subunit can't affect any other subunits, since there are no other subunits to affect. Consequently, cooperativity requires more than one subunit. Therefore, hemoglobin can exhibit cooperativity, but myoglobin can't. It is worth noting that simply having multiple

subunits does not mean cooperativity will exist. Hemoglobin is one protein that exhibits the characteristic, but many multisubunit proteins do not.



Interactive 2.2 – Hemoglobin in the presence (top) and absence (bottom) of oxygen

Storage vs. delivery

The lack of ability of myoglobin to adjust its affinity for oxygen according to the oxygen concentration (low affinity at low oxygen concentration, such as in tissues and high affinity at high oxygen concentration, such as in the lungs) means it is better suited for storing oxygen than for delivering it according to the varying

oxygen needs of and animal body. As we shall see, besides cooperativity, hemoglobin has other structural features that allow it to deliver oxygen precisely where it is needed most in the body.

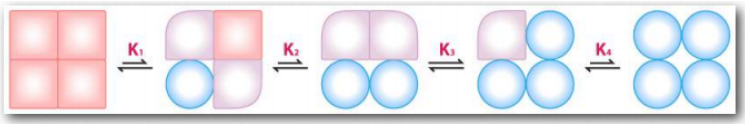


Figure 2.85 – Sequential model of binding. The sequential model is one way to explain hemoglobin's cooperativity. Squares represent no oxygen bound. Circles represent subunits bound with oxygen and rounded subunits correspond to units whose affinity for oxygen increases by interacting with a subunit that has bound oxygen. Image by Aleia Kim 131

Bohr effect

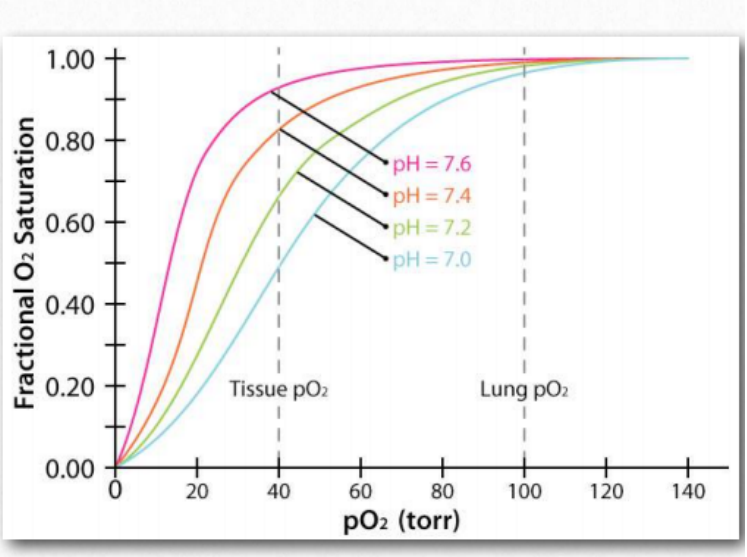


Figure 2.86 – The Bohr effect with respect to pH changes Image by Aleia Kim

The Bohr Effect was first described over 100 years ago by Christian Bohr, father of the famous physicist, Niels Bohr. Shown graphically (Figures 2.86, 2.87, and 2.88), the observed effect is that hemoglobin's affinity for oxygen decreases as the pH decreases

and as the concentration of carbon dioxide increases. Binding of the protons and carbon dioxide by amino Figure 2.85 – Sequential model of binding. The sequential model is one way to explain hemoglobin's cooperativity. Squares represent no oxygen bound. Circles represent subunits bound with oxygen and rounded subunits correspond to units whose affinity for oxygen increases by interacting with a subunit that has bound oxygen. Image by Aleia Kim acid side chains in the globin proteins helps to facilitate structural changes in them. Most commonly, the amino acid affected by protons is histidine #146 of the β strands. When this happens, the ionized histidine can form an ionic bond with the side chain of aspartic acid #94, which has the effect of stabilizing the T-state (reduced oxygen binding state) and releasing oxygen. Other histidines and the amine of the amino terminal amino acids in the α-chains are also binding sites for protons.

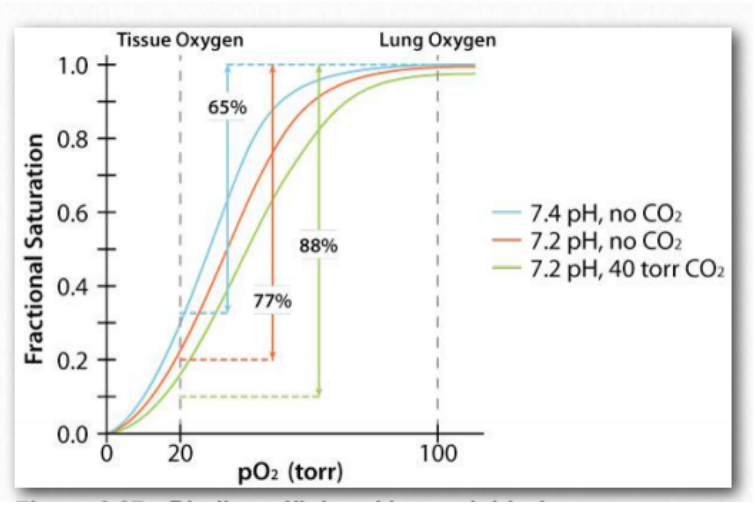


Figure 2.87 – Binding affinity of hemoglobin for oxygen under different conditions Image by Aleia Kim

2,3-BPG

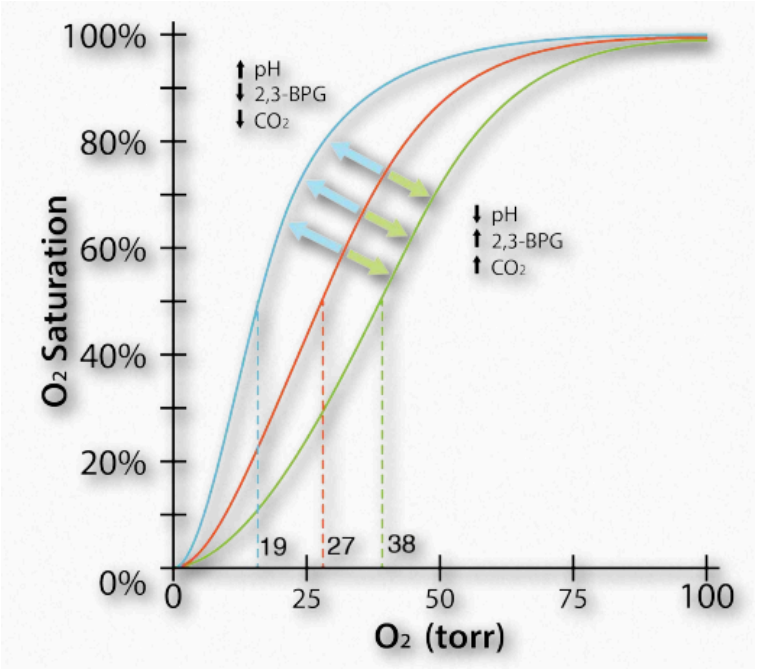


Figure 2.88 – The Bohr effect physiologically – oxygen binding curves for resting muscle (blue), active muscle (green) and reference muscle (orange) with respect to pH, 2,3-BPG, and CO2 Image by Aleia Kim

Another molecule favoring the release of oxygen by hemoglobin is 2,3- bisphosphoglycerate (also called 2,3-BPG or just BPG – Figure 2.89). Like protons and carbon dioxide, 2,3-BPG is produced by actively respiring tissues, as a byproduct of glucose metabolism. The 2,3-BPG mole cule fits into the 'hole of the donut' of adult hemoglobin (Figure 2.89). Such binding of 2,3-BPG favors the T-state (tight – low oxygen binding) of hemoglobin, which has a reduced affinity for oxygen. In the absence of 2,3-BPG, hemoglobin can more easily exist in the R-state (relaxed – higher oxygen binding), which has a high affinity for oxygen.

Smokers

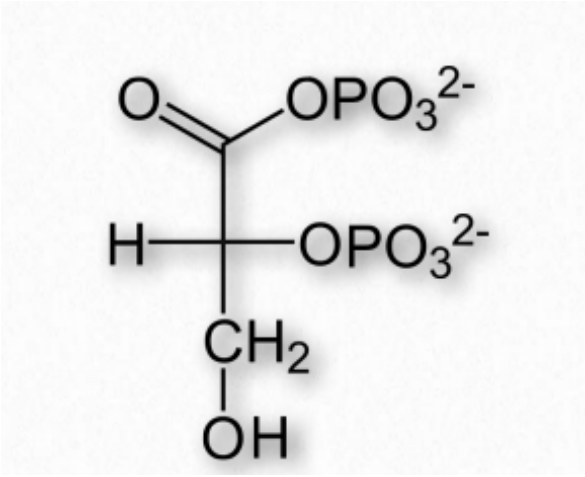


Figure 2.89 – The structure of 2,3 bisphosphoglycerate (2,3-BPG)

Notably, the blood of smokers is higher in the concentration of 2,3-BPG than non-smokers, so more of their hemoglobin remains in the T-state and thus the oxygen carrying capacity of smokers is lower than non-smokers.Another reason why smokers' oxygen carrying capacity is lower than that of non-smokers is that cigarette smoke contains carbon monoxide and this molecule, which has almost identical dimensions to molecular oxygen, effectively outcompetes with oxygen for binding to the iron atom of heme (Figure 2.90). Part of carbon monoxide's toxicity is due to its ability to bind hemoglobin and prevent oxygen from binding.

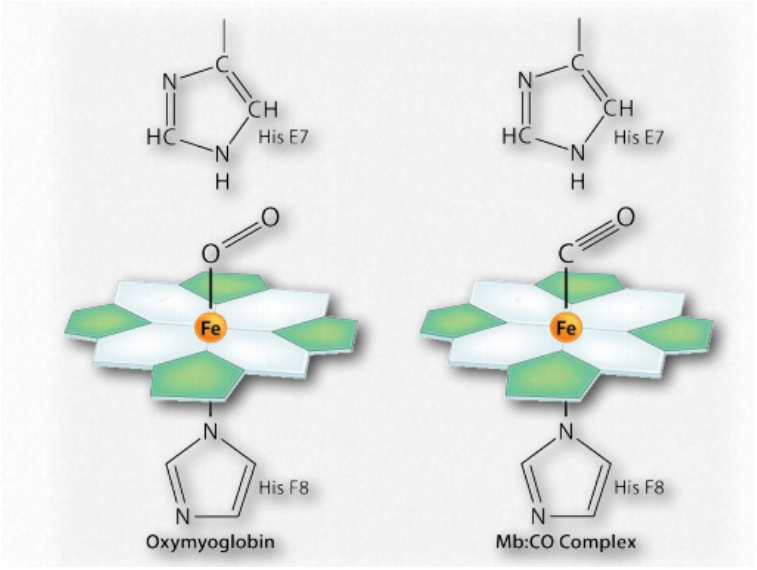


Figure 2.90 – Binding of oxygen (left) and carbon monoxide (right) by a heme group of hemoglobin Image by Aleia Kim

Carbon dioxide

Figure 2.91 – Hemoglobin's hole of the donut for binding 2,3-BPG Wikipedia

Carbon dioxide binds to form a carbamate when binding the α-amine of each globin chain. The process of forming this structure releases a proton, which helps to further enhance the Bohr effect. Physiologically, the binding of CO2 and H+ has significance because actively respiring tissues (such as contracting muscles) require oxygen and release protons and carbon dioxide. The higher the concentration of protons and carbon dioxide, the more oxygen is released to feed the tissues that need it most.

About 40% of the released protons and about 20% of the carbon dioxide are carried back to the lungs by hemoglobin. The remainder travel as part of the bicarbonate buffering system or as dissolved CO2. In the lungs, the process reverses itself. The lungs have a higher pH than respiring tissues, so protons are released from hemoglobin and CO2 too is freed to be exhaled.

Fetal hemoglobin

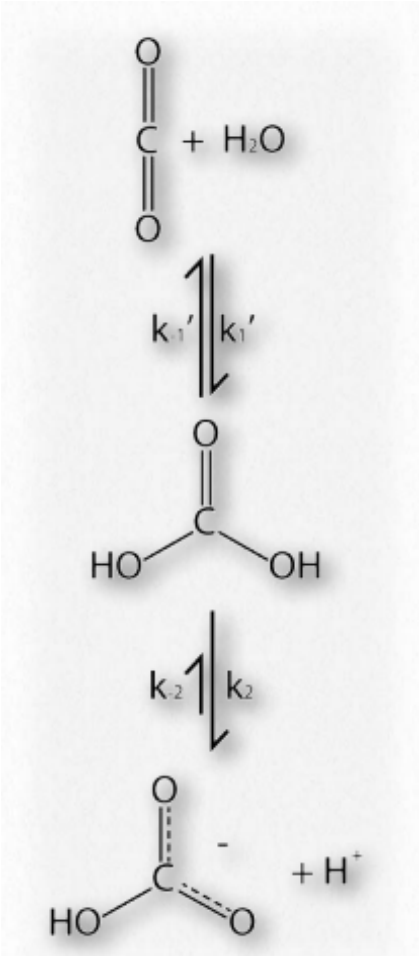


Figure 2.92 – Formation of bicarbonate from CO2 in blood

Adult hemoglobin releases oxygen when it binds 2,3- BPG. This is in contrast to fetal hemoglobin, which has a slightly different configuration (α2γ2) than adult hemoglobin (α2β2). Fetal hemoglobin has a greater affinity for oxygen than maternal hemoglobin, allowing the fetus to obtain oxygen effectively from the mother's blood. Part of the reason for fetal hemoglobin's greater affinity for oxygen is that it doesn't bind 2,3-BPG. Consequently, fetal hemoglobin remains in the R-state much more than adult hemoglobin and

because of this, fetal hemoglobin has greater affinity for oxygen than adult hemoglobin and can take oxygen away from adult hemoglobin. Thus, the fetus can get oxygen from the mother.

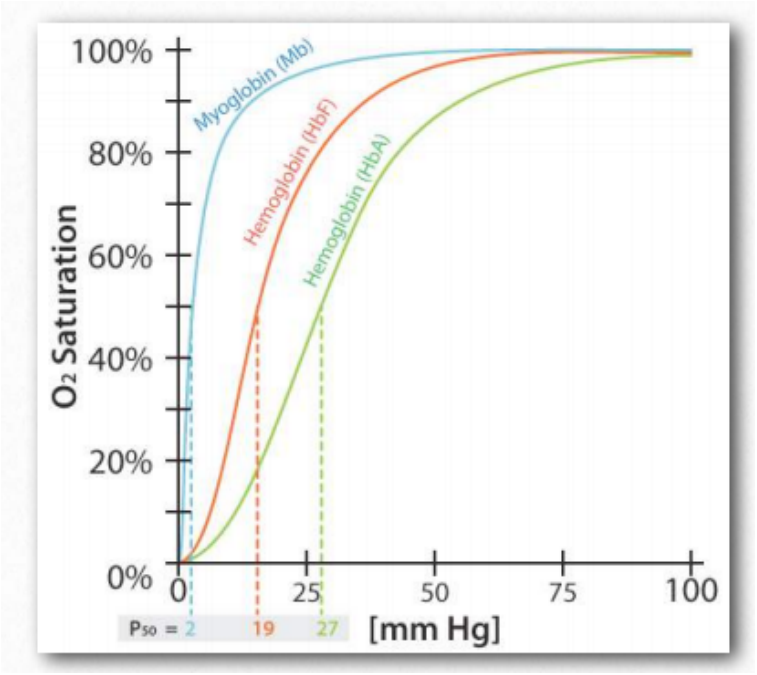


Figure 2.93 – Comparison of oxygen binding of myoglobin (blue), fetal hemoglobin (orange), and adult hemoglobin (green) Image by Aleia Kim

Sickle cell disease

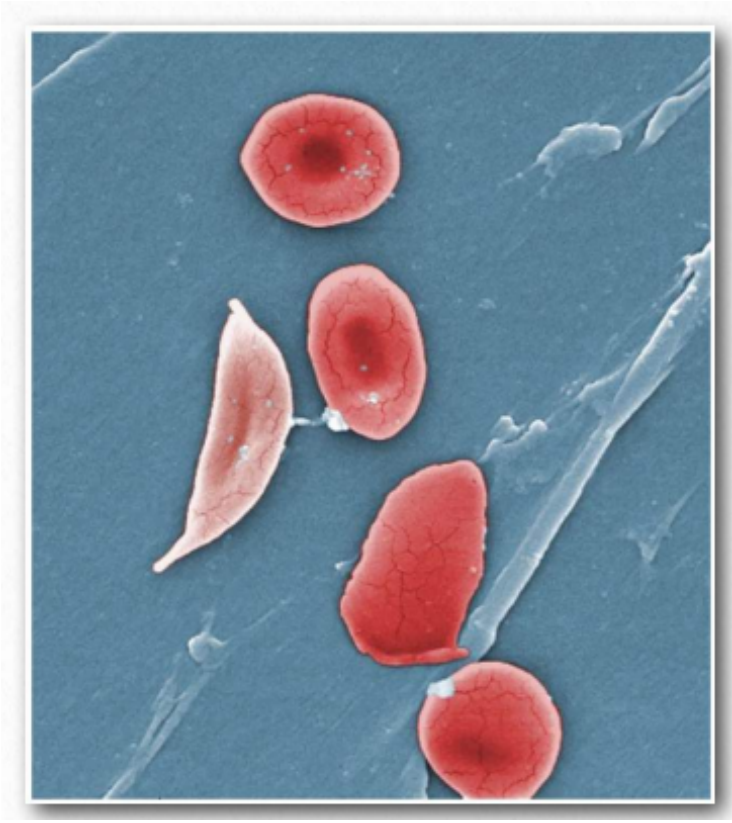


Figure 2.94 – Four normal red blood cells (right) and one sickled red blood cell (left) Wikipedia

Mutations to the globin genes coding for hemoglobin can sometimes have deleterious consequences. Sickle cell disease (also called sickle cell anemia) is a genetically transmitted disease that arises from such mutations. There are different forms of the disease. It is a recessive trait, meaning that to be afflicted with it, an individual must inherit two copies of the mutated gene.

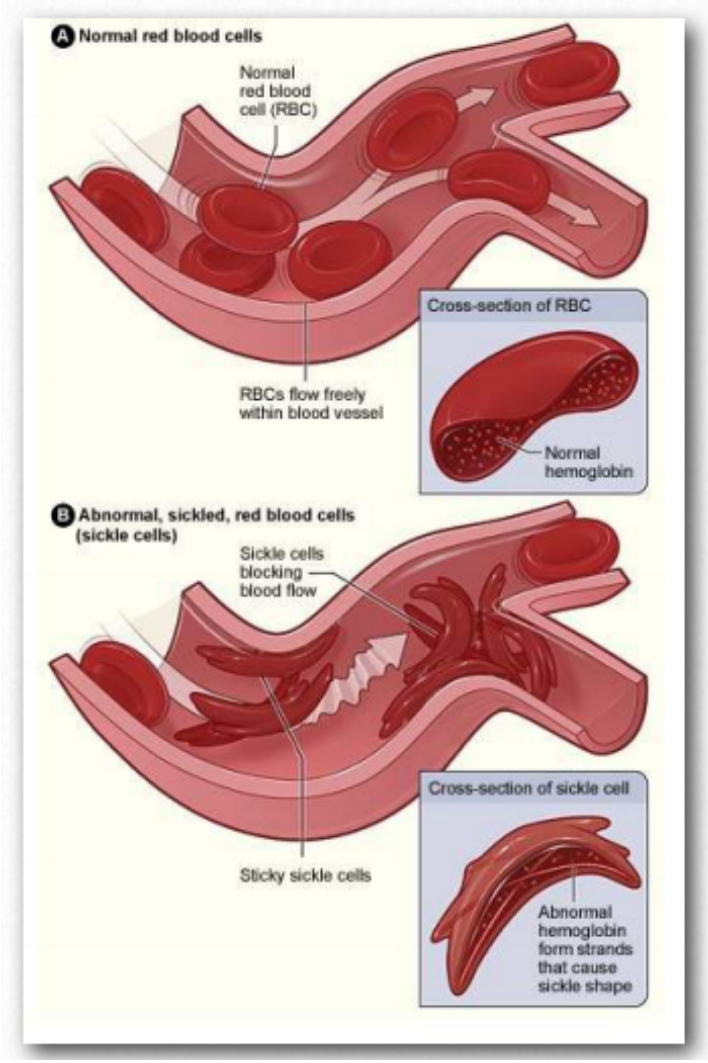


Figure 2.95 – Movement of blood in capillaries. Top – normal red blood cells. Bottom – sickled red blood cells

The predominant form of hemoglobin in adults is hemoglobin A, designated HbA (two α chains and two β chains). The mutant form is known as HbS. The most common mutation is an A to T mutation in the middle of the codon for the seventh amino acid (some counting

schemes call it the sixth amino acid) of the β-chain. This results in conversion of a GAG codon to GTG and thus changes the amino acid specified at that position from a glutamic acid to a valine. This minor change places a small hydrophobic patch of amino acids on the surface of the β-globin chains.

Polymerization

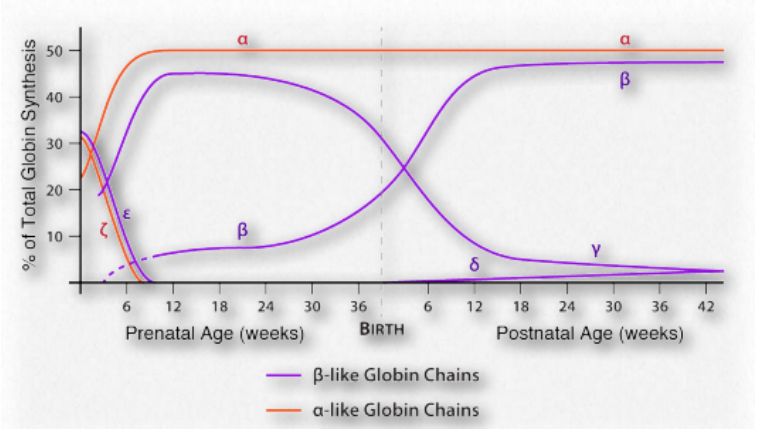


Figure 2.96 – Pattern of expression of six different globins of hemoglobin – α,β,γ,ε,δ, and ζ Image by Aleia Kim

Under conditions of low oxygen, these hydrophobic patches will associate with each other to make long polymers of hemoglobin molecules. The result is that the red blood cells containing them will change shape from being rounded to forming the shape of a sickle (Figure 2.94). Rounded red blood cells readily make it through tiny capillaries, but sickleshaped cells do not.

Worse, they block the flow of other blood cells. Tissues where these blockages occur are already low in oxygen, so stopping the flow of blood through them causes them to go quickly anaerobic, causing pain and, in some cases, death of tissue. In severe circumstances, sickled red blood cells death may result. The disease is referred to as an anemia because the sickling of the red blood cells targets them for removal by the blood monitoring system of the body, so a person with the disease has chronically reduced numbers of red blood cells.

Heterozygote advantage

Interestingly, there appears to be a selective advantage to people who are heterozygous for the disease in areas where malaria is prominent. Heterozygotes do not suffer obvious ill effects of the disease, but their red blood cells appear to be more susceptible to rupture when infected. As a consequence, the parasite gets less of a chance to reproduce and the infected person has a greater chance of survival.

The protective effect of the mutant gene, though, does not extend to people who suffer the full blown disease (homozygotes for the mutant gene). Treatments for the disease include transfusion, pain management, and avoidance of heavy exertion. The drug hydroxyurea has been linked to reduction in number and severity of attacks, as well as an increase in survival time1,2. It appears to work by reactivating expression of the fetal hemoglobin gene, which typically is not synthesized to any significant extent normally after about 6 weeks of age.

Oxygen binding

Animals have needs for oxygen that differ from all other organisms. Oxygen, of course, is the terminal electron acceptor in animals and is necessary for electron transport to work. When electron transport is functioning, ATP generation by cells is many times more efficient than when it is absent. Since abundant ATP is essential for muscular contraction and animals move around a lot – to catch prey, to exercise, to escape danger, etc., having an abundant supply of oxygen is important.

This is particularly a concern deep inside tissues where diffusion of oxygen alone (as occurs in insects) does not deliver sufficient quantities necessary for long term survival. The issue is not a problem for plants since, for the most part, their motions are largely related to growth and thus don't have rapidly changing needs/ demands for oxygen that animals have. Unicellular organisms have a variety of mechanisms for obtaining oxygen and surviving without it. Two other important oxygen binding proteins besides hemoglobin are myoglobin and hemocyanin.

Myoglobin

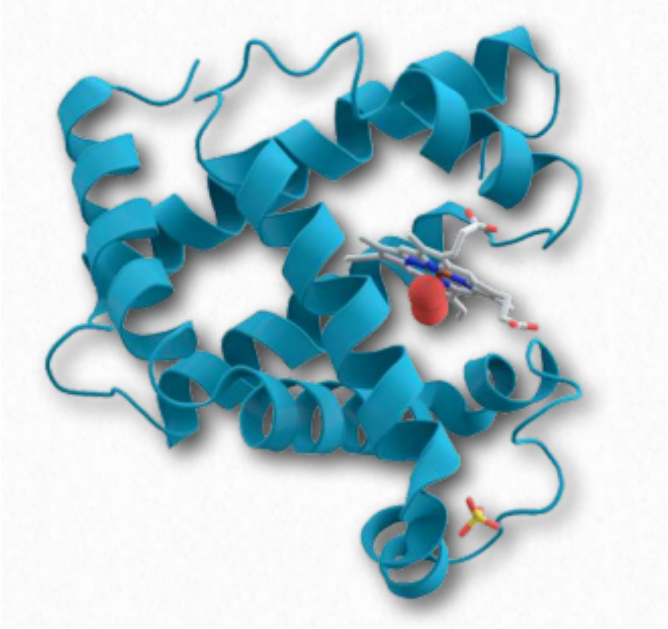


Figure 2.97 – Myoglobin bound to oxygen

Myoglobin is the primary oxygen-storage protein found in animal muscle tissues. In contrast to hemoglobin, which circulates throughout the body, myoglobin protein is only found in muscle tissue and appears in the blood only after injury. Like hemoglobin, myoglobin binds oxygen at a prosthetic heme group it contains.

The red color of meat arises from the heme of myoglobin and the browning of meat by cooking it comes from oxidation of the ferrous (Fe++) ion of myoglobin's heme to the ferric (Fe+++) ion via oxidation in the cooking process. As meat sits in our atmosphere (an oxygen-rich environment), oxidation of Fe++ to Fe+++ occurs, leaving the brown color noted above. If meat is stored in a carbon monoxide (CO) environment, CO binds to the heme group and reduces the amount of oxidation, keeping meat looking red for a longer period of time.

High affinity

Myoglobin (Figure 2.97) displays higher affinity for oxygen at low

oxygen concentrations than hemoglobin and is therefore able to absorb oxygen delivered by hemoglobin under these conditions. Myoglobin's high affinity for oxygen makes it better suited for oxygen storage than delivery. The protein exists as a single subunit of globin (in contrast to hemoglobin, which contains four subunits) and is related to the subunits found in hemoglobin. Mammals that dive deeply in the ocean, such as whales and seals, have muscles with particularly high abundance of myoglobin. When oxygen concentration in muscles falls to low levels, myoglobin releases its oxygen, thus functioning as an oxygen "battery" that delivers oxygen fuel when needed and holding onto it under all other conditions. Myoglobin holds the distinction of being the first protein for which the 3D structure was determined by X-ray crystallography by John Kendrew in 1958, an achievement for which he later won the Nobel Prize.

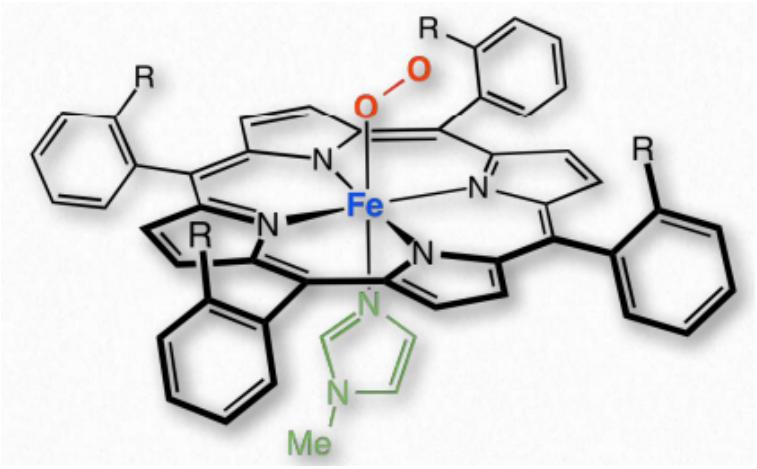


Figure 2.98 – Oxygen bound at heme of myoglobin Wikipedia

Hemocyanin

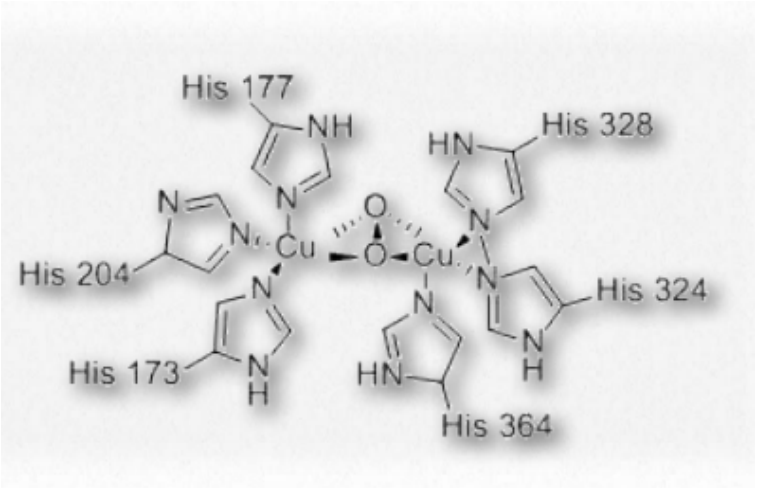


Figure 2.99 – Oxygen binding in hemocyanin Wikipedia

Hemocyanin is the protein transporting oxygen in the bodies of molluscs and arthropods. It is a coppercontaining protein found not within blood cells of these organisms, but rather is suspended in the circulating hemolymph they possess. The oxygen binding site of hemocyanin contains a pair of copper(I) cations directly coordinated to the protein by the imidazole rings of six histidine side chains.

Figure 2.100 – Hemocyanin (purple) in a red rock crab Wikipedia

Most, but not all hemocyanins bind oxygen non-cooperatively and are less efficient than hemoglobin at transporting oxygen. Notably, the hemocyanins of horseshoe crabs and some other arthropods do, in fact, bind oxygen cooperatively. Hemocyanin contains many subunit proteins, each with two copper atoms that can bind one oxygen molecule (O2). Subunit proteins have atomic masses of about 75 kilodaltons (kDa). These may be arranged in dimers or hexamers depending on species. Superstructures comprised of dimer or hexamer complexes are arranged in chains or clusters and have molecular weights of over 1500 kDa.

3.1: Principles of Catalysis

If there is a magical component to life, an argument can surely be made for it being catalysis. Thanks to catalysis, reactions that can take hundreds of years to complete in the uncatalyzed "real world," occur in seconds in the presence of a catalyst. Chemical catalysts, such as platinum, can speed reactions, but enzymes put chemical catalysts to shame (Figure 4.1).

Enzyme	Nonenzymatic Half-Life	Uncatalyzed Rate (k_{un} s^{-1})	Catalyzed Rate (k_{cat} s^{-1})	Rate Enhancement (k_{cat} s^{-1}/k_{un} s^{-1})
OMP decarboxylase	78,000,000 years	28×10^{-16}	39	1.4×10^{17}
Staphylococcal nuclease	130,000 years	1.7×10^{-13}	95	5.6×10^{14}
Carboxypeptidase A	7.3 years	3.0×10^{-9}	578	1.9×10^{11}
Ketosteroid isomerase	7 weeks	1.7×10^{-7}	66,000	3.9×10^{11}
Triose phosphate isomerase	1.9 days	4.3×10^{-6}	4,300	1.0×10^{9}
Chorismate mutase	7.4 hours	2.6×10^{-5}	50	1.9×10^{6}
Carbonic anhydrase	5 seconds	1.3×10^{-1}	1×10^{6}	7.7×10^{6}

Abbreviations:
OMP - Orotine monophosphate
AMP - Adenosine monophosphate

Figure 4.1 – Rate enhancement for several enzymes Image by Aleia Kim

How can enzymes accelerate reactions to this degree? Chemical reactions follow the universal trend of moving towards lower energy, but they often have an energy barrier in place that must be overcome. The secret to catalytic action is reducing the magnitude of that barrier. When the barrier is lowered, the reaction occurs more quickly.

To develop a fuller understanding of this phenomenon, it is worth taking a moment to discuss the concept of *equilibrium*.

When a biochemical reaction is at equilibrium the concentrations of reactants and products do not change over time. This does not mean that the reactions have stopped, but rather that the forward reaction and a reverse reactions have reached a point where the forward reaction occurs at the same rate as the reverse. : If you had 8 molecules of A, and 4 of B at the beginning, and 2 molecules of A were converted to B, while 2 molecules of B were simultaneously

converted back to A, the number of molecules of A and B would remain unchanged, i.e., the reaction is at equilibrium.

Note that the "equi" part of this word relates to equal, as one might expect, but it does not relate to absolute concentrations.

In Juggling it is possible to establish a state much like chemical equilibrium, where each of several jugglers has 2 pins but pins are always being tossed and caught. scragz, CC BY 2.0 <*https://creativecommons.org/licenses/by/2.0*>, *via Wikimedia Commons*

Biochemical reactions, when not at equilibrium, tend to move in the direction of equilibrium. So chemical reactions involving enzymes are often working in an environment where the concentrations of reactants are shifting, and the flow of reactants to products is influenced by other linked reactions in what is called a biochemical or metabolic pathway.

Activation energy

Figure 4.2 schematically depicts the energy changes that occur during the progression of a simple reaction. In order for the reaction to proceed, an activation energy must be overcome in order for the reaction to occur.

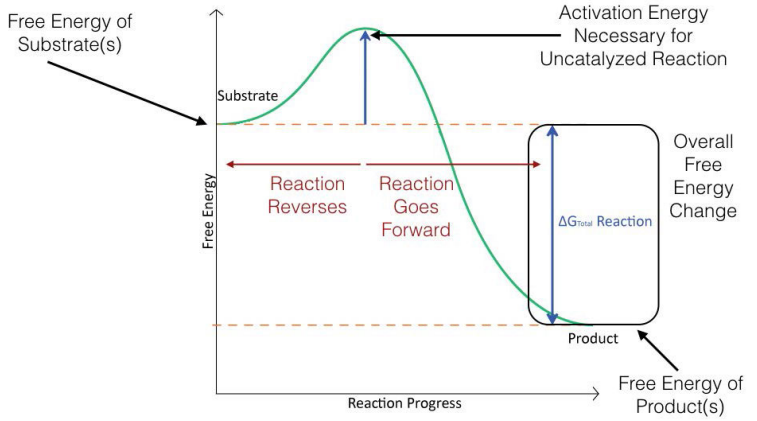


Figure 4.2 – Energy changes during the course of an uncatalyzed reaction. Image by Aleia Kim

In Figure 4.3, the activation energy for a catalyzed reaction is overlaid. As you can see, the reactants start at the same energy level (same place on the Y axis) for both catalyzed and uncatalyzed reactions, and the products end at the same energy for both catalyzed and uncatalyzed reactions as well. The catalyzed reaction, however, has a lower energy of activation (dotted line) than the uncatalyzed reaction.

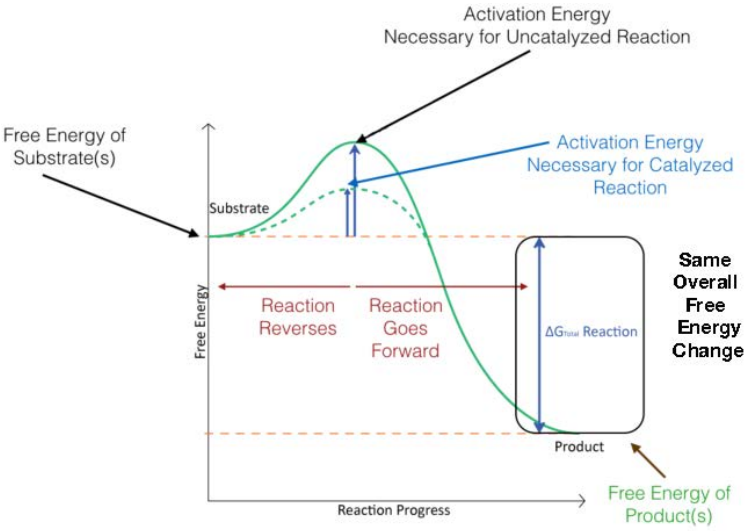


Figure 4.3 – Energy changes during the course of an uncatalyzed reaction (solid green line) and a catalyzed reaction (dotted green line). Image by Aleia Kim

Reversibility

The ratio of reactants to products at equilibrium is a function of the size of the energy difference between the product and reactant states, shown as the different placement on the Y axis on the graphs above. The lower the energy of the products compared to the reactants, the larger the percentage of molecules that will be products at equilibrium.

Since an enzyme lowers the activation energy for a reaction it can speed the reversal of a reaction just as it speeds a reaction in the forward direction. At equilibrium, however, no change in concentration of reactants and products occurs. Thus, enzymes

speed the time required to reach equilibrium, but do not affect the ratio of products to reactants at equilibrium.

The ratio of products to reactants can be thought of as a 'goal' of sorts for the reaction, though of course molecules and atoms are not living things and do not actually have desires. But it is a state that reactions tend toward. When there are alterations to a system at equilibrium, the reaction will proceed in such a way as to restore that equilibrium ratio.

In particular, in biochemical pathways multiple reactions are linked so that the product of one reaction is the reactant in another. As product is drawn away from the first reaction that system moves from its equilibrium state, and the reaction will proceed to convert additional reactant to product in order to work toward the equilibrium ratio.

Exceptions

A few reactions involve such a large energy release (downward drop on the graphs) as to make the reverse reactions nearly impossible. In these cases the equilibrium ratios are heavily skewed toward products.

They are related to the disappearance of a substrate or product of a reaction. Consider the first reaction below which is catalyzed by the enzyme carbonic anhydrase:

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons HCO_3^- + H^+$$

In the forward direction, carbonic acid is produced from water and carbon dioxide. It can either remain intact in the solution or ionize to produce bicarbonate ion and a proton. In the reverse direction, water and carbon dioxide are produced. Carbon dioxide, of course, is a gas and can leave the solution and escape.

When reaction molecules are removed, as they would be if carbon dioxide escaped, the reaction is pulled in the direction of the molecule being lost and reversal cannot occur unless the missing

molecule is replaced. In the second reaction occurring on the right, carbonic acid (H2CO3) is "removed" by ionization. This too would limit the reaction going back to carbon dioxide in water. This last type of "removal" is what occurs in metabolic pathways. In this case, the product of one reaction (carbonic acid) is the substrate for the next (formation of bicarbonate and a proton).

In the metabolic pathway of glycolysis, ten reactions are connected in this manner and reversing the process is much more complicated than if just one reaction was being considered.

General mechanisms of action

We have seen now that enzymatically catalyzed reactions are orders of magnitude faster than uncatalyzed and chemical-catalyzed reactions. And that process has been explained using concepts related to energy and equilibrium. But the magic of enzyme catalysis can also be considered from a different point of view: that is, what does an individual enzyme do in order to bring the activation energy barrier on a reaction down?

Changes

One of the remarkable things about catalysts is that they are not reactants in the normal sense: they are not consumed during a reaction. In other words, the catalyst ends up after a reaction just the way it started. Enzymes share this property, although in the midst of the catalytic action, an enzyme is temporarily changed. In fact, it is the ability of an enzyme to change that leads to its incredible efficiency.

These changes may be subtle electronic ones, more significant covalent (bonding) modifications, or structural changes arising from the flexibility inherent in enzymes. These kinds of adjustments are not possible for chemical catalysts such as the metal platinum.

Enzyme flexibility allows movement and movement facilitates alteration of electronic environments necessary for catalysis. Enzymes are, thus, much more efficient than rigid chemical catalysts as a result of their abilities to facilitate the changes needed to optimize the catalytic process.

Enzymes are large and complicated molecules, which is the structural basis behind this difference. That complexity also means that a huge variety of enzymes are possible, each with a carefully tuned ability to carry out particular reactions with great efficiency and with great selectivity, working only on certain very specific reactants to convert them to particular products.

Substrate binding and the Active Site

Enzymes have binding sites that not only selectively 'grab' a reactant (called a **substrate** when discussing enzymes), but also place it in a position to be electronically induced to react, either within itself or with another substrate.

The enzyme itself may play a role in the electron shifts or that process may occur as a result of substrates being oriented very close to each other.

The location on an enzyme where substrate binds (the **substrate-binding site**) is also usually the site of the reaction itself, called the **active site**.

Enzyme flexibility

As large and complex molecules, enzymes are somewhat flexible and responsive to their local environments. Slight changes in shape,

often arising from the binding of the substrate itself, help to optimally position substrates for reaction after they bind.

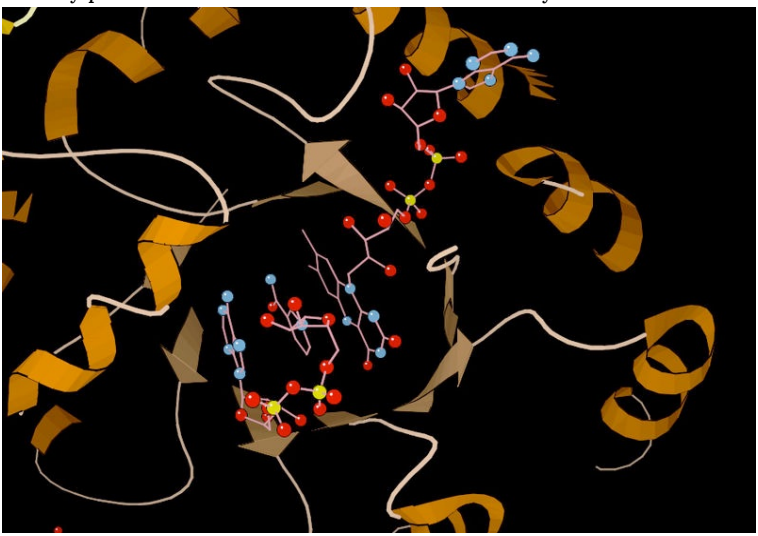


Figure 4.4 – Substrate binding by methylenetetrahydrofolate reductase

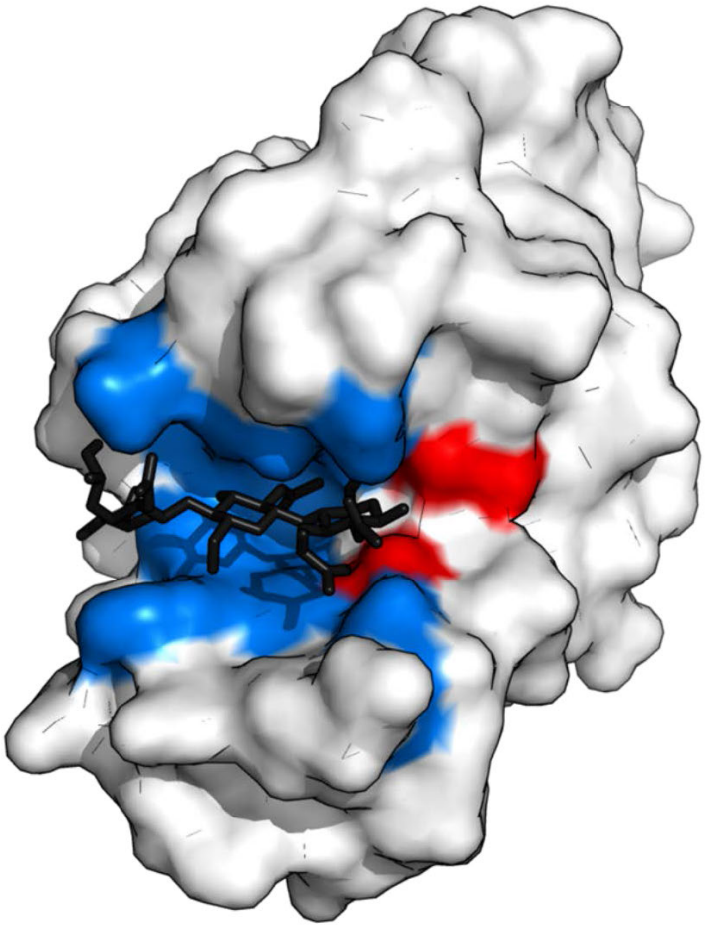


Figure 4.5 – Lysozyme with substrate binding site (blue), active site (red) and bound substrate (black). Wikipedia

Induced fit

These changes in shape are explained, in part, by the Induced Fit

Model of Catalysis (Figure 4.6), which illustrates that not only do enzymes change substrates, but that substrates also transiently change enzyme structure. At the end of the catalysis, the enzyme is returned to its original state. This model contrasts with the Lock and Key model, which presents enzymes with a fixed shape that is perfectly matched for binding its substrate(s). Evidence supports the Induced Fit model as a better, truer description of reality.

Enzyme flexibility also plays an important role in control of enzyme activity. Enzymes alternate between the T (tight) state, which is a lower activity state and the R (relaxed) state, which has greater activity.

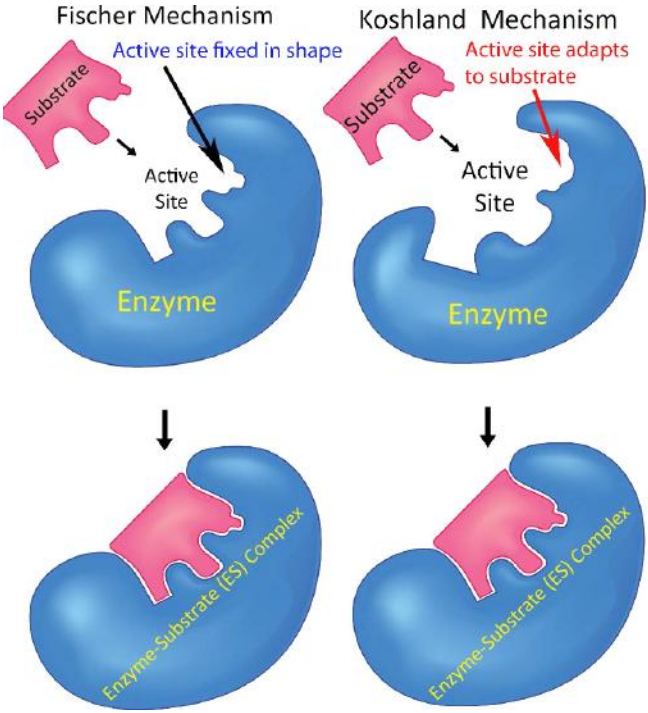


Figure 4.6 – Fischer's lock and key model (left) Vs. Koshland's induced fit model (right). Image by Aleia Kim

Ordered binding

Some enzymes need to bind multiple substrates, for instance if they are catalyzing a reaction where two substrates are covalently linked. For these systems, binding of the first substrate induces structural changes in the enzyme necessary for binding the second substrate. These mechanisms are aligned with the Induced Fit model, where enzymes are able to adjust to the initial binding and thus become more likely to bind the second substrate.

Consider lactate dehydrogenase, the enzyme which catalyzes the reaction below:

$$NADH + Pyruvate \rightarrow Lactate + NAD^{+}$$

This enzyme requires that NADH must bind prior to the binding of pyruvate. As noted earlier, this is consistent with an induced fit model of catalysis. In this case, binding of the NADH changes the enzyme shape/environment so that pyruvate can bind and without binding of NADH, the substrate cannot access the pyruvate binding site.

Reaction types

Enzyme-catalyzed reactions can be of several types, as shown in Figure 4.7. In one mechanism, called sequential reactions, at some point in the reaction, both substrates will be bound to the enzyme. There are, in turn, two different ways in which this can occur – random and ordered.

Figure 4.7 – Categories of enzymatic reactions

Types of Reactions	
Single Substrate – Single Product	$A \rightleftarrows B$
Single Substrate – Multiple Products	$A \rightleftarrows B + C$
Multiple Substrates – Single Products	$A + B \rightleftarrows C$
Multiple Substrates – Multiple Products	$A + B \rightleftarrows C + D$

Coenzymes

Organic molecules that assist enzymes and facilitate catalysis are **cofactors** called **coenzymes**. The term cofactor is a broad category usually subdivided into inorganic ions and coenzymes. If the coenzyme is very tightly or covalently bound to the enzyme, it is referred to as a **prosthetic group**. Enzymes without their cofactors are inactive and referred to as **apoenzymes**. Enzymes containing all of their cofactors are called **holoenzymes**.

Cofactor	Enzyme
Coenzyme	
5′-Deoxyadenosyl cobalamin	Methylmalonyl mutase
Biotin	Pyruvate carboxylase
Coenzyme A (CoA)	Acetyl CoA carboxylase
Flavin adenine nucleotide	Monoamine oxidase
Nicotinamide adenine dinucleotide	Lactate dehydrogenase
Pyridoxal phosphate	Glycogen phosphorylase
Tetrahydrofolate	Thymidylate synthase
Thiamine pyrophosphate	Pyruvate dehydrogenase
Metal	
K^+	Propionyl CoA carboxylase
Mg^{2+}	Restriction Endonucleases; Hexokinase
Mn	Superoxide dismutase
Mo	Nitrate reductase
Ni^{2+}	Urease
Se	Glutathione peroxidase
Zn^{2+}	Carbonic anhydrase; Carboxypeptidase

Figure 4.27 – Enzyme cofactors. Image by Aleia Kim

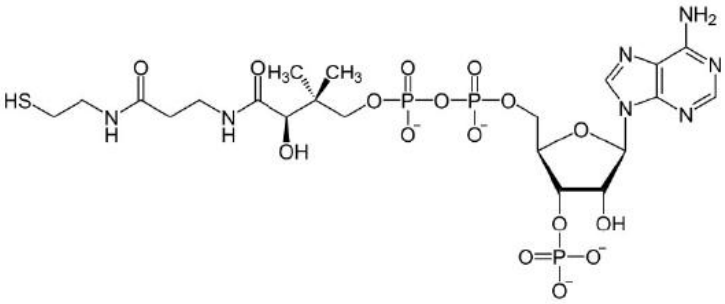


Figure 4.28 – Coenzyme A (CoA)

Some coenzymes are derivatives of vitamins. Vitamins are organic molecules necessary in the diet for normal biochemical function. The chemical structures of vitamins are frameworks that an organism is unable to construct on its own, in adequate quantity. Vitamins as we know them (Vitamin A, vitamin D, etc.) are frequently names that refer to small groups of closely-related chemical structures, any of which can serve the role in our bodies.

Vitamins may undergo chemical modification before playing a particular role in a body. For instance, niacin (Vitamin B3) can be

chemically converted to nicotinamide adenine dinucleotide (NAD+), a key molecule in energy metabolism.

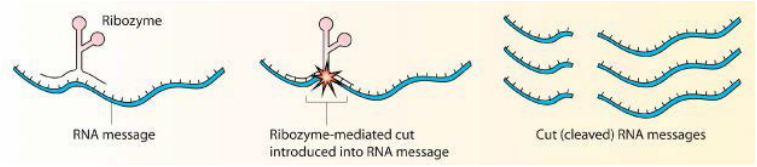
Vitamins are frequently incorporated into coenzymes such as this. However not all vitamins serve as coenzymes, and not all coenzymes come from vitamins. The term vitamin is derived from 'vital amine,' which is a reference to the need for intake in order to live ('vital') and their chemical composition as an amine-another reference which includes many but not all of the chemicals in this group.

Among the cofactors listed in the table Fig. 4.27 are many substances derived from vitamins: every substances named in the list of coenzymes contains a large fragment which comes from one of the B vitamins.

Ribozymes

Nearly all enzymes are proteins. However some RNA molecules are also capable of speeding reactions.

The most famous of these molecules was discovered by Tom Cech in the early 1980s. Cech was puzzled at his inability to find any proteins catalyzing a process his group was studying in a microorganism named Tetrahymena. Ultimately, the catalysis was recognized as coming from RNA itself. Since this discovery, many other examples of catalytic RNAs have been found. Catalytic RNA molecules are known as ribozymes.



359

Figure 4.31 – Cleavage of an RNA by a ribozyme. Wikipedia

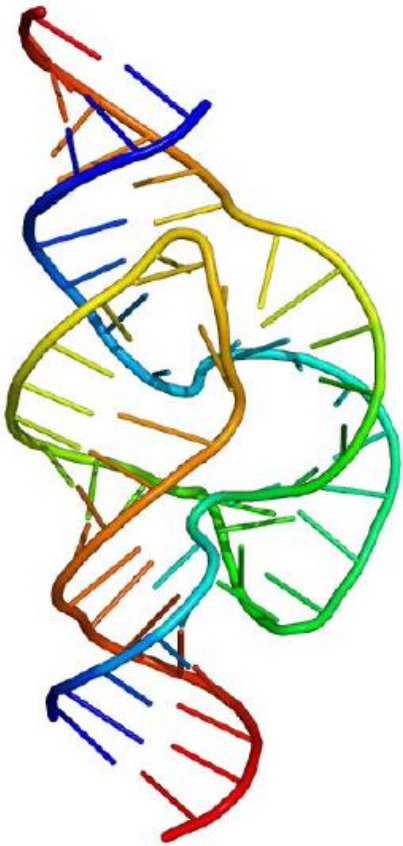


Figure 4.32 – Hammerhead ribozyme. Wikipedia

Not unusual

Ribozymes, however, are not rarities of nature. The protein-making ribosomes of cells are essentially giant ribozymes. The 23S rRNA of the prokaryotic ribosome and the 28S rRNA of the eukaryotic ribosome catalyze the formation of peptide bonds.

Ribozymes are also important in our understanding of the evolution of life on Earth. They have been shown to be capable via selection to evolve self-replication. Indeed, ribozymes actually answer a chicken/egg dilemma – which came first, enzymes that do the work of the cell or nucleic acids that carry the information required to produce the enzymes. As both carriers of genetic information and catalysts, ribozymes are likely both the chicken and the egg in the origin of life.

3.2: Control of Enzymatic Activity

Regulation of enzyme activity

Apart from their ability to greatly speed the rates of chemical reactions in cells, enzymes have another property that makes them valuable. This property is that their activity can be regulated, allowing them to be activated and inactivated, as necessary. This is tremendously important in maintaining homeostasis, permitting cells to respond in controlled ways to changes in both internal and external conditions.

Inhibition of specific enzymes by drugs can also be medically useful. Understanding the mechanisms that control enzyme activity is, therefore, of considerable importance.

Inhibition

We will first discuss four types of enzyme inhibition – competitive, non-competitive, uncompetitive, and suicide inhibition. Of these, the first three types are reversible. The last one, suicide inhibition, is not.

Competitive inhibition

Probably the easiest type of enzyme inhibition to understand is competitive inhibition. It is the one most commonly exploited pharmaceutically. Molecules that are competitive inhibitors of enzymes resemble one of the normal substrates of an enzyme. An example is methotrexate, which resembles the folate substrate of the enzyme named dihydrofolate reductase (DHFR). This enzyme normally catalyzes the chemical reduction of folate, an important reaction in the metabolism of nucleotides.

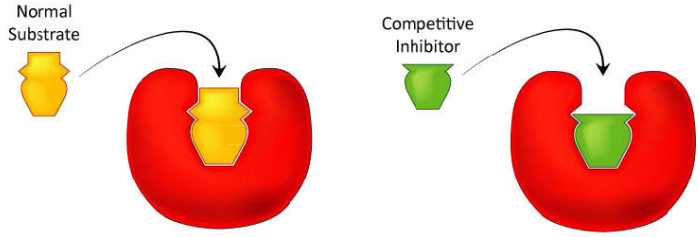


Figure 4.33 – Competitive inhibitors resemble the normal substrate and compete for binding at the active site. Image by Aleia Kim

Inhibitor binding

When the drug methotrexate is present, some of the DHFR enzyme binds to it, instead of to folate, and during the time methotrexate is bound, the enzyme is inactive and unable to bind folate. Thus, the enzyme is inhibited. Notably, the binding site on DHFR for methotrexate is the active site, the same place that folate would normally bind. As a result, methotrexate 'competes' with folate for binding to the enzyme. The more methotrexate there is, the more effectively it competes with folate for the enzyme's active site. Conversely, the more folate there is, the less of an effect methotrexate has on the enzyme because folate outcompetes it.

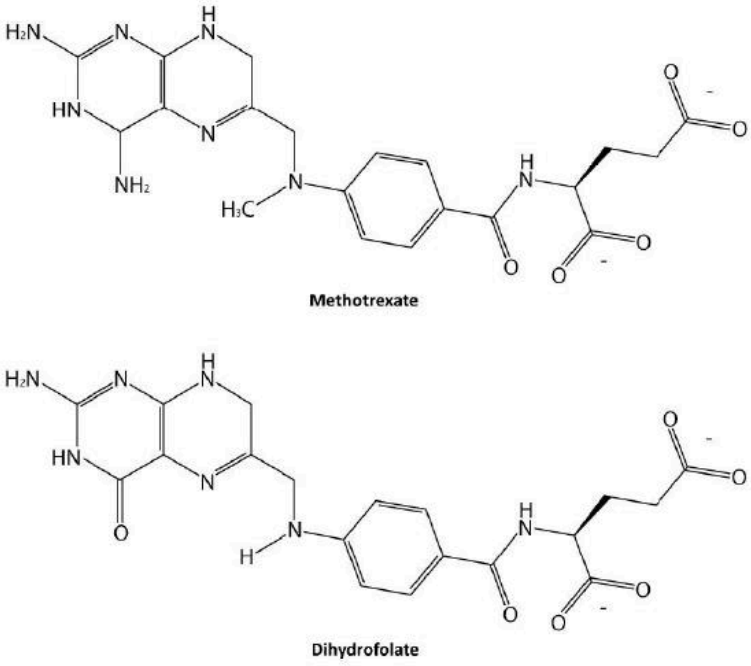


Figure 4.34 – Methotrexate and dihydrofolate. Image by Ben Carson

Non-competitive inhibition

A second type of inhibition employs inhibitors that do not resemble the substrate and bind not to the active site, but rather to a separate site on the enzyme (Figure 4.37). The effect of binding a non-competitive inhibitor is significantly different from binding a competitive inhibitor because there is no direct competition. In the case of competitive inhibition, the effect of the inhibitor can be reduced and eventually overwhelmed by flooding the enzyme with increasing amounts of substrate. With non-competitive inhibition, increasing the amount of substrate has no effect on the percentage

of enzyme that is able to act. Indeed, in non-competitive inhibition, the percentage of enzyme inhibited remains the same through all ranges of substrate concentrations.

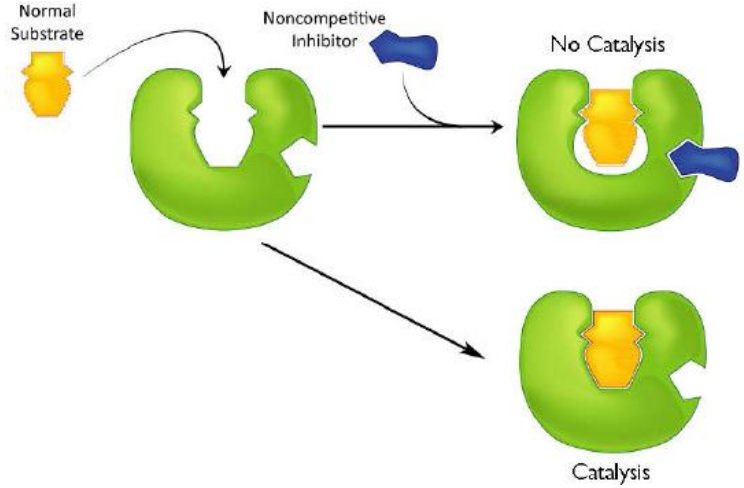


Figure 4.37 – Non-competitive inhibition – inhibitor does not resemble the substrate and binds to a site other than the active site. Image by Aleia Kim

Uncompetitive inhibition

A third type of enzymatic inhibition is that of uncompetitive inhibition. In this case the uncompetitive inhibitor (I) binds only to the enzyme-substrate (ES) complex (Figure 4.40). The inhibitor-bound complex forms mostly under concentrations of high substrate and the ES-I complex cannot release product while the inhibitor is bound.

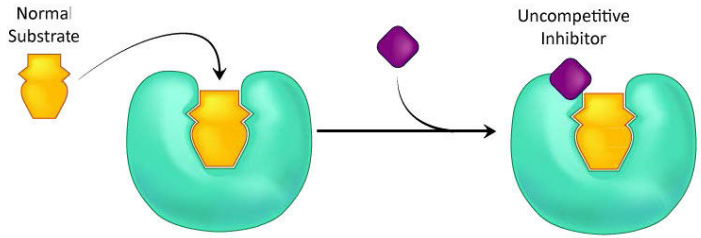


Figure 4.40 – Uncompetitive inhibition. Image by Aleia Kim

How can we tell what is happening between inhibitors and enzymes? Experimental evidence is obtained by measuring the rate of the enzyme-catalyzed reaction under various conditions. As substrate concentrations are increased, a competitive inhibitor will gradually 'lose out' to substrate and the rate of product formation will rise. But a similar effect will not be seen in the other forms of inhibition.

Additional experimental approaches are standard methods for such studies. These approaches have been enormously helpful to scientists studying natural systems as well as pharmaceutical interventions that target enzyme activity.

Suicide inhibition

In contrast to the first three types of inhibition, which involve reversible binding of the inhibitor to the enzyme, suicide inhibition is irreversible, because the inhibitor becomes covalently bound to the enzyme during the inhibition. Suicide inhibition rather closely resembles competitive inhibition because the inhibitor generally resembles the substrate and binds to the active site of the enzyme. The primary difference is that the suicide inhibitor is chemically reactive in the active site and makes a bond with it that precludes its removal. Such a mechanism is that employed by penicillin (Figure 4.43), which covalently links to the bacterial enzyme, DD

transpeptidase and stops it from functioning. Since the normal function of the enzyme is to make a bond necessary for the peptidoglycan complex of the bacterial cell wall, the cell wall cannot properly form and bacteria cannot reproduce.

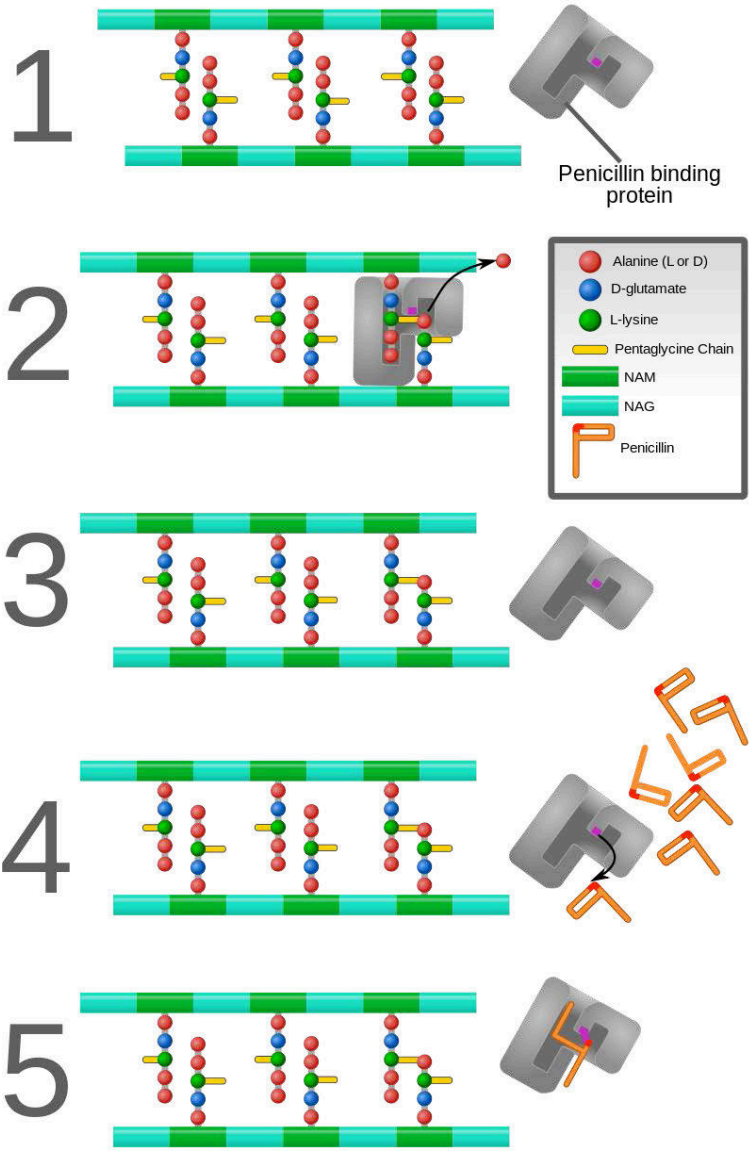


Figure 4.43 – Action of penicillin. DD-transpeptidase builds peptidoglycan layer of bacterial cell wall (1-3). Binding of penicillin by DD-transpeptidase stops peptidoglycan synthesis (4-5). Wikipedia

Control of enzymes

Enzymes are subject to controls from outside, but cells themselves also moderate and control enzyme activity. This allows an organism or cell to be responsive to its surroundings and physiological state.

There are four general ways this occurs:

1. allosterism,
2. covalent modification,
3. access to substrate, and
4. control of enzyme synthesis/breakdown.

Some enzymes are controlled by more than one of these methods.

Allosterism

The term allosterism refers to the fact that the activity of certain enzymes can be affected by the binding of small molecules at sites away from the active site.

When these effector molecules bind they cause the changes in the enzyme shape, as a result of binding.

Homotropic effectors, which are the same as the substrate molecules, usually are activators of the enzymes they bind to. from the less active T-state to the more active R-state, as the enzyme shifts its shape when a lot of substrate is around.

Away from the active site, enzymes can be either activated or inhibited when substances different than the substrate (heterotropic effectors) bind and induce different physical states

(shapes, as it were) in an enzyme. These shifts alter the ability of the binding site to collect and hold substrate. On the other hand, an enzyme can be activated by effector binding as well.

Feedback inhibition

A special kind of allosteric control is exhibited by the enzyme HMG-CoA reductase, which catalyzes an important reaction in the pathway leading to the synthesis of cholesterol. Binding of cholesterol to the enzyme reduces the enzyme's activity significantly. Cholesterol is not a substrate for the enzyme, so it is therefore a heterotropic effector.

Notably, though, cholesterol is the end-product of the pathway that HMG-CoA reductase catalyzes a reaction in. When enzymes are inhibited by an end-product of the pathway in which they participate, they are said to exhibit feedback inhibition.

Feedback inhibition always operates by allosterism and further, provides important and efficient control of an entire biochemical pathway. By inhibiting an early enzyme in a pathway, the flow of materials (and ATP hydrolysis required for their processing) for the entire pathway is stopped or reduced, assuming there are not alternate supply methods.

Pathway control

In the cholesterol biosynthesis pathway, stopping this one enzyme has the effect of shutting off (or at least slowing down) the entire pathway. This is significant because after catalysis by HMG-CoA reductase, there are over 20 further reactions necessary to make cholesterol, many of them requiring ATP energy. Shutting down one reactions stops all of them.

Covalent modification of enzymes

Some enzymes are synthesized in a completely inactive form and their activation requires covalent bonds in them to be cleaved. Such inactive forms of enzymes are called zymogens. Examples include

the proteins involved in blood clotting and protein-destroying enzymes of the digestive system, such as trypsin, chymotrypsin, and pepsin. These enzymes can not be active where they are made (inside the pancreas) but only where they end up (inside the GI tract). When they become active in the tissues of the pancreas, pancreatitis occurs.

Cascades

For both the blood clotting enzymes and the digestive enzymes, the zymogens are activated in a protease cascade. This occurs when activation of one enzyme activates others in a sort of spreading chain reaction. In such a scheme the first enzyme activated proteolytically cleaves the second zymogen, causing it to be activated, which in turn activates a third and this may proceed through several levels of enzymatic action (Figure 4.50).

The advantage of cascades is that they allow a large amount of zymogens to become activated fairly quickly, since there is an amplification of the signal at each level of catalysis.

Blood clotting involves polymerization of a protein known as fibrin. Since random formation of fibrin is extremely hazardous because it can block the flow of blood, potentially causing heart attack/stroke, the body synthesizes fibrin as a zymogen (fibrinogen) and its activation results from a "cascade" of activations of proteases that arise when a signal is received from a wound. Similarly, the enzyme catalyzing removal of fibrin clots (plasmin) is also synthesized as a zymogen (plasminogen), since random clot removal would also be hazardous (see below also).

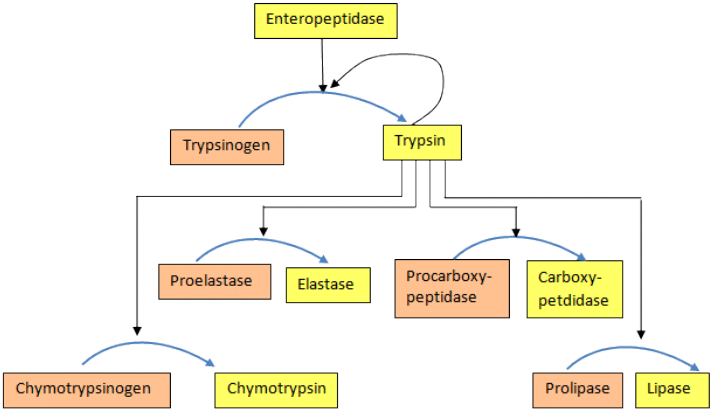


Figure 4.50 – Protease activation scheme. Wikipedia

Phosphorylation/dephosphorylation

Another common mechanism for control of enzyme activity by covalent modification is phosphorylation. The phosphorylation of enzymes (on the side chains of serine, threonine or tyrosine residues) is carried out by protein kinases. Enzymes activated by phosphorylation can be regulated by the addition of phosphate groups by kinases or their removal by phosphatases. Thus, this type of covalent modification is readily reversible, in contrast to proteolytic cleavage.

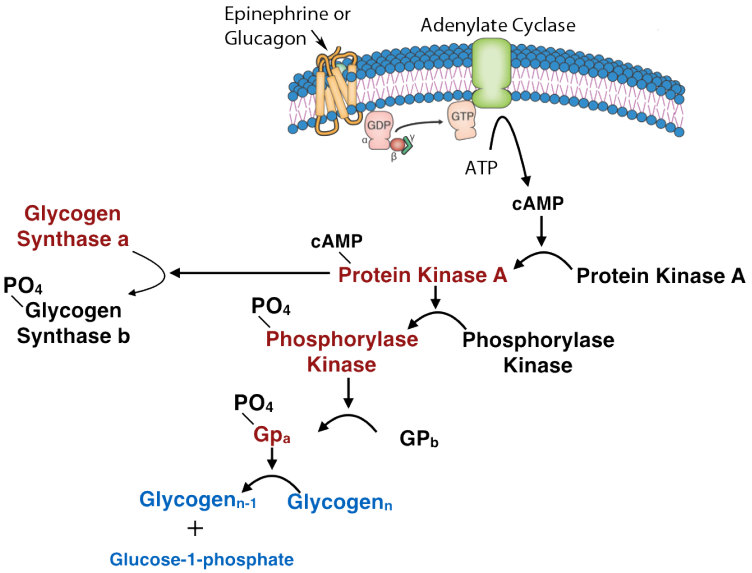


Figure 4.51 – Regulation by covalent modification of glycogen catabolism enzymes

Reduction/oxidation

An interesting covalent control of enzymes using reduction/oxidation is exhibited in photosynthetic plants. In the light phase of photosynthesis, electrons are excited by light and flow through carriers to NADP+, forming NADPH. Thus, in the light, the NADPH concentration is high. When NADPH concentration is high, the concentration of reduced ferredoxin (a molecule donating electrons to NADP+) is also high.

Reduced ferredoxin can transfer electrons to thioredoxin, reducing it. Reduced thioredoxin can, in turn, transfer electrons to proteins to reduce their disulfide bonds. Four enzymes related to the Calvin cycle can receive electrons from thioredoxin and become activated, as a result.

These include sedoheptulose 1,7-bisphosphatase, ribulose-5-phosphate kinase, fructose 1,6-bisphosphatase, and glyceraldehyde 3-phosphate dehydrogenase. Thus, in the light, electrons flow, causing NADPH to accumulate and ferredoxin to

push electrons in the direction of these enzymes above, activating them and favoring the Calvin cycle. In the dark, the concentration of reduced NADPH, reduced ferredoxin, and reduced thioredoxin fall, resulting in loss of electrons by the Calvin cycle enzymes (oxidations that re-form disulfide bonds) and the Calvin cycle inactivates.

Other enzyme control mechanisms

Other means of controlling enzymes relate to access to substrate (substrate-level control) and control of enzyme synthesis from DNA. Hexokinase is an enzyme that is largely regulated by availability of its substrate, glucose. When glucose concentration is low, the product of the enzyme's catalysis, glucose-6-phosphate, inhibits the enzyme's function.

Regulation of enzymes by controlling their synthesis is covered later in the book in the discussion relating to control of gene expression.

3.3: Mechanisms of Catalysis

Enzymes as a group are amazing in their ability to catalyze and provide control of chemical reactions in biological systems. There is a wealth of knowledge about specific enzymes and enzyme families, too, and each of these is both important and fascinating. While including many of these would be overwhelming, considering just a few helps put the general knowledge into context. This section provides some examples and goes into detail on one of them – the mechanism of action of serine proteases.

Chymotrypsin

We will begin with mechanism of action of one enzyme – chymotrypsin. Found in our digestive system, chymotrypsin's catalytic activity is cleaving peptide bonds in proteins and it uses the side chain of a serine in its mechanism of catalysis. Many other protein-cutting enzymes employ a very similar mechanism and they are known collectively as serine proteases (Figure 4.52).

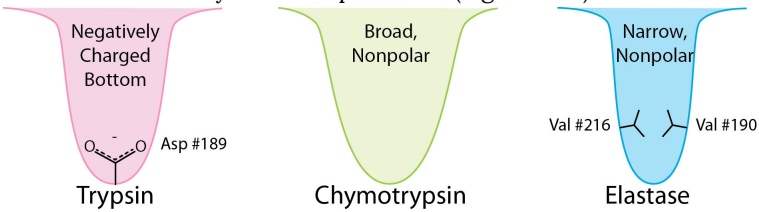


Figure 4.52 – Substrate binding sites (S1 pockets) of three serine proteases. Image by Aleia Kim

These enzymes are found in prokaryotic and eukaryotic cells and all use a common set of three amino acids in the active site called a catalytic triad (Figure 4.53). It consists of aspartic acid, histidine, and serine. The serine is activated in the reaction mechanism to form a nucleophile in these enzymes and gives the class their name.

With the exception of the recognition that occurs at the substrate binding site, the mechanism shown here for chymotrypsin would be applicable to any of the serine proteases.

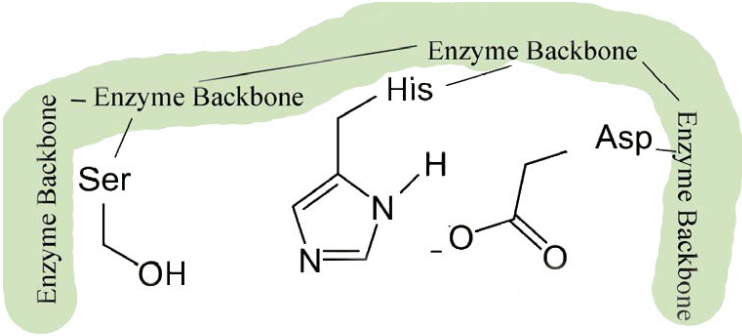


Figure 4.53 – 1. Active site of chymotrypsin showing the catalytic triad of serine – histidine-aspartic acid

Specificity

As a protease, chymotrypsin acts fairly specifically, cutting not all peptide bonds, but only those that are adjacent to relatively non-polar amino acids in the protein. One of the amino acids it cuts adjacent to is phenylalanine. The enzyme's action occurs in two phases – a fast phase that occurs first and a slower phase that follows. The enzyme has a substrate binding site that includes a region of the enzyme known as the S1 pocket. Let us step through the mechanism by which chymotrypsin cuts adjacent to phenylalanine.

Substrate binding

The process starts with the binding of the substrate in the S1 pocket (Figure 4.54). The S1 pocket in chymotrypsin has a hydrophobic hole in which the substrate is bound. Preferred substrates will include amino acid side chains that are bulky and hydrophobic, like phenylalanine. If an ionized side chain, like that of glutamic acid binds in the S1 pocket, it will quickly exit, much like water would avoid an oily interior.

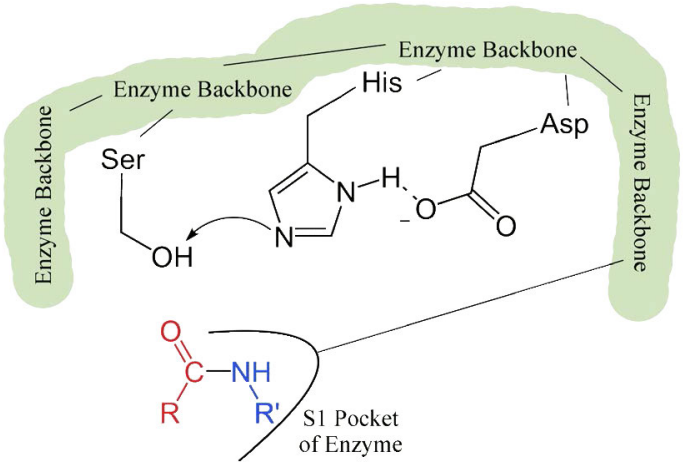


Figure 4.54 – 2. Binding of substrate to S1 pocket in the active site

Shape change on binding

When the proper substrate binds in the S1 pocket, its presence induces an ever so slight change in the shape of the enzyme. This subtle shape change on the binding of the proper substrate starts the steps of the catalysis. Since the catalytic process only starts

when the proper substrate binds, this is the reason that the enzyme shows specificity for cutting at specific amino acids in the target protein. Only amino acids with the side chains that interact well with the S1 pocket start the catalytic wheels turning.

The slight changes in shape involve changes in the positioning of three amino acids (aspartic acid, histidine, and serine) in the active site known as the *catalytic triad*.

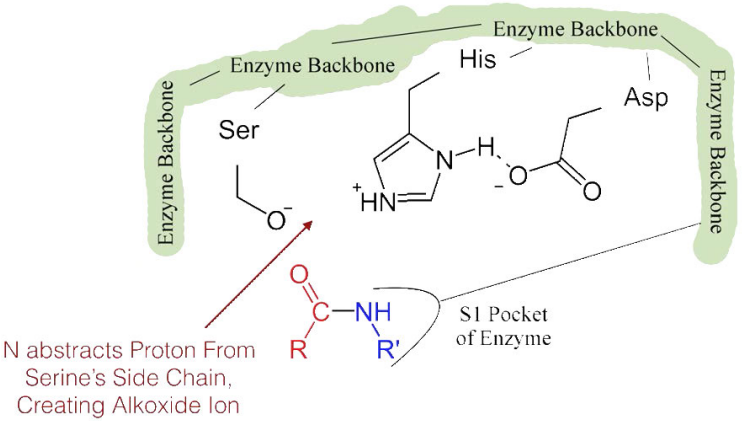


Figure 4.55 – 3. Formation of alkoxide ion

The shift of the negatively charged aspartic acid towards the electron rich histidine ring favors the abstraction of a proton by the histidine from the hydroxyl group on the side chain of serine, resulting in production of a very reactive alkoxide ion in the active site (Figure 4.55).

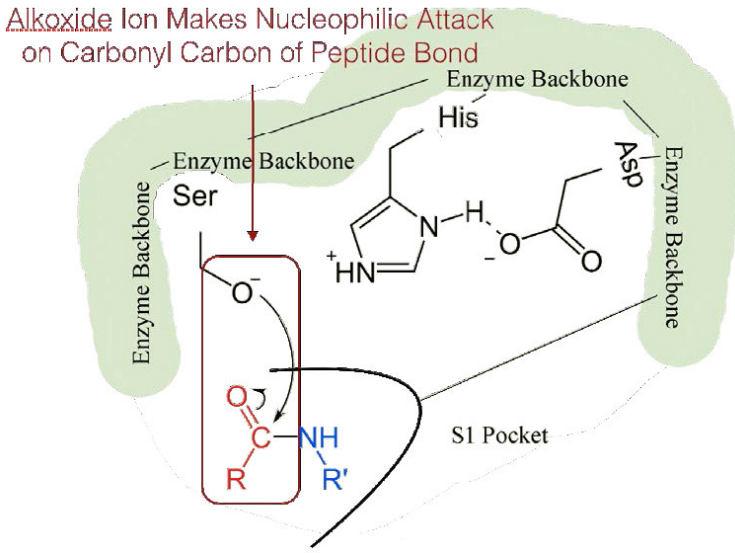


Figure 4.56 – 4. Nucleophilic attack

Since the active site at this point also contains the polypeptide chain positioned with the phenylalanine side chain embedded in the S1 pocket, the alkoxide ion performs a nucleophilic attack on the peptide bond on the carboxyl side of phenylalanine sitting in the S1 pocket (Figure 4.56). This reaction breaks the peptide bond (Figure 4.57) and causes two things to happen.

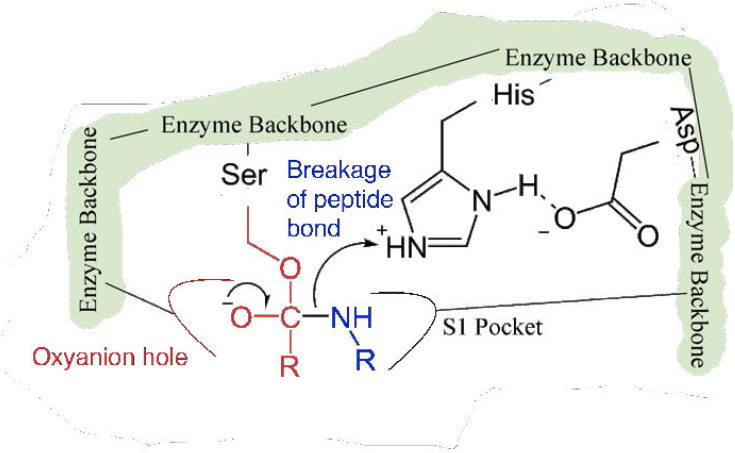


Figure 4.57 – 5. Stabilization by oxyanion hole. Breakage of peptide bond.

First, one end of the original polypeptide is freed and exits the active site (Figure 4.58). The second is that the end containing the phenylalanine is covalently linked to the oxygen of the serine side chain. At this point we have completed the first (fast) phase of the catalysis.

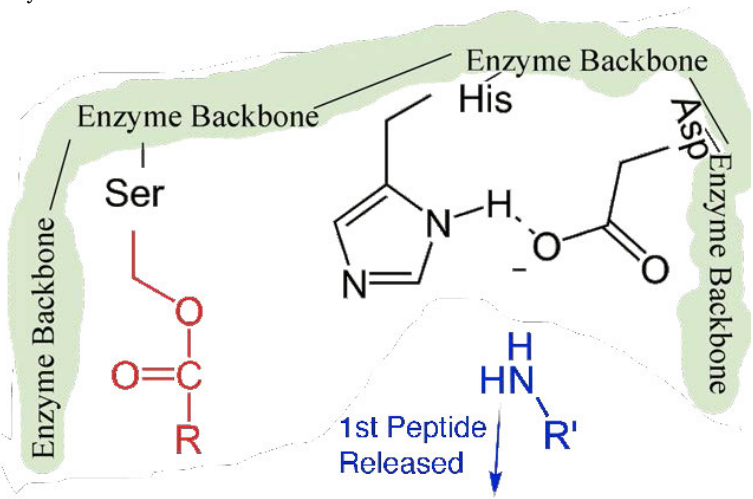


Figure 4.58 – 6. First peptide released. Other half bonded to serine.

Slower second phase

The second phase of the catalysis by chymotrypsin is slower. It requires that the covalent bond between phenylalanine and serine's oxygen be broken so the peptide can be released and the enzyme can return to its original state. The process starts with entry of water into the active site. Water is attacked in a fashion similar to that of the serine side chain in the first phase, creating a reactive hydroxyl group (Figure 4.59) that performs a nucleophilic attack on the phenylalanine-serine bond (Figure 4.60), releasing it and replacing the proton on serine. The second peptide is released in the process and the reaction is complete with the enzyme back in its original state (Figure 4.61).

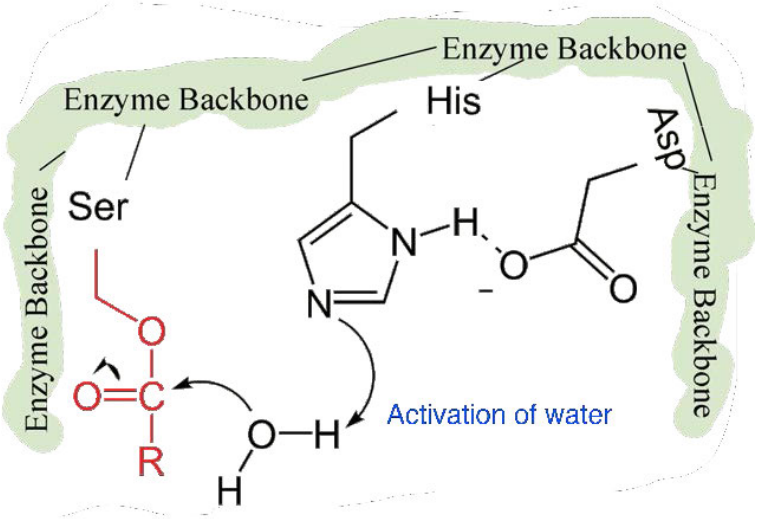


Figure 4.59 – 7. Activation of water by histidine

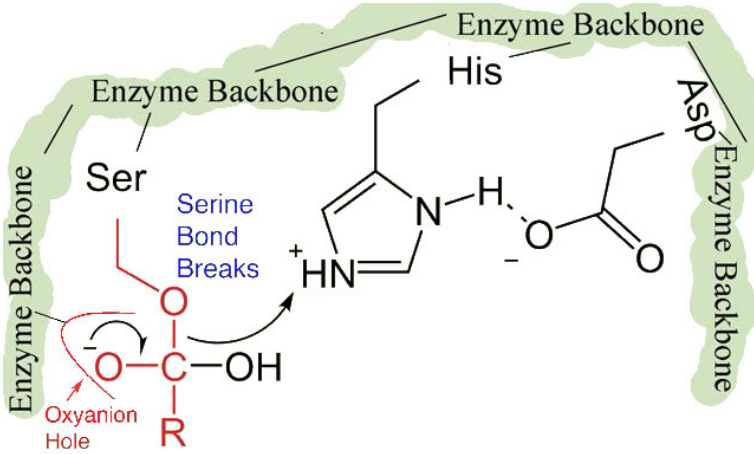


Figure 4.60 – 8. Nucleophile attack by hydroxyl creates tetrahydryl intermediate stabilized by oxyanion hole. Bond to serine breaks.

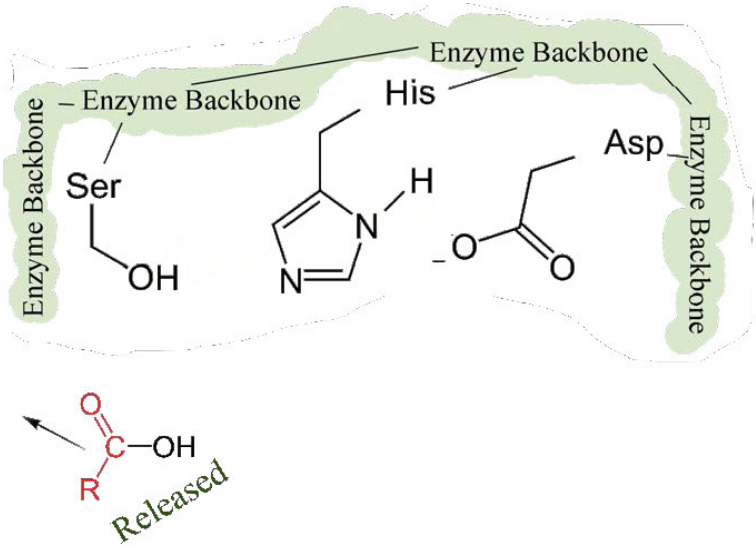


Figure 4.61 – 9. Second half of peptide released. Enzyme active site restored.

Serine proteases

The list of serine proteases is quite long. They are grouped in two broad categories – 1) those that are chymotrypsin-like and 2) those that are subtilisin-like. Though subtilisin-type and chymotrypsin-like enzymes use the same mechanism of action, including the catalytic triad, the enzymes are otherwise not related to each other by sequence and appear to have evolved independently. They are, thus, an example of convergent evolution – a process where evolution of different forms converge on a structure to provide a common function.

The serine protease enzymes cut adjacent to specific amino acids and the specificity is determined by the size/shape/charge of amino acid side chain that fits into the enzyme's S1 binding pocket (Figure 4.62).

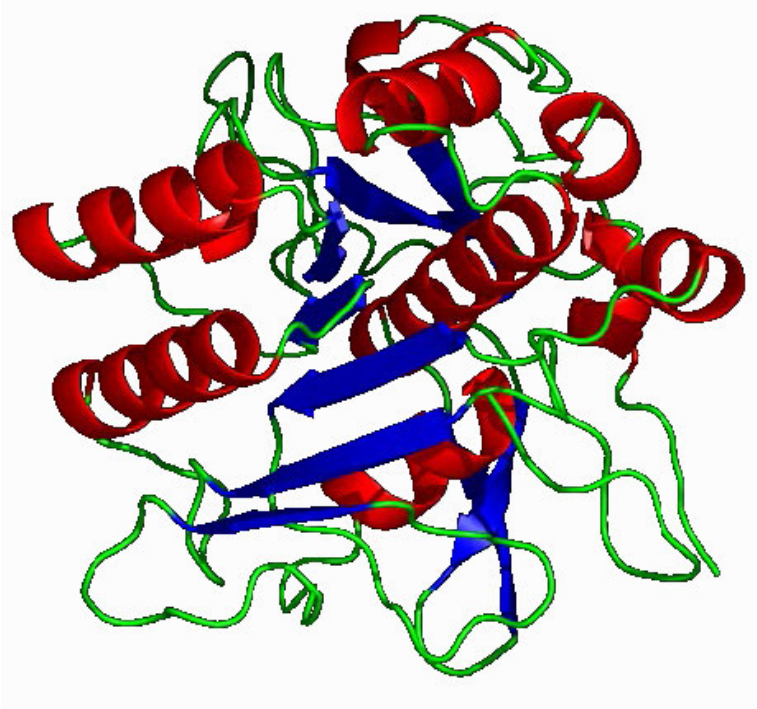


Figure 4.62 – Subtilisin – A serine protease

Examples of serine proteases include trypsin, chymotrypsin, elastase, subtilisin, signal peptidase I, and nucleoporin. Serine proteases participate in many physiological processes, including blood coagulation, digestion, reproduction, and the immune response.

Metalloproteases

Metalloproteases (Figure 4.64) are enzymes whose catalytic mechanism for breaking peptide bonds involves a metal. Most metalloproteases use zinc as their metal, but a few use cobalt, coordinated to the protein by three amino acid residues with a labile water at the fourth position. A variety of side chains are used – histidine, aspartate, glutamate, arginine, and lysine. The water is the target of action of the metal which, upon binding of the proper substrate, abstracts a proton to create a nucleophilic hydroxyl group that attacks the peptide bond, cleaving it (Figure 4.64). Since the nucleophile here is not attached covalently to the enzyme, neither of the cleaved peptides ends up attached to the enzyme during the catalytic process. Examples of metalloproteases include carboxypeptidases, aminopeptidases, insulinases and thermolysin.

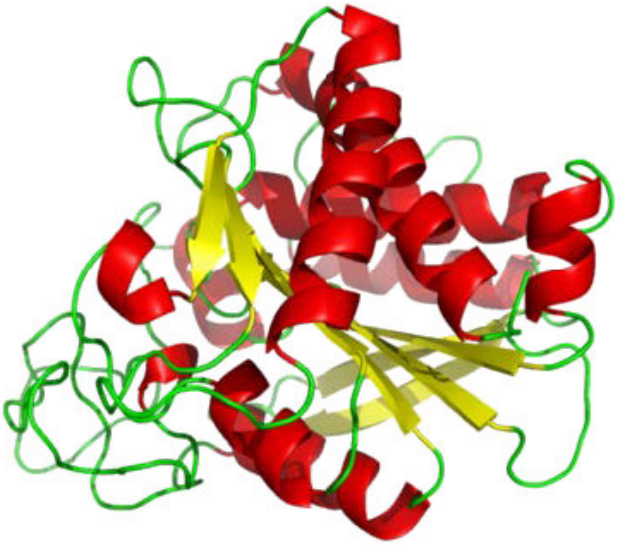


Figure 4.64 – Carboxypeptidase – A metalloprotease

Protease inhibitors

Molecules which inhibit the catalytic action of proteases are known as protease inhibitors. These come in a variety of forms and have biological and medicinal uses. Many biological inhibitors are proteins themselves. Protease inhibitors can act in several ways, including as a suicide inhibitor, a transition state inhibitor, a denaturant, and as a chelating agent. Some work only on specific classes of enzymes. For example, most known aspartyl proteases are inhibited by pepstatin. Metalloproteases are sensitive to anything that removes the metal they require for catalysis. Zinc-containing metalloproteases, for example, are very sensitive to EDTA, which chelates the zinc ion.

One category of proteinaceous protease inhibitors is known as

the serpins. Serpins inhibit serine proteases that act like chymotrypsin. 36 of them are known in humans.

Serpins are unusual in acting by binding to a target protease irreversibly and undergoing a conformational change to alter the active site of its target. Other protease inhibitors act as competitive inhibitors that block the active site.

Serpins can be broad in their specificity. Some, for example, can block the activity of cysteine proteases. One of the best known biological serpins is α-1-anti-trypsin (A1AT – Figure 4.66) because of its role in lungs, where it functions to inhibit the elastase protease. Deficiency of A1AT leads to emphysema. This can arise as a result of genetic deficiency or by cigarette smoking. Reactive oxygen species produced by cigarette smoking can oxidize a critical methionine residue (#358 of the processed form) in A1AT, rendering it unable to inhibit elastase. Uninhibited, elastase can attack lung tissue and cause emphysema. Most serpins work extracellularly. In blood, for example, serpins like antithrombin can help to regulate the clotting process.

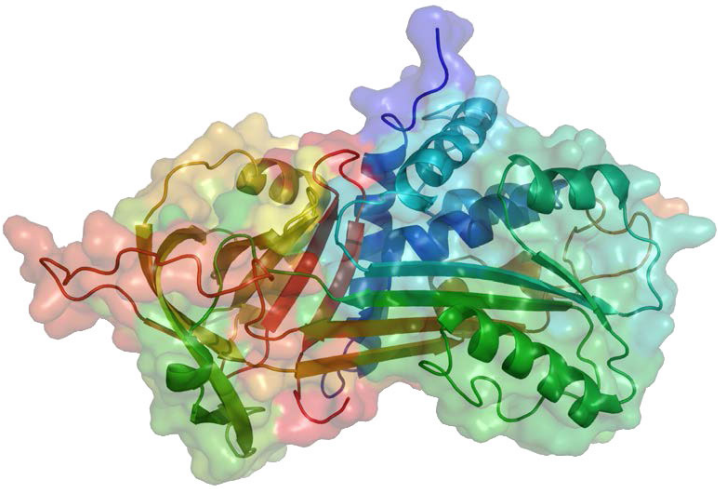


Figure 4.66 – α-1-antitrypsin

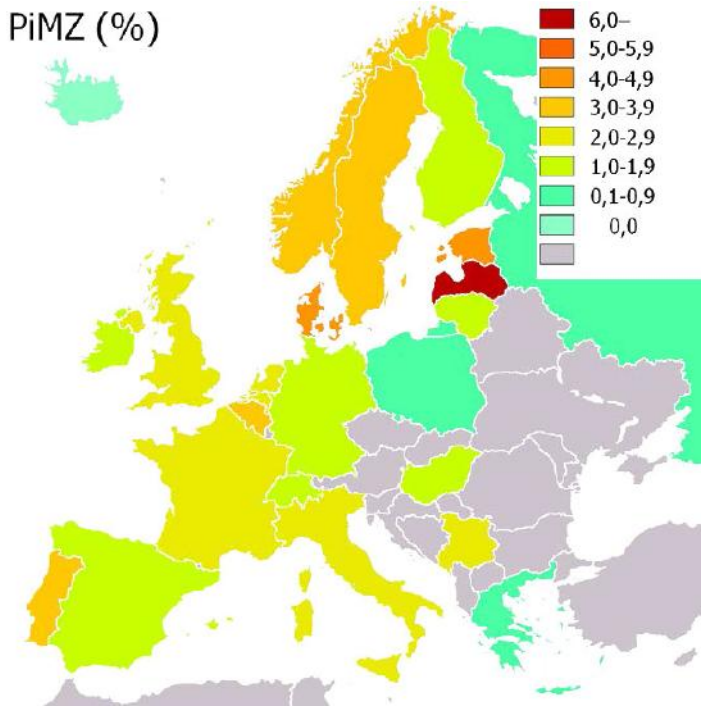


Figure 4.67 – Incidence of α-1-antitrypsin (PiMZ) deficiency in Europe by percent. Wikipedia

Anti-viral Agents

Protease inhibitors are used as anti-viral agents to prohibit maturation of viral proteins – commonly viral coat proteins.

They are part of drug "cocktails" used to inhibit the spread of HIV in the body and are also used to treat other viral infections, including hepatitis C. They have also been investigated for use in treatment of malaria and may have some application in anti-cancer therapies as well.

4.1: Prelude to Information Processing

"The blueprints for the construction of one human being requires only a meter of DNA and one tiny cell. … even Mozart started out this way." – L.L. Larison Cudmore

As creatures used to regarding ourselves as exceptional, humans must surely be humbled to realize that the instructions, for making one of our own, reside in a molecule so simple that scientists, for a very long time, did not believe could possibly contain enough information to build even a simple cell. But a large body of evidence, built up over the past century, supports Larison Cudmore's assertion that the information for making you and me (and all the other kinds of living things in the world) is encoded in DNA. Tying in with classical genetic observations about how characteristics are passed on from one generation to the next, the discovery that there was a molecule that carried this information altered forever how people thought about heredity.

The elucidation of the structure of DNA provided greater insights into how traits might be encoded in a molecule, and the ways in which the information is used by cells. As we learn more about this topic, scientists have remarked on how the information in our DNA resembles the programs that drive computers. While this analogy is a simplification, there is definitely it is useful, with information in our DNA directly serving as stored information that determines the properties of the proteins that operate in cells and whole bodies. The analogy from stored information to accessed and 'read' information to functional machines is one many of us can relate to.

If this sounds strange, it is even more intriguing to realize DNA is copied and passed on from cell to cell, from one generation to the next. There is an unbroken line of inheritance from the first cell to every organism alive today.

The subject of biochemistry includes extensive consideration of structures in biology and also reactions (metabolism). But it also, importantly, includes informational considerations. In this way biochemistry is different from the other subdisciplines of chemistry, because it involves the essence of life: reproduction, the propagation of itself. For this to occur information must be passed around within and between cells.

Genetic information, its storage, how it is read and interpreted, and gives rise to the cellular activities that we can observe is a major subject of this chapter. Another kind of information is also considered, which is the molecular information that cells receive from, and send to, each other. The interplay of these two kinds of information is responsible for the form and behavior of all living organisms.

4.2: Structure and Function - Nucleic Acids

The nucleic acids, DNA and RNA, may be thought of as the information storage molecules of the cell. In this section, we will examine the structures of DNA and RNA, and how these structures are related to the functions these molecules perform.

We will begin with DNA, which is the hereditary information in every cell, that is copied and passed on from generation to generation.

DNA was discovered in 1869 by Friedrich Miescher, and was identified as the genetic material in experiments in the 1940s led by Oswald Avery, Colin MacLeod, and Maclyn McCarty.

In 1953, the international Science journal Nature published 3 papers on the structure of DNA. The authors included James Watson and Francis Crick, Rosalind Franklin and Erwin Chargaff. Arguably,

the Watson and Crick paper has had more scientific impact per word than any other research article ever published. Today, instruction about the double helical structure of DNA is foundational to biology classes in high schools and colleges. It launched the study of genetics forward by providing information that linked structure at the atomic and molecular level to heredity and the basis of inherited disease.

The double helix structure presented in the Watson and Crick paper and justified by the X-ray data from Rosalind Franklin, includes a 2-strand structure which has at its core, fused ring structures called nitrogenous bases joined by hydrogen bonds to form base pairs. The base adenine always pairs with thymine, and the base guanine invariably pairs with cytosine. Two hydrogen bonds form between adenine and thymine, and three hydrogen bonds hold together guanine and cytosine (Figure 2.127).

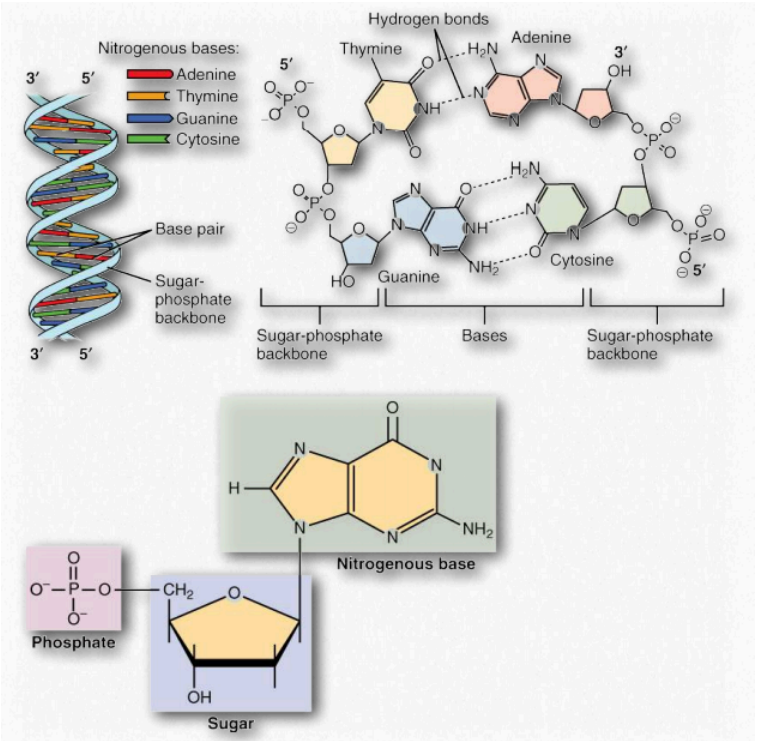


Figure 2.127 – A DNA duplex with base pairs, a closeup of base pairing, and a closeup of a nucleotide Wikipedia

The complementary structure of these bases revealed to Watson and Crick how DNA might be (and in fact, is) replicated. Further, it indicates how information is DNA is transmitted to RNA for the synthesis of proteins. Both of these important biological processes rely on the ability of either strand of the double helix being 'readable' in isolation from the other. Separation of the strand exposes hydrogen bonding sites where a complementary strand could be built up, purely by knowing the basis of the pairing: A pairs with T, G with C, etc..

Molecular machinery does essentially this, as it uses a single strand as a template for reproducing DNA molecules with the proper base sequences. Separate molecular machinery also transcribes this information to mRNA, where the code can then be translated into an amino acid sequence. Strings of amino acids fold into proteins, and thus the stored information finds expression as functioning proteins.
In addition to the hydrogen bonds between bases of each strand, the double helix molecule is held together by hydrophobic interactions of the stacked, non-polar bases.

Structure

A single DNA strand (half of a double helix) can be described as a polymer of nucleoside monophosphates held together by phosphodiester bonds. It is a long and narrow molecule. Two such paired strands make up the DNA molecule, which is then twisted into a helix. In the most common B form, the DNA helix twists with a repeat of 10.5 base pairs per full helical turn, with sugars and phosphate forming the covalent phosphodiester "backbone" of the

molecule and the adenine, guanine, cytosine, and thymine bases oriented in the middle where they form the now familiar base pairs that look like the rungs of a ladder.

Building blocks

The term nucleotide refers to the building blocks of both DNA (deoxyribonucleoside triphosphates, dNTPs) and RNA (ribonucleoside triphosphates, NTPs). In order to discuss this important group of molecules, it is necessary to define some terms.

Nucleotides contain three primary structural components. These are a nitrogenous base, a 5-carbon pentose sugar, and at least one phosphate. The nitrogenous bases found in nucleic acids include adenine and guanine (called purines) and cytosine, uracil, or thymine (called pyrimidines).

There are two sugars found in nucleotides – deoxyribose and ribose (Figure 2.128). Deoxyribose differs from ribose at carbon 2, with ribose having an OH group, where deoxyribose has H.

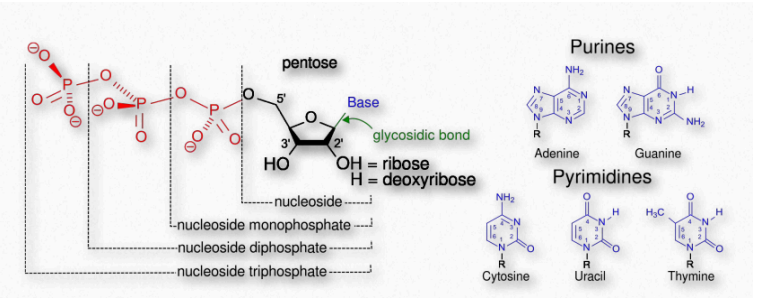


Figure 2.128 – Nucleotides, nucleosides, and bases

Nucleotides containing deoxyribose are called deoxyribonucleotides and are the forms found in DNA.

Nucleotides containing ribose are called ribonucleotides and are found in RNA.

Both DNA and RNA contain nucleotides with adenine, guanine, and cytosine, but with very minor exceptions, RNA contains uracil nucleotides, whereas DNA contains thymine nucleotides. When a base is attached to a sugar, the product gains a new name.

- **uracil-containing** = uridine (attached to ribose) / deoxyuridine (attached to deoxyribose)
- **thymine-containing** = ribothymidine (attached to ribose) / thymidine (attached to deoxyribose)
- **cytosine-containing** = cytidine (attached to ribose – Figure 2.129) / deoxycytidine (attached to deoxyribose)
- **guanine-containing** = guanosine (attached to ribose) / deoxyguanosine (attached to deoxyribose)
- **adenine-containing** = adenosine (attached to ribose) / deoxyadenosine (attached to deoxyribose)

The addition of one or more phosphates to a nucleoside makes it a nucleotide. Nucleotides are often referred to as nucleoside phosphates, for this reason. The number of phosphates in the nucleotide is indicated by the appropriate prefixes (mono, di or tri).

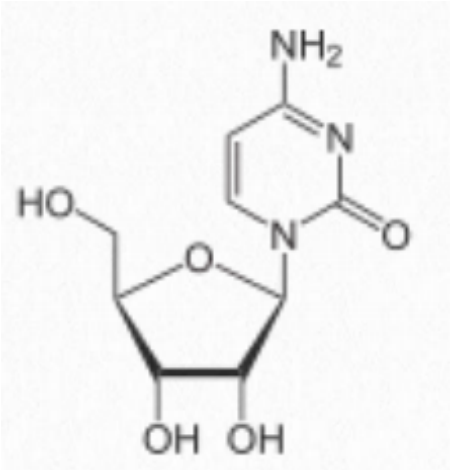


Figure 2.129 Cytidine

Addition of second and third phosphates to a nucleoside

monophosphate requires an input of energy, due to the repulsion of negatively charged phosphates. This chemical energy can be released in the reverse reaction. This conversion of triphosphate nucleotides (such as ATP) to diphosphates is a reaction that is employed for energy storage and utilization by cells.

Note: Ribonucleotides as Energy Sources

Though ATP is the most common and best known cellular energy source, each of the four ribonucleotides plays important roles in providing energy. GTP, for example, is the energy source for protein synthesis (translation) as well as for a handful of metabolic reactions.

Hydrogen bonds

Hydrogen bonds between the base pairs hold a nucleic acid duplex together, with two hydrogen bonds per A-T pair (or per A-U pair in RNA) and three hydrogen bonds per G-C pair. The B-form of DNA has a prominent major groove and a minor groove tracing the path of the helix (Figure 2.132). Proteins, such as transcription factors bind in these grooves and access the hydrogen bonds of the base pairs to "read" the sequence therein.

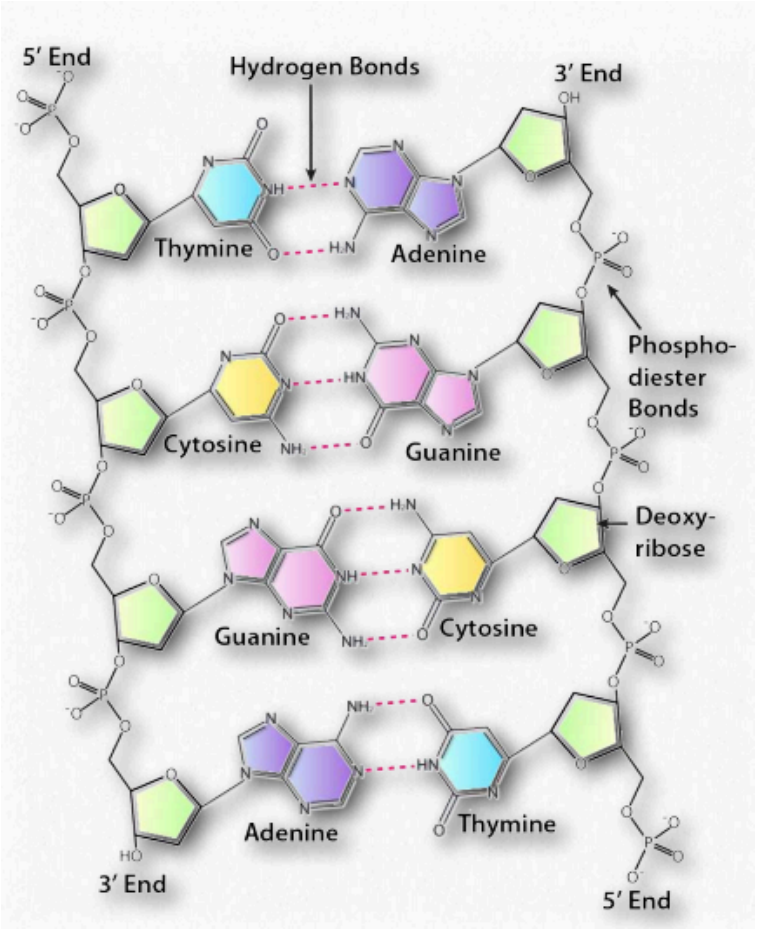


Figure 2.131 – Anti-parallel orientation of a DNA duplex, phosphodiester backbone, and base pairing Image by Aleia Kim

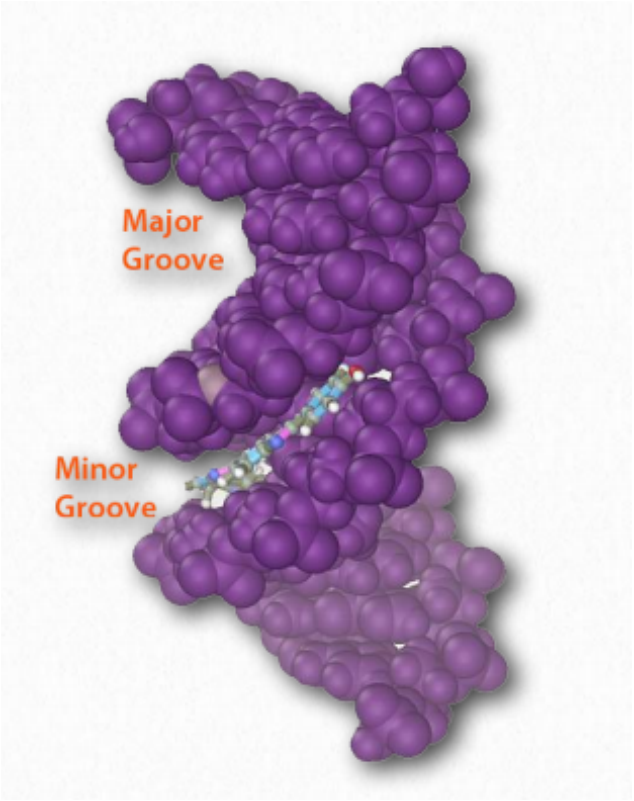
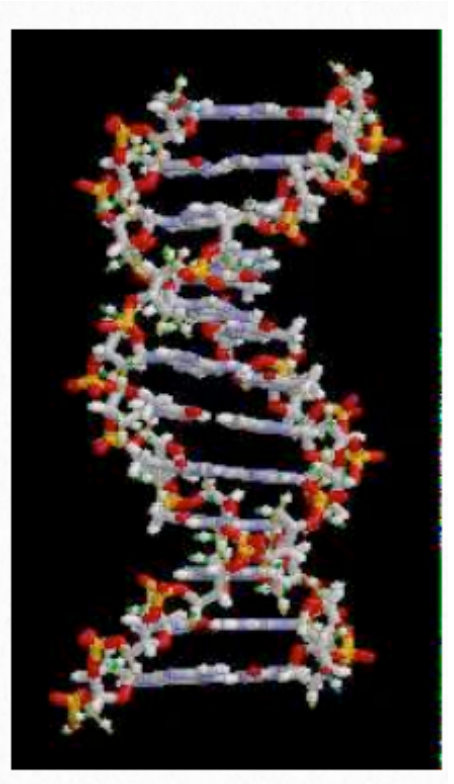


Figure 2.132 – Major and minor grooves of DNA. The minor groove has been bound by a dye Wikipedia



Movie 2.5 – B-form DNA duplex rotating in space Wikipedia

RNA

The structure of RNA (Figure 2.137) is very similar to that of a single strand of DNA. Built of ribonucleotides, joined together by the same sort of phosphodiester bonds as in DNA, RNA uses uracil in place of thymine. The building of messenger RNAs by copying a DNA template is a crucial step in the transfer of the information in DNA to a form that directs the synthesis of protein. Additionally,

ribosomal and transfer RNAs serve important roles in "reading" the information in the mRNA codons and in polypeptide synthesis. RNAs are also known to play important roles in the regulation of gene expression.

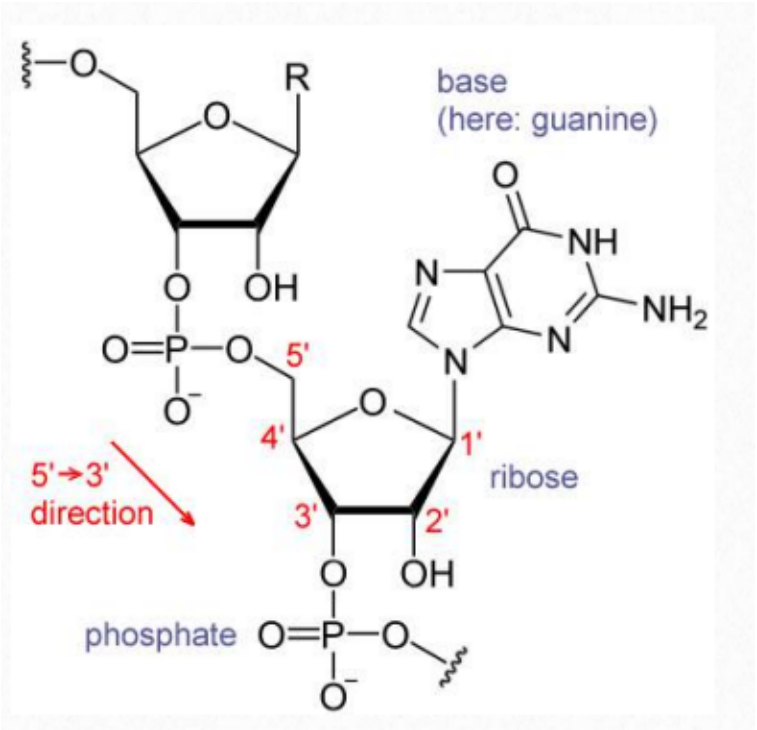


Figure 2.137 A section of an RNA molecule Wikipedia

RNA world

The discovery, in 1990, that RNAs could play a role in catalysis, a function once thought to be solely the domain of proteins, was followed by the discovery of many more so-called ribozymes- RNAs that functioned as enzymes. This suggested the answer to a long-standing chicken or egg puzzle – if DNA encodes proteins, but the replication of DNA requires proteins, how did a replicating system come into being? This problem could be solved if the first replicator was RNA, a molecule that can both encode information and carry out catalysis. This idea, called the "RNA world" hypothesis, suggests

that DNA as genetic material and proteins as catalysts arose later, and eventually prevailed because of the advantages they offer. The lack of a 2'OH on deoxyribose makes DNA more stable than RNA. The double-stranded structure of DNA also provides an elegant way to easily replicate it. RNA catalysts, however, remain, as remnants of that early world. In fact, the formation of peptide bonds, essential for the synthesis of proteins, is catalyzed by RNA.

Secondary structure

Figure 2.138 – tRNA Images – 3D projection (left) and 2D projection (inset) Wikipedia

With respect to structure, RNAs are more varied than their DNA cousin. Created by copying regions of DNA, cellular RNAs are synthesized as single strands, but they often have self-complementary regions leading to "foldbacks" containing duplex regions. These are most easily visualized in the ribosomal RNAs

(rRNAs) and transfer RNAs (tRNAs) (Figure 2.138), though other RNAs, including messenger RNAs (mRNAs), small nuclear RNAs (snRNAs), microRNAs (Figure 2.139), and small interfering RNAs (siRNAs) may each have double helical regions as well.

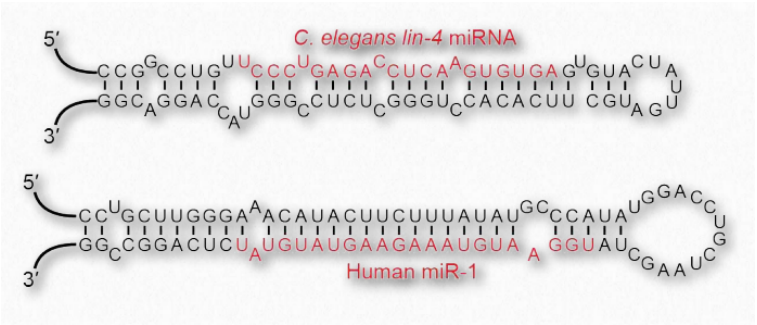


Figure 2.139 – MicroRNA stem loops. Wikipedia

Base pairing

Base pairing in RNA is slightly different than DNA. This is due to the presence of the base uracil in RNA in place of thymine in DNA. Like thymine, uracil forms base pairs with adenine, but unlike thymine, uracil can, to a limited extent, also base pair with guanine, giving rise to many more possibilities for pairing within a single strand of RNA.

These additional base pairing possibilities mean that RNA has many ways it can fold upon itself that single-stranded DNA cannot. Folding, of course, is critical for protein function, and we now know that, like proteins, some RNAs in their folded form can catalyze reactions just like enzymes. Such RNAs are referred to as ribozymes. It is for this reason scientists think that RNA was the first genetic material, because it could not only carry information, but also catalyze reactions. Such a scheme might allow certain RNAs to make copies of themselves, which would, in turn, make more copies of themselves, providing a positive selection.

Stability

RNA is less chemically stable than DNA. The presence of the 2' hydroxyl on ribose makes RNA much more prone to hydrolysis than DNA, which has a hydrogen instead of a hydroxyl. Further,

RNA has uracil instead of thymine. It turns out that cytosine is the least chemically stable base in nucleic acids. It can spontaneously deaminate and in turn is converted to a uracil. This reaction occurs in both DNA and RNA, but since DNA normally has thymine instead of uracil, the presence of uracil in DNA indicates that deamination of cytosine has occurred and that the uracil needs to be replaced with a cytosine. Such an event occurring in RNA would be essentially undetectable, since uracil is a normal component of RNA. Mutations in RNA have much fewer consequences than mutations in DNA because they are not passed between cells in division.

Catalysis

RNA structure, like protein structure, has importance, in some cases, for catalytic function. Like random coils in proteins that give rise to tertiary structure, single-stranded regions of RNA that link duplex regions give these molecules a tertiary structure, as well. Catalytic RNAs, called ribozymes, catalyze important cellular reactions, including the formation of peptide bonds in ribosomes (Figure 2.114). DNA, which is usually present in cells in strictly duplex forms (no tertiary structure, per se), is not known to be involved in catalysis.

RNA structures are important for reasons other than catalysis. The 3D arrangement of tRNAs is necessary for enzymes that attach amino acids to them to do so properly. Further, small RNAs called siRNAs found in the nucleus of cells appear to play roles in both gene regulation and in cellular defenses against viruses. The key to the mechanisms of these actions is the formation of short foldback RNA structures that are recognized by cellular proteins and then chopped into smaller units. One strand is copied and used to base pair with specific mRNAs to prevent the synthesis of proteins from them.

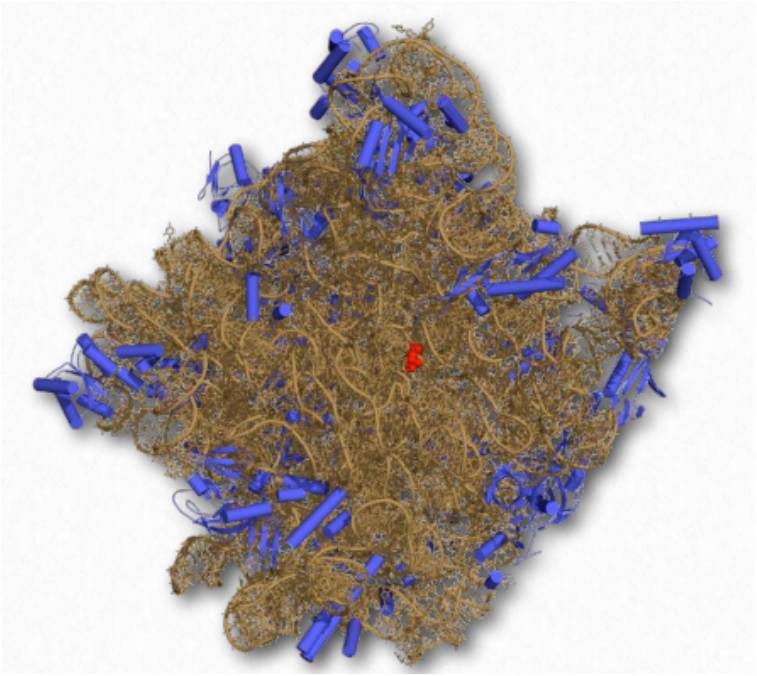


Figure 2.140 – Structure of the 50S ribosomal subunit. rRNA shown in brown. Active site in red Wikipedia

Ames test

The Ames test (Figure 2.147) is an analytical method that allows one to determine whether a compound causes mutations in DNA (is mutagenic) or not. The test is named for Dr. Bruce Ames, a UC Berkeley emeritus professor who was instrumental in creating it. In the procedure, a single base pair of a selectable marker of an organism is mutated in a plasmid to render it nonfunctional. In the example, a strain of Salmonella is created that lacks the ability to grow in the absence of histidine. Without histidine, the organism will not grow, but if that one base in the plasmid's histidine gene gets changed back to its original base, a functional gene will be made and the organism will be able to grow without histidine.

A culture of the bacterium lacking the functional gene is grown with the supply of histidine it requires. It is split into two vials. To

one of the vials, a compound that one wants to test the mutagenicity of is added. To the other vial, nothing is added. The bacteria in each vial are spread onto plates lacking histidine. In the absence of mutation, no bacteria will grow. The more colonies of bacteria that grow, the more mutation happened. Note that even the vial without the possible mutagenic compound will have a few colonies grow, as a result of mutations unlinked to the potential mutagen.

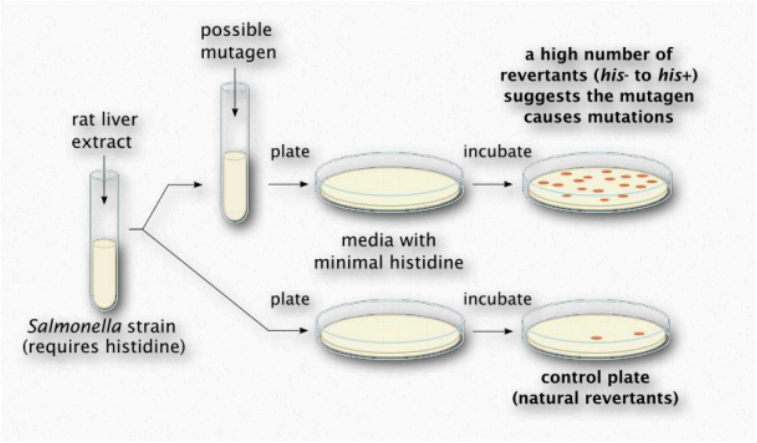


Figure 2.147 – The Ames test Wikipedia

Mutation happens in all cells at a low level. If the plate with the cells from the vial with the compound has more colonies than the cells from the control vial (no compound), then that would be evidence that the compound causes more mutations than would normally occur and it is therefore a mutagen. On the other hand, if there was no significant difference in the number of colonies on each plate, then that would suggest it is not mutagenic. The test is not perfect – it identifies about 90% of known mutagens – but its simplicity and inexpensive design make it an excellent choice for an initial screen of a compound.

4.3: Gene Expression

Gene expression is the process that takes the information encoded in DNA and converts it into functional biomolecules, primarily proteins. The information in DNA, encoded in the sequence of nitrogenous bases along its length, is used as the instructions for protein production. While there are some important exceptions, nearly all of the coding DNA is used in this way, directing the cell to construct amino acid chains in particular sequences. The exceptions include the production of some biologically active RNA molecules, such as transfer-RNA molecules, which are not made into protein.

This Central Dogma of Molecular Biology describes this general pathway of information flow. A generally-true version which applies to eukaryotic organisms can be communicated quickly:

DNA → RNA → protein

This sequence of events describes in basic terms the information flow that occurs as our genes, inherited in the DNA we get from our biological parents, are expressed in our bodies.

The two-step sequence here includes, first, the transcription of information in DNA to RNA. The DNA molecules contained in a eukaryotic cell nucleus contains portions of nucleic acid sequence that are coding sequences. These sequences, called 'genes,' are used as a template to produce RNA that leaves the nucleus. The RNA carries the information with it in the form of the sequence of its bases. After exiting the nucleus, messenger RNA (mRNA) interacts with cellular components called Ribosomes, where the information contained within its sequence is translated into a specific sequence of amino acids. The ribosome reads the mRNA and engages in the assembly of the polypeptide chain with the selection of amino acids guided by small transfer RNA molecules (tRNAs) that are able to translate the code. Each sequence of 3 nitrogenous bases on the mRNA are converted to a single, select

amino acid based on the genetic code.
Transcription is a process whereby the information in DNA is copied into the RNA molecules that can take it out of the nucleus. The 'language' of the code is unchanged in this process, as each nucleotide in DNA is copied into a nucleotide of RNA.
Translation is a process involving a change of 'language,' from the nucleotide sequence in the mRNA to the amino acid sequence of the nascent protein.
Transcription occurs within the nucleus, where the coding DNA remains. The translation process is carried outside of the nucleus, in the **cytoplasm** of the cell, by ribosomes.
Ribosomes are large protein-nucleic acid complexes that contain ribosomal RNA (rRNA) and proteins. The proteins and rRNA are organized into two subunits, one large and another small. Ribosomes function by binding to mRNAs and holding them in a way that allows the amino acids encoded by the RNA to be joined sequentially to form a polypeptide. Transfer RNAs are the carriers of the appropriate amino acids to the ribosome.
The ribosome is an important example of a non-protein catalyst, as it catalyzes the reactions that form peptide bonds, linking amino acids together.

The Genetic Code

A code can be thought of as a system for storing or communicating information. A familiar example is the use of letters to represent the names of airports (e.g., PDX for Portland, Oregon and ORD for Chicago's O'Hare). When a tag on your luggage shows PDX as the destination, it conveys information that your bag should be sent to Portland, Oregon. To function well, such a code must have unique identifiers for each airport and people who can decode the identifiers correctly. That is, PDX must stand only for Portland, Oregon and no other airport. Also, luggage handlers must be able

to correctly recognize what PDX stands for, so that your luggage doesn't land in Phoenix, instead.

How does this relate to genes and the proteins they encode?

Genes are first transcribed into mRNA, as we have already discussed. The sequence of an mRNA, copied from a gene, directly specifies the sequence of amino acids in the protein it encodes. Each amino acid in the protein is specified by a sequence of 3 bases called a codon in the mRNA (Figure 7.81). For example, the amino acid tryptophan is encoded by the sequence UGG on an mRNA. All of the twenty amino acids used to build proteins have, likewise, 3-base sequences that encode them.

Degeneracy

Given that there are 4 bases in RNA, the number of different 3-base combinations that are possible is 43, or 64. There are, however, only 20 amino acids that are used in building proteins in cells. This discrepancy in the number of possible codons and the actual number of amino acids they specify is explained by the fact that the same amino acid may be specified by more than one codon. In fact, with the exception of the amino acids methionine and tryptophan, all the other amino acids are encoded by multiple codons. Codons for the same amino acid are often related, with the first two bases the same and the third being variable. An example would be the codons for alanine: GCU, GCA, GCC and GCG all stand for alanine. This sort of redundancy in the genetic code is termed degeneracy.

Additionally, several 3-letter codons are read as 'start' or 'stop' messages that spur the intiation or the end of the process of translation.

Translating the code

While the ribosomes are literally the factories that join amino acids together using the instructions in mRNAs, another class of RNA molecules, the transfer RNAs (tRNAs) are also needed for translation (Figure 7.83 and Interactive 7.1). Transfer RNAs are small RNA molecules, about 75-90 nucleotides long, that function to 'interpret' the instructions in the mRNA during protein synthesis. Transfer RNAs are extensively modified post-transcriptionally and contain a large number of unusual bases. The sequences of tRNAs have several self complementary regions, where the single-stranded tRNA folds on itself and base-pairs to form what is sometimes described as a clover leaf structure.

This structure is crucial to the function of the tRNA, providing both the sites for attachment of the appropriate amino acid and for recognition of codons in the mRNA. In terms of the bead analogy above, someone or something has to be able to bring a red bead in when the instructions indicate UGG, and a green bead when the instructions say UUU. This, then, is the function of the tRNAs. They must be able to bring the amino acid corresponding to the instructions to the ribosome.

A given transfer RNA is specific for a particular amino acid. Assemblies of charged tRNAs, loaded up with their respective amino acids, await use near the ribosome. The base-pairing of the anticodon on a charged tRNA with the codon on the mRNA is what brings the correct amino acids in to the ribosome to be added on to the growing protein chain (Figure 7.85).

Three steps

Having considered the steps of translation in broader terms, we can now look at them in greater detail. We will consider the three steps of translation (bel0w) individually.

Initiation (binding of the ribosomal subunits to the transcript and initiator tRNA)
Elongation (repeated addition of amino acids to the growing polypeptide, based on the sequence of the mRNA – Figure 7.87)
Termination (release of the completed polypeptide and dissociation of the ribosome into its subunits).

We already know that processed mRNAs are sent from the nucleus to the cytoplasm in eukaryotic cells, while in prokaryotic cells, transcription and translation occur in a single cellular compartment. The small and large subunits of ribosomes, each composed of characteristic rRNAs and proteins, are found in the cytoplasm and assemble on mRNAs to form complete ribosomes that carry out translation. Both prokaryotic and eukaryotic ribosomal subunits are made up of one or more major rRNAs together with a large number of ribosomal proteins. The small subunits of prokaryotic cells are called the 30S ribosomal subunits, while their counterparts in eukaryotes are the 40S subunits. The large ribosomal subunits in prokaryotes are the 50S subunits, while those in eukaryotic cells are 60S. These differences reflect the larger mass of eukaryotic ribosomes. The rRNA components of ribosomes are important for the recognition of the 5' end of the mRNA, and also play a catalytic role in the formation of peptide bonds.

Polypeptide processing

What happens to the newly synthesized polypeptide after it is

released from the ribosome? Functional proteins are not simply strings of amino acids. The polypeptide must fold properly in order to perform its function in the cell. It may also undergo a variety of modifications such as the addition of phosphate groups or sugars, etc. Some proteins are produced as inactive precursors that must be cleaved by proteases to be functional.

Proper folding of a protein into its 3-dimensional conformation is necessary for it to function effectively. As described in an earlier chapter (HERE), the folding of a protein is largely influenced by hydrophobic interactions that result in folding of the protein in such a way as to position hydrophobic residues in the interior, or core, of the protein, away from the aqueous environment of the cell.

Proper folding may also involve the interaction of regions of the polypeptide that are distant from each other, so that portions of the N-terminal region of the polypeptide may be in close proximity to parts of the C-terminus of the final folded molecule.

As a polypeptide emerges from the ribosome, protein chaperones bind to and shield regions of polypeptides and keep them from improperly interacting with one another or with other proteins in the vicinity until they can fold into their correct final shape (Figure 7.98). In addition, other chaperones that are able sequester proteins in such a way as to permit unfolding and refolding of misfolded polypeptides. These proteins ensure that the vast majority of proteins in cells are folded into their correct, functional 3-dimensional shapes.

Sorting and Delivery

An additional challenge in eukaryotic cells is the presence of internal, membrane-bounded compartments. Each compartment contains different proteins with different functions. But the vast majority of proteins in eukaryotic cells are made by ribosomes in the cytoplasm of the cell.

Each of the thousands of proteins made in the cytoplasm must, therefore, be delivered to the appropriate cellular compartment in which it functions. Some proteins are delivered to their destinations in an unfolded state, and are folded within the compartment in which they function. Others are fully folded and may be post-translationally modified before they are sent to their cellular (or extracellular) destinations.

All this sorting and delivering (frequently across membrane barriers) is a complex and amazing process. But the information necessary to guide proteins to their final destinations is built into their structure, and recognized by cellular machinery which guides the process.

Information Processing: Translation

779

780

Figure 7.81 – The standard genetic code

Wikipedia

Figure 7.80 – The central dogma in a bacterial cell

Wikipedia

781

Figure 7.82 – Coding in DNA, transcribed to RNA, translated to protein

782

Figure 7.83 – tRNA – 3D projection (left) and 2D projection (inset)

Wikipedia

783

Interactive 7.1 – Phenylalanyl-tRNA

PDB

Interactive 7.1 – Phenylalanyl-tRNA PDB

Figure 7.84 – Charging of a tRNA by aminoacyl tRNA synthetase

Wikipedia

784

Figure 7.85 – Codons in mRNA pair with anticodons on tRNA to

bring the appropriate amino acid to the ribosome for polypeptide assembly

785

Figure 7.86 – The A, P, and E sites in a ribosome

Image by Martha Baker

Figure 7.87 – Overview of elongation

Wikipedia

786

Movie 7.1 – 30S ribosomal subunit

Wikipedia

Movie 7.1 – 30S ribosomal subunit Wikipedia

787

Figure 7.89 – Conserved sequences adjacent to start codons for various bacterial genes

Image by Martha Baker

Figure 7.88 – Structure of 5S rRNA

788

Figure 7.90 – Base pairing between the Shine-Dalgarno sequence in the mRNA and the 16S rRNA

Image by Martha Baker

Movie 7.2 – Large ribosomal subunit

Wikipedia

Movie 7.2 – Large ribosomal subunit Wikipedia

789

Figure 7.91 – Initiation – assembly of the ribosomal translation complex

Image by Martha Baker

790

Figure 7.92 – Kozak sequence plot showing relative abundance of bases surrounding the AUG (ATG) start codon of human genes

Figure 7.93 – EF-Tu (blue) bound to tRNA (red) and GTP (yellow)

791

Figure 7.94 – The process of elongation

Image by Martha Baker

792

Figure 7.95 – 50S ribosomal subunit. RNA in brown. Protein in blue. Peptidyl transferase site in red.

Wikipedia

793

Figure 7.96 – The process of translation

Wikipedia

Movie 7.3 Translation of a protein secreted into the endoplasmic reticulum. Small subunit in yellow. Large subunit in green. tRNAs in blue.

Wikipedia

Movie 7.3 Translation of a protein secreted into the endoplasmic reticulum. Small subunit in yellow. Large subunit in green. tRNAs in blue. Wikipedia

794

Figure 7.97 – Another perspective of translation. The 3' end of the mRNA is on the left and the ribosome is moving from right to left

795

Figure 7.98 – Action of chaperone to facilitate proper folding of a protein (orange)

Image by Aleia Kim

796

Figure 7.99 – Rough (ribosome bound) and smooth endoplasmic reticulum

Wikipedia

797

Figure 7.100 – N-terminal signal sequence (green) emerging from the ribosome.

798

Figure 7.101 – Translation of a protein into the endoplasmic reticulum

Image by Aleia Kim

799

801

4.4: Signaling

Since different cells take on specialized functions in a multicellular organism, they must be able to coordinate activities. Cells grow, divide, or differentiate in response to specific signals. They may change shape or migrate to another location. At the physiological level, cells in a multicellular organism, must respond to everything from a meal just eaten to injury, threat, or the availability of a mate. They must know when to divide, when to undergo apoptosis (programmed cell death), when to store food, and when to break it down. A variety of mechanisms have arisen to ensure that cell-cell communication is not only possible, but astonishingly swift, accurate and reliable.

How are signals sent between cells? Like pretty much everything that happens in cells, signaling is dependent on molecular recognition. The basic principle of cell-cell signaling is simple. A particular kind of molecule, sent by a signaling cell, is recognized and bound by a receptor protein in (or on the surface of) the target cell. The signal molecules are chemically varied- they may be proteins, short peptides, lipids, nucleotides or the substances called catecholamines, to name a few.

Signal properties

The chemical properties of the signal determine whether its receptors are on the cell surface or intracellular. If the signal is small and hydrophobic it can cross the cell membrane and bind a receptor inside the cell. If, on the other hand, the signal is charged, or very large, it would not be able to diffuse through the plasma membrane. Such signals need receptors on the cell surface, typically

transmembrane proteins that have an extracellular portion that binds the signal and an intracellular part that passes on the message within the cell (Figure 7.130).

Receptors are specific for each type of signal, so each cell has many different kinds of receptors that can recognize and bind the many signals it receives. Because different cells have different sets of receptors, they respond to different signals or combinations of signals. The binding of a signal molecule to a receptor sets off a chain of events in the target cell. These events could cause change in various ways, including, but not limited to, alterations in metabolic pathways or gene expression in the target cell.

How the binding of a signal to a receptor brings about change in cells is the topic of this section. We will examine a few of the major receptor types and the consequences of signal binding to these receptors. Although the specific molecular components of the various signal transduction pathways differ, they all have some features in common (Figure 7.131):

- The binding of a signal to its receptor is usually followed by the generation of a new signal(s) within the cell. The process by which the original signal is converted to a different form and passed on within the cell to bring about change is called signal transduction.
- Most signaling pathways have multiple signal transduction steps by which the signal is relayed through a series of molecular messengers that can amplify and distribute the message to various parts of the cell.
- The last of these messengers usually interacts with a target protein(s) and changes its activity, often by phosphorylation.
- When a signal sets a particular pathway in motion, it is acting like an ON switch. This means that once the desired result has been obtained, the cell must have a mechanism that acts as an OFF switch.

Understanding this underlying similarity is helpful, because learning

the details of the different pathways becomes merely a matter of identifying which molecular component performs a particular function in each individual case.

In the remainder of this section we consider several different signal transduction pathways, each mediated by a different kind of receptor.

Ligand-gated ion channel receptors

The simplest and fastest of signal pathways is seen in the case of signals whose receptors are gated ion channels (Figure 7.132). Gated ion channels are made up of multiple transmembrane proteins that create a pore, or channel, in the cell membrane. Depending upon its type, each ion channel is specific to the passage of a particular ionic species. The term "gated" refers to the fact that the ion channel is controlled by a "gate" which must be opened to allow the ions through. The gates are opened by the binding of an incoming signal (ligand) to the receptor, allowing the almost instantaneous passage of millions of ions from one side of the membrane to the other. Changes in the interior environment of the cell are thus brought about in microseconds and in a single step.

Swift response

This type of swift response is seen, for example, in neuromuscular junctions, where muscle cells respond to a message from the neighboring nerve cell (Figure 7.133). The nerve cell releases a neurotransmitter signal into the synaptic cleft, which is the space between the nerve cell and the muscle cell it is "talking to". An example of such a neurotransmitter signal is acetylcholine. When the acetylcholine molecules are released into the synaptic cleft,

they diffuse rapidly till they reach their receptors on the membrane of the muscle cell. The binding of the acetylcholine to its receptor, an ion channel on the membrane of the muscle cell, causes the gate in the ion channel to open. The resulting ion flow through the channel can immediately change the membrane potential of the cell. This, in turn, can trigger other changes in the cell.

The speed with which changes are brought about in neurotransmitter signaling is evident when you think about how quickly you remove your hand from a hot surface. Sensory neurons carry information to the brain from your hand on the hot surface and motor neurons signal to your muscles to move the hand, in less time than it took you to read this sentence!

Nuclear hormone receptors

The receptors for signals like steroid hormones are part of a large group of proteins known as the nuclear hormone receptor superfamily. These receptors recognize and bind not only steroid hormones, but also retinoic acid, thyroid hormone, vitamin D and other signals. The subset of the nuclear hormone receptors that bind steroid hormones are intracellular proteins. Steroid hormones (Figure 7.135), as you are aware, are related to cholesterol, and as hydrophobic molecules, they are able to cross the cell membrane by themselves. This is unusual, as most signals coming to cells are incapable of crossing the plasma membrane, and thus, must have cell surface receptors.

Once within the cell, steroid hormones bind to their receptors, which may reside in the cytoplasm or in the nucleus. Steroid hormone receptors are proteins with a double life: they are actually dormant transcription regulators that are inactive till a steroid hormone binds and causes a conformational change in them. When this happens, the receptors, with the hormone bound, bind to regulatory sequences in the DNA and modulate gene expression.

Because steroid hormone receptors act by modulating gene expression, the responses to steroid hormones are relatively slow. (There are also some effects of steroid hormones that do not involve transcriptional regulation, but the majority work through changing gene expression.) Like other transcriptional activators, steroid receptors have a DNA-binding domain (DBD) and an activation domain. They also have a ligand-binding domain (LBD) that binds the hormone.

Glucocorticoid receptor

Examples of such signaling pathways are those mediated by the glucocorticoid receptor (Figures 7.136 & 7.137). Glucocorticoids, sometimes described as stress hormones, are made and secreted by the adrenal cortex. Physiologically, they serve to maintain homeostasis in the face of stress and exhibit strong anti-inflammatory and immunosuppressive properties. Because of these effects, synthetic glucocorticoids are used in the treatment of a number of diseases from asthma and rheumatoid arthritis to multiple sclerosis. All of these effects are mediated through the signaling pathway which starts with the binding of a glucocorticoid hormone to its receptor. Recall that steroids can cross the plasma membrane, so glucocorticoids can diffuse into the cell and bind their receptors which are in the cytoplasm.

In the absence of the signal, glucocorticoid receptors are found bound to a protein chaperone, Hsp90 (Figure 7.137). This keeps the receptors from being transported to the nucleus. When a glucocorticoid molecule binds the receptor, the receptor undergoes a conformational change and dissociates from the Hsp90. The receptor, then, with the hormone bound, translocates into the nucleus. In the nucleus, it can increase the transcription of target genes by binding to specific regulatory sequences (labeled HRE for hormone-response elements). The binding of the hormone-

receptor complex to the regulatory elements of hormone-responsive genes modulates their expression. Many of these genes encode anti-inflammatory proteins, and their increased production accounts for the physiological effect of corticosteroid therapies.

The steroid receptor pathways are relatively simple and have only a couple of steps (Figure 7.138). Most other signaling pathways involve multiple steps in which the original signal is passed on and amplified through a number of intermediate steps, before the cell responds to the signal.

Cell surface receptors

We will now take a look at two signaling pathways, each mediated by a major class of cell surface receptor- the G-protein coupled receptors (GPCRs) and the receptor tyrosine kinases (RTKs). While the specific details of the signaling pathways that follow the binding of signals to each of these receptor types are different, it is easier to learn them when you can see what the pathways have in common, namely, interaction of the signal with a receptor, followed by relaying and amplification of the signal through a variable number of intermediate molecules, with the last of these molecules interacting with a target or target proteins and modifying their activity in the cell.

G-protein coupled receptors

G-protein coupled receptors (GPCRs) are involved in responses of cells to many different kinds of signals, from epinephrine, to odors, to light. In fact, a variety of physiological phenomena including

vision, taste, smell, and the fight-or-flight response are mediated by GPCRs. What are G-protein coupled receptors?

G-protein coupled receptors are cell surface receptors that pass on the signals that they receive with the help of guanine nucleotide binding proteins (a.k.a. G-proteins). Before thinking any further about the signaling pathways downstream of GPCRs, it is necessary to know a few important facts about these receptors and the G-proteins that assist them.

Though there are hundreds of different G-protein coupled receptors, they all have the same basic structure (Figure 7.139):

They all consist of a single polypeptide chain that threads back and forth seven times through the lipid bilayer of the plasma membrane. For this reason, they are sometimes called seven-pass transmembrane (7TM) receptors. One end of the polypeptide forms the extracellular domain that binds the signal while the other end is in the cytosol of the cell.

When a ligand (signal) binds the extracellular domain of a GPCR, the receptor undergoes a conformational change, on its cytoplasmic side, that allows it to interact with a G-protein that will then pass the signal on to other intermediates in the signaling pathway.

G-proteins

What is a G-protein? As noted above, a G-protein is a guanine nucleotide-binding protein that can interact with a G-protein linked receptor. G-proteins are associated with the cytosolic side of the plasma membrane, where they are ideally situated to interact with the tail of the GPCR, when a signal binds to the GPCR. There are many different G-proteins, all of which share a characteristic structure- they are composed of three subunits called α, β and γ (Figure 7.140). Because of this, they are sometimes called heterotrimeric G proteins (hetero=different, trimeric= having three parts).

Ligand binding

The guanine nucleotide binding site is on the α subunit of the G-protein. This site can bind GDP or GTP. The α subunit also has a GTPase activity, i.e., it is capable of hydrolyzing a GTP molecule bound to it into GDP.

In the unstimulated state of the cell, that is, in the absence of a signal bound to the GPCR, the G-proteins are found in the trimeric form (α-β-γ bound together) and the α subunit has a GDP molecule bound to it. In this form, the α subunit is inactive. With this background on the structure and general properties of the GPCRs and the G-proteins, we can now look at what happens when a signal arrives at the cell surface and binds to a GPCR (Figure 7.141).

The signaling pathway

The binding of a signal molecule by the extracellular part of the G-protein linked receptor causes the cytosolic tail of the receptor to interact with, and alter the conformation of, a G-protein associated with the inner face of the plasma membrane.
This has two consequences. First, the α subunit of the G-protein loses its GDP and binds a GTP, instead. Second, the G-protein breaks up into the GTP-bound α part and the β-γ part.

The binding of GTP to the α subunit and its dissociation from the β-γ subunits activate the α subunit. The activated α subunit can diffuse freely along the cytosolic face of the plasma membrane and act upon its targets. (The β-γ unit is also capable of activating its own targets.)

What happens when G-proteins interact with their target proteins? That depends on what the target is. G-proteins interact with different kinds of target proteins, of which we will examine two major categories:

Ion channels

We have earlier seen that some gated ion channels can be opened or closed by the direct binding of neurotransmitters to a receptor that is an ion-channel protein. In other cases, ion channels are regulated by the binding of G-proteins. That is, instead of the signal directly binding to the ion channel, it binds to a GPCR, which activates a G-protein that then may cause opening of the ion channel, either directly, by binding to the channel, or indirectly, through activating other proteins that can bind to the channel. The change in the distribution of ions across the plasma membrane causes a change in the membrane potential.

Enzyme activation

The interaction of G-proteins with their target enzymes can regulate the activity of the enzyme, either increasing or decreasing its activity. The change in activity of the target enzyme, in turn, results in downstream changes in other proteins in the cell, and alters the metabolic state of the cell. This is best understood by examining the well-studied response of cells to epinephrine, mediated through the β-adrenergic receptor, a type of G-protein coupled receptor.

Epinephrine (Figure 7.142), also known as adrenaline, is a catecholamine that plays an important role in the body's 'fight or flight' response. In response to stressful stimuli, epinephrine is secreted into the blood, to be carried to target organs whose cells will respond to this signal. If you were walking down a dark alley in an iffy neighborhood, and you heard footsteps behind you, your brain would respond to potential danger by sending signals that ultimately cause the adrenal cortex to secrete epinephrine into the blood stream. The epinephrine circulating in your system has many effects, including increasing your heart rate, but among its prime

targets are your muscle cells. The reason for this is that your muscle cells store energy in the form of glycogen, a polymer of glucose. If you need to run or fight off an assailant, your cells will need energy in the form of glucose.

But how does epinephrine get your cells to break down the glycogen into glucose? Binding of epinephrine to the β-adrenergic receptor on the surface of the cells causes the receptor to activate a G-protein associated with its cytoplasmic tail. As described above, this leads to the α subunit exchanging its GDP for GTP and dissociating from the β-γ subunits. The activated α subunit then interacts with the enzyme adenylate cyclase (also known as adenylyl cyclase) stimulating it to produce cyclic AMP (cAMP) from ATP. Cyclic AMP is often described as a "second messenger", in that it serves to spread the signal received by the cell. How does cAMP accomplish this?

cAMP molecules bind to, and activate an enzyme, protein kinase A (PKA – Figure 7.145). PKA is composed of two catalytic and two regulatory subunits that are bound tightly together. Upon binding of cAMP, the catalytic subunits are released from the regulatory subunits, allowing the enzyme to carry out its function, namely phosphorylating other proteins. Thus, cAMP can regulate the activity of PKA, which in turn, by phosphorylating other proteins can change their activity. In this case, the relevant protein that is activated is an enzyme, phosphorylase kinase. This enzyme can then phosphorylate and activate glycogen phosphorylase, the enzyme ultimately responsible for breaking glycogen down into glucose-1-phosphate – readily converted to glucose. The activation of glycogen phosphorylase supplies the cells with the glucose they need, allowing you to fight or flee, as you might see fit. Simultaneously, PKA also phosphorylates another enzyme, glycogen synthase. In the case of glycogen synthase, phosphorylation inactivates it, and prevents free glucose from being used up for glycogen synthesis, ensuring that your cells are amply supplied with glucose (Figure 7.146).

Common pattern

Although the steps described above seem complicated, they follow the simple pattern outlined at the beginning of this section:

• Binding of signal to receptor

• Several steps where the signal is passed on through intermediate molecules (G-proteins, adenylate cyclase, cAMP, and finally, PKA)

• Phosphorylation of target proteins by the kinase, leading to changes in the cell. The specific changes depend on the proteins that are phosphorylated by the PKA.

Why so many steps? If you need to activate glycogen phosphorylase to break down glucose in a hurry, why not have a system in which binding of a signal to the receptor directly activated the target enzyme?

The answer to this puzzle is simple: there is amplification of the signal at every step of the pathway. A single signal molecule binding to a receptor sets in motion a cascade of reactions, with the signal getting larger at each step, so that binding of one epinephrine molecule to its receptor results in the activation of a million glycogen phosphorylase enzyme molecules!

Turning signals off

If the signal binding to the receptor serves as a switch that sets these events in motion, there must be mechanisms to turn the pathway off. The first is at the level of the receptor itself. A kinase called G-protein receptor kinase (GRK) phosphorylates the cytoplasmic tail of the receptor. The phosphorylated tail is then bound by a protein called arrestin, preventing further interaction with a G-protein.

The next point of control is at the G-protein. Recall that the α subunit of the G-protein is in its free and activated state when it has GTP bound, and that it associates with the β-γ subunits and has a GDP bound when it is inactive. We also know that the α subunit has an activity that enables it to hydrolyze GTP to GDP. This GTP-hydrolyzing activity makes it possible for the α subunit, once it has completed its task, to return to its GDP bound state, re-associate with the β-γ part and become inactive again.

A third "off switch" is further down the signaling pathway, and controls the level of cAMP. We just noted that cAMP levels increase when adenylate cyclase is activated. When its job is done, cAMP is broken down by an enzyme called phosphodiesterase (Figure 7.147). When cAMP levels drop, PKA returns to its inactive state, putting a halt to the changes brought about by the activation of adenylate cyclase by an activated G-protein.

Yet another way that the effects of this pathway can be turned off is at the level of the phosphorylated target proteins. These proteins, which are activated by phosphorylation, can be returned to their inactive state by the removal of the phosphates by phosphatases.

Receptor tyrosine kinases

Another major class of cell surface receptors are the receptor tyrosine kinases or RTKs. Like the GPCRs, receptor tyrosine kinases bind a signal, then pass the message on through a series of intracellular molecules, the last of which acts on target proteins to change the state of the cell.

As the name suggests, a receptor tyrosine kinase is a cell surface receptor that also has a tyrosine kinase activity. The signal binding domain of the receptor tyrosine kinase is on the cell surface, while the tyrosine kinase enzymatic activity resides in the cytoplasmic part of the protein (Figure 7.148). A transmembrane α helix connects these two regions of the receptor.

What happens when signal molecules bind to receptor tyrosine kinases? Binding of signal molecules to the extracellular domains of receptor tyrosine kinase proteins causes two receptor molecules to dimerize (come together and associate – Figure 7.149). This brings the cytoplasmic tails of the receptors close to each other and causes the tyrosine kinase activity of these tails to be turned on. The activated tails then phosphorylate each other on several tyrosine residues (Figure 7.150). This is called autophosphorylation.

The phosphorylation of tyrosines on the receptor tails triggers the assembly of an intracellular signaling complex on the tails. The newly phosphorylated tyrosines serve as binding sites for a variety of signaling proteins that then pass the message on to yet other

proteins to bring about changes in the cell. Receptor tyrosine kinases mediate responses to a large number of signals, including peptide hormones like insulin and growth factors like epidermal growth factor (EGF). We will examine how insulin and EGF act on cells by binding to receptor tyrosine kinases.

Insulin receptor

Insulin plays a central role in the uptake of glucose from the bloodstream. It increases glucose uptake by stimulating the movement of glucose receptor GLUT4 to the plasma membrane of cells.

How does insulin increase GLUT4 concentrations in the cell membrane? The binding of insulin to the insulin receptor (IR – Figure 7.151), results in dimerization of the receptor monomers and subsequent autophosphorylation of the cytosolic kinase domains. The activated tyrosine kinase domains also phosphorylate intracellular proteins called Insulin Receptor Substrates or IRS proteins. These proteins interact with, and activate another kinase called the PI3-kinase. PI3-kinase then catalyzes the formation of the lipid molecule PIP3, which serves to activate yet another kinase, PDK1, which in turn, activates the Akt group of kinases. It is this group of enzymes that appears to increase the translocation of the GLUT4 to the plasma membrane (Figure 7.152), as cells that lack functional Akts exhibit poor glucose uptake and insulin resistance.

EGFR pathway

Epidermal growth factor, EGF, is an important signaling molecule involved in growth, proliferation and differentiation in mammalian cells. EGF acts through the EGF receptor, EGFR, a receptor tyrosine kinase (Figure 7.153). Because of its role in stimulating cell proliferation and because overexpression of EGFR is associated with some kinds of cancers, EGFR is the target for many anti-cancer therapies. We can trace the signal transduction pathway from the binding of EGF to its receptor to the stimulation of cell division.

EGF binding to the EGFR is followed by receptor dimerization and stimulation of the tyrosine kinase activity of the cytosolic domains of the EGFR. Autophosphorylation of the receptor tails is followed

by the assembly of a signaling complex nucleated by the binding of proteins that recognize phosphotyrosine residues. An important protein that is subsequently activated by the signaling complexes on the receptor tyrosine kinases is called Ras (Figure 7.154). The Ras protein is a monomeric guanine nucleotide binding protein that is associated with the cytosolic face of the plasma membrane (in fact, it is a lot like the α subunit of trimeric G-proteins). Just like the α subunit of a G-protein, Ras is active when GTP is bound to it and inactive when GDP is bound to it. Also, like the α subunit, Ras can hydrolyze the GTP to GDP.

Ras activation

Activation of Ras accompanies the exchange of the GDP bound to the inactive Ras for a GTP. Activated Ras triggers a phosphorylation cascade of three protein kinases, which relay and distribute the signal. These protein kinases are members of a group called the MAP kinases (Mitogen Activated Protein Kinases). The final kinase in this cascade phosphorylates various target proteins, including enzymes and transcriptional activators that regulate gene expression.

The phosphorylation of various enzymes can alter their activities, and set off new chemical reactions in the cell, while the phosphorylation of transcriptional activators can change which genes are expressed. The combined effect of changes in gene expression and protein activity alter the cell's physiological state and promote cell division.

Once again, in following the path of signal transduction mediated by RTKs, it is possible to discern the same basic pattern of events: a signal is bound by the extracellular domains of receptor tyrosine kinases, resulting in receptor dimerization and autophosphorylation of the cytosolic tails, thus conveying the message to the interior of the cell.

The message is then passed on via a signaling complex to proteins that stimulate a series of kinases. The terminal kinase in the cascade acts on target proteins and brings about in changes in protein activities.

What is the OFF switch for RTKs? It turns out that RTKs with the

signal bound can be endocytosed into the cell and broken down. That is, the region of the plasma membrane that the RTK is on can be internally pinched off into a vesicle containing the ligand-bound receptor which is then targeted for degradation.

Ras, which is activated by GTP binding, can also be deactivated by hydrolysis of the GTP to GDP. The importance of this mechanism for shutting down the pathway is evident in cells that have a mutant ras gene encoding a Ras protein with defective GTPase activity. Unable to shut off Ras, the cells continue to receive a signal to proliferate. The National Cancer Institute estimates that more than 30% of human cancers are driven by mutations in ras genes.

The descriptions above provide a very simple sketch of some of the major classes of receptors and deal primarily with the mechanistic details of the steps by which signals received by various types of receptors bring about changes in cells. A major take-home lesson is the essential similarity of the different pathways. Another point to keep in mind is that while we have looked at each individual pathway in isolation, a cell, at any given time receives multiple signals that set off a variety of different responses at once (Figure 7.155). The pathways described above show a considerable degree of "cross-talk" and the response to any given signal is affected by the other signals that the cell receives simultaneously. The multitude of different receptors, signals, and the combinations thereof are the means by which cells are able to respond to an enormous variety of different circumstances.

RTKs, cancer and cancer therapies

As described above, binding of EGF to its receptor triggers a signaling pathway that results in the activation of a series of Mitogen Activated Protein Kinases (MAP kinases). These kinases are so-called because they are activated by a mitogen, a molecule, like EGF and other growth factors, that stimulates mitosis or cell division. The final kinase in the MAP kinase cascade phosphorylates a number of target proteins, many of them transcription factors, that when activated, increase the expression of genes associated with cell proliferation.

Given that the EGF-receptor pathway normally functions to stimulate cell division, it is not surprising that malfunctions in the pathway could lead to uncontrolled cell proliferation, or cancer. Next, we will take a brief look at some examples of such defects.

HER2

The human EGF receptor (HER) family has four members, HER1, HER2, HER3 and HER4. These are all receptor tyrosine kinases, cell surface receptors that bind EGF (Figure 7.157) and stimulate cell proliferation.

A crucial step in the signal transduction pathway is the dimerization of the receptors following binding of the signal, EGF, to the receptor. While HER1, HER3 and HER4 must bind the signal to dimerize, the structure of the HER2 receptor can, apparently, allow the receptor monomers to dimerize independently of EGF binding.

This means that the downstream events of the signaling pathway can be triggered even in the absence of a growth signal. In normal cells, only a few HER2 receptors are expressed at the cell surface, so this property of HER2 plays a relatively minor role in stimulating cell division. However, in about a quarter of breast cancer patients, HER2 receptors are overexpressed, leading to increased dimerization and subsequent uncontrolled cell proliferation.

Breast cancers that are HER2-positive can be more aggressive with a greater tendency to metastasize (spread) so therapy that blocks HER2 signaling is key in successful treatment of such cancers. Herceptin, a monoclonal antibody against the HER2 receptor, has been shown to be an effective treatment against Her2-positive breast cancers. Herceptin works by binding specifically to the extracellular domain of the HER2 receptor (Figure 7.158). This prevents dimerization of the receptor and thus blocks downstream signaling. Additionally, the binding of the Herceptin antibody to the receptor signals the immune system to destroy the HER2-positive cells.

Bcr-abl

Another example of a cancer caused by defects in an RTK signaling pathway is chronic myeloid leukemia (CML). Patients with

CML have an abnormal receptor tyrosine kinase that is the product of a hybrid gene called bcr-abl, formed by the breakage and rejoining of chromosomes 9 and 22. This abnormal tyrosine kinase is constitutively dimerized, even when no signal is bound. As a result, it continuously signals cells to divide, leading to the massive proliferation of a type of blood cells called granulocytes.

As with HER2, the problem in CML is a receptor tyrosine kinase that dimerizes in the absence of a growth signal. The approach in this case was to target the next step in the signaling pathway. As you know, dimerization of RTKs activates the tyrosine kinase domain of the receptor, which results in the autophosphorylation of the cytoplasmic domains of both monomers. The phosphorylated tyrosines serve to recruit a number of other signaling proteins that pass the signal on within the cell.

In the case of the bcr-abl RTK, the drug Gleevec (imatinib) was designed to bind near the ATP-binding site of the tyrosine kinase domain. This "locks" the site in a conformation that inhibits the enzymatic activity of the tyrosine kinase and thus blocks downstream signaling. With no "grow" signal passed on, cells stop proliferating.

Information Processing: Signaling

827

828

Figure 7.130 – Schematic representation of a transmembrane receptor protein. E = extracellular; P = plasma membrane; I = intracellular

Wikipedia

Figure 7.129 – Some examples of signal molecules

829

Figure 7.132 – Ligand-gated ion channel receptor opening in response to a signal (ligand)

Wikipedia

Figure 7.131 -General features of signal transduction pathways

830

Figure 7.133 – Neuromuscular signaling – A = motor neuron axon;

B = axon terminal; C = synaptic cleft; D = muscle cell; E = myofibril . Steps in the process – 1) action potential reaches the axon terminal; 2) voltage-dependent calcium gates open; (3) neurotransmitter vesicles fuse with the presynaptic membrane and acetylcholine (ACh) released into the synaptic cleft; (4) ACh binds to postsynaptic receptors on the sarcolemma; (5) ACh binding causes ion channels to open and allows sodium ions to flow across the membrane into the muscle cell; 6) flow of sodium ions across the membrane into the muscle cell generates action potential which travels to the myofibril and results in muscle contraction.

Wikipedia

831

Figure 7.134 – Nerve systems

Wikipedia

832

Figure 7.135 – Steroid hormones structures, with the names of their receptors

Wikipedia

833

Figure 7.136 – Glucocorticoid receptor with its three domains – DNA binding (left), activator domain (top), and ligand binding domain (boxed).

Wikipedia

Figure 7.137 – Glucocorticoid signaling pathway

Wikipedia

834

Figure 7.138 – Steroid hormone signaling

Image by Aleia Kim

835

Figure 7.139 – Structure of a G-protein linked receptor

Wikipedia

836

Figure 7.140- A heterotrimeric G-protein: α subunit in blue, βγ subunits red and green

837

Figure 7.141 – Cycle of G-protein activation – 1) binding of ligand; 2) change of receptor structure; 3) stimulation of α-subunit; 4) binding of GTP, release of GDP; 5) separation of α-subunit from β-γ; 6) hydrolysis of GTP by α-subunit and return to inactive state.

Wikipedia

838

Figure 7.142 – β2-adrenergic receptor embedded in membrane (gray)

Wikipedia

Figure 7.143 – Epinephrine

Wikipedia

839

Figure 7.144 – G-protein coupled receptor. Signal starts with ligand binding (orange circle). Gs = G-protein; AC = adenylate cyclase.

Wikipedia

Figure 7.145 – Activation of Protein Kinase A by cAMP

Image by Martha Baker

840

Figure 7.146 – Simultaneous activation of glycogen breakdown and inhibition of glycogen synthesis by epinephrine's binding of b-adrenergic receptor. Red enzyme names = activated forms; black enzyme names = inactivated forms; GPb = glycogen phosphorylase b; GPa = glycogen phosphorylase a.

Image by Penelope Irving

841

β-Adrenergic Signaling Off Switches

1. GRK Phosphorylates Receptor Tail

Receptor Tail Bound by Arrestin

2. α Subunit G-protein Cleaves GTP to GDP

β-γ subunits Reassociate with α Subunit

3. cAMP Hydrolyzed by Phosphodiesterase

PKA Becomes Inactive

4. Dephosphorylation of Phosphorylated Proteins by Phosphoprotein Phosphatase

β-Adrenergic Signaling On Switches

1. Binding of Signal Molecule to Receptor

2. Passage of Signal Through Several Molecules (G-proteins, Adenylate Cyclase, cAMP, PKA)

3. Phosphorylation of Target Proteins

842

Figure 7.147 – Cyclic AMP is broken down by phosphodiesterase

Figure 7.148 – Structure of a receptor tyrosine kinase

843

Figure 7.149 – Signal binding results in receptor dimerization and activation of tyrosine kinase activity

Figure 7.150 – Activated tyrosine kinases phosphorylate tyrosines on the receptor tails.

Figure 7.151 -The insulin receptor, a receptor tyrosine kinase

Wikipedia

Figure 7.152 – Effects of insulin binding to its receptor tyrosine kinase: 1) insulin binding; 2) activation of protein activation cascades. These include: 3) translocation of Glut-4 transporter to plasma membrane and influx of glucose; 4) glycogen synthesis; 5) glycolysis; and 6) fatty acid synthesis.

Wikipedia

844

845

Figure 7.153 – EGFR signaling beginning at top with binding of EGF, dimerization of receptor, transmission of signal through proteins, activation of kinases, phosphorylation of transcription factors and effects on transcription

Image by Aleia Kim

Figure 7.154 – Ras with GTP bound

Wikipedia

5.1: Energy in Biological Systems

Living organisms are made up of cells, and cells contain an enormous collection of biochemical components. Living cells, though, are not random collections of these molecules. They are extraordinarily organized or "ordered". By contrast, in the nonliving world, there is a universal tendency to increasing disorder. Maintaining and creating the order required for functioning cells takes the input of energy. Without energy, life is not possible.

Oxidative Energy

The primary mechanism used by non-photosynthetic organisms to obtain energy is oxidation chemistry. Reduced carbon in molecules is the most commonly oxidized energy source. The energy released during the oxidative steps is "captured" in the formation of ATP and can be used later for energy-requiring processes.

The more reduced a carbon atom is, the more energy can be realized from its oxidation. Fatty acids are highly reduced, whereas carbohydrates are moderately so. Complete oxidation of both leads to carbon dioxide, which has the lowest potential energy state.

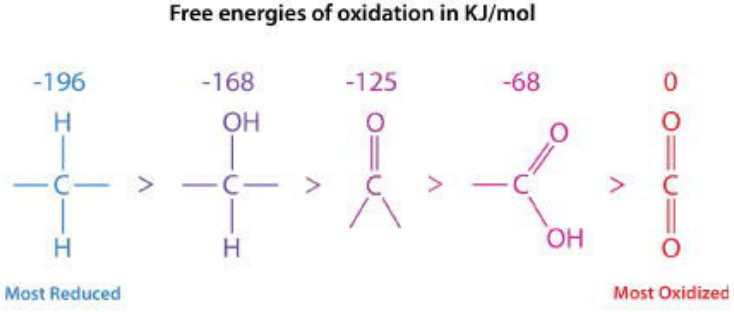


Figure 1: *Five oxidation states of carbon. Image by Pehr Jacobson*

In the series shown in Figure 1, the most reduced form of carbon is on the left. The energy of oxidation of each form is shown above it. Consider the formulas for fatty acids and carbohydrates:

- Palmitic acid: $C_{16}H_{34}O_2$
- Glucose: $C_6H_{12}O_6$

Palmitic acid only contains two oxygens per sixteen carbons, whereas glucose has six oxygen atoms per six carbons. The remaining bonding to carbon in taken up by hydrogens. Consequently, when palmitic acid is fully oxidized, it generates more ATP per carbon (128/16) than glucose (38/6). This is one of two main reasons our bodies use fat (contains fatty acids) as our primary

energy storage material. (The other reason is that carbohydrates are stored with associated water molecules, which adds lots of weight but no extra energy).

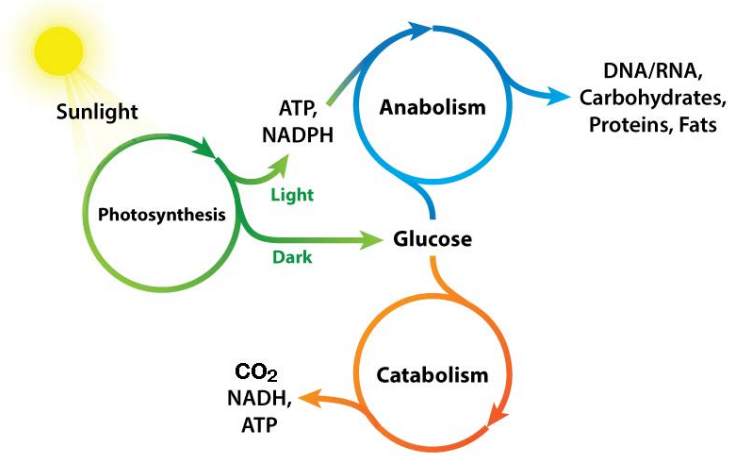


***Figure 2**: Photosynthesis: The primary source of biological energy. Image by Aleia Kim*

Oxidation vs. Reduction in Metabolism

Biochemical processes that break things down from larger to smaller are called *catabolic processes*. Catabolic processes are often oxidative in nature and energy releasing. Some, but not all, of that energy is captured as ATP. If not all of the energy is captured as ATP, what happens to the rest of it? The answer is simple. It is released as heat and it is for this reason our bodies generate warmth, especially when we exert ourselves.

By contrast, synthesizing large molecules from smaller ones (for example, making proteins from amino acids) is referred to as

anabolism. Anabolic processes often involve the reduction of carbon or other atoms (Figures 5.3 & 5.4) and require energy input. By themselves, they would not occur, as they are reversing oxidation and decreasing entropy (making many small things into a larger and more ordered one). To overcome this energy barrier, cells must expend energy. For example, if one wishes to reduce CO_2 to carbohydrate, energy must be used to do so. Plants do this during the dark reactions of photosynthesis (Figure 3). The energy source for the reduction is ultimately the sun. The electrons for the reduction come from water, and the CO_2 is removed from the atmosphere and gets incorporated into a sugar.

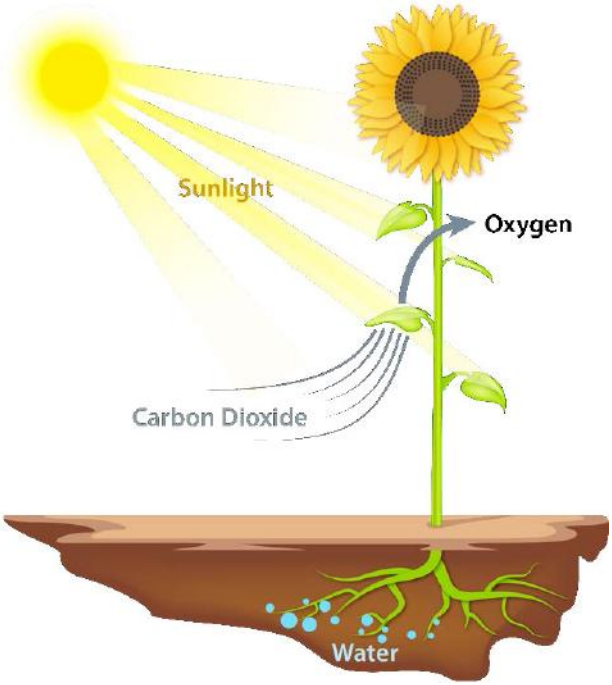


Figure 3: Movement of biological energy. Image by Aleia Kim

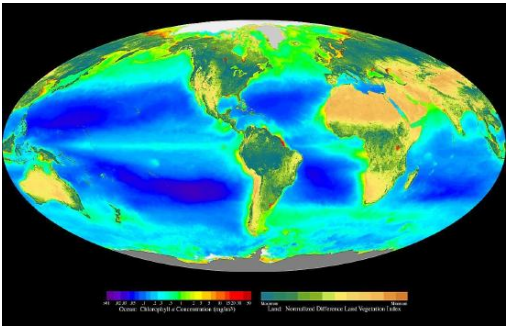


Figure 4: Photosynthesis as measured by chlorophyll concentration

Energy Coupling

The synthesis of the many molecules in a functioning cell creates a need for energy in the cell. Cells overcome this energy obstacle by using ATP to "drive" energy-requiring reactions (Figure 6). The energy needed to drive reactions is harvested in very controlled conditions in enzymes. This involves a process called 'coupling'.

Coupled reactions rely on linking an energetically favorable reaction (i.e., one with a negative $\Delta G°'$) with the reaction requiring an energy input. As long as the overall energy change of the two reactions combined is negative, the reaction will proceed. Hydrolysis of ATP is an energetically favorable reaction that is commonly linked to many energy requiring reactions in cells. Without the hydrolysis of ATP (or GTP, in some cases), these reactions would not occur.

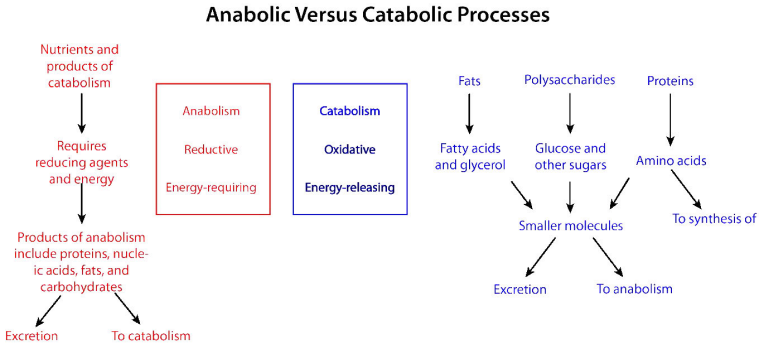


Figure 5: *Synthesis and breakdown pathways in metabolism. Image by Pehr Jacobson*

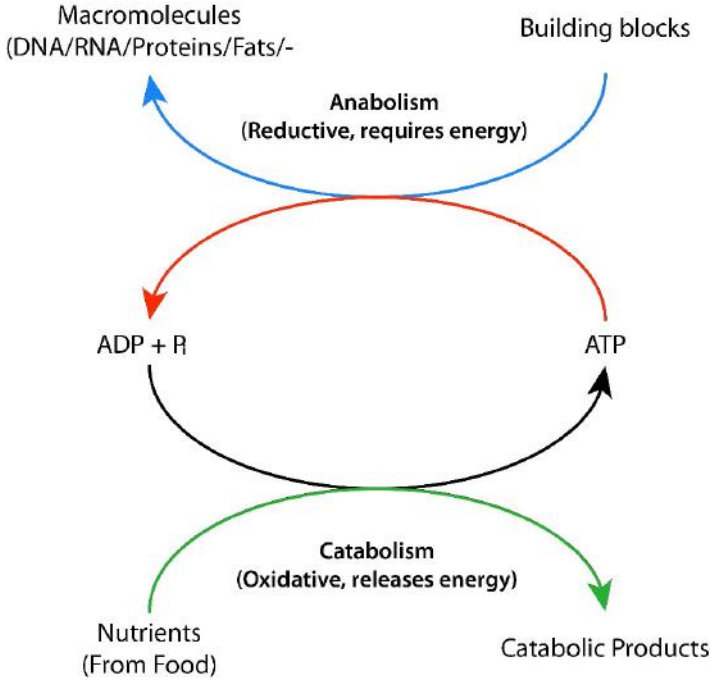


Figure 6: *Cycling of biological energy via ADP and ATP. Image by Pehr Jacobson*

Entropy and energy

Cells are very organized or ordered structures, leading some to mistakenly conclude that life somehow violates the laws of physics. In fact, that notion is incorrect. The second law of thermodynamics doesn't say that entropy always increases, just that, left alone, it tends to do so, in an isolated system. Cells are not isolated systems, though, in that they obtain energy, either from the sun, if they are autotrophic, or food, if they are heterotrophic.

To counter the universal tendency towards disorder on a local scale requires energy. As an example, take a fresh deck of cards which is neatly aligned with Ace-King-Queen . . . . 4,3,2 for each suit. Throw the deck into the air, letting the cards scatter. When you pick them up, they will be more disordered than when they started. However, if you spend a few minutes (and expend a bit of energy), you can reorganize the same deck back to its previous, organized state. If entropy always increased everywhere, you could not do this. However, with the input of energy, you overcame the disorder. This illustrates an important concept: the cost of fighting disorder is energy.

Biological energy

There are, of course, many reasons that organisms need energy. Muscular contraction, synthesis of molecules, neurotransmission, signaling, thermoregulation, and subcellular movements are all energy-requiring processes. Where does this energy come from? The currencies of energy are generally phosphate-containing molecules. ATP is the best known and most abundant of these, but GTP is also an important energy source (energy source for protein synthesis). CTP is involved in synthesis of glycerophospholipids and

UTP is used for synthesis of glycogen and other sugar compounds. In each of these cases, the energy is in the form of potential chemical energy stored in the multi-phosphate bonds of a nucleotide triphosphate. Hydrolyzing those bonds releases the energy in them.

Of the triphosphates, ATP is the primary energy source, acting to facilitate the synthesis of the others by action of the enzyme NDPK. ATP is produced in cells from its precursor ADP (adenosine diphosphate) by three distinct types of phosphorylation –

- oxidative phosphorylation (in mitochondria),
- photophosphorylation (in chloroplasts of plants), and
- substrate level phosphorylation (in enzymatically catalyzed reactions).

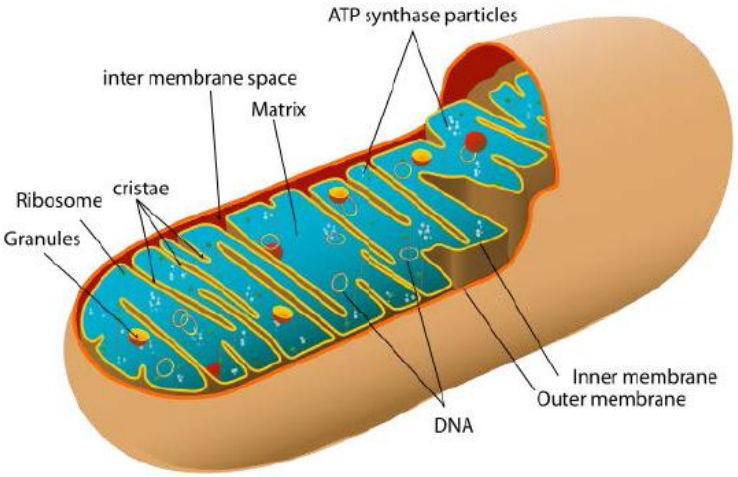


Figure 7: *Mitochondrion*

Chemical and electrical potential

Impermeable to most ions and polar compounds, biological

membranes are essential for processes that generate cellular energy. Consider Figure 5.8. A lipid bilayer separates two solutions with different concentrations of a solute. There is a greater concentration of negative ions in the bottom and a greater concentration of positive ions on the top.

Whenever there is a difference in concentration of molecules across a membrane, there is said to be a concentration gradient across it. A difference in concentration of ions across a membrane also creates a charge (or electrical) gradient. Because there is a difference in both the chemical concentration of the ions and in the charge on the two sides of the membrane, this is described as an electrochemical gradient (Figures 5.8 -5.10).

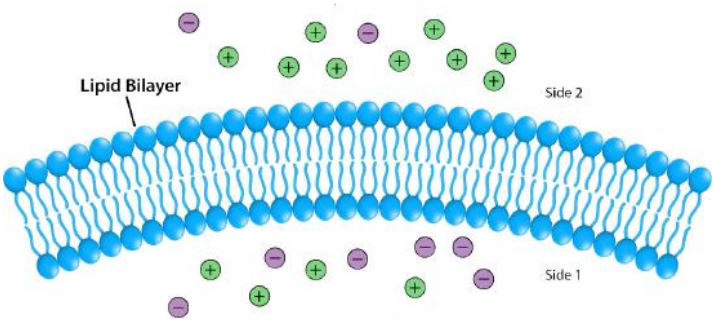


Figure 5.8: Differences in ion concentration across a membrane give rise to chemical and electrical gradients. Image by Pehr Jacobson

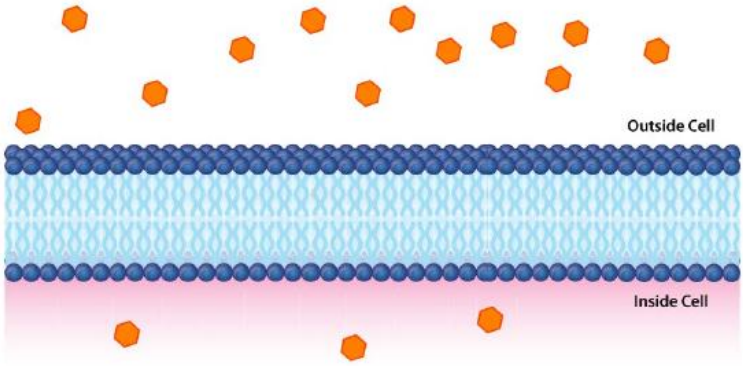


Figure 9: *A chemical gradient. Image by Aleia Kim*

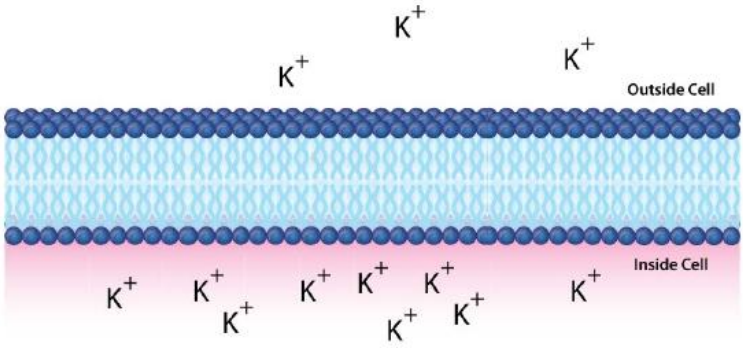


Figure 10*: An electrochemical gradient of potassium ions. Image by Aleia Kim*

Potential energy

Such gradients function like batteries and contain potential energy. When the potential energy is harvested by cells, they can create ATP, transmit nerve signals, pump molecules across membranes, and more. That release of the energy comes when the gradient is somehow relieved and the concentrations equalize.

Reduction Potential

In discussing chemical potential, we must also consider reduction potential. Reduction potential measures the tendency of a chemical to be reduced by electrons. It is also called several other names/variables: including redox potential, oxidation/reduction potential, or ORP.

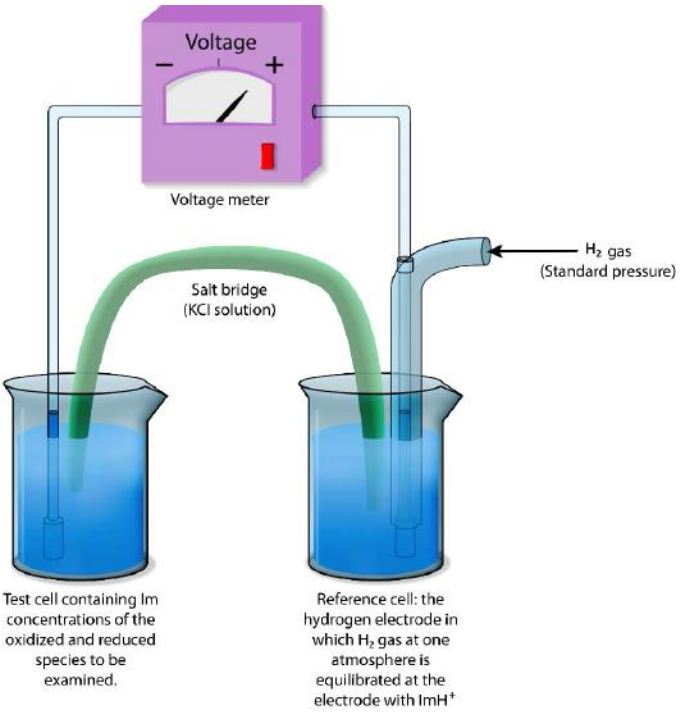
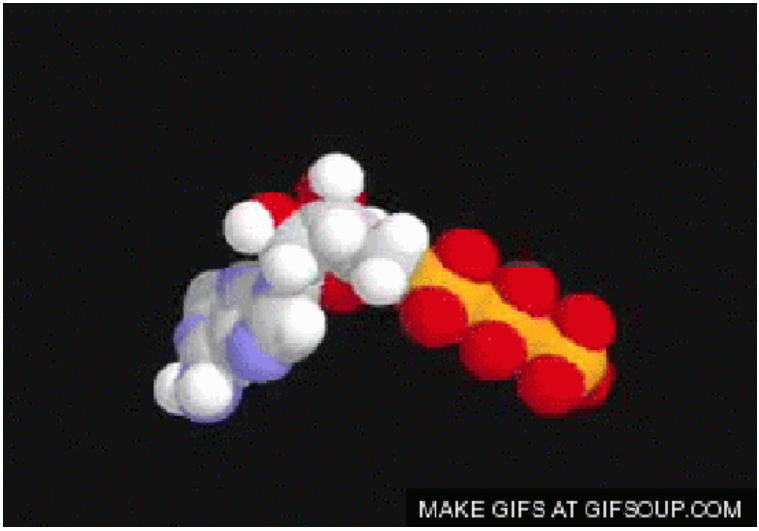


Figure 11: Reference electrode for measuring reduction potentials. Image by Pehr Jacobson

Reduction potential is measured in volts, or millivolts. A substance with a higher reduction potential will have a greater tendency to accept electrons and be reduced. Sometimes it is useful to use the lens of reduction potential to understand the energy changes associated with biochemical reactions, and the driving forces behind them.

Energy Storage in Triphosphates



Movie 5.1: ATP: *The fuel of the cell*

Formation of triphosphates, like ATP, is essential to meeting the cell's energy needs for synthesis, motion, and signaling. In a given day, an average human body makes and breaks down more than its weight in triphosphates. This is especially remarkable considering that there is only about 250 g of the molecule present in the body at any given time. Energy in ATP is released by hydrolysis of a phosphate from the molecule.

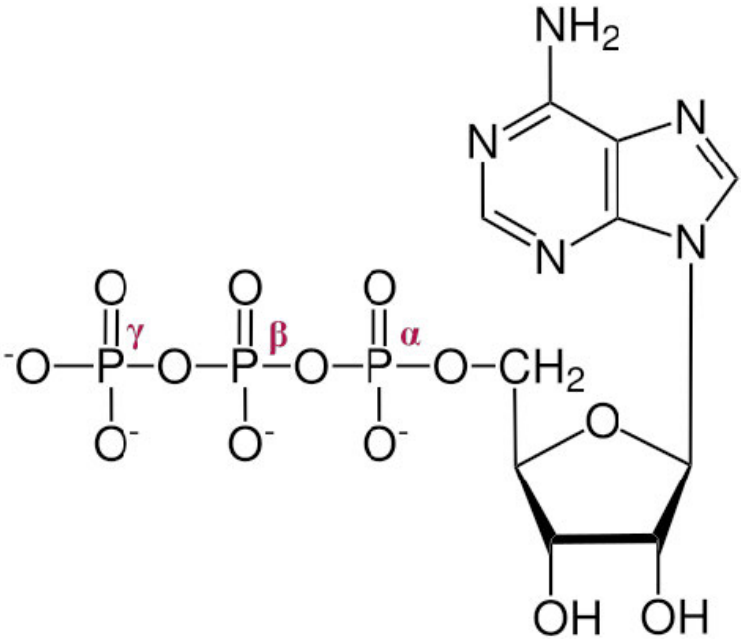


Figure 12 ATP showing α, β and γ phosphates

The three phosphates, starting with the one closest to the sugar are referred to as α, β, and γ (Figure 12). It is the γ phosphate that is cleaved in hydrolysis and the product is ADP. This hydrolysis is an energy-releasing process.

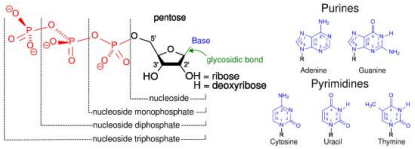


Figure 13: Nucleotides, nucleosides, and bases

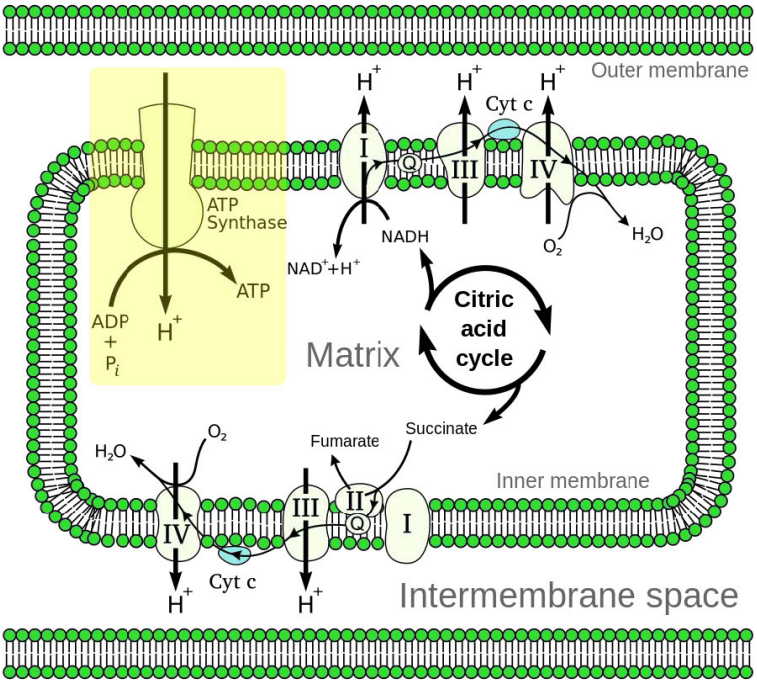
5.2: Electron Transport and Oxidative Phosphorylation

In eukaryotic cells, the vast majority of ATP synthesis occurs in the mitochondria in a process called oxidative phosphorylation. Even plants, which generate ATP by photophosphorylation in chloroplasts, contain mitochondria for the synthesis of ATP through oxidative phosphorylation.

Oxidative phosphorylation is linked to a process known as electron transport (Figure 5.14). The electron transport system (also called the Electron Transport Chain, or ETC), located in the inner mitochondrial membrane, transfers electrons donated by the reduced molecules NADH and FADH2 through a series of electrons acceptors, to oxygen. As we shall see, movement of electrons through system essentially sets up a gradient across a membrane that is then used to make ATP.

The NADH and FADH2 that acts as the fuel for this process comes

from other parts of metabolism for energy. The end products are ATP and water.



Chemiosmotic model

Dr. Peter Mitchell introduced a radical proposal in 1961 to explain the mechanism by which mitochondria make ATP. It is known as the chemiosmotic hypothesis and has been shown over the years to be correct.

Mitchell proposed that synthesis of ATP in mitochondria depends on an electrochemical gradient, across the mitochondrial inner membrane, that arises ultimately from the energy of reduced electron carriers, NADH and FADH2.

There are three general processes that work together to accomplish this feat:

- **Electrons are passed through the ETC:** A pH gradient (also described as a proton gradient or hydrogen ion gradient) is created when electrons from NADH and FADH2 are passed through the electron transport chain (ETC) located in the inner mitochondrial membrane.
- **Protons are pumped across a membrane, setting up a gradient:** Movement of electrons through the series of of electron carriers is coupled to the pumping of protons out of the mitochondrial matrix across the inner mitochondrial membrane.
- **ATP Synthase catalyzes phosphorylation of ADP:** protons re-enter the mitochondrial matrix via the transmembrane ATP synthase protein complex, which combines ADP with inorganic phosphate to make ATP.

Tight coupling

Under normal metabolic conditions, tight coupling is said to exist between electron transport and the synthesis of ATP (called oxidative phosphorylation). Chemicals which permeabilize the inner mitochondrial membrane disallow the gradient to form, causing uncoupling of the ETC to ATP synthesis. That is, they allow the protons to leak back into the mitochondrial matrix, rather than through the ATP synthase, so that the movement of electrons through the ETS can no longer support the synthesis of ATP.

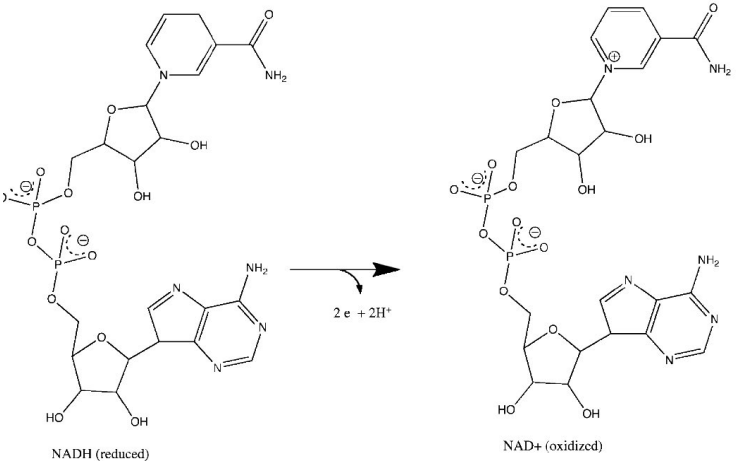
Such uncoupling occurs in situations such as:

- Brown fat in animals uses uncoupling to generate heat but not ATP when heat generation is a priority, such as during hibernation or in infants unable to maintain their temperature alone.
- Some plants use uncoupling to heat their tissues and volatilize

odor chemicals that attract pollinators.
- Some poisons, including insecticides, cause uncoupling and starve an organism from the production of ATP, leading to death.

Power plants

Mitochondria are called the power plants of the cell because most of a cell's ATP is produced there via oxidative phosphorylation. The mechanism by which ATP is made in oxidative phosphorylation is one of the most interesting in all of biology.

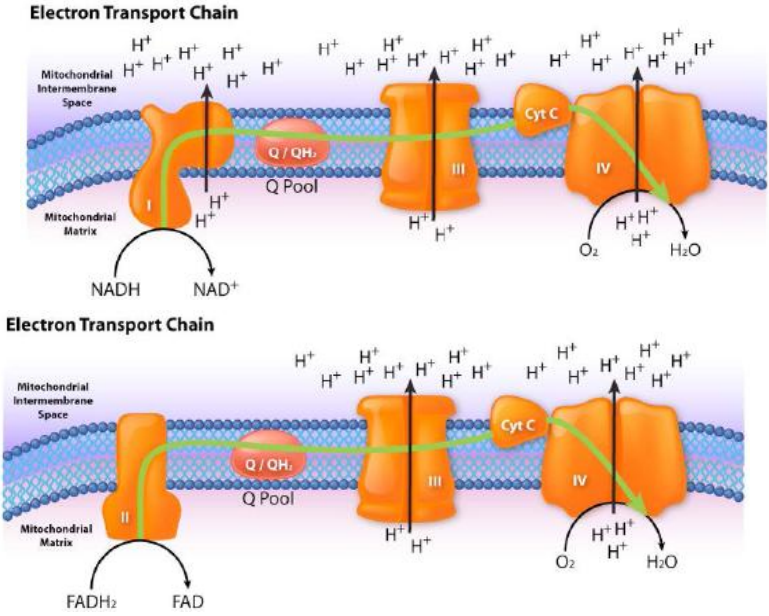


NADH (reduced) NAD+ (oxidized)

Considerations

The process has three primary considerations. The first is electrical

– electrons from reduced electron carriers, such as NADH and FADH2, enter the electron transport system via Complex I and II, respectively. As seen in Figure 5.16 and Figure 5.17, electrons move from one complex to the next, not unlike the way they move through an electrical circuit. Such movement occurs a a result of a set of reduction-oxidation (redox) reactions with electrons moving from a more negative reduction potential to a more positive one.



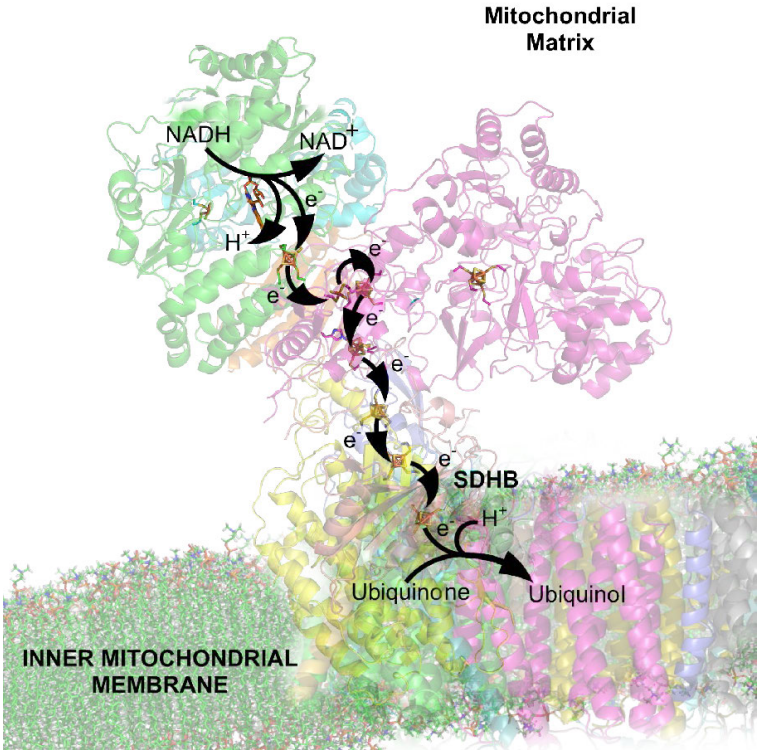
One can think of this occurring as a process where carriers "take" electrons away from complexes with lower reduction potential, much the way a bully takes lunch money from a smaller child. In this scheme, the biggest "bully" is oxygen in Complex IV. Electrons gained by a carrier cause it to be reduced, whereas the carrier giving up the electrons is oxidized.

Entry of electrons to system

Movement of electrons through the chain begins either by 1) transfer from NADH to Complex I (Figure 5.16) or 2) movement of electrons through a covalently bound FADH2 (Figure 5.17) in the membrane-bound succinate dehydrogenase (Complex II). (An alternate entry point for electrons from FADH2 is the Electron Transferring Flavoprotein via the electron-transferring-flavoprotein dehydrogenase, not shown).

Traffic cop

Both Complex I and II pass electrons to the inner membrane's coenzyme Q (CoQ – Figures 5.18 & 5.19). In each case, coenzyme Q accepts electrons in pairs and passes them off to Complex III (CoQH2-cytochrome c reductase) singly. Coenzyme Q thus acts as a traffic cop, regulating the flow of electrons through the ETS.



Docking station

Complex III is a docking station or interchange for the incoming electron carrier (coenzyme Q) and the outgoing carrier (cytochrome c). Movement of electrons from Coenzyme Q to Complex III and then to cytochrome C occurs as a result of what is referred to as the Q-cycle (see below).

Complex III acts to ferry electrons from CoQ to cytochrome c. Cytochrome c takes one electron from Complex III and passes it to Complex IV (cytochrome oxidase). Complex IV is the final protein

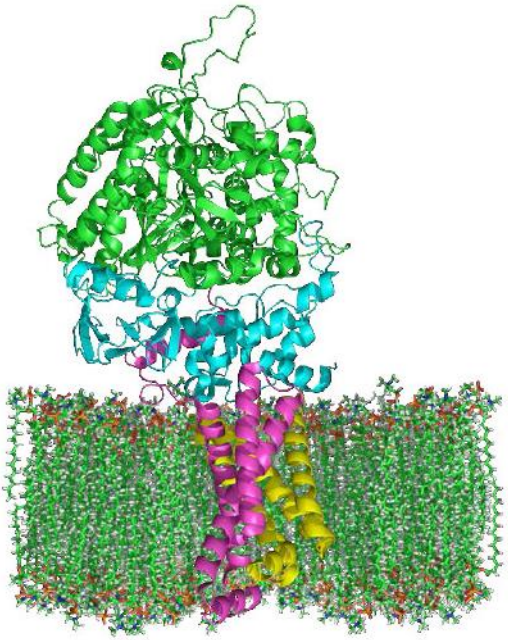
recipient of the electrons. It passes them to molecular oxygen (O2) to make two molecules of water. Making two water molecules requires four electrons, so Complex IV must accept, handle, and pass to molecular oxygen four separate electrons, causing the oxidation state of oxygen to be sequentially changed with addition of each electron.

Proton pumping

As electrons pass through complexes I, III, and IV, there is a release of a small amount of energy at each step, which is used to pump protons from the mitochondrial matrix (inside of mitochondrion) and deposit them in the intermembrane space (between the inner and outer membranes of the mitochondrion). The effect of this redistribution is to increase the electrical and chemical potential across the membrane.

Potential energy

As discussed earlier, electrochemical gradients have potential energy. Students may think of the process as "charging the battery." Just like a charged battery, the potential arising from the proton differential across the membrane can be used to do things. In the mitochondrion, what the proton gradient does is facilitate the production of ATP from ADP and Pi. This process is known as oxidative phosphorylation, because the phosphorylation of ADP to ATP is dependent on the chemical oxidations occurring in the mitochondria.

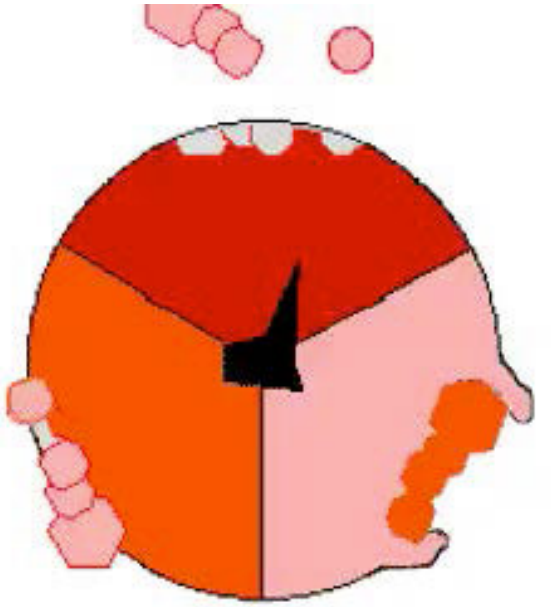


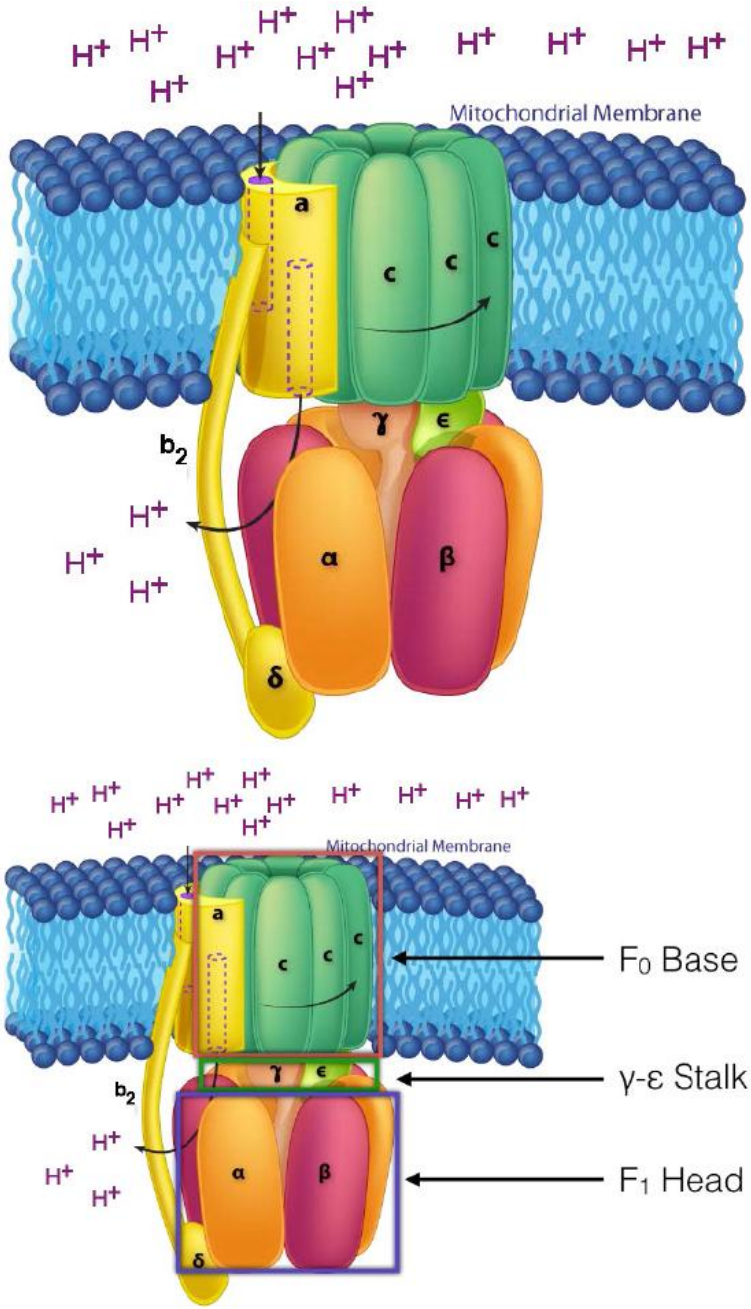
ATP synthase

The protein complex harvesting energy from the proton gradient and using it to make ATP from ADP is an enzyme that has several names – Complex V, PTAS (Proton Translocating ATP Synthase), and ATP synthase (Figure 5.29). Central to its function is the movement of protons through it (from the intermembrane space back into the matrix). Protons will only provide energy to make ATP if their concentration is greater in the intermembrane space than in the matrix and if ADP is available.

In summary, the electron transport system charges the battery for oxidative phosphorylation by pumping protons out of the mitochondrion. The intact inner membrane of the mitochondrion keeps the protons out, except for those that re-enter through ATP Synthase. The ATP Synthase allows protons to re-enter the mitochondrial matrix and harvests their energy to make ATP.

ATP synthase mechanism




H+
Mitochondrial Membrane
a
c
c
c
γ
ϵ
b2
α
β
δ
Mitochondrial Membrane
F0 Base
γ-ε Stalk
F1 Head

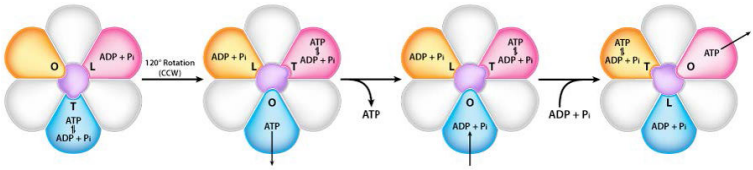
The ATP Synthase itself is an amazing nanomachine that makes ATP using a gradient of protons flowing through it from the intermembrane space back into the matrix. It is not easy to depict in a single image what the synthase does. Figure 5.31 and Figure 5.32 illustrate the multi-subunit nature of this membrane protein, which acts like a turbine at a hydroelectric dam. The movement of protons through the ATP Synthase c-ring causes it and the γ-ε stalk attached to it to turn. It is this action that is necessary for making ATP.

In ATP Synthase, the spinning components, or rotor, are the membrane portion (c ring) of the F0 base and the γ-ε stalk, which is connected to it. The γ-ε stalk projects into the F1 head of the mushroom structure. The F1 head contains the catalytic ability to make ATP. The F1 head is hexameric in structure with paired α and β proteins arranged in a trimer of dimers. ATP synthesis occurs within the β subunits.

Rotation of γ unit

Turning of the γ shaft (caused by proton flow) inside the α-β trimer of the F1 head causes each set of β proteins to change structure slightly into three different forms called Loose, Tight, and Open (L,T,O – Figure 5.31). Each of these forms has a function.

The Loose form binds ADP + Pi. The Tight form "squeezes" them together to form the ATP. The Open form releases the ATP into the mitochondrial matrix. Thus, as a result of the proton flow through the ATP synthase, from the intermembrane space into the matrix, ATP is made from ADP and Pi.



Energy efficiency

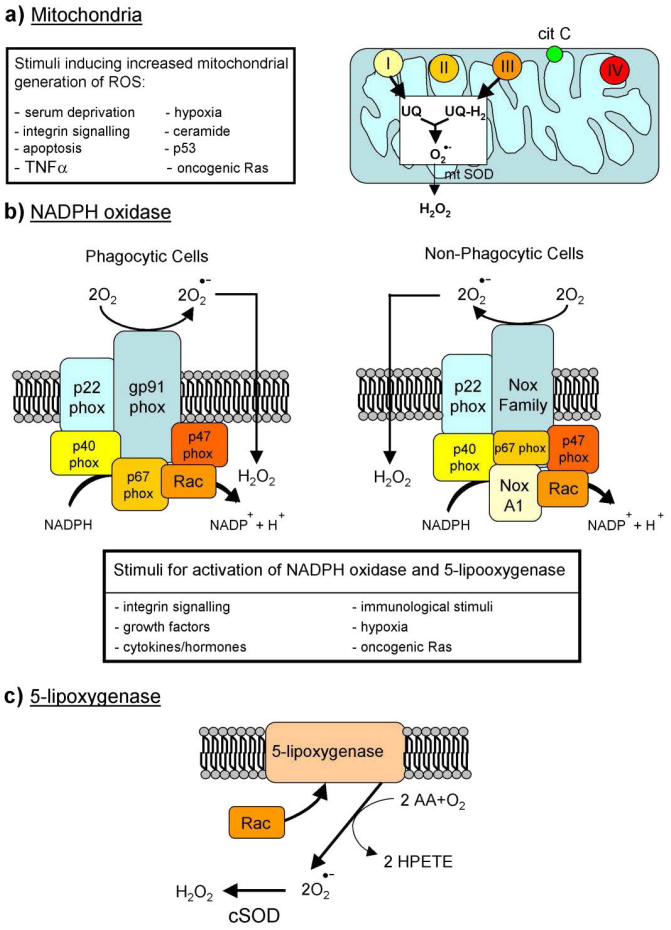
Cells are not 100% efficient in energy use. Nothing we know is.

Consequently, cells do not get as much energy out of catabolic processes as they put into anabolic processes. A good example is the synthesis and breakdown of glucose, something liver cells are frequently doing. The complete conversion of glucose to pyruvate in glycolysis (catabolism) yields two pyruvates plus 2 NADH plus 2 ATPs. Conversely, the complete conversion of two pyruvates into glucose by gluconeogenesis (anabolism) requires 4 ATPs, 2 NADH, and 2 GTPs. Since the energy of GTP is essentially equal to that of ATP, gluconeogenesis requires a net of 4 ATPs more than glycolysis yields. This difference must be made up in order for the organism to meet its energy needs. It is for this reason that we eat. In addition, the inefficiency of our capture of energy in reactions results in the production of heat and helps to keep us warm.

Metabolic controls of energy

It is also noteworthy that cells do not usually have both catabolic and anabolic processes for the same molecules occurring simultaneously inside of them (for example, breakdown of glucose and synthesis of glucose) because the cell would see no net production of anything but heat and a loss of ATPs with each turn of the cycle. Such cycles are called futile cycles and cells have controls in place to limit the extent to which they occur. Since futile cycles can, in fact, yield heat, they are used as sources of heat in some types of tissue. Brown adipose tissue of mammals uses this strategy, as described earlier. See also HERE for more on heat generation with a futile cycle.

Reactive oxygen species



Reactive oxygen species (ROS – Figure 5.37) are oxygen containing molecules, such as peroxides, hydroxyl radical, superoxide, peroxynitrite, and others that are very chemically reactive. Though some ROS, such as peroxide and nitric oxide have important biological functions in signaling, increases in reactive oxygen

species in times of stress can cause significant damage in the cell. Exogenous sources of ROS, such as pollution, tobacco, smoke, radiation or drugs can also cause significant problems.

Endogenous production of ROS is directed towards intracellular signaling (H2O2 and nitric oxide, for example) and defense. Many cells, for example, have NADPH oxidase (Figure 5.38) embedded in the exterior portion of the plasma membranes, in peroxisomes, and endoplasmic reticulum. It produces superoxides in the reaction below to kill bacteria .

In the immune system, cells called phagocytes engulf foreign cells and then use ROS to kill them. ROS can serve as signals for action. In zebrafish, damaged tissues have increased levels of H2O2 and this is thought to be a signal for white blood cells to converge on the site. In fish lacking the genes to produce hydrogen peroxide, white blood cells do not converge at the damage site. Sources of hydrogen peroxide include peroxisomes, which generate it as a byproduct of oxidation of long chain fatty acids.

Aging

Reactive oxygen species are at the heart of the free radical theory of aging, which states that organisms age due to the accumulation of damage from free radicals in their cells. In yeast and Drosophila, there is evidence that reducing oxidative damage can increase lifespan. In mice, increasing oxidative damage decreases life span, though in Caenorhabditis, blocking production of superoxide dismutase actually increases lifespan, so the role of ROS in aging is not completely clear.

It is clear, though, that accumulation of mitochondrial damage is problematic for individual cells. Bcl-2 proteins on the surface of mitochondria monitor damage and if they detect it, will activate

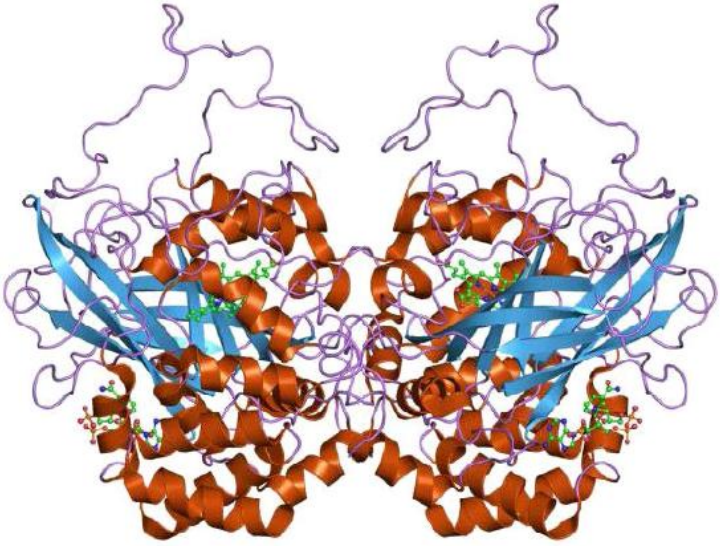
proteins called Bax to stimulate the release of cytochrome c from the mitochondrial membrane, stimulating apoptosis (programmed cell death). Eventually the dead cell will be phagocytosed.

A common endogenous source of superoxide is the electron transport chain. Superoxide can be produced when movement of electrons into and out of the chain don't match well. Under these circumstances, semi-reduced CoQ can donate an electron to O2 to form superoxide (O2-). Superoxide can react with many molecules, including DNA where it can cause damage leading to mutation. If it reacts with the aconitase enzyme, ferrous iron (Fe++) can be released which, in turn, can react in the Fenton reaction to produce another reactive oxygen species, the hydroxyl radical (Figure 5.39) .

Countering the effects of ROS are enzymes, such as catalase, superoxide dismutase, and anti-oxidants, such as glutathione and vitamins C and E.

Glutathione protects against oxidative damage by being a substrate for the enzyme glutathione peroxidase. Glutathione peroxidase catalyzes the conversion of hydrogen peroxide to water (next page).

Catalase



Catalase (Figure 5.40) is an important enzyme for cells of all types that live in an oxygen environment. A first line of defense against reactive oxygen species, catalase catalyzes the breakdown of hydrogen peroxide into water and oxygen.

2 H2O2 <=> 2 H2O + O2

The enzyme, which employs four heme groups in its catalysis, works extremely rapidly, converting up to 40,000,000 molecules of hydrogen peroxide to water and oxygen per enzyme per second. It is abundantly found in peroxisomes.

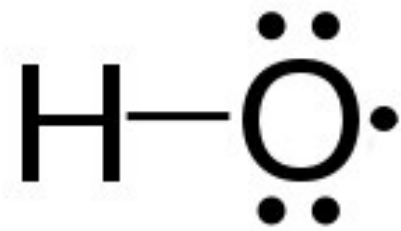
In addition to catalase's ability to break down hydrogen peroxide, the enzyme can also use hydrogen peroxide to oxidize a wide variety organic compounds, including phenols, formic acid, formaldehyde, acetaldehyde, and alcohols, but with much lower efficiency.

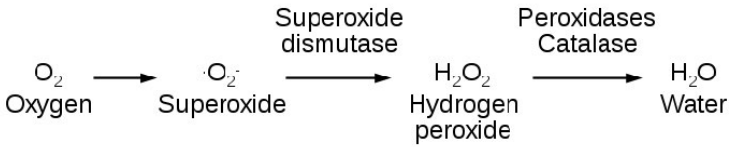
Health

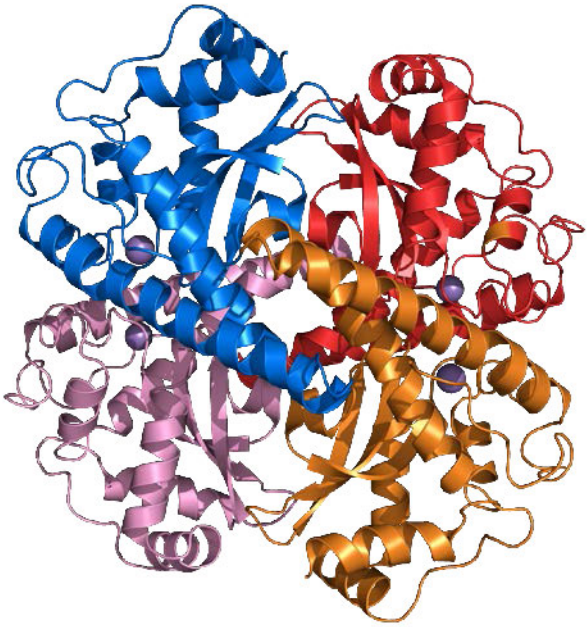
The importance of catalase for health is uncertain. Mice deficient in the enzyme appear healthy and humans with low levels of the enzyme display few problems. On the other hand, mice engineered

to produce higher levels of catalase, in at least one study, lived longer. The ability of organisms to live with lower levels or no catalase may arise from another group of enzymes, the peroxiredoxins, which also act on hydrogen peroxide and may make up for lower quantities of catalase. Last, there is evidence that reduced levels of catalase with aging may be responsible for the graying of hair. Higher levels of H2O2 with reduced catalase results in a bleaching of hair follicles.

Superoxide dismutase



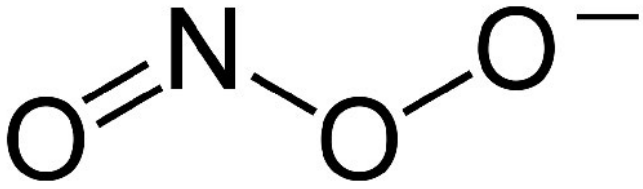




Another important enzyme for protection against reactive oxygen species is superoxide dismutase (SOD), which is found, like catalase, in virtually all organisms living in an oxygen environment. Superoxide dismutase, also like catalase, has a very high Kcat value and, in fact, has the highest Kcat/Km known for any known enzyme. It catalyzes the reactions at the top of the next column (superoxides shown in red):

The enzyme thus works by a ping-pong (double displacement) mechanism (see HERE), being converted between two different forms.

The hydrogen peroxide produced in the second reaction is easily handled by catalase and is also less harmful than superoxide, which can react with nitric oxide (NO) to form very toxic peroxynitrite ions (Figure 5.43). Peroxynitrite has negative effects on cells, as shown in Figure 5.45.



In addition to copper, an ion of Zn++ is also bound by the enzyme and likely plays a role in the catalysis. Forms of superoxide dismutase that use manganese, nickel, or iron are also known and are mostly found in prokaryotes and protists, though a manganese SOD is found in most mitochondria. Copper/zinc enzymes are common in eukaryotes.

Three forms of superoxide dismutase are found in humans and localized to the cytoplasm (SOD1 – Figure 5.45), mitochondria (SOD2 – Figure 5.46), and extracellular areas (SOD3 – Figure 5.47). Mice lacking any of the three forms of the enzyme are more sensitive to drugs, such as paraquat. Deficiency of SOD1 in mice leads to hepatocellular carcinoma and early loss of muscle tissue related to aging. Drosophila lacking SOD2 die before birth and those lacking SOD1 prematurely age.

In humans, superoxide dismutase mutations are associated with the genetically-linked form of Amyotrophic Lateral Sclerosis (ALS) and over-expression of the gene is linked to neural disorders associated with Down syndrome.

Mixed function oxidases

Other enzymes catalyzing reactions involving oxygen include the mixed function oxidases. These enzymes use molecular oxygen for two different purposes in one reaction. The mixed function part of the name is used to indicate reactions in which two different

substrates are being oxidized simultaneously. Monooxygenases are examples of mixed function oxidases. An example of a mixed function oxidase reaction is shown below.

AH + BH2 + O2 <=> AOH + B + H2O

In this case, the oxygen molecule has one atom serve as an electron acceptor and the other atom is added to the AH, creating an alcohol.

Cytochrome P450 enzymes

Cytochrome P450 enzymes (called CYPs) are family of heme-containing mixed function oxidase enzymes found in all domains of life. Over 21,000 CYP enzymes are known. The most characteristic reaction catalyzed by these enzymes follows

Monooxygenase reactions such as this are relatively rare in the cell due to their use of molecular oxygen. CYPs require an electron donor for reactions like the one shown here and frequently require a protein to assist in transferring electrons to reduce the heme iron. There are six different classes of P450 enzymes based on how they get electrons

1. Bacterial P450 – electrons from ferredoxin reductase and ferredoxin
2. Mitochondrial P450 – electrons from adrenodoxin reductase and adrenodoxin
3. CYB5R/cyb5 – electrons come from cytochrome b5
4. FMN/Fd – use a fused FMN reductase
5. Microsomal P450 – NADPH electrons come via cytochrome P450 reductase or from cytochrome b5 and cytochrome b5 reductase
6. P450 only systems – do not require external reducing power

The CYP genes are abundant in humans and catalyze thousand

of reactions on both cellular and extracellular chemicals. There are 57 human genes categorized into 18 different families of enzymes. Some CYPs are specific for one or a few substrates, but others can act on many different substrates.

CYP enzymes are found in most body tissues and perform important functions in synthesis of steroids (cholesterol, estrogen, testoterone, Vitamin D, e.g.), breakdown of endogenous compounds (bilirubin), and in detoxification of toxic compounds including drugs. Because they act on many drugs, changes in CYP activity can produce unexpected results and cause problems with drug interactions.

Bioactive compounds, for example, in grapefruit juice, can inhibit CYP3A4 activity, leading to increased circulating concentrations of drugs that would normally have been acted upon by CYP3A4. This is the reason that patients prescribed drugs that are known to be CYP3A4 substrates are advised to avoid drinking grapefruit juice while under treatment. St. Johns Wort, an herbal treatment, on the other hand, induces CYP3A4 activity, but inhibits CYP1A1, CYP1B1, and CYP2D6. Tobacco smoke induces CYP1A2 and watercress inhibits CYP2E1.

Cytochromes

Cytochromes are heme-containing proteins that play major roles in the process of electron transport in the mitochondrion and in photosynthesis in the chloroplast. They exist either as monomers (cytochrome c) or as subunits within large redox complexes (Complex III and Complex IV of electron transport. An atom of iron at the center of the heme group plays a central role in the process, flipping between the ferrous (Fe++) and ferric (Fe+++) states as a result of the movement of electrons through it.

There are several different cytochromes. Cytochrome c (Figure 5.47) is a soluble protein loosely associated with the mitochondrion. Cytochromes a and a3 are found in Complex IV. Complex III has cytochromes b and c1 and the plastoquinol-plastocyanin reductase of the chloroplast contains cytochromes b6 and f. Another important class of enzymes containing cytochromes is the cytochrome P450 oxidase group (see above). They get their name from the fact that they absorb light at 450 nm when their heme iron is reduced.

Iron-Sulfur Proteins

Iron-sulfur proteins contain iron-sulfur clusters in a variety of formats, including sulfide-linked di-, tri-, and tetrairon centers existing in different oxidation states (Figures 5.48 & 5.49). The clusters play a variety of roles, but the best known ones are in electron transport where they function in the redox reactions involved in the movement of electrons.

Complexes I and Complex II contain multiple Fe-S centers. Iron-sulfur proteins, though, have many other roles in cells. Aconitase uses an iron-sulfur center in its catalytic action and the ability of the enzyme to bind iron allows it to function as a barometer of iron concentration in cells. Iron-sulfur centers help to generate radicals in enzymes using S-Adenosyl Methionine (SAM) and can serve as a source of sulfur in the synthesis of biotin and lipoic acid. Some iron-sulfur proteins even help to regulate gene expression.

Ferredoxin

Ferredoxins are iron-sulfur containing proteins performing electron transfer in a wide variety of biological systems and processes. They include roles in photosynthesis in chloroplasts. Ferredoxins are classified structurally by the iron-sulfur clustered centers they contain. Fe2S2 clusters (Figure 5.50) are found in chloroplast membranes and can donate electrons to glutamate synthase, nitrate reductase, and sulfite reductase and serve as electron carriers between reductase flavoproteins and bacterial dioxygenase systems. Adrenodoxin is a soluble human Fe2S2 ferredoxin (also called ferredoxin 1) serving as an electron carrier (to cytochrome P450) in mitochondrial monooxygenase systems. Fe4S4 ferredoxins are subdivided as low and high potential ferredoxins, with the latter ones functioning in anaerobic electron transport chains.

Ferritin

Ferritin is an intracellular iron-storage protein found in almost all living organisms, from bacteria to higher plants and animals. It is a globular protein complex with 24 subunits and is the primary intracellular iron-storage protein in eukaryotes and prokaryotes. Ferritin functions to keep iron in a soluble and non-toxic form. Its ability to safely store iron and release it in a controlled fashion allow it to act like the prime iron buffer and solubilizer in cells – keeping the concentration of free iron from going to high or falling too low. Ferritin is located in the cytoplasm in most tissues, but it is also found in the serum acting as an iron carrier. Ferritin that doesn't contain any iron is known as apoferritin.

Monoamine oxidases

Monoamine oxidases are enzymes that catalyze the oxidative deamination of monoamines, such as serotonin, epinephrine, and dopamine. Removal of the amine with oxygen results in the production of an aldehyde and ammonia. The enzymes are found inside and outside of the central nervous system.

There are two types of monoamine oxidase enzymes – MAO-A and MAO-B. MAO-A is particularly important for oxidizing monoamines consumed in the diet. Both MAO-A and MAO-B play important roles in inactivating monoaminergic neurotransmitters. Both enzymes act on dopamine, tyramine (Figure 5.50), and tryptamine. MAO-A is the primary enzyme for metabolizing melatonin, serotonin, norepinephrine, and epinephrine, while MAO-B is the primary enzyme for phenethylamine (Figure 5.51) and benzylamine. MAO-B levels have been reported to be considerably reduced with tobacco usage.

Actions of monoamine oxidases thus affects levels of neurotransmitters and consequently are thought to play roles in neurological and/or psychiatric disorders. Aberrant levels of MAOs have been linked to numerous psychological problems, including depression, attention deficit disorder (ADD), migraines, schizophrenia, and substance abuse. Medications targeting MAOs are sometimes used to treat depression as a last resort – due to potential side effects. Excess levels of catecholamines, such as epinephrine, norepinephrine, and dopamine, can result in dangerous hypertension events.

DNA damage theory of aging

The DNA Damage Theory of Aging is based on the observation that, over time, cells are subject to extensive oxidative events. As already noted, these afford opportunities for the formation of ROS that can damage cellular molecules, and it follows that accumulation of such damage, especially to the DNA would be deleterious to the cell. The build-up of DNA damage could, thus, be responsible for the changes in gene expression that we associate with aging.

Numerous damage events

The amount of DNA damage that can occur is considerable. In mice, for example, it is estimated that each cell experiences 40,000 to 150,000 damage events per day. The damage, which happens to nuclear as well as to mitochondrial DNA, can result in apoptosis and/or cellular senescence. DNA repair systems, of course, protect against damage to DNA, but over time, unrepairable damage may accumulate.

Oxidative damage

DNA damage can occur in several ways. Oxidation can damage nucleotides and alter their base-pairing tendencies. Oxidation of guanine by reactive oxygen species, for example, can produce 8-oxo-guanine (Figures 5.52 and 5.53). This oxidized nucleobase commonly produced lesion in DNA arising from action of reactive oxygen species like superoxides. 8-oxoguanine is capable of forming a stable base pairing interaction within a DNA duplex with adenine, potentially giving rise to a mutation when DNA replication proceeds. 8-oxoguanine can be repaired if recognized in time by a DNA glycosylase, which acts to clip out the damaged base and it can then be replaced by the proper one. Polycyclic aromatic hydrocarbons from cigarette smoke, diesel exhaust, or overcooked meat can covalently bind to DNA and, if unrepaired, lead to mutation. Chemical damage to DNA can result in broken or cross-linked DNAs.

Diseases of DNA repair

The importance of DNA repair in the aging process is made clear

by diseases affecting DNA repair that lead to premature aging. These include Werner syndrome, for whom the life expectancy is 47 years. It arises as a result of loss of two enzymes in base excision repair. People suffering from Cockayne syndrome have a life expectancy of 13 years due to mutations that alter transcription-coupled nucleotide excision repair, which is an important system for fixing oxidative damage.

Further, the life expectancies of 13 species of mammalian organisms correlates with the level of expression of the PARP DNA repair-inducing protein. Interestingly, people who lived past the age of 100 had a higher level of PARP than younger people in the population.

Antioxidants

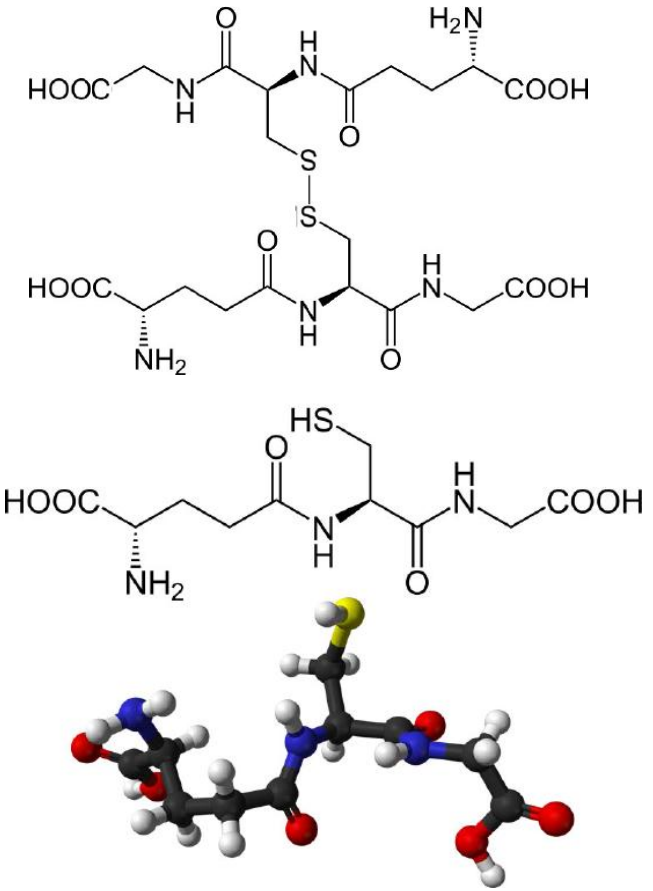
There is a growing interest in the subject of antioxidants because of health concerns raised by our knowledge of problems created as a result of spontaneous oxidation of biomolecules by Reactive Oxygen Species (ROS), such as superoxide. Antioxidants have the chemical property of protecting against oxidative damage by being readily oxidized themselves, preferentially to other biomolecules.

Biologically, cells have several lines of antioxidant defense. They include molecules, such as vitamins C, A, and E, glutathione, and enzymes that destroy ROS such as superoxide dismutase, catalase, and peroxidases.

Health effects

Oxidation by ROS is mutagenic and has been linked to atherosclerosis. Nonetheless, randomized studies of oral supplementation of various vitamin combinations have shown no protective effect against cancer and supplementation of Vitamin E and selenium has revealed no decrease in the risk of cardiovascular disease. Further, no reduction in mortality rates as a result of supplementation with these materials has been found, so the protective effects, if any, of antioxidants on ROS in human health remain poorly understood.

Glutathione



The major endogenous antioxidant found in cells spanning most living systems, glutathione is a tripeptide protecting cells against damage caused by reactive oxygen species and heavy metals (Figures 5.55 & 5.56). The three amino acids in glutathione (glutamate, cysteine, and glycine) are joined in an unusual fashion. The glutamate is joined to the center cysteine by a peptide bond between the R-group carboxyl of glutamate and the α-amine of cysteine. The bond between cysteine and glycine is a normal peptide bond between the α-carboxyl of cysteine and the α-amine of glycine.

The thiol group of cysteine is a reducing agent that reduces

disulfide bonds to sulfhydryls in cytoplasmic proteins. This, in turn, is the bridge when two glutathiones get oxidized and form a disulfide bond with each other (Figure 5.56). Glutathione's two oxidative states are abbreviated as follows: GSH (reduced) and GSSG (oxidized).

Disulfide-joined glutathiones can be separated by reduction of their bonds with glutathione reductase, using electrons from NADPH for the reduction.

Non-ribosomal synthesis

Glutathione is not made by ribosomes. Rather, two enzymes catalyze its synthesis. The enzyme γ-glutamylcysteine synthetase catalyzes the joining of the glutamate to the cysteine and then glutathione synthetase catalyzes the peptide bond formation between the cysteine and the glycine. Each step requires energy from ATP.

Essential for life

Glutathione is important for life. Mice lacking the first enzyme involved in its synthesis in the liver die in the first month after birth. In healthy cells, 90% of glutathione is in the GSH state. Higher levels of GSSG correspond to cells that are oxidatively stressed.

Besides reducing disulfide bonds in cells, glutathione is also important for the following:

- Neutralization of free radicals and reactive oxygen species.
- Maintenance of exogenous antioxidants such as vitamins C and E in their reduced forms.
- Regulation of the nitric oxide cycle

References

1. Winge, D.R., Mol Cell Biol. 2012 Jul; 32(14): 2647–2652. doi: 10.1128/MCB.00573-12

Energy: Electron Transport & Oxidative Phosphorylation

430

Figure 5.14 – Overview of electron transport (bottom left and top

right) and oxidative phosphorylation (top left – yellow box) in the mitochondrion

431

Figure 5.15 – Loss of electrons by NADH to form NAD+. Relevant reactions occur in the top ring of the molecule.

432

Figure 5.16 – Flow of electrons from NADH into the electron transport system. Entry is through complex I

Image by Aleia Kim

Figure 5.17 – Flow of electrons from FADH2 into the electron transport chain. Entry is through complex II.

Image by Aleia Kim

433

Figure 5.18 – Complex I embedded in the inner mitochondrial membrane. The mitochondrial matrix at at the top

Wikipedia

434

Figure 5.19 – Complex II embedded in inner mitochondrial membrane. Matrix is up.

Wikipedia

435

Figure 5.20 – Movement of electrons through complex I from NADH to coenzyme Q. The mitochondrial matrix is at the bottom

Image by Aleia Kim

Figure 5.21 – Movement of electrons from succinate through complex II (A->B->C->D->Q). Mitochondrial matrix on bottom.

Image by Aleia Kim

436

Figure 5.22 – Complex II in inner mitochondrial membrane showing electron flow. Matrix is up.

Wikipedia

Figure 5.23 – Coenzyme Q

437

Movie 5.2 – The Q-cycle

Wikipedia

Figure 5.24 – The Q-Cycle Image by Aleia Kim

Figure 5.24 – Complex III

Wikipedia

Figure 5.25 – The Q-cycle. Matrix is down.

Image by Aleia Kim

439

Figure 5.26 – Movement of electrons and protons through complex IV. Matrix is down

Image by Aleia Kim

Figure 5.25 – Cytochrome c with bound heme Group

Wikipedia

440

Figure 5.27 – Mitochondrial anatomy. Electron transport complexes and ATP synthase are embedded in the inner mitochondrial membrane

Image by Aleia Kim

441

Figure 5.28 – ATP synthase. Protons pass from intermembrane space (top) through the complex and exit in the matrix (bottom).

Image by Aleia Kim

442

Movie 5.3 – ATP Synthase – ADP + Pi (pink) and ATP (red). The view is end-on from the cytoplasmic side viewing the β subunits

Movie 5.3 – ATP Synthase – ADP + Pi (pink) and ATP (red). The view is end-on from the cytoplasmic side viewing the β subunits

443

Figure 5.29 – Important structural features of the ATP synthase

Image by Aleia Kim

444

Figure 5.30 – Loose (L), Tight (T), and Open (O) structures of the F1 head of ATP synthase. Change of structure occurs by rotation of γ-protein (purple) in center as a result of proton movement. Individual α and β units do not rotate

Image by Aleia Kim

445
Figure 5.31 – Respiration overview in eukaryotic cells
Wikipedia
446
Rest
ATP High / ADP Low
Oxidative Phosphorylation Low
Electron Transport Low
Oxygen Use Low
NADH High / NAD+ Low
Citric Acid Cycle Slow
Exercise
ATP Low / ADP High
Oxidative Phosphorylation High
Electron Transport High
Oxygen Use High
NADH Low / NAD+ High
Citric Acid Cycle Fast
447
Figure 5.32 – Three inhibitors of electron transport
Image by Aleia Kim
448
Figure 5.33 – Oligomycin A – An inhibitor of ATP synthase
Figure 5.34 – 2,4 DNP – an uncoupler of respiratory control
449
In Cells With Tight Coupling
O2 use depends on metabolism
NAD+ levels vary with exercise
Proton gradient high with no exercise
Catabolism depends on energy needs
ETS runs when OxPhos runs and vice versa
In Cells That Are Uncoupled
O2 use high
NAD+ Levels high
Little or no proton gradient

Catabolism high
OxPhos does not run, but ETS runs rapidly
450
451
Figure 5.35 – Alternative oxidase (AOX) of fungi, plants, and protozoa bypasses part of electron transport by taking electrons from CoQ and passing them to oxygen.
452
Figure 5.36 – Structure of an oxygen free radical
Wikipedia
NADPH + 2O2
NADP+ + 2O2– + H+
Figure 5.37 – Three sources of reactive oxygen species (ROS) in cells
Wikipedia
453
454
Figure 5.38 A hydroxyl radical
Wikipedia
455
Reduced Glutathione (GSH) + H2O2
Oxidized Glutathione (GSSG) + H2O
Figure 5.40 – Detoxifying reactive oxygen species
Figure 5.39 – Catalase
456
1. O2- + Enzyme-Cu++
O2 + Enzyme-Cu+
2. O2- + Enzyme-Cu+ + 2H+
H2O2 + Enzyme-Cu++
Figure 5.41 – SOD2 of humans
Figure 5.42 3 – Peroxynitrite Ion
Figure 5.44 – SOD1 of humans
Wikipedia
Figure 5.45 – SOD3 of humans
457

Figure 5.43 – Peroxynitrite's effects on cells lead to necrosis or apoptosis

Wikipedia

458

RH + O2 + NADPH + H+

ROH + H2O + NADP+

459

Figure 5.46 – Cytochrome c with its heme group

460

Figure 5.47 – Fe2S2 Cluster

Figure 5.48 – Redox reactions for Fe4S4 clusters

461

Figure 5.49 – Tyramine

Figure 5.50 – Phenethylamine

462

Figure 5.51 – Guanine and 8-oxo-guanine

Figure 5.52 – Adenine-8-oxo-guanine base pair. dR = deoxyribose

463

Figure 5.53 – Good antioxidant sources

464

Figure 5.55 – Oxidized glutathiones (GSSG) joined by a disulfide bond

Wikipedia

Figure 5.54 – Structure of reduced glutathione (GSH)

465

Figure 5.56 – Resveratrol

466

5.3: The Citric Acid Cycle

The primary catabolic pathway in the body is the citric acid cycle because it is here that oxidation to carbon dioxide occurs for breakdown products of the cell's major building blocks – sugars, fatty acids, and amino acids. The pathway is cyclic (Figure 6.63) and thus, doesn't really have a starting or ending point. All of the reactions occur in the cellular organelle called the mitochondria, though one enzyme is embedded in the organelle's inner membrane. As needs change, cells may use a subset of the reactions of the cycle to produce a desired molecule rather than to run the entire cycle.

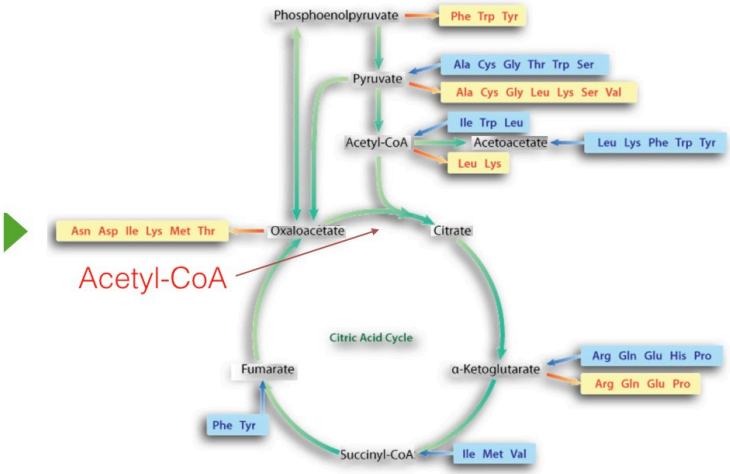


Figure 6.63 – Amino acid metabolism and the citric acid cycle. Amino acids boxed in yellow are made from the indicated intermediate. Amino acids in blue are made into the intermediate in catabolism. Image by Aleia Kim

Acetyl-CoA

The molecule "feeding" the citric acid cycle by bringing in new carbons is acetyl-CoA. It can be obtained from pyruvate (from the catabolism of carbohydrates via a pathway called glycolysis), from the fatty acid β-oxidation pathway, from ketone bodies, and from amino acid metabolism. It is worth noting that acetyl-CoA has very different fates, depending on the cell's energy status/needs. In other words, the pathway has branches out as well as multiple materials that can feed in to it. The description below describes oxidation (catabolism) in citric acid cycle.

Before discussing the citric acid cycle, it is important to first describe one important enzyme complex that is a major source of acetyl-CoA for the cycle.

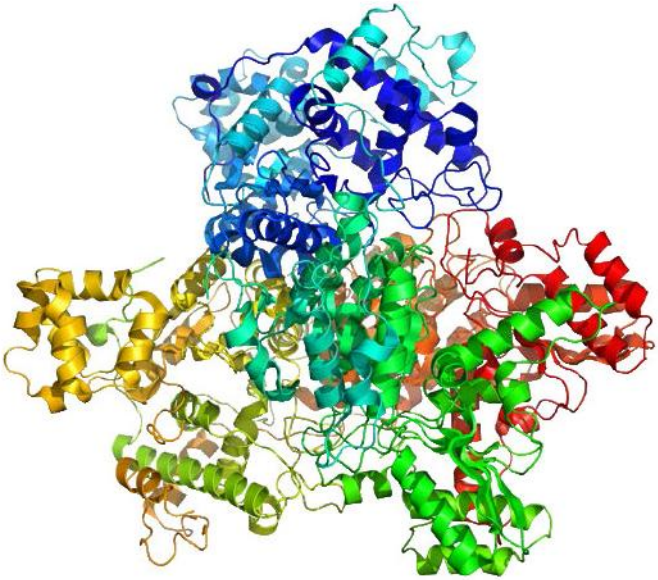


Figure 6.64 – E1 Subunit of Pyruvate Dehydrogenase. Wikipedia

Pyruvate decarboxylation

The pyruvate dehydrogenase enzyme is a complex of multiple copies of three subunits that catalyze the decarboxylation (removal of CO_2 from) of pyruvate to form acetyl-CoA. The reaction mechanism requires use of five coenzymes, including thiamin pyrophosphate which is derived from the B vitamin named thiamin. Pyruvate dehydrogenase is an enormous complex in mammals with a size five times greater than ribosomes.

Puruvate dehydrogenase produces the reduced nucleotide NADH as well as Acetyl-CoA. As with other sources of NADH or FADH2, the reduced nucleotide can be used as a source of electrons for the Electron Transport Chain, leading to the formation of ATP via oxidative phosphorylation.

Pyruvate dehydrogenase regulation

Pyruvate deyhdrogenase is regulated both allosterically and by covalent modification – phosphorylation / dephosphorylation. Regulation of pyruvate dehydrogenase, whether by allosteric or covalent mechanisms has the same strategy. Indicators of high energy availability in a cell shut down the enzyme, whereas indicators of low energy stimulate it.

For allosteric regulation, the high energy indicators affecting the enzyme are ATP, acetyl-CoA, NADH, and fatty acids, which inhibit it. AMP, Coenzyme A, NAD+, and calcium, on the other hand, stimulate it (Figure 6.67).

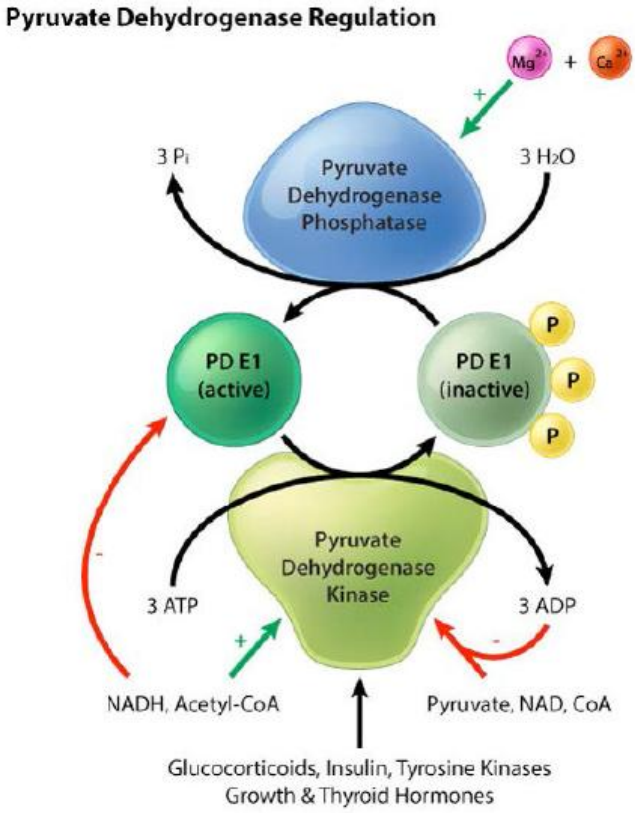


Figure 6.67 – Regulation scheme for pyruvate dehydrogenase (PD). Image by Aleia Kim

Covalent modification regulation of pyruvate dehydrogenase is a bit more complicated. It occurs as a result of phosphorylation by pyruvate dehydrogenase kinase (PDK – Figure 6.67) or dephosphorylation by pyruvate dehydrogenase phosphatase (PDP). PDK is allosterically activated in the mitochondrial matrix when NADH and acetyl-CoA concentrations rise. While the mechanism is a bit complex, the activation or deactivation of the complex by phosphorylation is aligned with energy needs of the cell: when

energy is very available, the complex is deactivated, and when more energy is needed, the complex is activated.

Other signals also affect the activity of the complex. For instance the hormone insulin also also activates pyruvate kinase and the glycolysis pathway, spurring cellular use of internalized glucose.

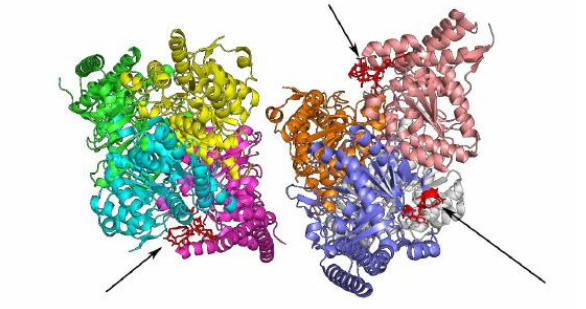


Figure 6.68 – Pyruvate dehydrogenase complex with three phosphorylation sites in red marked by arrows.Wikipedia

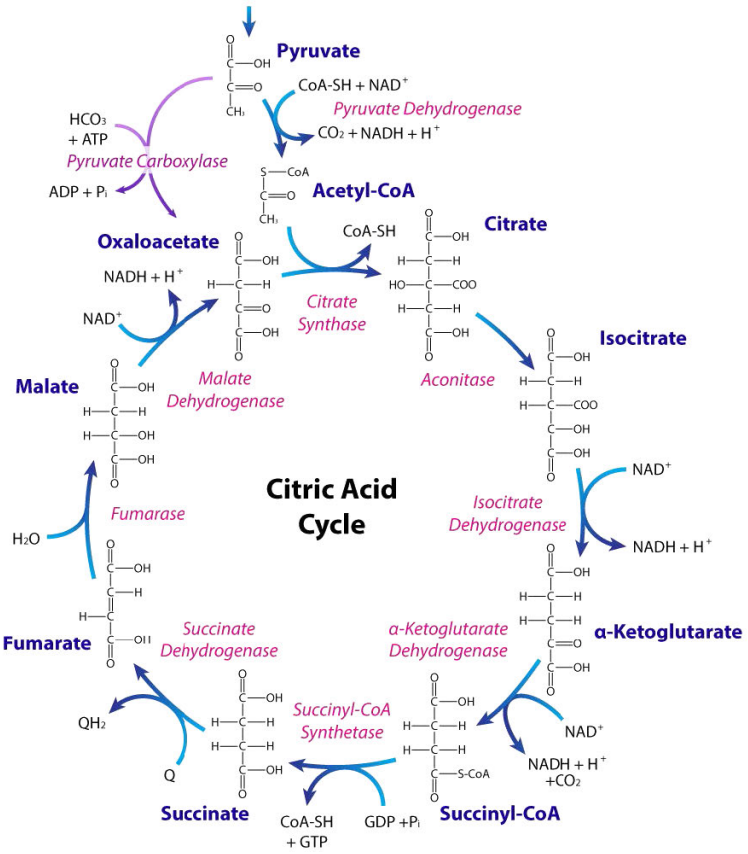


Figure 6.69 – The citric acid cycle. Image by Aleia Kim

Citric acid cycle reactions

The usual starting point point for consideration of the Citric Acid Cycle itself is the addition of acetyl-CoA to oxaloacetate (OAA) to form citrate.

Acetyl-CoA for the pathway can come from a variety of sources.

The reaction joining it to OAA is catalyzed by citrate synthase and is quite energetically downhill. This, in turn, helps to "pull" the malate dehydrogenase reaction preceding it in the cycle.

In the next reaction, citrate is isomerized to isocitrate by action of the enzyme called aconitase.

Isocitrate is a branch point in plants and bacteria for the glyoxylate cycle.

Oxidative decarboxylation of isocitrate by isocitrate dehydrogenase produces the first NADH and yields α-ketoglutarate.

This five carbon intermediate is a branch point for synthesis of the amino acid glutamate. In addition, glutamate can also be made easily into this intermediate in the reverse reaction. Decarboxylation of α-ketoglutarate produces succinyl-CoA and is catalyzed by α-ketoglutarate dehydrogenase.

The enzyme α-ketoglutarate dehydrogenase is structurally very similar to pyruvate dehydrogenase and employs the same five coenzymes – NAD+, FAD, CoA-SH, thiamine pyrophosphate, and lipoamide.

These coenzymes include derivatives of the vitamins Niacin, Riboflavin, and Thiamin.

Note that at this point the cycle has produced 2 NADH molecules (reducing NAD+) during oxidation-reduction reactions. Two carbons have been 'lost' from the cycle as carbon dioxide, leaving four carbons from the 6-carbon citrate.

The reduced NADH can feed the Electron Transport Chain, contributing to the formation of ATP via oxidative phosphorylation.

Regeneration of oxaloacetate

The remainder of the citric acid cycle involves conversion of the four carbon succinyl-CoA into oxaloacetate. Succinyl-CoA is a branch point for the synthesis of heme. Succinyl-CoA is converted to succinate in a reaction catalyzed by succinyl-CoA synthetase

(named for the reverse reaction) and a GTP is produced, as well – the only substrate level phosphorylation in the cycle.

The energy for the synthesis of the GTP comes from energy released during hydrolysis of the thioester bond between succinate and the CoA-SH.

Succinate Oxidation

Oxidation of succinate occurs in the next step, catalyzed by succinate dehydrogenase. This interesting enzyme both catalyzes this reaction and participates in the electron transport system, funneling electrons from the FADH2 it gains in the reaction to coenzyme Q. The product of the reaction, fumarate, gains a water across its trans double bond in the next reaction, catalyzed by fumarase to form malate.

Fumarate is also a byproduct of nucleotide metabolism and of the urea cycle. Malate is important also for transporting electrons across membranes in the malate-aspartate shuttle and in ferrying carbon dioxide from mesophyll cells to bundle sheath cells in C4 plants.

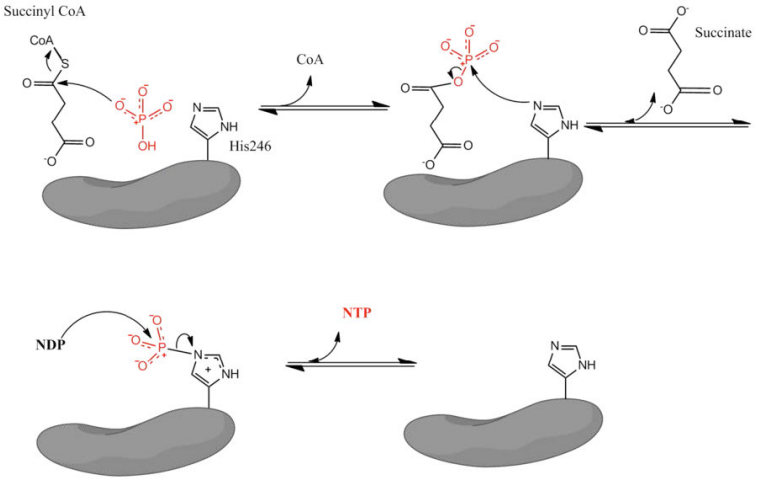


Figure 6.70 – Succinyl-CoA synthetase mechanism

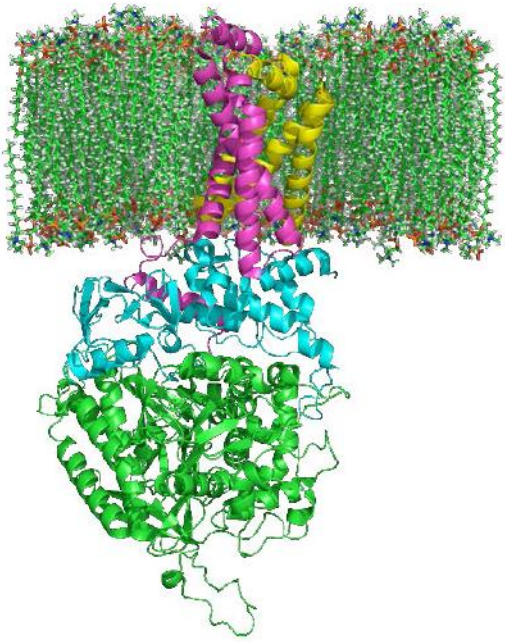


Figure 6.71 – Succinate dehydrogenase embedded in the mitochondrial inner membrane (top). Wikipedia

Rare oxidation

Conversion of malate to oxaloacetate by malate dehydrogenase is a rare biological oxidation that is energetically unfavorable on its own.

The reaction is 'pulled' by the energetically favorable conversion of oxaloacetate to citrate in the citrate synthase reaction described above. Oxaloacetate intersects other important pathways, including amino acid metabolism (readily converted to aspartic acid), transamination (nitrogen movement) and gluconeogenesis.

Regulation of the citric acid cycle

Allosteric regulation of the citric acid cycle occurs as substrates and products of the pathway, or molecules involved in energy transfer, activate or inhibit enzymes in ways that allow for appropriate responses by the cell to energy states. Substrates/products that regulate or affect the pathway include acetyl-CoA and succinyl-CoA .

High energy molecular indicators, such as ATP and NADH will tend to inhibit the cycle and low energy indicators (NAD+, AMP, and ADP) will tend to activate the cycle.

Regulated enzymes

Regulated enzymes in the cycle include citrate synthase (inhibited by NADH, ATP, and succinyl-CoA), isocitrate dehydrogenase (inhibited by ATP, activated by ADP and NAD+), and α-ketoglutarate dehydrogenase (inhibited by NADH and succinyl-CoA and activated by AMP).

As tends to occur in metabolism, regulation occurs on enzymes that catalyze reactions with large energy drops and which are at key branch points, shutting off entire paths that do not need to be active.

The citric acid cycle is a major catabolic pathway producing a considerable amount of energy for cells. Organisms that have the biochemical machinery to perform oxidative phosphorylation use the cycle to extract far more energy from their food than those that do not have this ability.

Since the electron transport chain uses molecular oxygen (O2) as its final electron acceptor, the citrate cycle is only beneficial as an energy source for organisms that can metabolize aerobically (with oxygen). Anaerobic organisms do not have mitochondria, do not carry out these reactions and are at a disadvantage energetically as a result. However they often can exist in ecological niches where oxygen is scarce or non-existent, and where they do not need to compete with aerobic life.

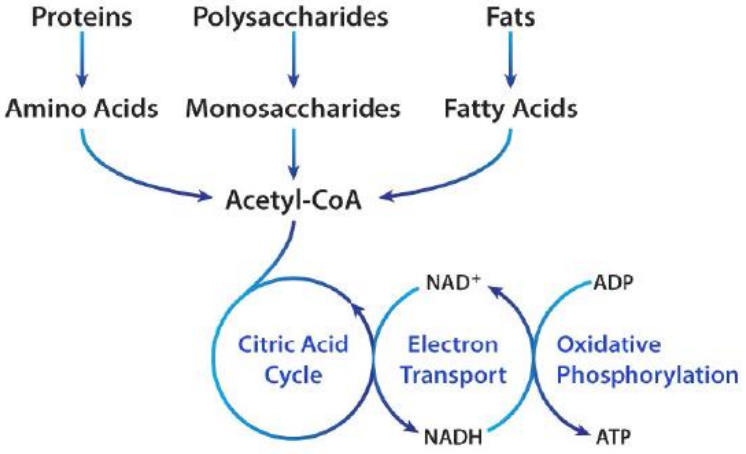


Figure 6.77 – Acetyl-CoA metabolism. Image by Aleia Kim

Acetyl-CoA metabolism

Acetyl-CoA is one of the most "connected" metabolites in biochemistry, appearing in fatty acid oxidation/synthesis, pyruvate oxidation, the citric acid cycle, amino acid anabolism/catabolism, ketone body metabolism, steroid/bile acid synthesis, and (by extension from fatty acid metabolism) prostaglandin synthesis .

These many different metabolic pathways converge at this key molecule, feeding into the cycle to provide organisms energy from a variety of food and stored chemical sources.

6.1: Structure and Function-Carbohydrates

Carbohydrates are commonly described as sugars, or saccharides, from the Greek word for sugar. The simplest carbohydrates are called monosaccharides. An example is glucose. Monosaccharides can be joined to make larger molecules. Disaccharides contain two monosaccharides. Sucrose is a disaccharide, containing both fructose and glucose. Mono and disaccharides are sometimes referred to as simple sugars. Polysaccharides are chains of many sugar subunits. Examples include glycogen and cellulose, both of which are polymers of glucose (configured differently).

Carbohydrates are literally "hydrates of carbon." This name

derives from the generalized formula of simple monosaccharides, which can be written in the form of $C_x(H_2O)_x$, where x is a digit typically between 3 and 8. Not all sugars have this formula, however. Deoxyribose, the sugar found in every nucleotide in a DNA molecule lacks one oxygen and thus has the formula $C_5H_{10}O_4$.

Carbohydrates are important in cells as energy sources (especially glucose, glycogen, and amylose), as markers of cellular identity (oligosaccharides on the surface of cells of multicellular organisms), as structural components (cellulose holding up plants), and as constituents of nucleotides (ribose in RNA, deoxyribose in DNA).

The building blocks of all carbohydrates are the monosaccharides.

Shown below are Fischer projection formulas for a group of common monosaccharides. Fischer projection formulas are similar but not identical to organic structural formulas. Carbons in the sugar are represented with the elemental symbol C at the end of the chain, but also are represented by vertices (such as carbon 1 in D-Ribose below) and by intersecting perpendicular lines (carbons 2, 3, and 4 in D-Ribose).

The tetrahedral arrangement around the carbons in the chain of a monosaccharide are represented as flat, with 90 degree bond angles, in the Fischer projection. This is not an accurate representation of the three-dimensional molecules. To interpret these structures as 3D models, each carbon within the chain can be considered in sequence. The bonds shown vertically in the Fischer projection are oriented back, away from the viewer, while the horizontal bonds (to H and OH) emerge forward, out of the plane of view.

Fischer projections make for easy drawing and comparison of carbohydrate structure but their interpretation is prone to error.

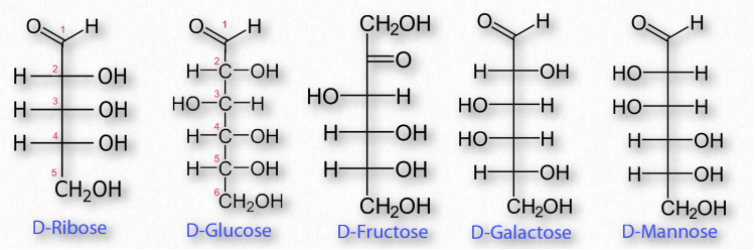


Figure 2.148 – Common sugar structures

Monosaccharides

The most common monosaccharides include glucose, fructose, galactose, ribose, and mannose. Of these sugars, all but one (fructose) exists as an aldehyde. Fructose and some other less well known sugars are ketones. Figure 2.148 shows the structure of these sugars.

By convention, the letters 'ose' at the end of a biochemical name flags a molecule as a sugar. Thus, there are glucose, galactose, sucrose, and many other '-oses'. Other descriptive **nomenclature** involves use of a prefix that tells how many carbons the sugar contains. For example, glucose, which contains six carbons, is described as a hexose. The following list shows the prefixes for numbers of carbons in a sugar:

- Tri- = 3
- Tetr- = 4
- Pent- = 5
- Hex- =6
- Hept- = 7
- Oct- = 8

Other prefixes identify whether the sugar contains an aldehyde group (aldo-) or a ketone (keto-) group. Prefixes may be combined. Glucose, which is a 6-carbon sugar with an aldehyde group, can be

described as an aldohexose. The list that follows gives the common sugars and their descriptors. •Ribose = aldo-pentose

•Glucose = aldo-hexose

•Galactose = aldo-hexose

•Mannose = aldo-hexose

•Fructose = keto-hexose

Diastereomers

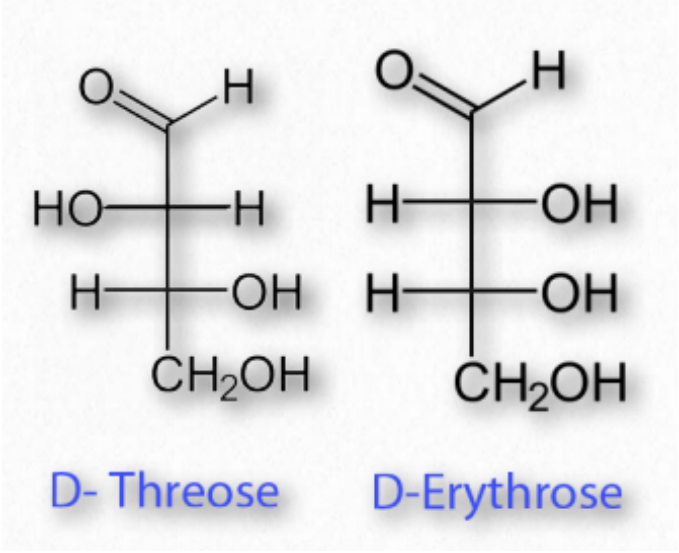


Figure 2.149 – Diastereomers

Sugars may have multiple chiral carbons and thus differ from each other in the configuration of groups around those asymmetric carbons. Two sugars having the same chemical form (aldoses, for example) and the same number of carbons, but that differ only in the stereochemical orientations of their carbons are referred to as diastereomers (Figure 2.149). For example, glucose, galactose, and mannose all have the formula of C6H12O6, but are chemically

distinct from each other in the orientation of groups around the carbons within them.

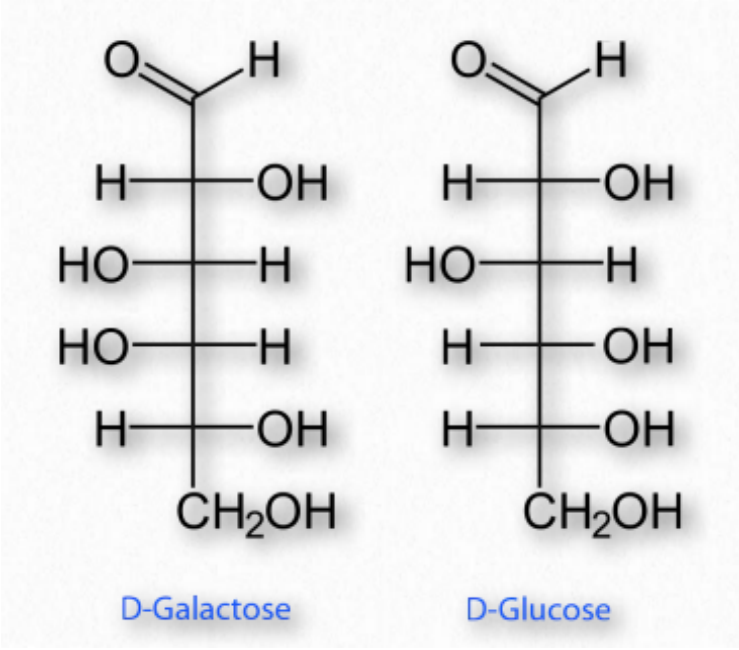


Figure 2.150 – Epimers – D-Galactose and D-Glucose differ only in the configuration of carbon #4

Enantiomers and epimers

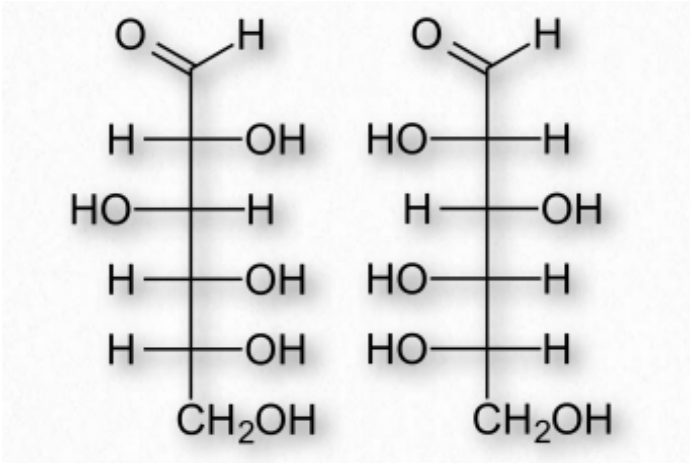
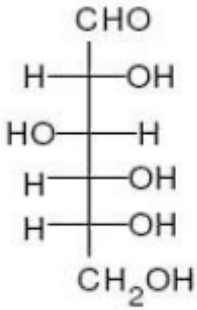


Figure 2.151 – Enantiomers – D-Glucose (left) and L-Glucose (right) are mirror

If two sugars are identical except for having one chiral carbon arranged differently (such as images glucose and galactose – Figure 2.150), they are considered **epimers** of one another. If two sugars are mirror images of each other, they are enantiomers (Figure 2.151). Biochemical notation uses the letters D and L to describe monosaccharide stereochemistry in a very particular way. As a result, one enantiomer will be given an L designation while the other is D. So L-glucose is the mirror image of D-glucose.



Movie 2.6 – Conversion of glucose from a straight chain form to a ring form

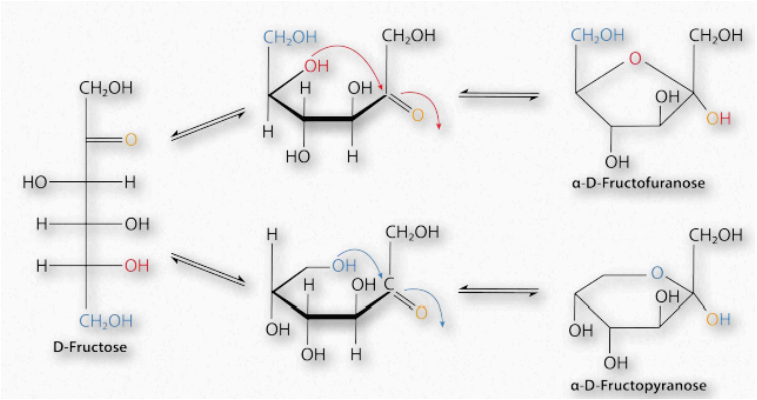


Figure 2.152 – Conversion of D-fructose between furanose (top right), linear (left), and pyranose (bottom right) forms Image by Pehr Jacobson

Sugars with five and six carbons can readily cyclize (Figure 2.152, Movie 2.6) in solution. When they do, a new asymmetric carbon is created that didn't exist in the same sugars when they were in the straight chain form, as the carbon to oxygen double bond converts to an alcohol. This carbon has a special name – it is called the anomeric carbon and (like the other asymmetric carbons in sugars) it can have the hydroxyl in two different positions. These positions are referred to as α and β . Sugars, such as α-D-glucose and β-D-

glucose that differ only in the configuration of the anomeric carbon are referred to as anomers (Figure 2.153).

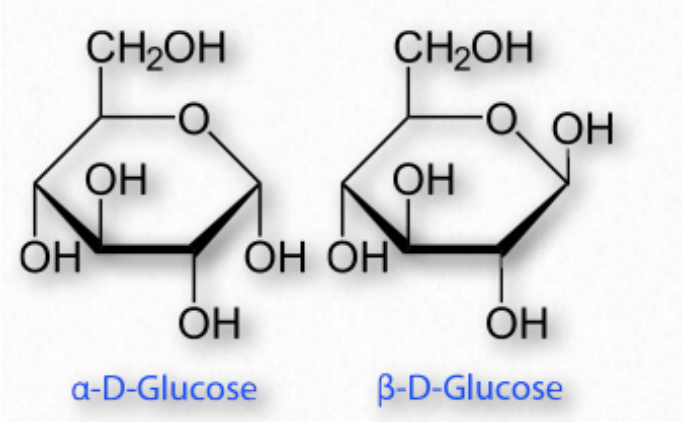


Figure 2.153 – Anomers – α-DGlucose and β-D-Glucose differ only in the configuration of the anomeric carbon #1

Sugars cyclizing to form rings with five atoms in them (see fructose in Figure 2.128) are referred to as furanoses (named for furan) and those forming rings with six atoms, such as glucose in the same figure, are called pyranoses (named for pyran). The carbonyl carbon becomes the anomeric carbon in the ring by binding to the oxygen of a hydroxyl elsewhere in the chain. α- and β- forms of a given sugar can readily "flip" between each form in solution, so long as the anomeric hydroxyl is free, because the bonding in cyclic forms is unstable, so molecules interconvert in solution. Most pentoses and hexoses can form both furanose and pyranose structures (Figure 2.152).

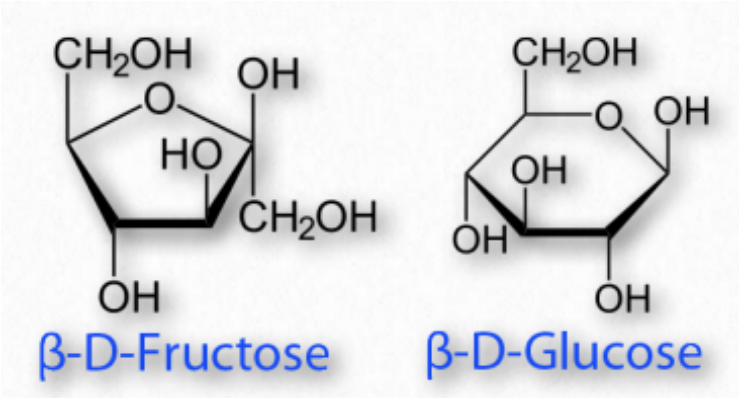


Figure 2.154 – A furanose (left) and a pyranose (right)

Linking the anomeric hydroxyl to another group will create a structure called a glycoside which will remain locked in whichever α- or β- configuration they were in when the anomeric hydroxyl was altered.

Boat/chair conformations

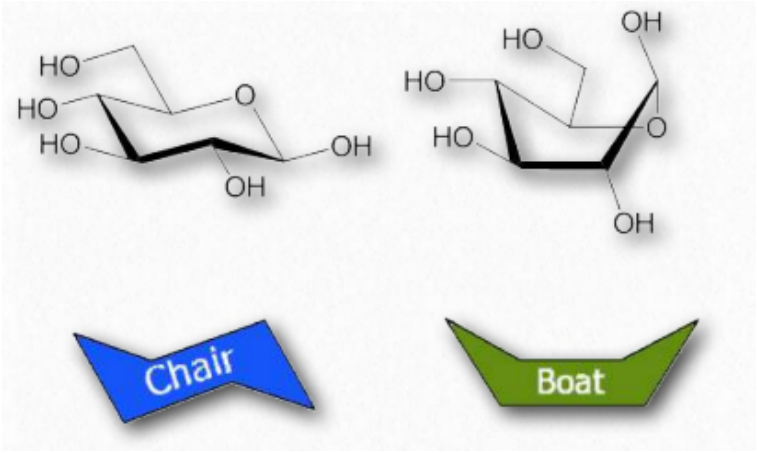


Figure 2.155 – Chair and boat forms of glucose

Orbitals of carbon prefer to be in tetrahedral conformations and

this means that the bonds between carbons in a ring do not lie flat. Indeed, rings "pucker" to try to accommodate this tendency, giving rise to different 3D forms for any given sugar. Some of these forms resemble boat structures, which others resemble chairs or envelopes (Figure 2.155). The stablest (and thus most abundant) of these forms have all of the hydoxyls in the equatorial positions, resulting in less steric hindrance.

Modified monosaccharides

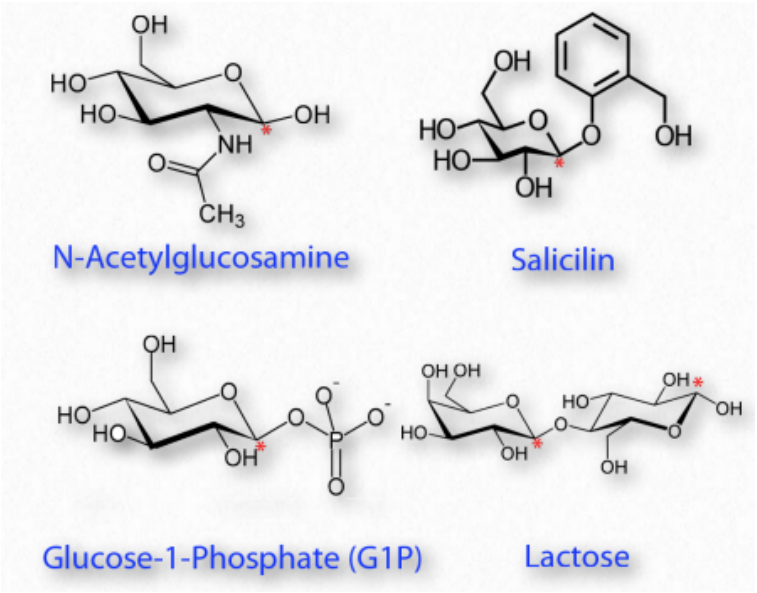


Figure 2.156 – Modified sugars. Locations of glycosidic carbon indicated with red asterisks. All are glycosides except N-acetylglucosamine

Many chemical modifications can occur on sugar residues (Figure 2.156). Common ones include oxidation, reduction, phosphorylation, and substitution of an amine or an acetylamine for a hydroxyl. The ones that affect the anomeric hydroxyl group make glycosides (Figure 2.157), whereas modifications that don't affect the anomeric hydroxyl, (glucose-6-phosphate, for example), do not.

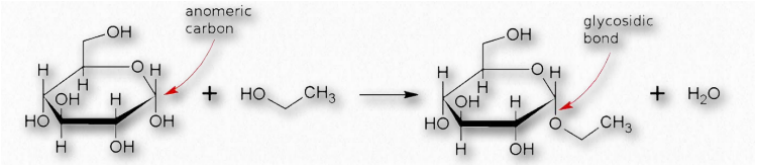


Figure 2.157 – Formation of a glycosidic bond

Oxidation/reduction

Figure 2.158 – A positive Benedict's test starting at left and moving right Wikipedia

The last considerations for simple sugars relative to their structure are their chemical reactivity and modification. Sugars that are readily oxidized are called 'reducing sugars' because their oxidation causes other reacting molecules to be reduced. A test for reducing sugars is known as Benedict's test. In it, sugars are mixed and heated with an alkaline solution containing Cu++. Reducing sugars will donate an electron to Cu++, converting it to Cu+, which will produce cuprous oxide Cu2O, as an orange precipitate (Figure 2.158). Since Cu++ solution is blue, the change of color provides an easy visual indication of a reducing sugar.

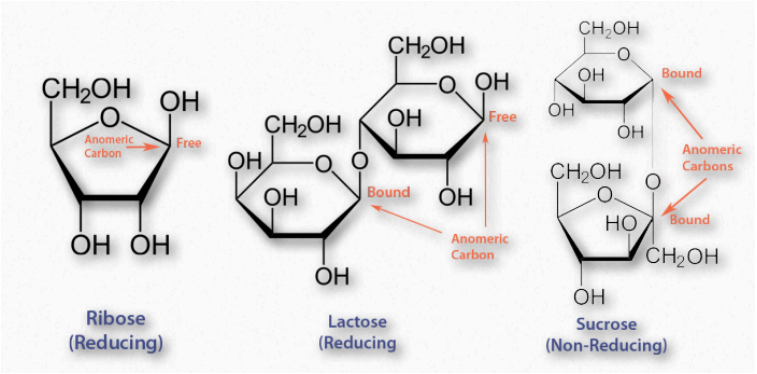


Figure 2.159 – Reducing and non-reducing sugars

The aldehyde group of aldoses is very susceptible to oxidation, whereas ketoses are less so, but can easily be oxidized if, like fructose, they contain an α-hydroxyl and can tautomerize to an aldose. Most monosaccharides are reducing sugars. This includes all of the common ones galactose, glucose, fructose, ribose, xylose, and mannose. Some disaccharides, such as lactose and maltose are reducing sugars since they have at least one anomeric carbon free, allowing that part of the sugar to linearize and yield an aldose. Sucrose, on the other hand has no anomeric carbons free – both are involved in a glycosidic linkage, so they cannot linearize and thus it is not a reducing sugar.

Oxidation and reduction of sugars can occur in cells. As we will see, phosphorylation of sugars occurs routinely during metabolism.

Glucuronic acid

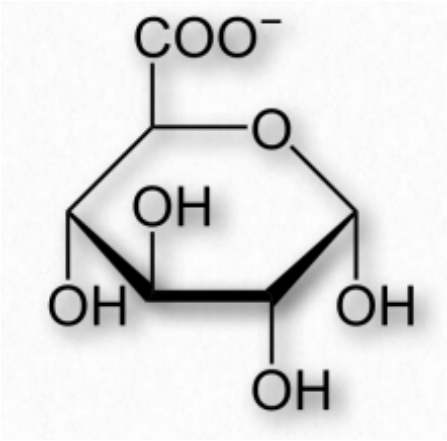


Figure 2.160 – Glucuronic acid

One oxidation product of glucose is glucuronic acid, a six carbon molecule where the CH_2OH on carbon six is oxidized to a carboxylic acid (Figure 2.160). Related oxidized sugars include galacturonic acid and mannuronic acid. Glucuronic acid is commonly conjugated to other molecules in the liver/bile by UDP-glucuronyltransferase enzymes to make the molecules more water soluble for excretion, since the carboxyl group of glucoronic acid ionizes readily at physiological pH. The reactions are usually done starting with glucuronic acid linked to UDP (UDPGlucuronic Acid). In addition, glucuronic acid is made from a UDP-glucose precursor. Glucuronic acid is a common constituent of glycosaminoglycans, proteoglycans, and glycoglycerolipids. Glucuronic acid is found in heparin, dermatan sulfate, chondroitin sulfate, hyaluronic acid, and keratan sulfate. Glucuronic acid is also a precursor of ascorbic acid (Vitamin C) in organisms that synthesize this compound.

Sugar alcohols

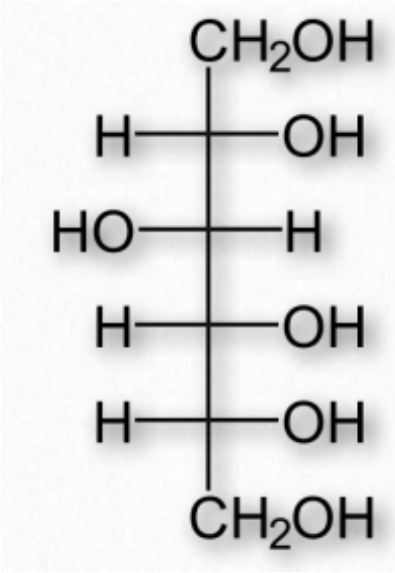


Figure 2.161 – Sorbitol (also called glucitol)

Reduction of aldoses or ketoses by hydrogenation produces the corresponding sugar alcohols. The compounds are widely used as thickeners of food or as artificial sweeteners, due to their ability to stimulate sweet receptors on the tongue. Common sugar alcohols (sugar progenitor in parentheses) include glycerol (glyceraldehyde), xylitol (xylose), sorbitol (Figure 2.161 – from glucose), galactitol (galactose), arabitol (arabinose), and ribitol (ribose). Most of these compounds have a sweetness of between 0.4 and 1.0 times as sweet as sucrose, but provide considerably fewer calories per weight. Xylitol is the sweetest of them with a sweetness equal to that of sucrose.

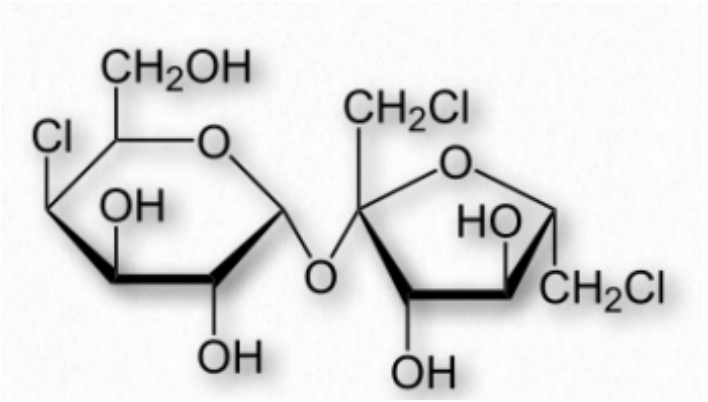


Figure 2.162 – Structure of sucralose

Sugar alcohols are used sometimes to mask the aftertaste of other artificial sweeteners. Many of them also produce a cooling sensation upon dissolving, due to that being an endothermic process for them, resulting in a pleasant mouth sensation. Last, they are poorly absorbed by intestines, and so have a low glycemic index.

Artificial sweeteners

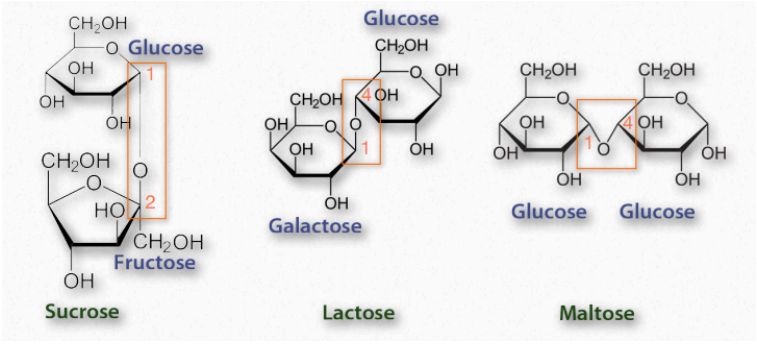


Figure 2.163 – Common disaccharides – glycosidic bonds in rectangles

Artificial sweeteners are compounds that stimulate taste receptors for sweetness, but are metabolized for energy

inefficiently at best. Such compounds frequently are many times sweeter than table sugar (sucrose) on a weight/weight basis and are referred to as "intensely sweet." Most of the artificial sweeteners are not carbohydrates, but rather are able to stimulate the same sweet receptors that sugar does. Seven such compounds are approved for use in the U.S. – stevia, aspartame, sucralose, neotame, acesulfame potassium, saccharin, 1 1 1 2 4 4 and advantame. The sugar alcohol known as sorbitol is also sometimes used as an artificial sweetener.

Disaccharides

Disaccharides (Figure 2.163) are made up of two monosaccharides. The most common ones include sucrose (glucose and fructose), lactose (galactose and glucose), and maltose (glucose and glucose). All of the common disaccharides contain at least one glycosidic bond. We name the disaccharides according to which carbons are linked to each other and the how the anomeric carbon of the glycosidic bond is configured. Lactose, for example, is described as β-Dgalactopyranosyl-(1→4)-D-glucose, or more succinctly as having an α-1,4 glycosidic bond.

Oligosaccharides

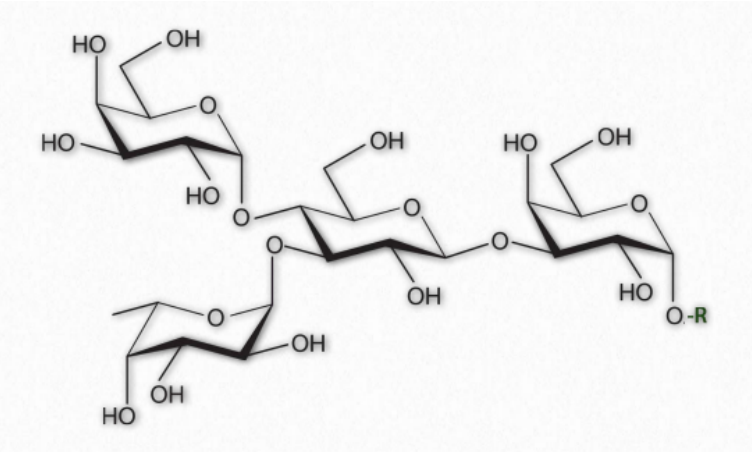


Figure 2.164 – A Branched oligosaccharide attached to an RGroup

As their name implies, oligosaccharides (Figure 2.164) are comprised of a few (typically 3 to 9) sugar residues. These often, but not always contain modified sugars. Unlike all of the other saccharides, oligosaccharides are not typically found unattached to other cellular structures. Instead, oligosaccharides are found bound, for example, to sphingolipids (making cerebrosides or gangliosides) or proteins (making glycoproteins).

Oligosaccharides in membrane glycoproteins play important roles in cellular identity/ recognition. The patterns of oligosaccharides displayed on the extracellular face of the plasma membrane acts as a sort of barcode that identifies specific cell types. The immune system recognizes these identity tags in the body. “Foreign” oligosaccharide structures trigger the immune system to attack them. While this provides a very good defense against invading cells of an organism, it also can pose significant problems when organs are transplanted from one individual into another, with rejection of donated organs, in some cases.

Organelle targeting

The oligosaccharides that are attached to proteins may also determine their cellular destinations. Improper glycosylation or errors in subsequent sugar modification patterns can result in the failure of proteins to reach the correct cellular compartment.

For example, inclusion cell disease (also called I-cell disease) arises from a defective phosphotransferase in the Golgi apparatus. This enzyme normally catalyzes the addition of a phosphate to a mannose sugar attached to a protein destined for the lysosome. In the absence of a functioning enzyme, the unphosphorylated glycoprotein never makes it to the lysosome and is instead exported out of the cell where it accumulates in the blood and is excreted in the urine. Individuals with Icell disease suffer developmental delays, abnormal skeletal development, and restricted joint movement.

Glycosylation

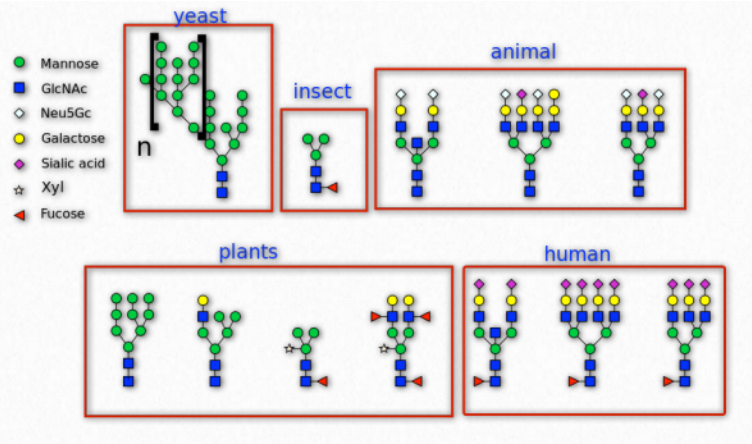


Figure 2.165 – N-linked glycosylation in various organisms Wikipedia

Sugars are commonly attached to proteins in a process called glycosylation. Typically the attachment is to a hydroxyl or other functional group. The majority of proteins synthesized in the endoplasmic reticulum are glycosylated.

N-glycans on cell surfaces play roles in the immune system. The immunoglobulin types (IgG, IgA, IgE, IgD, and IgM) have distinct glycosylation patterns that confer unique functions by affecting their affinities for immune receptors. Glycans also are important in self/non-self identity is tissue rejection and autoimmune diseases.

Glycoproteins

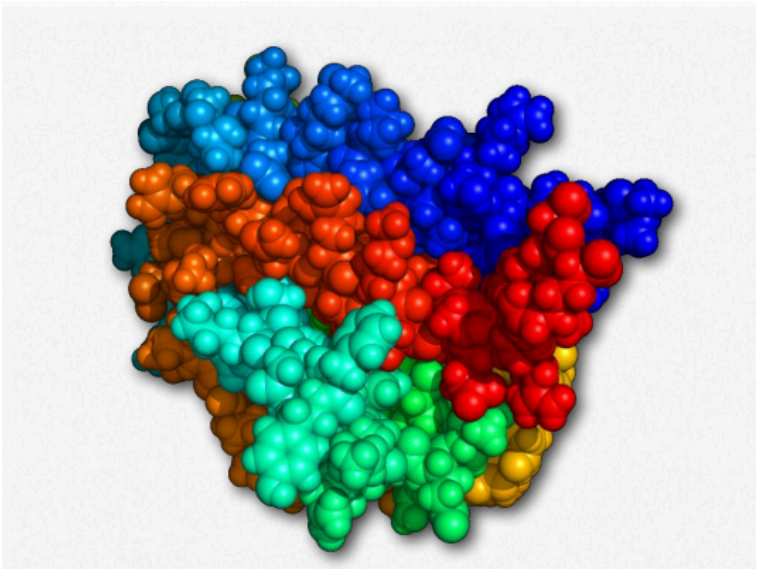


Figure 2.170 – Erythropoietin Wikipedia

Glycoproteins are a very diverse collection of saccharide-containing proteins with many functions. Attachment of the saccharide to the protein is known as glycosylation. Secreted extracellular proteins and membrane proteins with exposed

extracellular regions are often glycosylated. Saccharides attached to these may be short (oligosaccharides) or very large (polysaccharides). Glycoproteins play important roles in the immune system in antibodies and as components of the major histocompatibility complex (MHC). They are important for interactions between sperms and eggs, in connective tissues and are abundant in egg whites and blood plasma. Two glycoproteins (gp41 and gp120) are part of the HIV viral coat and are important in the infection process. Some hormones, such as erythropoietin, human chorionic gonadotropin, follicle-stimulating hormone and luteinizing hormone are also glycoproteins.

Glycation

Glycation is a chemical process (nonenzymatic) that occurs when a protein or lipid covalently binds to a sugar, such as glucose or fructose. Glycation differs from glycosylation in that the latter process is controlled by enzymes and results in specific attachment of specific sugars to biomolecules. Glycation, by contrast, is driven by two properties of monosaccharides 1) their chemistry and 2) their concentration. Glycations may be endogenous (occurring in an organism) or exogenous (occurring external to an organism).

Exogenous glycation arises most commonly as a result of cooking of food and this results in attachment of sugars to lipids and/or proteins to form advanced glycation endproducts (AGEs). At temperatures above 120°C, AGE production occurs readily and contributes to the taste and the appearance of the food we eat.

Cooking

Browning of food, for example, is a product of glycation and is

enhanced as the sugar content of a food increases. Browning of french fries is often enhanced, for example, by adding sugar to them. The formation of a crust of bread or the toasting of bread are other examples. These glycations are products of the Maillard reaction in which a reactive sugar carbonyl group combines with a nucleophilic amine of an amino acid. The process is favored in an alkaline environment, when amines are less protonated. The formation of the harder shell of a pretzel, for example, results from addition of lye to the exterior. At higher temperatures, though, a carcinogen known as acrylamide can be formed by reactions involving asparagine.

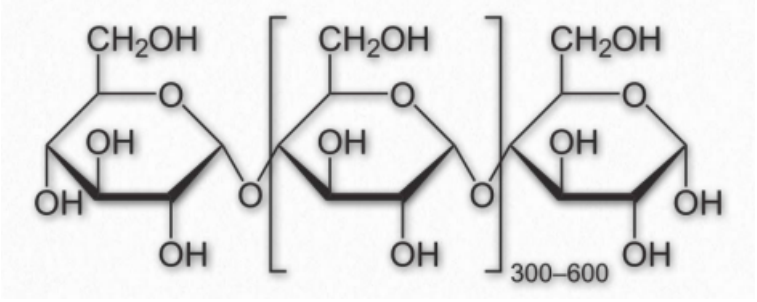


Figure 2.171 – Repeating unit of amylose

Endogenous glycation, on the other hand, arises with a frequency that is proportional to the concentration of free sugar in the body. These occur most frequently with fructose, galactose, and glucose in that decreasing order and are detected in the bloodstream. Both proteins and lipids can be glycated and the accumulation of endogenous advanced glycation endproducts (AGEs) is associated with Type 2 diabetes, as well as in increases in cardiovascular disease (damage to endothelium, cartilage, and fibrinogen), peripheral neuropathy (attack of myelin sheath), and deafness (loss of myelin sheath).

The formation of AGEs increases oxidative stress, but is also thought to be exacerbated by it. Increased oxidative stress, in turn causes additional harm. Damage to collagen in blood cells causes them to stiffen and weaken and is a factor in hardening of the arteries and formation of aneurysms, respectively. One indicator of

diabetes is increased glycation of hemoglobin in red blood cells, since circulating sugar concentration are high in the blood of diabetics. Hemoglobin glycation is measured in testing for blood glucose control in diabetic patients.

Homopolymer	Monomeric Unit
Glycogen	Glucose
Cellulose	Glucose
Amylose	Glucose
Callose	Glucose
Chitin	N-acetylglucosamine
Xylan	Xylose
Mannan	Mannose
Chrysolaminarin	Glucose

Polysaccharides

Long polymers of sugar residues are called polysaccharides and can be up to many thousands of units long. Polysaccharides are found free (not attached to other molecules) or bound to other cellular structures such as proteins. Some polysaccharides are homopolymers (contain only one kind of sugar). Others are heteropolymers (glycosaminoglycans, hemicellulose). Polysaccharides function in energy storage (nutritional polysaccharides, such as glycogen, amylose, amylopectin, e.g.), structure enhancement (chitin, cellulose, e.g.), and lubrication (hyaluronic acid, e.g.). These individual categories of polysaccharides are discussed below.

Nutritional polysaccharides

This group of polysaccharides is used exclusively for storage of sugar residues. They are easily easily broken down by the organism making them, allowing for rapid release of sugar to meet rapidly changing energy needs.

Amylose

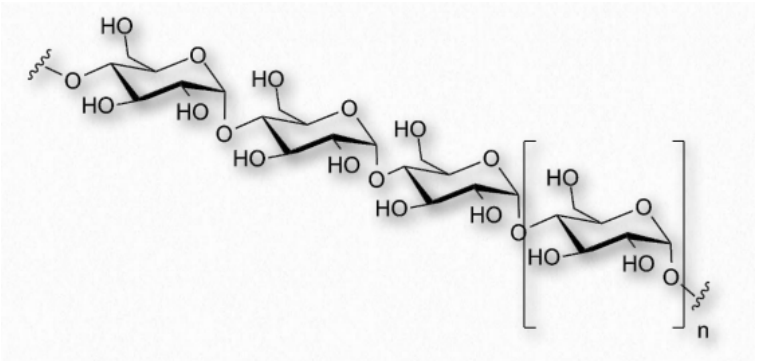


Figure 2.172 – Another view of amylose

Amylose has the simplest structure of any of the nutritional polysaccharides, being made up solely of glucose polymers linked only by α-1,4 bonds (Figure 2.171 & 2.172). (Note that the term 'starch' is actually a mixture of amylose and amylopectin). Amylose is insoluble in water and is harder to digest than amylopectin (see below). The complexing of amylopectin with amylose facilitates its water Figure 2.172 – Another view of amylose solubility and its digestion. Amylose is produced in plants for energy storage and since plants don't have rapidly changing demands for glucose (no muscular contraction, for example), its compact structure and slow breakdown characteristics are consistent with plants' needs.

Amylopectin and glycogen

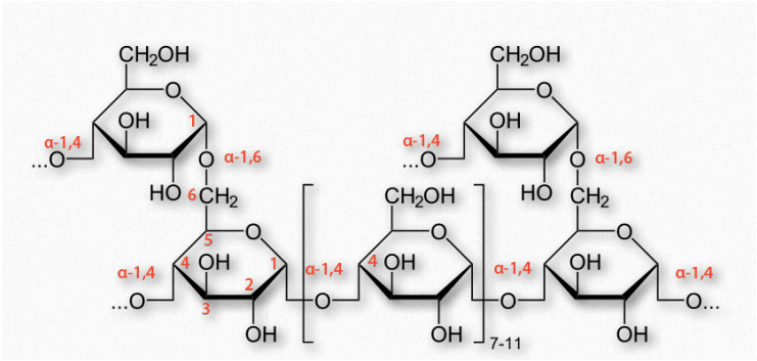


Figure 2.173 – Structure of glycogen

More complicated homopolymers of glucose are possessed by amylopectin in plants and glycogen (Figure 2.173) in animals. Both compounds contain long glucose chains with α-1,4 bonds like amylose, but unlike amylose, these long chains have branches of α-1,6 bonds. Amylopectin is the less-branched of the two, having such bonds about every 25-30 residues, whereas glycogen has branches about every 8-12 residues.

Branching plays important roles in increasing water solubility and in providing more “ends” to the polymer. In animals, glycogen is broken down starting at the ends, so more ends means more glucose can be released quickly. Again, plants, which have a lower need for quick release of glucose than animals get by with less branching and fewer ends.

Structural polysaccharides

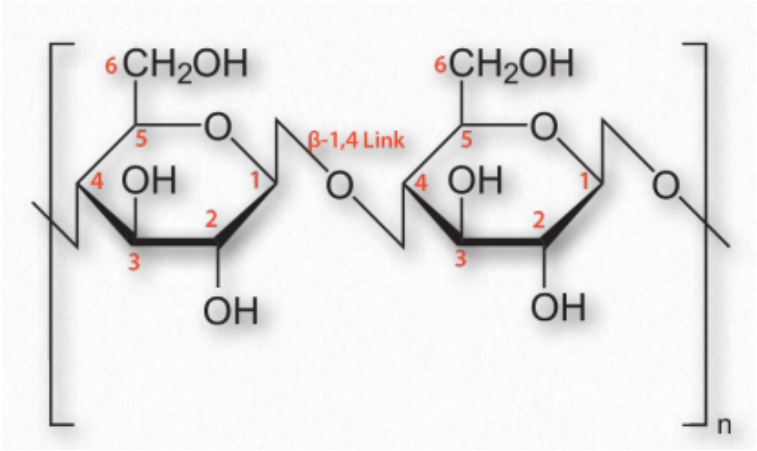


Figure 2.174 – Cellulose with β-1,4 links between glucose sugars

An additional function of polysaccharides in cells relates to structure. Cellulose, which is a polymer of glucose with exclusive β-1,4 linkages between the units (Figure 2.174) is an important structural component of plants and fungi cells. Notably, most non-ruminant animals are unable to digest this polymer, as they lack the enzyme known as cellulase.

Ruminants, such as cattle, however, contain in their rumen a bacterium that possesses this enzyme and allows them to obtain glucose energy from plants. Another group of polysaccharides found in plant cell walls is the hemicelluloses. This class of molecules encompasses several branched heteropolymers of (mostly) D-pentose sugars along with a few hexoses and L-sugars as well. Hemicelluloses are shorter than cellulose (500-3000 sugars versus 7000-15,000 sugars).

Monomer sugars of polysaccharides besides glucose include xylose, mannose, galactose, rhamnose, and arabinose. Xylose is usually present in the greatest amount (Figure 2.175).

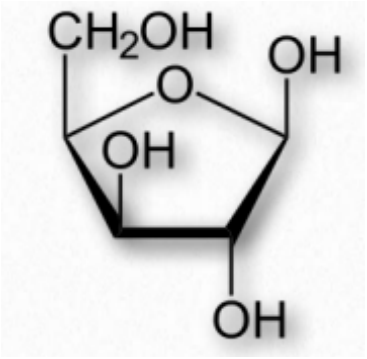


Figure 2.175 Xylose

Chitin

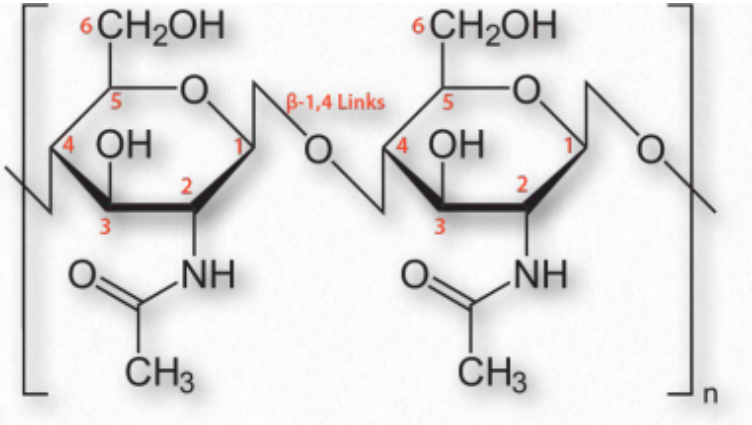


Figure 2.176 – Chitin with β-1,4 links between N-acetylglucosamine sugars

Chitin (Figure 2.176) is another structural polysaccharide, being comprised of N-acetylglucosamine units joined by β-1,4 linkages. It is a primary component of the cell walls of fungi and is also prominent in the exoskeletons of arthropods and insects, as well as the beaks and internal shells of cephalopods (Figure 2.177). Chitin's structure was solved by Albert Hofmann in 1929. It is like cellulose

except for the acetylamine group replacing the hydroxyl on position 2. This change allows hydrogen bonding to occur between adjacent polymers, thus providing greater strength.

Pectins

Another group of structural polysaccharides is the pectins (Figure 2.178). These com pounds are present in most primary plant cell walls and are abundant in non-woody parts of terrestrial plants. They are rich in galacturonic acid (α-1,4 links with no branches – Figure 2.179) and are used commercially as a gelling agent in jams/ jellies, as well as a stabilizer in fruit juices and milk drinks. Pectin consumption may result in reduced blood cholesterol levels due to its tendency to 1) bind cholesterol and 2) to increase viscosity in the intestinal tract, thus reducing absorption of cholesterol from food. Pectins also trap carbohydrates in the digestive system and reduce their rate of absorption.

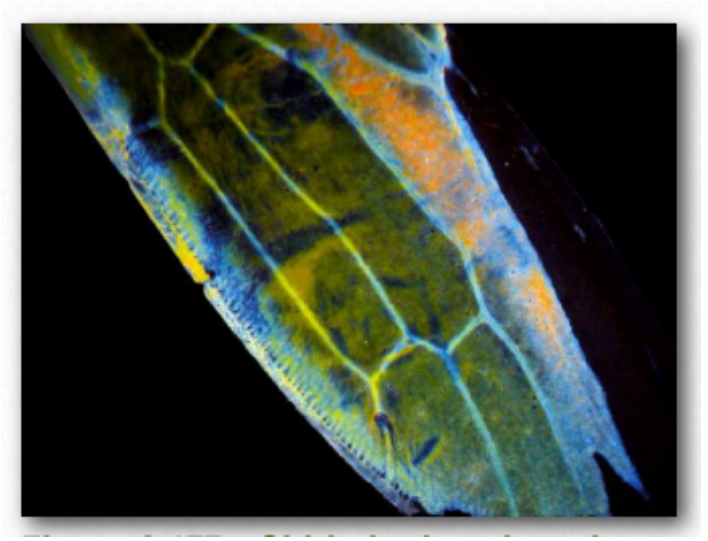


Figure 2.177 – Chitin in the wing of a sap beetle Wikipedia

Figure 2.178 – A powdered form of pectin

Lectins

Lectins are not carbohydrates, but proteins that specifically bind to carbohydrate molecules found in animals and plants (where they are known as phytohemagglutinins) and are each highly specific for certain sugars. They function in cellular and molecular recognition, as well as cell adhesion. One lectin recognizes hydrolytic enzymes containing mannose-6-phosphate and targets them to be delivered to lysosomes. In the innate immune system, a mannose binding lectin helps defend against invading microbes. Other lectins have roles in inflammation and autoimmune disorders.

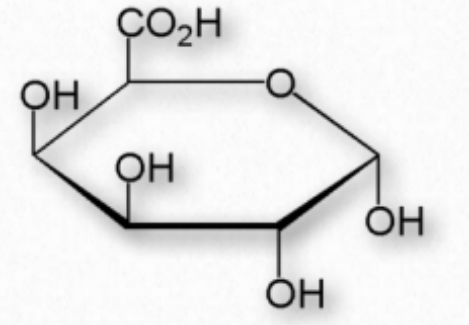


Figure 2.179 – α-DGalacturonic acid – An important component of pectin polymers

Some viruses and bacteria use lectins to recognize and bind specific carbohydrate residues on the surface of target cells. Flu virus, for example, carries a lectin known as hemagglutinin (Figure 2.180) that binds to sialic acid and is essential for entrance of the virus into the target cell. After binding, the viral particle enters by endocytosis after the hemagglutinin has been cleaved by a protease.

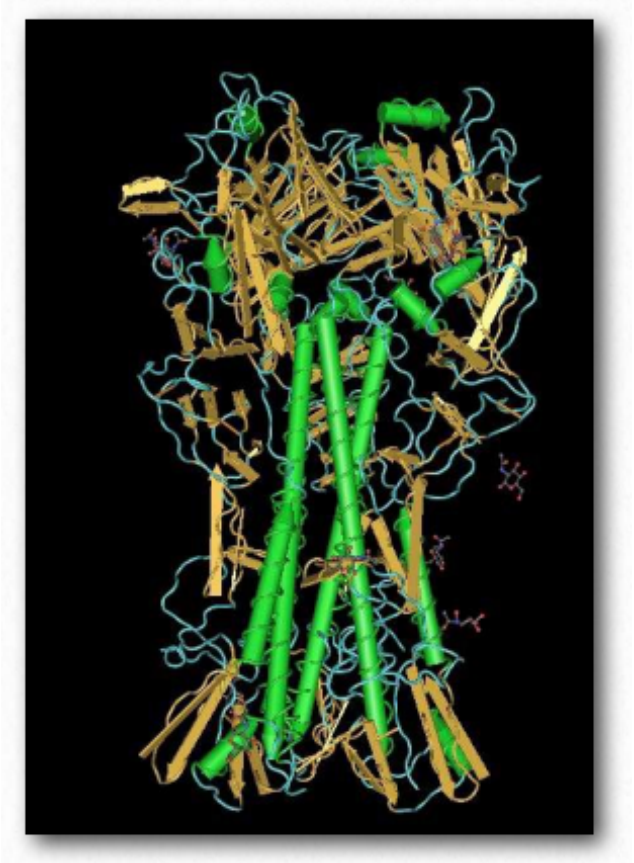


Figure 2.180 – Hemagglutinin

After replication of the virus inside of the cell, hemagglutinin and and a viral enzyme known as neuraminidase cluster in the cell membrane. Viral RNA and associated viral proteins cluster near this

membrane site and new viruses bud off in a portion of the cell's membrane after the hemagglutinin-sialic acid link to the infected cell is released by the neuraminidase cutting the bond between the sialic acid and the rest of the cell surface carbohydrate. Drugs, such as tamiflu, that interfere with neuraminidase work by preventing release of the viral particle. Unreleased particles will tend to aggregate and not function.

Some viral glycoproteins from hepatitis C virus may attach to lectins on the surface of liver cells in their infectious cycle. The bacterium Helicobacter pylori uses a cell surface lectin to bind oligosaccharides on epithelial cells lining the stomach. One lectin known as ricin is a very powerful toxin. It is produced in the endosperm of seeds of the castor oil plant and is of concern as a bioterrorism weapon as a result of its acute toxicity when inhaled or ingested.

Lectins were discovered originally in plants and have been most studied in legumes, but lectins are now known to be widely dispersed in nature. In the immune system, a mannan binding lectin (MBL) helps mediate the first defenses against microorganisms. Other immune system lectins are thought to modulate inflammatory processes and probably play a role in self/non-self recognition that is at the root of rejection of transplanted organs.

Heparin

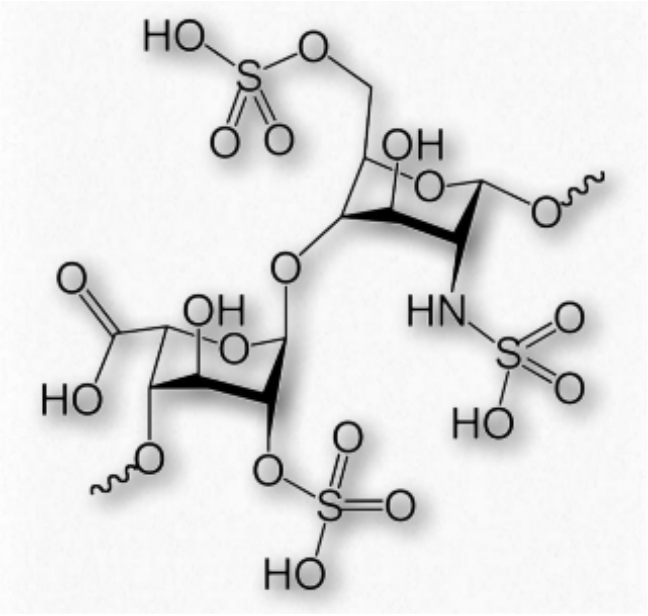


Figure 2.186 – Repeating sulfated disaccharide in heparin

Heparin (Figures 2.186 & 2.187) is a modified polysaccharide whose biological function is unclear, but whose ability to prevent clotting of blood is used for medical purposes. Heparin does not dissolve blood clots. Rather, it acts to prevent conversion of fibrinogen to fibrin.

Whether or not heparin is actually used by the body for its anticoagulation property is uncertain. It is stored in the secretory granules of mast cells and released at the point of injury and it has been proposed it is a protection against bacteria and other foreign materials.

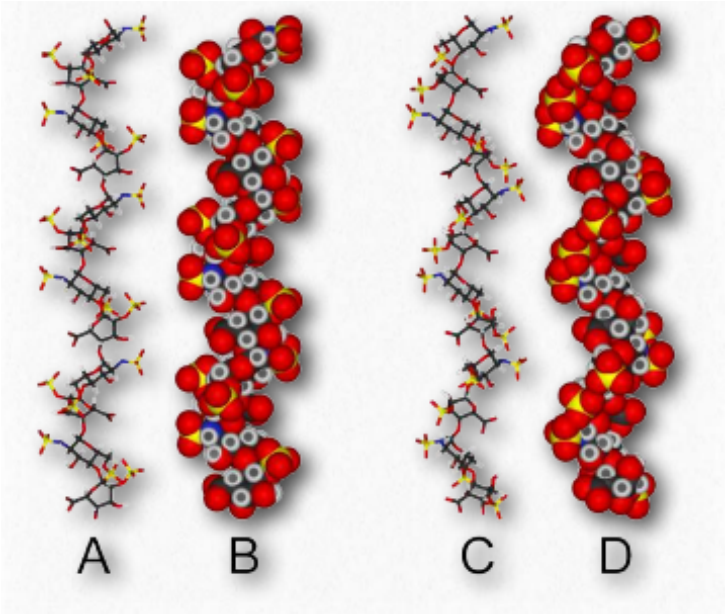


Figure 2.187 – Two structures for heparin

Heparin has abundant sulfates and is, in fact, the molecule with the highest negative charge density known. Its size varies from 3 kDa to 30 kDa, with an average of about 15 kDa. The repeating disaccharide of 2-Osulfated iduronic acid and 6-O-sulfated, N-sulfated glucosamine, occupies about 85% of the molecule. Copper salts of heparin help stimulate the synthesis of blood vessels (angiogenic).

Hyaluronic acid

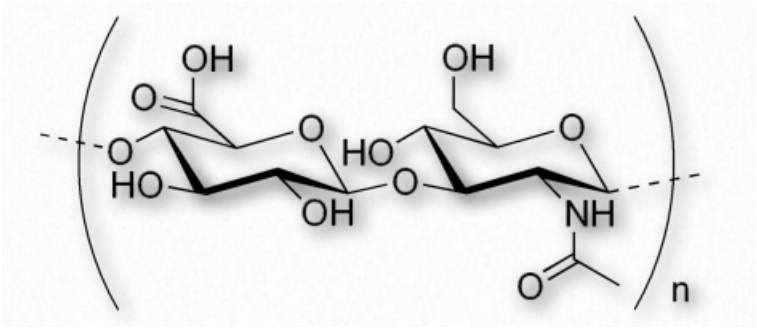


Figure 2.188 – Repeating disaccharide of hyaluronic acid

Hyaluronic acid (also known as hyaluronan or hyaluronate) is a glycosaminoglycan found in connective, epithelial, and nerve tissues. It is an unusual glycosaminoglycan (Figure 2.188), lacking sulfate, is made by hyaluronan synthases on the inner face of the plasma membrane and has a molecular weight in the millions. An average adult body contains about 15 grams of HA, one third of which is replaced every day. The repeating unit in hyaluronic acid is a disaccharide structure of D-glucuronic acid joined to D-N-acetylglucosamine. The compound, which can have upwards of 25,000 units of the disaccharide, is delivered directly into the extracellular matrix by enzymes from its plasma membrane site of synthesis.It is an important component of the extracellular matrix, where it assists in cell proliferation and migration. The polymer provides an open hydrated matrix to facilitate general cell migration whereas directed cell migration occurs via the interaction between hyaluronic acid and specific cell surface receptors. HA interaction with the receptor RHAMM (Receptor for Hyaluronan Mediated Motility) has been shown to be involved in wound repair as well as tumor progression.

Synovial Fluid

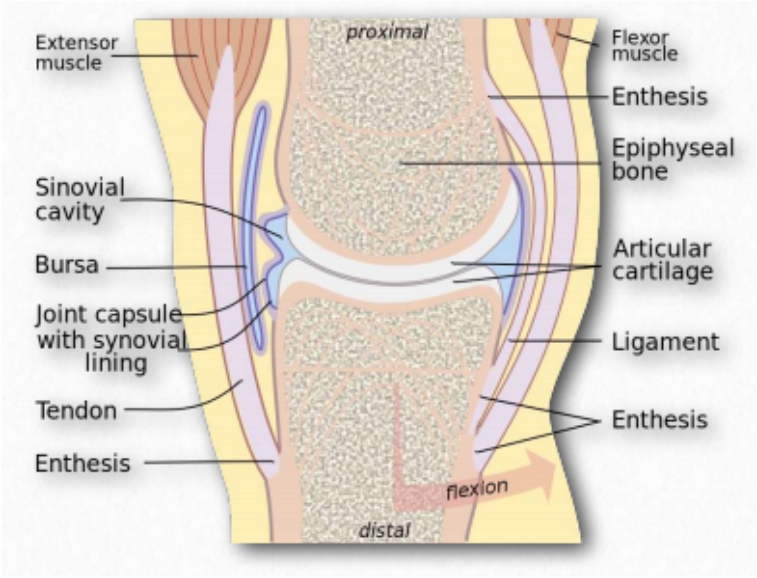


Figure 2.189 – Synovial fluid in joint lubrication Wikipedia

The function of hyaluronic acid has traditionally been described as providing lubrication in synovial fluid (the lubricating material in animal joints – Figure 2.189). Along with the proteoglycan called lubricin, hyaluronic acid turns water into lubricating material. Hyaluronic acid is present as a coat around each cell of articular cartilage and forms complexes with proteoglycans that absorb water, giving resilience (resistance to compression) to cartilage. Aging causes a decrease in size of hyaluronans, but an increase in concentration.

Function in skin

Hyaluronic acid is a major component of skin and has functions in

tissue repair. With exposure to excess UVB radiation, cells in the dermis produce less hyaluronan and increase its degradation.

For some cancers the plasma level of hyaluronic acid correlates with malignancy. Hyaluronic acid levels have been used as a marker for prostate and breast cancer and to follow disease progression. The compound can to used to induce healing after cataract surgery. Hyaluronic acid is also abundant in the granulation tissue matrix that replaces a fibrin clot during the healing of wounds. In wound healing, it is thought that large polymers of hyaluronic acid appear early and they physically make room for white blood cells to mediate an immune response.

Breakdown

Breakdown of hyaluronic acid is catalyzed by enzymes known as hyaluronidases. Humans have seven types of such enzymes, some of which act as tumor suppressors. Smaller hyaluronan fragments can induce inflammatory response in macrophages and dendritic cells after tissue damage. They can also perform proangiogenic functions.

6.2: Sugar Metabolism

Glycolysis

Carbohydrates, whether synthesized by photosynthetic organisms, stored in cells as glycogen, or ingested by heterotrophs, must be broken down to obtain energy for the cell's activities as well as to synthesize other molecules required by the cell.

Starch and glycogen are the main energy storage forms of carbohydrates in plants and animals, respectively. To use these sources of energy, cells must break down the polymers to their component monomers: glucose. The glucose is then taken up by cells through transporter proteins in cell membranes. The metabolism of glucose, as well as other six carbon sugars (hexoses) begins with the catabolic pathway called glycolysis. Glycolysis occurs in the cytosol of cells, not on or in the mitochrondria.

The end metabolic products of glycolysis are two molecules of ATP, two molecules of NADH and two molecules of pyruvate (Figure 6.3), which, in turn, can be oxidized further in the citric acid cycle.

Just one step of the glycolysis pathway involves the loss/gain of electrons, but the end product of the pathway, pyruvate, can be completely oxidized by aerobic organisms to carbon dioxide (Figure 6.2). Indeed, without production of pyruvate from glucose in glycolysis, a major energy source for aerobic cells would not be available.

Glucose is the most abundant hexose in nature and is traditionally used to illustrate the reactions of glycolysis, but fructose (in the form of fructose-6- phosphate) is also readily metabolized, while galactose can easily be converted into glucose for catabolism in the pathway as well. The end metabolic products of glycolysis are two

molecules of ATP, two molecules of NADH and two molecules of pyruvate (Figure 6.3), which can feed into the Citric Acid Cycle.

Anaerobic organisms do not have the metabolic machinery to completely oxidize pyruvate to carbon dioxide. In these organisms, or in organisms that are experiencing a lack of available oxygen, glycolysis provides some metabolic energy and pyruvate is converted to ethanol or lactic acid.

Ethanol fermentation performed by yeasts and bacteria have and have had a huge influence on the production of food and beverage products. Ethanol beverages (beer, wine, and mash for distillled spirits), breads, cheeses, pickles, and cultured dairy products such as yogurt and kefir can all be produced with the help of these fermentation pathways.

Glycolysis includes ten reactions linked in an almost completely linear pathway:

Reaction 1

Glucose gets a phosphate from ATP to make glucose-6-phosphate (G6P) in a reaction catalyzed by the enzyme hexokinase, a transferase enzyme.

Glucose + ATP $\rightleftarrows$ G6P + ADP + H+

Hexokinase is one of three regulated enzymes in glycolysis and is inhibited by one of the products of its action – G6P. Hexokinase has flexibility in its substrate binding and is able to phosphorylate a variety of hexoses, including fructose, mannose, and galactose.

Why phosphorylate glucose?

Phosphorylation of glucose serves two important purposes. First, the addition of a phosphate group to glucose effectively traps it in the cell, as G6P cannot diffuse across the lipid bilayer. Second, the reaction decreases the concentration of free glucose, favoring additional import of the molecule.

Reaction 2

Next, G6P is converted to fructose-6-phosphate (F6P), in a reaction catalyzed by the enzyme phosphoglucose isomerase:

$$\mathrm{G_6P} \rightleftarrows \mathrm{F_6P}$$

Reaction 3

$$ceF6P + ATP \rightleftarrows F1,6BP + ADP + H+$$

The second input of energy occurs when F6P gets another phosphate from ATP in a reaction catalyzed by the enzyme phosphofructokinase-1 (PFK-1 – another transferase) to make fructose-1,6- bisphosphate (F1,6BP). PFK-1 is a very important enzyme regulating glycolysis, with several allosteric activators and inhibitors.

Like the hexokinase reaction the energy from ATP is needed to make the reaction energetically favorable. PFK-1 is the most important regulatory enzyme in the pathway and this reaction is the rate-limiting step. It is also one of three essentially irreversible reactions in glycolysis.

Reaction 4

$$\mathrm{F_1,6\,BP} \rightleftarrows \mathrm{D{-}GLYAL_3P + DHAP}$$

With the glycolysis pump thus primed, the pathway proceeds to split the F1,6BP into two 3-carbon intermediates. This reaction catalyzed by aldolase is energetically a “hump” to overcome in the glycolysis direction. In order to get over the energy hump, cells

must increase the concentration the reactant (F1,6BP) and decrease the concentration of the products, which are D-glyceraldehyde-3-phosphate (D-GLYAL3P) and dihydroxyacetone phosphate (DHAP).

Reaction 5

$$\mathrm{DHAP} \rightleftarrows \mathrm{D{-}GLYAL_3P}$$

In the next step, DHAP is converted to DGLYAL3P in a reaction catalyzed by the enzyme triosephosphate isomerase. At this point, the six carbon glucose molecule has been broken down to two identical units of three carbons each – D-GLYAL3P.

From this point forward each reaction of glycolysis occurs twice for each glucose that has fed into the pathway.

Reaction 6

$$\mathrm{D{-}GLYAL_3P + NAD^+ + PiD{-}1\,,3\,BPG + NADH + H^+}$$

Figure 6.9 – Reaction #5 – Triose phosphate isomerase with unstable, toxic intermediate (methyl glyoxal) Image by Ben Carson

In this reaction, D-GLYAL3P is oxidized in the only oxidation step of glycolysis catalyzed by the enzyme glyceraldehyde-3- phosphate dehydrogenase, an oxidoreductase. The aldehyde in this reaction is oxidized, then linked to a phosphate to make an ester – D-1,3-bisphospho-glycerate (D- 1,3BPG). Electrons from the oxidation are donated to NAD+, creating NADH.

NAD+ is a critical constituent in this reaction and is the reason that cells need a fermentation option at the end of the pathway (see below).

Note here that ATP energy was not required to put the phosphate onto the oxidized D-GLYAL3P. The reason for this is because the

energy provided by the oxidation reaction is sufficient for adding the phosphate.

Reaction 7

$$\mathrm{D{-}1,3\,BPG + ADP \rightleftarrows 3\,PG + ATP}$$

The two phosphates in the tiny 1,3BPG molecule repel each other and give the molecule high potential energy. This energy is utilized by the enzyme phosphoglycerate kinase (another transferase) to phosphorylate ADP and make ATP, as well as the product, 3-phosphoglycerate (3-PG). This is an example of a substrate-level phosphorylation: ATP is produced as substrate reacts. Such mechanisms for making ATP require an intermediate with a high enough energy to phosphorylate ADP to make ATP.

Since there are two 1,3 BPGs produced for every glucose, the two ATPs produced in this reaction replenish the two ATPs used to start the cycle and the net ATP count at this point of the pathway is zero.

Reaction 8

$$\mathrm{3{-}PG \rightleftarrows 2{-}PG}$$

Conversion of the 3-PG intermediate to 2-PG (2-phosphoglycerate) occurs by an important mechanism. An intermediate in this readily reversible reaction (catalyzed by phosphoglycerate mutase – a mutase enzyme) is 2,3-BPG. This intermediate, which is stable, is released with low frequency by the enzyme instead of being con- Figure 6.13 – Two routes to formation of 2,3-BPG Figure 6.14 – 2,3- Bisphosphoglycerate (2,3-BPG) Figure 6.12 – Reaction #8 – Conversion of 3-PG to 2-PG verted to 2-PG. 2,3BPG is important because it binds to hemoglobin and stimulates release of oxygen. The molecule can also be made from 1,3-BPG as a product of a reaction catalyzed by bisphophglycerate mutase (Figure 6.13).

Why does this make sense? Cells which are metabolizing glucose rapidly release more 2,3-BPG and, as a result, get more oxygen,

supporting their needs. Notably, cells which are metabolizing rapidly are using oxygen more rapidly and are more likely to be deficient in it.

Reaction 9

$$2-\mathrm{PG} \rightleftarrows \mathrm{PEP} + \mathrm{H_2O}$$

2-PG is converted by enolase (a lyase) to phosphoenolpyruvate (PEP) by removal of water, creating a very high potential energy product.

Reaction 10

$$\mathrm{PEP} + \mathrm{ADP} + \mathrm{H^+} \rightleftarrows \mathrm{PYR} + \mathrm{ATP}$$

Conversion of PEP to pyruvate by pyruvate kinase is the second substrate level phosphorylation of glycolysis, creating ATP. This reaction is what some refer to as the "Big Bang" of glycolysis because there is almost enough energy in PEP to stimulate production of a second ATP, but it is not used. Consequently, this energy is lost as heat. If you wonder why you get hot when you exercise, the heat produced in the breakdown of glucose is a prime contributor and the pyruvate kinase reaction is a major source.

Pyruvate kinase is the third and last enzyme of glycolysis that is regulated (see below). The primary reason this is the case is to be able to prevent this reaction from occurring when cells are making PEP while going through a different process, building glucose via the pathway called gluconeogenesis.

Catabolism of other sugars

Though glycolysis is a pathway focused on the metabolism of glucose and fructose, the fact that other sugars can be readily metabolized into glucose means that glycolysis can be used for

extracting energy from them as well. Galactose is a good example. It is commonly produced in the produced in the body as a result of hydrolysis of lactose, catalyzed by the enzyme known as lactase (Figure 6.17). Deficiency of lactase is the cause of lactose intolerance.

Galactose enters into glycolysis after a few reactions that convert it into glucose-6-phosphate.

Deficiency of galactose conversion enzymes results in accumulation of galactose (from breakdown of lactose). Excess galactose is converted to galactitol, a sugar alcohol. Galactitol in the human eye lens causes it to absorb water and this may be a factor in formation of cataracts.

Pyruvate metabolism

As described in an earlier chapter, pyruvate produced in glycolysis can be oxidized to acetyl-CoA, which is itself oxidized in the citric acid cycle to carbon dioxide. That is not the only metabolic fate of pyruvate, though (Figure 6.23).

Among its other possible fates, pyruvate can be processed anaerobically by fermentation to ethanol (in bacteria and yeasts) or lactic acid (in animals).

Specific glycolysis controls

Control of glycolysis and gluconeogenesis is unusual for metabolic pathways, in that regulation occurs at multiple points. For glycolysis, this involves three enzymes:

1. Hexokinase (Glucose $\rightleftarrows$ G6P)
2. Phosphofructokinase-1 (F6P $\rightleftarrows$ F1,6BP)

3. Pyruvate kinase (PEP ⇄ Pyruvate).

Polysaccharide metabolism

Sugars are metabolized rapidly in the body and that is one of the primary reasons they are used. Managing levels of glucose in the body is very important – too much leads to complications related to diabetes and too little gives rise to hypoglycemia (low blood sugar). Sugars in the body are maintained by three processes – 1) diet; 2) synthesis (gluconeogenesis); and 3) storage. The storage forms of sugars are, of course, the polysaccharides and their metabolism is our next topic of discussion.

Amylose and amylopectin

The energy needs of a plant are much less dynamic than those of animals. Muscular contraction, nervous systems, and information processing in the brain require large amounts of quick energy. Because of this, the polysaccharides stored in plants are somewhat less complicated than those of animals. Plants store glucose for energy in the form of amylose (Figure 6.34) and amylopectin and for structural integrity in the form of cellulose. These structures differ in that cellulose contains glucose units solely joined by β-1,4 bonds, whereas amylose has only α-1,4 bonds and amylopectin has α-1,4 and α-1,6 bonds.

Glycogen

Animals store glucose primarily in liver and muscle in the form of a compound related to amylopectin known as glycogen. The structural differences between glycogen and amylopectin are solely due to the frequency of the α-1,6 branches of glucoses. In glycogen they occur about every 10 residues instead of every 30-50, as in amylopectin (Figure 6.35).

Because glycogen contains so many glucoses, it acts like a battery backup for the body, providing a quick source of glucose when needed and providing a place to store excess glucose when glucose concentrations in the blood rise.

The branching of glycogen is an important feature of the molecule

metabolically as well. Since glycogen is broken down from the "ends" of the molecule, more branches translate to more ends, and more glucose that can be released at once.

7.1 Lipids

On July 11, 2003, the Food and Drug Administration amended its food labeling regulations to require that manufacturers list the amount of *trans* fatty acids on Nutrition Facts labels of foods and dietary supplements, effective January 1, 2006. This amendment was a response to published studies demonstrating a link between the consumption of *trans* fatty acids and an increased risk of heart disease. *Trans* fatty acids are produced in the conversion of liquid oils to solid
fats, as in the creation of many commercial margarines and shortenings. They have been shown to increase the levels of low-density lipoproteins (LDLs)—complexes that are often referred to as bad
cholesterol—in the blood. In this chapter, you will learn about fatty acids and what is meant by a *trans* fatty acid, as well as the difference between fats and oils. You will also learn what cholesterol is and why it is an important molecule in the human body.

Fats and oils, found in many of the foods we eat, belong to a class of biomolecules known as
lipids. Gram for gram, they pack more than twice the caloric content of carbohydrates: the oxidation of fats and oils supplies about 9 kcal of energy for every gram oxidized, whereas the oxidation of carbohydrates supplies only 4 kcal/g. Although the high caloric content of fats may be bad news for the dieter, it says something about the efficiency of nature's designs. Our bodies use carbohydrates, primarily in the form of glucose, for our *immediate* energy needs. Our capacity for storing carbohydrates for later use is limited to tucking away a bit of glycogen in the liver or in muscle tissue. We store our *reserve* energy in lipid form, which requires far less space than the same amount of energy stored in carbohydrate form.

Lipids have other biological functions besides energy storage. They are a major component of the membranes of the 10 trillion cells in our bodies. They serve as protective padding and insulation for vital organs. Furthermore, without lipids in our diets, we would be deficient in the fat-soluble
vitamins A, D, E, and K.
Lipids are not defined by the presence of specific functional groups, as
carbohydrates, nucleic acids and proteins are, but by a physical property—solubility. Compounds isolated from body tissues are classified as lipids if they are more soluble in organic solvents, such as dichloromethane, than in water. By this criterion, the lipid category includes not only fats and oils, which are esters of the trihydroxy alcohol glycerol and fatty acids, but also compounds that incorporate functional groups derived from phosphoric acid, carbohydrates, or amino alcohols, as well as steroid compounds such as cholesterol.

(Figure 17.0.1 presents one scheme for classifying the various kinds of
lipids). We will discuss the various kinds of lipids by considering one subclass at a time and pointing out structural similarities and differences as we go.

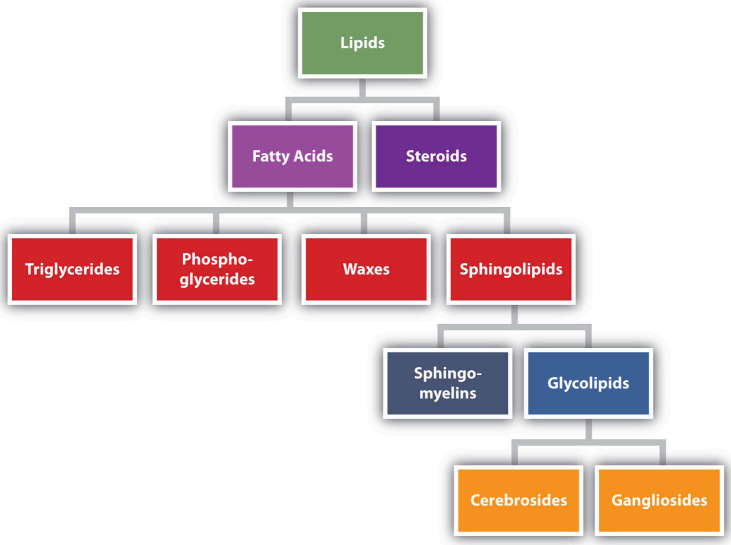
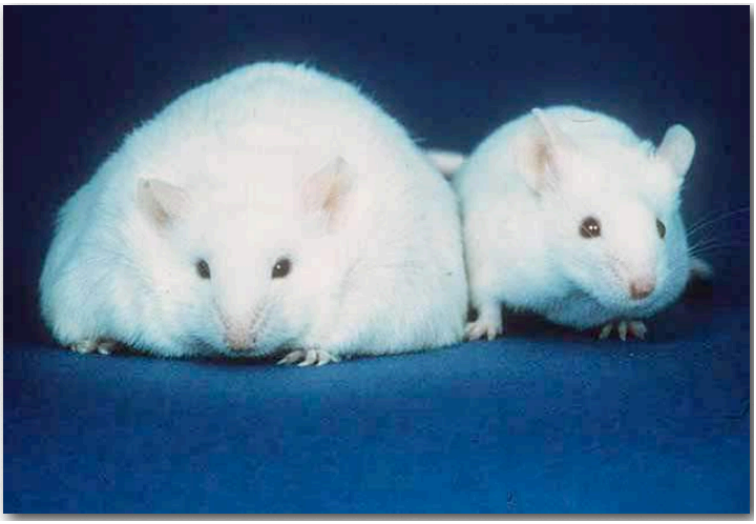


Figure 17.0.1 17.0.1

7.2: Structure and Function - Lipids and Membranes



Lipids are a diverse group of molecules that all share the characteristic that at least a portion of them is hydrophobic. Lipids play many roles in cells, including serving as energy storage (fats/ oils), constituents of membranes (glycerophospholipids, sphingolipids, cholesterol), hormones (steroids), vitamins (fat soluble), oxygen/ electron carriers (heme), among others. For lipids that are very hydrophobic, such as fats/ oils, movement and storage in the aqueous environment of the body requires special structures. Other, amphipathic lipids, such as glycerophospholipids and sphingolipids spontaneously organize themselves into lipid bilayers when placed in water. Interestingly, major parts of many lipids can be constructed from acetyl-CoA, so they can be made from other nutrients including carbohydrates. That said, some lipids are dietarily essential.

Fatty acids

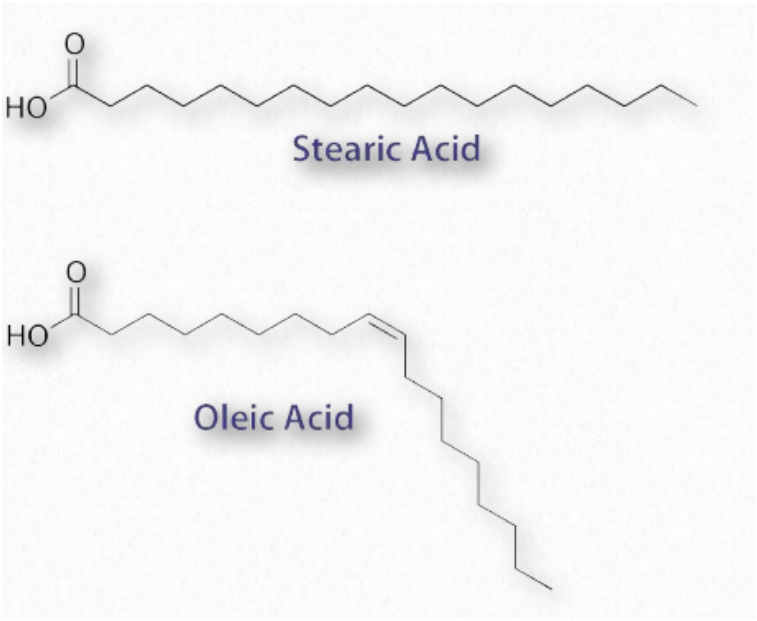


Figure 2.190 – Saturated fatty acid (stearic acid) and unsaturated fatty acid (oleic acid)

The most ubiquitous lipids in cells are the fatty acids. Found in triglycerides and glycerophospholipids, and serving as as membrane anchors for proteins and other biomolecules, fatty acids are important for energy storage, membrane structure, and as precursors of most classes of lipids. Fatty acids, as can be seen from Figure 2.190 are characterized by a polar or charged, hydrophilic head group and a long, hydrocarbon, and thus hydrophobic tail.

Fatty acids with hydrocarbon tails that lack any double bonds are described as saturated, while those with one or more double bonds in their tails are known as unsaturated fatty acids. The effect of double bonds on the fatty acid tail is to introduce a kink, or bend, in the tail, as shown for oleic acid.

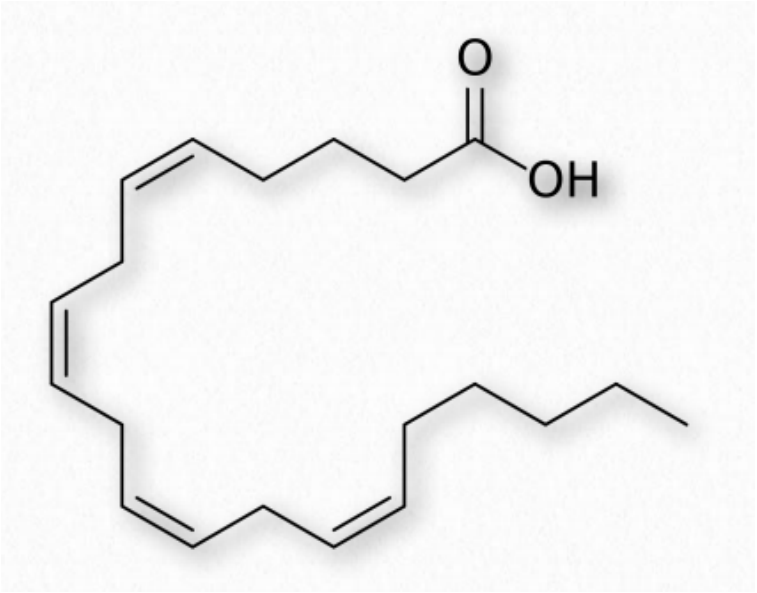


Figure 2.191 – Arachidonic acid – A polyunsaturated fatty acid Wikipedia

The saturated fatty acid named stearic acid, by contrast, has a straight hydrocarbon tail. Figures 2.190-2.194 show the most common saturated and unsaturated fatty acids. Fatty acids with unsaturated tails pack less tightly together, have weaker London forces between them, and thus have a lower melting temperature than those with saturated tails of the same length. Shorter tails also decrease melting temperature. These properties carry over to the fats/oils containing them.

Common name	Chemical structure	
Caprylic acid	$CH_3(CH_2)_6COOH$	8
Capric acid	$CH_3(CH_2)_8COOH$	10
Lauric acid	$CH_3(CH_2)_{10}COOH$	12
Myristic acid	$CH_3(CH_2)_{12}COOH$	14
Palmitic acid	$CH_3(CH_2)_{14}COOH$	16
Stearic acid	$CH_3(CH_2)_{16}COOH$	18
Arachidic acid	$CH_3(CH_2)_{18}COOH$	20
Behenic acid	$CH_3(CH_2)_{20}COOH$	22
Lignoceric acid	$CH_3(CH_2)_{22}COOH$	24
Cerotic acid	$CH_3(CH_2)_{24}COOH$	26

Figure 2.192 – Saturated fatty acids. Number of carbons in right column Wikipedia

Fatty acids with more than one double bond are called polyunsaturated. Plants are excellent sources of unsaturated and polyunsaturated fatty acids. The position of the double bond(s) in fatty acids has important considerations both for their biosynthesis and for their actions in the body. Biochemically, the double bonds found in fatty acids are nearly always in the cis configuration. So-called trans fats arise as a chemical by-product of partial hydrogenation of vegetable oil, and very occasionally naturally.

Name	Double Bond Info	
Myristoleic acid	*cis*-Δ^9	14:1
Palmitoleic acid	*cis*-Δ^9	16:1
Sapienic acid	*cis*-Δ^6	16:1
Oleic acid	*cis*-Δ^9	18:1
Elaidic acid	*trans*-Δ^9	18:1
Vaccenic acid	*trans*-Δ^{11}	18:1
Linoleic acid	*cis,cis*-Δ^9,Δ^{12}	18:2
Linoelaidic acid	*trans,trans*-Δ^9,Δ^{12}	18:2
α-Linolenic acid	*cis,cis,cis*-$\Delta^9,\Delta^{12},\Delta^{15}$	18:3
Arachidonic acid	*cis,cis,cis,cis*-$\Delta^5\Delta^8,\Delta^{11},\Delta^{14}$	20:4
Eicosapentaenoic acid	*cis,cis,cis,cis,cis*-$\Delta^5,\Delta^8,\Delta^{11},\Delta^{14},\Delta^{17}$	20:5
Erucic acid	*cis*-Δ^{13}	22:1
Docosahexaenoic acid	*cis,cis,cis,cis,cis,cis*-$\Delta^4,\Delta^7,\Delta^{10},\Delta^{13},\Delta^{16},\Delta^{19}$	22:6

Figure 2.193 – Unsaturated fatty acids. Right column Indicates number of carbons and double bonds Wikipedia

In humans, consumption of trans is associated with higher low density lipoprotein (LDL) levels and lower high density lipoprotein (HDL) levels. Each of these effects is thought to contribute to the risk of developing coronary artery disease.

The most common fatty acids in our body include palmitate, stearate, oleate, linolenate, linoleate, and arachidonate.

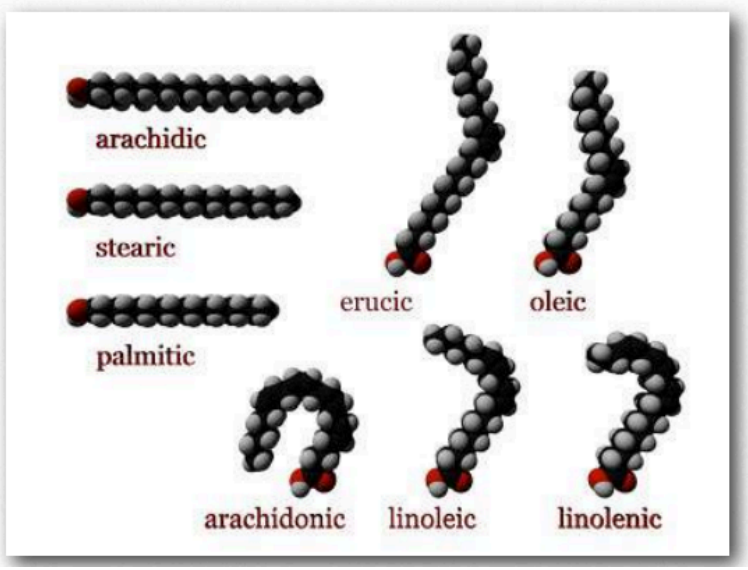


Figure 2.194 – Fatty acid models. Carboxyl end labeled in red Wikipedia

Numbering

Figure 2.195 shows two different systems for locating double bonds in a fatty acid. The ω system counts carbons starting with the methyl end (shown in red) while the Δ system counts from the carboxyl end (shown in blue). For example, an ω-3 (omega 3) fatty acid would have a double bond at the third carbon from the methyl end. In the Δ system, a fatty acid that has a cis double bond at carbon 6, counting from the carboxyl end, would be written as cis-Δ6.

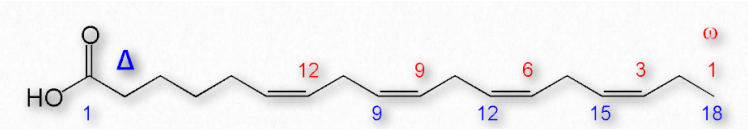


Figure 2.195 – Δ and ω numbering systems for fatty acids Image by Pehr Jacobson

Fatty acids are described as essential fatty acids if they must be in

the diet and nonessential fatty acids if the organism can synthesize them. Humans and other animals lack the desaturase enzymes necessary to make double bonds at positions greater than Δ-9, so fatty acids with double bonds beyond this position must be obtained in the diet. Linoleic acid and linolenic acid both fall in this category. Related unsaturated fatty acids can be made from these fatty acids, so the presence of linoleic and linolenic acids in the diet eliminates the need to have all unsaturated fatty acids in the diet. Both linoleic and linolenic acid contain 18 carbons, but linoleic acid is an ω-6 fatty acid, whereas linolenic acid is an ω-3 fatty acid. Notably, ω-6 fatty acids tend to be proinflammatory, whereas ω-3 fatty acids are lesser so.

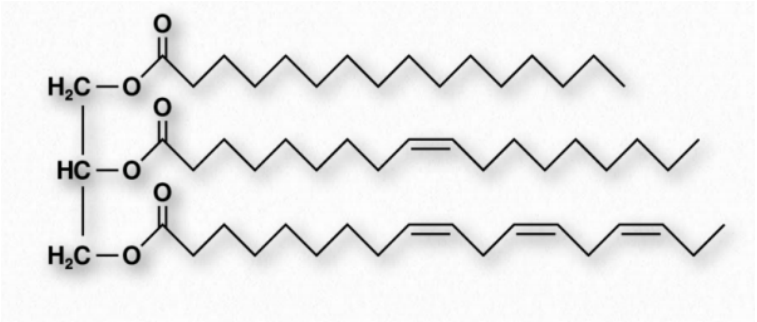


Figure 2.196 – Structure of a triglyceride

Triglycerides

Fats and oils are the primary energy storage forms of animals and are also known as triacylglycerols and triglycerides, since they consist of a glycerol molecule linked via ester bonds to three fatty acids (Figure 2.196). Fats and oils have the same basic structure. We give the name fat to those compounds that are solid at room temperature and the name oil to those that are liquid at room temperature. Note that biological oils are not the same as petroleum oils.

Increasing the amount of unsaturation in a fat decreases its melting temperature. Organisms like fish, which live in cool environments, have fats with more unsaturation. So fish oil from

cold-water species, for instance, has more unsaturation and is marketed as a rich source of healthy fats.

Adipocytes

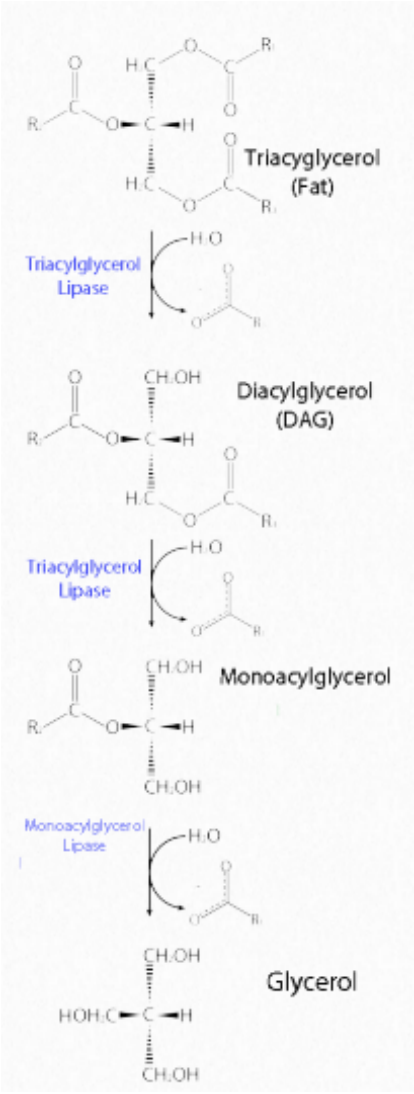


Figure 2.197 – Lipase action on a fat Image by Aleia Kim

Fats are stored in the body in specialized cells known as adipocytes. Enzymes known as lipases release fatty acids from fats

by hydrolysis reactions (Figure 2.197). Triacylglycercol lipase (pancreatic – Figure 2.198) is able to cleave the first two fatty acids from the fat. A second enzyme, monoacylglycerol lipase, cleaves the last fatty acid.

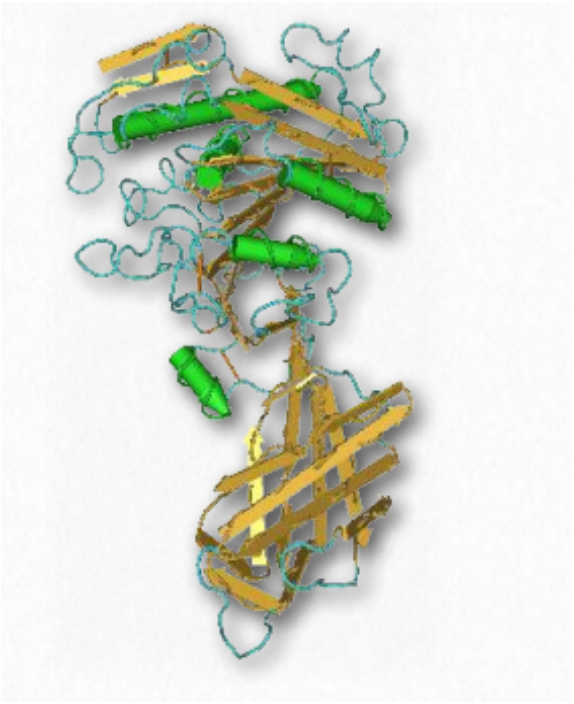


Figure 2.198 – Pancreatic lipase

Glycerophospholipids

Glycerophospholipids (phosphoglycerides) are important components of the lipid bilayer of cellular membranes. Phosphoglycerides are structurally related to fats, as both are derived from phosphatidic acid (Figure 2.199). In these structures

various groups such as ethanolamine, serine, choline, inositol, and others (Figure 2.200) are linked to the phosphate of phosphatidic acid and then to one carbon of glycerol, via an ester linkage. The second and third positions on the glycerol contain ester links to fatty acid, in the same manner as for triglycerides.

All of these compounds form lipid bilayers in aqueous solution , due to their amphiphilic nature. The phosphate-containing head group is hydrophilic, while the acyl chains from the fatty acids are hydrophobic. These molecules spontaneously form bilayer structures in watery environments with the hydrophilic portions interacting strongly with water, pushing the fatty acyl chains to the interior of the structure.

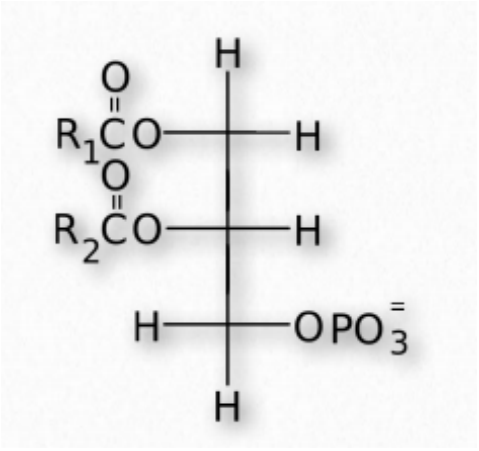


Figure 2.199 – Structure of phosphatidic acid. R1 and R2 are alkyl groups of fatty acids.

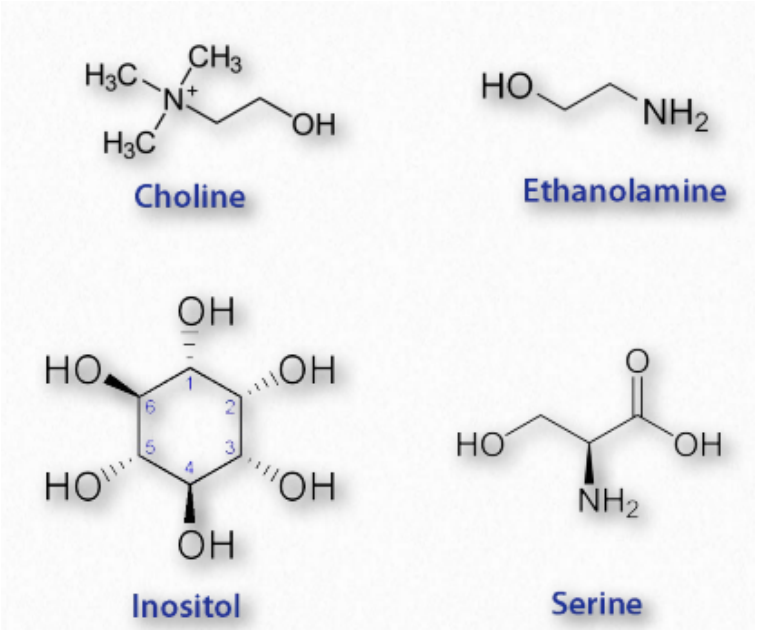


Figure 2.200 – Four common components of phosphatides Wikipedia

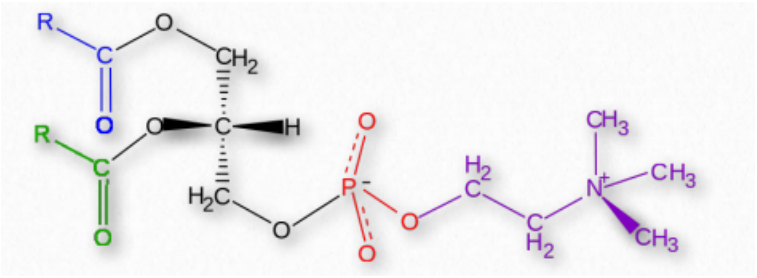


Figure 2.201 – Phosphatidylcholine

Sphingolipids

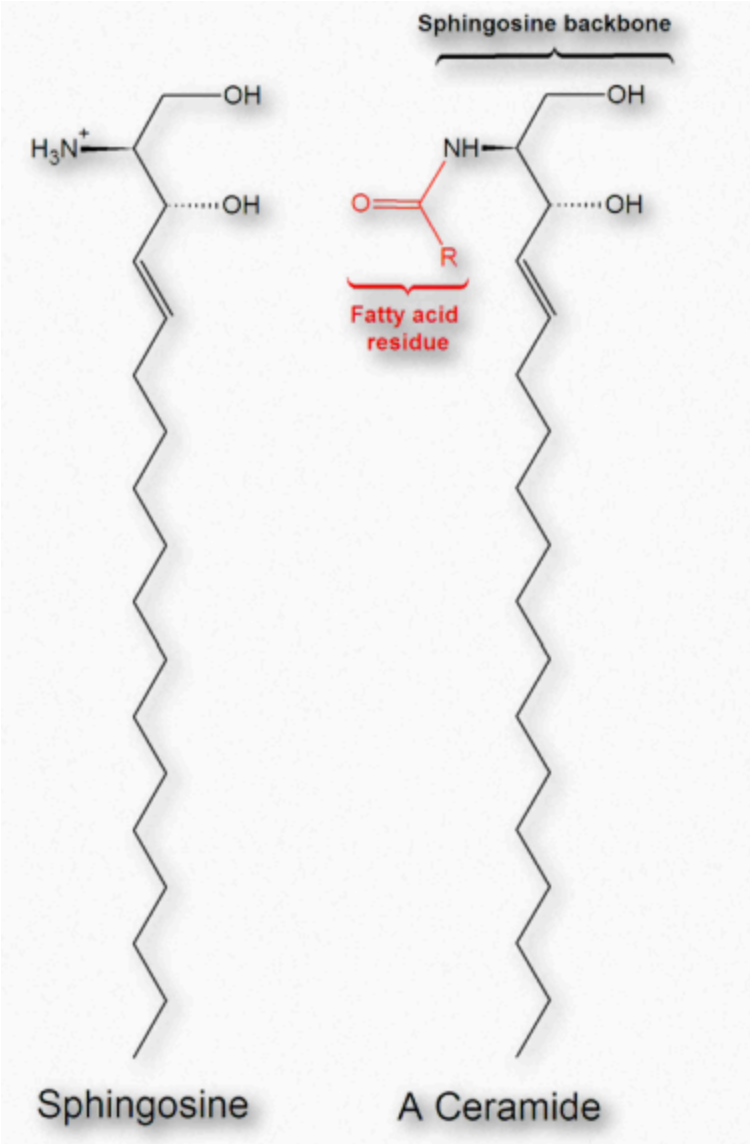


Figure 2.210 – Sphingosine and a ceramide made from it Wikipedia

Fatty acids are also components of a broad class of molecules called sphingolipids. Sphingolipids are structurally similar to glycerophospholipids, though they are synthesized completely

independently of them starting with palmitic acid and the amino acid serine. Sphingolipids are named for the amino alcohol known as sphingosine (Figure 2.210), though they are not directly synthesized from it. Figure 2.211 shows the generalized structure of sphingolipids.

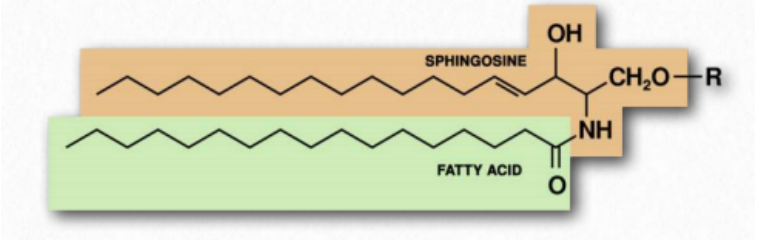


Figure 2.211 Schematic structure of a sphingolipid

If the R-group is a hydrogen, the molecule is called a ceramide. When the R-group is phosphoethanolamine the resulting molecule is sphingomyelin, an important component of the myelin sheath and lipid membranes. If a single, simple sugar is instead added, a cerebroside is created (Figure 2.212). Addition of a complex oligosaccharide creates a ganglioside.

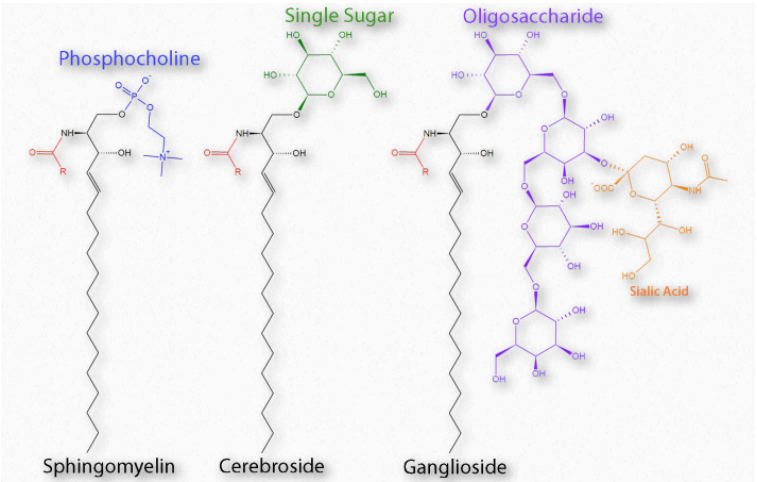


Figure 2.212 – Categories of Sphingolipid Wikipedia

Eicosanoids

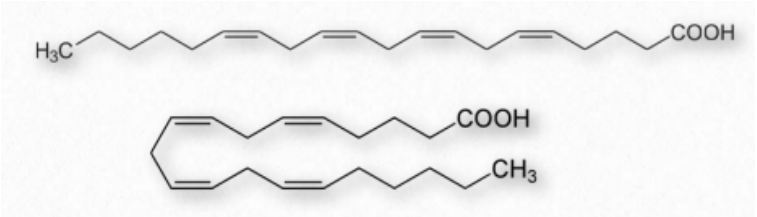


Figure 2.213 – Arachidonic acid drawn as straight (top) and bent (bottom)

Fatty acids made from omega-6 and omega-3 fatty acids include three important fatty acids containing 20 carbons. They include arachidonic acid (an ω-6 fatty acid with four double bonds (Δ-5,8,11,14) – Figure 2.213), eicosapentaenoic acid (an ω-3 fatty acid with five double bonds, and dihomo-γ-linolenic acid (an ω-6 fatty acid with three double bonds). The class of compounds known as eicosanoids is made by oxidation of these compounds. Subclasses include include prostaglandins, prostacyclins, thromboxanes, lipoxins, leukotrienes, and endocannabinoids (Figures 2.214-2.219). Eicosanoids play important roles affecting inflammation, immunity, mood, and behavior.

Prostaglandins

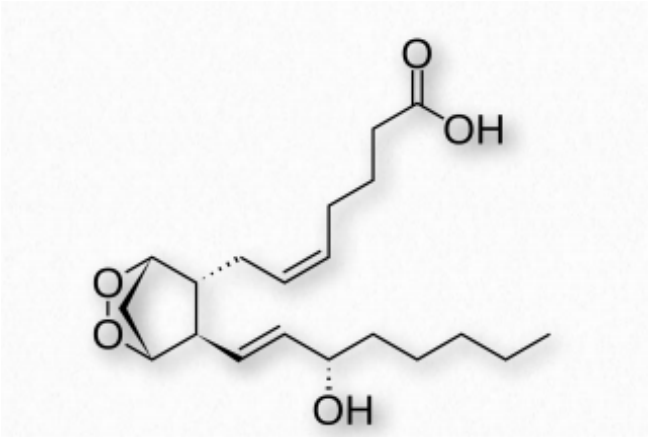


Figure 2.214 – Prostaglandin PGH_2

A collection of molecules acting like hormones, prostaglandins are derived from arachidonic acid and have many differing (even

conflicting) physiological effects. These include constriction or dilation of vascular smooth muscle cells, induction of labor, regulation of inflammation, and action on the thermoregulatory center of the hypothalamus to induce fever, among others.

Prostaglandins are grouped with the thromboxanes (below) and prostacyclins (below), as prostanoids. The prostanoids, which all contain 20 carbons are a subclass of the eicosanoids. Prostaglandins are found in most tissues of higher organisms. They are autocrine or paracrine compounds produced from essential fatty acids.

Interesting prostaglandins

PGD_2 – inhibits hair follicle growth, vasodilator, causes bronchial constriction, higher in lungs of asthmatics than others.

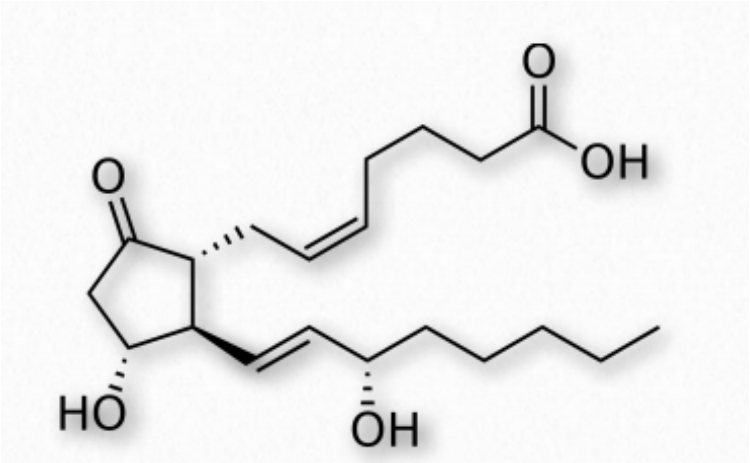


Figure 2.215 Prostaglandin E

PGE_2 (Figure 2.215) – exerts effects in labor (soften cervix, uterine contraction), stimulates bone resorption by osteoclasts, induces fever, suppresses T-cell receptor signaling, vasodilator, inhibits release of noradrenalin from sympathetic nerve terminals. It is a potent activator of the Wnt signaling pathway.

A prostaglandin can have opposite effects, depending on which receptor it binds to. Binding of PGE2 to the EP1 receptor causes bronchoconstriction and smooth muscle contraction, whereas

binding of the same molecule to the EP2 receptor causes bronchodilation and smooth muscle relaxation.

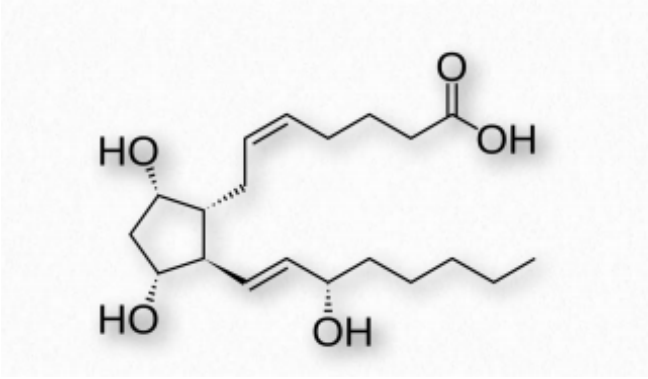


Figure 2.216 – Prostaglandin $F_{2\alpha}$

$PGF_{2\alpha}$ (Figure 2.216)- uterine contractions, induces labor, bronchoconstriction.

PGI_2 – vasodilation, bronchodilation, inhibition of platelet aggregation.

Thromboxanes

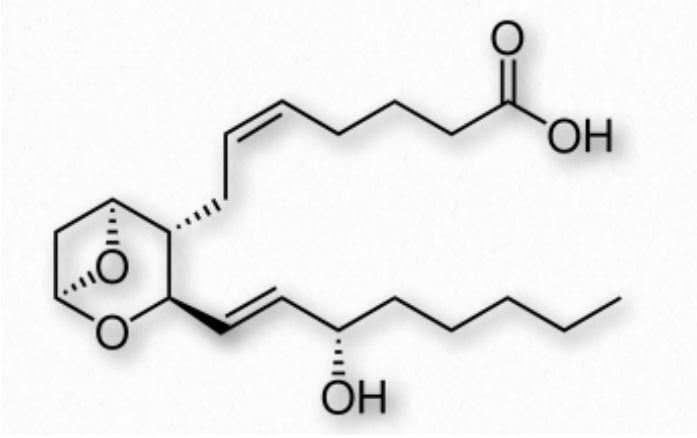


Figure 2.217 Thromboxane A2 [2]

Thromboxanes play roles in blood clot formation and named for their role in thrombosis. They are potent vasoconstrictors and facilitate platelet aggregation. They are synthesized in platelets, as well. The anti-clotting effects of aspirin have their roots in the

inhibition of synthesis of PGH2, which is the precursor of the thromboxanes. The most common thromboxanes are A2 (Figure 2.217) and B2.

Prostacyclin

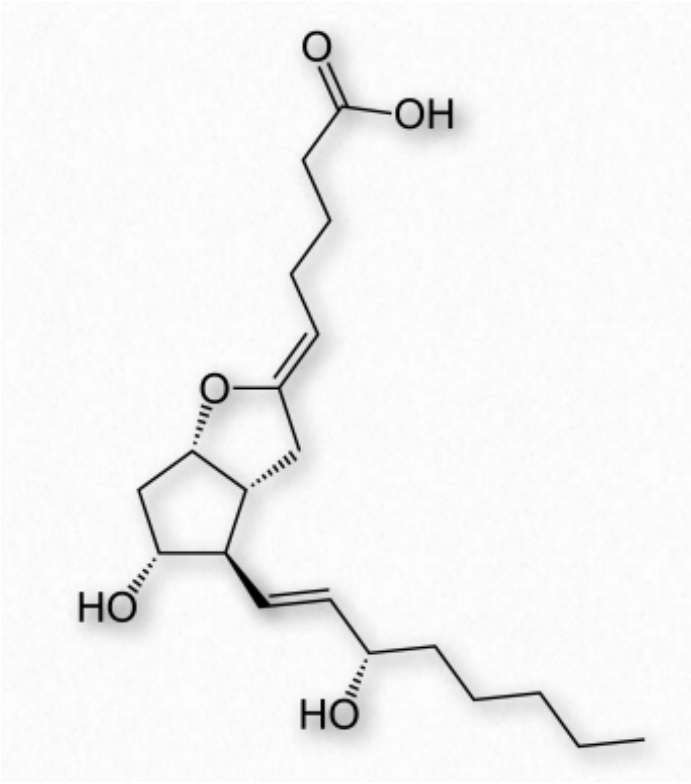


Figure 2.218 – Prostacyclin

Prostacyclin (also known as prostaglandin I2 or PGI2 – Figure 2.218) counters the effects of thromboxanes, inhibiting platelet activation and acting as vasodilators. It is produced from PGH2 by action of the enzyme prostacyclin synthase.

Leukotrienes

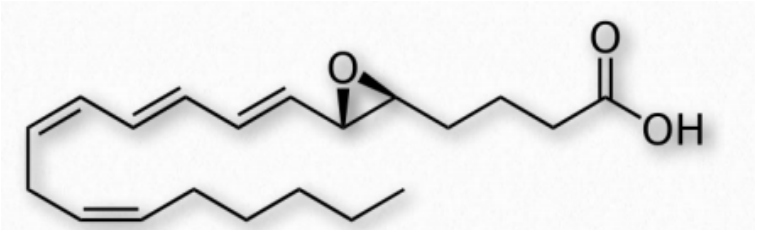


Figure 2.219 – Leukotriene A_4 (LTA_4)

Another group of eicosanoid compounds are the leukotrienes (Figure 2.219). Like prostaglandins, leukotrienes are made from arachidonic acid. The enzyme catalyzing their formation is a dioxygenase known as arachidonate 5-lipoxygenase. Leukotrienes are involved in regulating immune responses. They are found in leukocytes and other immunocompetent cells, such as neutrophils, monocytes, mast cells, eosinophils, and basophils. Leukotrienes are associated with production of histamines and prostaglandins, which act as mediators of inflammation. Leukotrienes also trigger contractions in the smooth muscles of the bronchioles. When overproduced, they may pay a role in asthma and allergic reactions. Some treatments for asthma aim at inhibiting production or action of leukotrienes.

Cholesterol

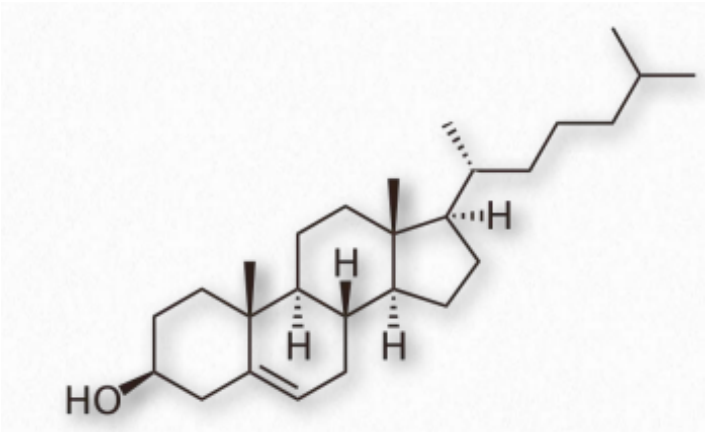


Figure 2.220 – Structure of cholesterol

Arguably, no single biomolecule has generated as much

discussion and interest as has cholesterol (Figure 2.220). Certainly, from the perspective of the Nobel Prize committee, no small molecule even comes close, with 13 people having been awarded prizes for work on it. Evidence for cholesterol's importance comes from the study of brain tissue where it comprises 10-15% of the dry mass.

Membrane flexibility

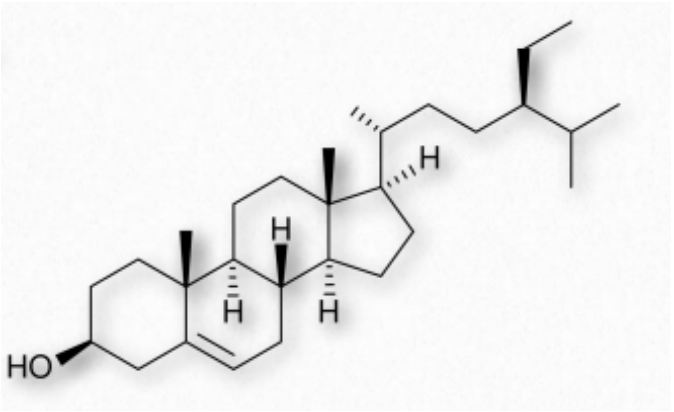


Figure 2.221 – Sitosterol – A phytosterol

In animal cells, cholesterol provides for membrane flexibility that allows for cellular movement that is in contrast to plant and bacterial cells with fixed structures. Cholesterol is made in many cells of the body, with the liver making the greatest amount. The anabolic pathway leading to synthesis of cholesterol is known as the isoprenoid pathway and branches of it lead to other molecules including other fat-soluble vitamins.

Figure 2.222 – Margarine – A common source of trans fat Wikipedia

Cholesterol is only rarely found in prokaryotes and is found in only trace amounts in plants. Instead, plants synthesize similar compounds called phytosterols (Figure 2.221). On average, the body of a 150 pound adult male makes about 1 gram of cholesterol per day, with a total content of about 35 grams.

Packaging

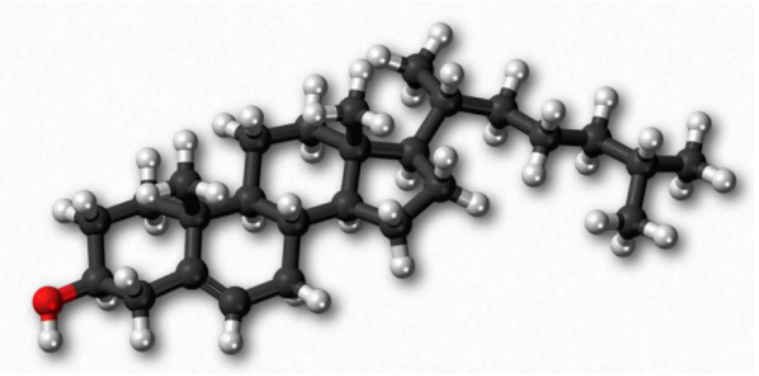


Figure 2.223 – Cholesterol – Ball and stick model

Cholesterol's (and other lipids') hydrophobicity requires special packaging into lipoprotein complexes (called chylomicrons, VLDLs,

IDLs, LDLs, and HDLs) for movement in the lymph system and bloodstream. Though cholesterol can be made by cells, they also take it up from the blood supply by absorbing cholesterol-containing LDLs directly in a process called receptor-mediated endocytosis.

Oxidative damage to LDLs can lead to formation of atherosclerotic plaques. This is why cholesterol has gotten such a negative image in the public eye. The liver excretes cholesterol through the bile into the digestive system, but the compound is recycled from there.

Reducing cholesterol levels

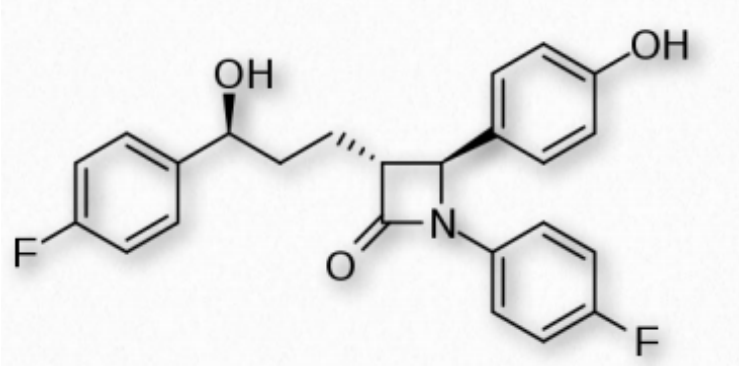


Figure 2.224 – Ezetimibe – An inhibitor of cholesterol absorption

Strategies for reducing cholesterol in the body focus primarily on three areas – reducing consumption, reducing endogenous synthesis, and reducing the recycling via the digestive tract. Dietary considerations about consumption of fats are currently debated. Dietary trans fats, though, quite clearly correlate with incidence of coronary heart disease. Consumption of vegetables may provide some assistance with reducing levels of cholesterol recycled in the digestive system, because plant phytosterols compete with cholesterol for reabsorption and when this happens, a greater percentage of cholesterol exits the body in the feces. Drugs related to penicillin are also used to inhibit cholesterol recycling. One of these is ezetimibe, shown in Figure 2.224.

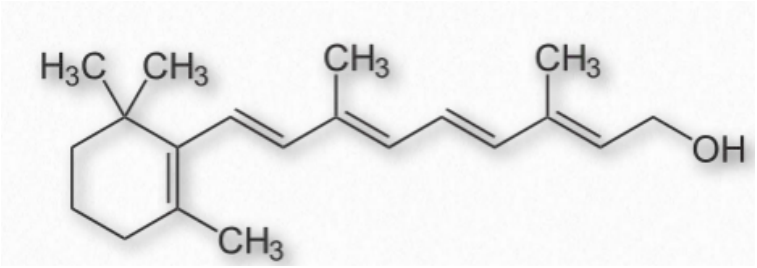


Figure 2.225 – All-trans retinol

Genetic defects in the cholesterol movement system are a cause of the rare disease known as familial hypercholesterolemia in which the blood of afflicted individuals contains dangerously high levels of LDLs. Left untreated, the disease is often fatal in the first 10-20 years of life. While LDLs have received (and deserve) a bad rap, another group of lipoprotein complexes known as the HDLs (high density lipoprotein complexes) are known as "good cholesterol" because their levels correlate with removal of debris (including cholesterol) from arteries and reduce inflammation.

Membrane function

In membranes, cholesterol is important as an insulator for the transmission of signals in nerve tissue. It also helps manage fluidity of membranes over a wide range of temperatures. Stacked in the lipid bilayer, cholesterol decreases a membrane's fluidity and its permeability to neutral compounds, as well as protons and sodium ions. Cholesterol may play a role in signaling by helping with construction of lipid rafts within the cell membrane.

Lipoprotein complexes and lipid movement in the body

Lipoprotein complexes are combinations of apolipoproteins and lipids bound to them that solubilize fats and other non-polar molecules, such as cholesterol, so they can travel in the bloodstream between various tissues of the body. The apolipoproteins provide the emulsification necessary for this. Lipoprotein complexes are formed in tiny "balls" with the water soluble apolipoproteins on the outside and non-polar lipids, such as fats, cholesteryl esters, and fat soluble vitamins on the inside.

They are categorized by their densities. These include (from highest density to the lowest) high density lipoproteins (HDLs), Low Density Lipoproteins (LDLs), Intermediate Density Lipoproteins (IDLs), Very Low Density Lipoproteins (VLDLs) and the chylomicrons. These particles are synthesized in the liver and small intestines.

Apolipoproteins

Name	Lipoprotein Complex(es)	Function
ApoA-I	HDL	Promotes Fat Movement to Liver
ApoA-II	HDL	Inhibit LCAT
ApoA-IV	Chylomicrons / HDL	Activate LCAT
AboB-48	Chylomicrons	Cholesterol Transport
AboB-100	VLDL / LDL	Bind LDL Receptor
ApoC-I	VLDL / LDL	Unknown
ApoC-II	AII	Activate Lipoprotein Lipase
ApoC-III	AII	Inhibit Lipoprotein Lipase
ApoD	HDL	Unknown
ApoE	VLDL / Chylomicrons / HDL	Clearance of Chylomicrons Remnants and VLDLs

Figure 2.256 – Apolipoproteins

Each lipoprotein complex contain a characteristic set of apolipoproteins, as shown in Figure 2.256. ApoC-II and ApoC-III are notable for their presence in all the lipoprotein complexes and the roles they play in activating (ApoC-II) or inactivating (ApoC-III) lipoprotein lipase. Lipoprotein lipase is a cellular enzyme that

catalyzes the breakdown of fat from the complexes. ApoE (see below) is useful for helping the predict the likelihood of the occurrence of Alzheimer's disease in a patient.

Gene editing

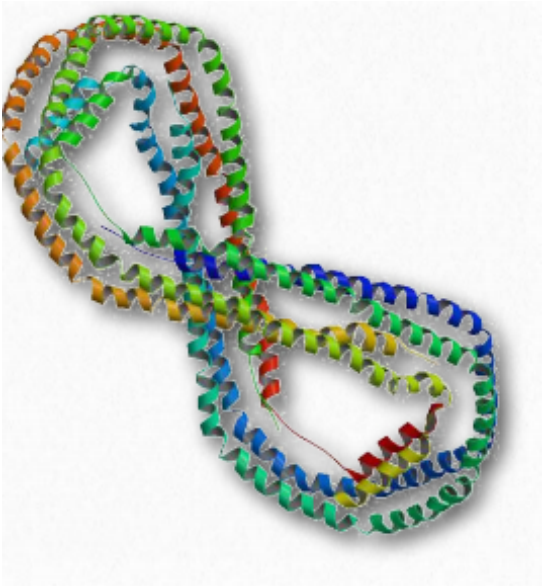


Figure 2.257 – ApoA-I

ApoB-48 and ApoB-100 are interesting in being coded by the same gene, but a unique mRNA sequence editing event occurs that converts one into the other. ApoB-100 is made in the liver, but ApoB-48 is made in the small intestine. The small intestine contains an enzyme that deaminates the cytidine at nucleotide #2153 of the common mRNA. This changes it to a uridine and changes the codon it is in from CAA (codes for glutamine) to UAA (stop codon). The liver does not contain this enzyme and does not make the change in the mRNA. Consequently, a shorter protein is synthesized in the intestine (ApoB-48) than the one that is made in the liver (ApoB-100) using the same gene sequence in the DNA.

ApoE and Alzheimer's disease

ApoE is a component of the chylomicrons and is also found in

brain, macrophages, kidneys, and the spleen. In humans, it is found in three different alleles, E2, E3, and E4. The E4 allele (present at about 14% of the population) is associated with increased likelihood of contracting Alzheimer's disease. People heterozygous for the allele are 3 times as likely to contract the disease and those homozygous for it are 15 times as likely to do so. It is not known why this gene or allele is linked to the disease. The three alleles differ only slightly in amino acid sequence, but the changes do cause notable structural differences. The E4 allele is associated with increased calcium ion levels and apoptosis after injury. Alzheimer's disease is associated with accumulation of aggregates of the β-amyloid peptide. ApoE does enhance the proteolytic breakdown of it and the E4 isoform is not as efficient in these reactions as the other isoforms.

7.3: Notable Lipids

Fat-soluble Vitamins

Vitamin A

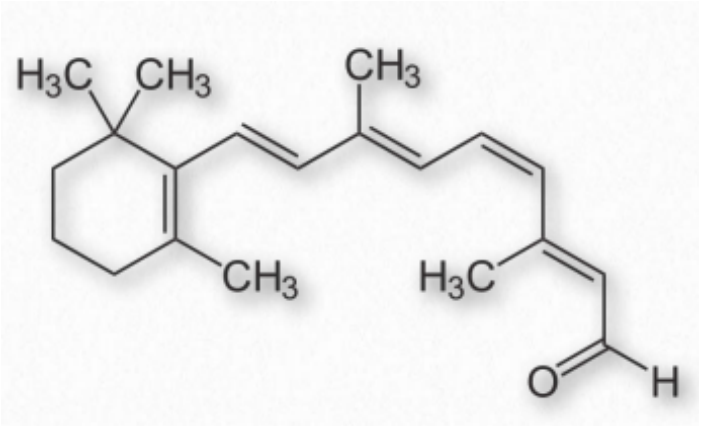


Figure 2.226 – 11-cis retinal

Vitamin A comes in three primary chemical forms, retinol (storage in liver – Figure 2.225), retinal (role in vision – Figure 2.226), and retinoic acid (roles in growth and development). All vitamin A forms are diterpenoids and differ only in the chemical form of the terminal group. Retinol is mostly used as the storage form of the vitamin.

Retinol is commonly esterified to a fatty acid and kept in the liver. In high levels, the compound is toxic. Retinol is obtained in the body by hydrolysis of the ester or by reduction of retinal. Importantly, neither retinal nor retinol can be made from retinoic acid. Retinoic acid is important for healthy skin and teeth, as well as bone growth. It acts in differentiation of stem cells through a specific cellular retinoic acid receptor.

Sources

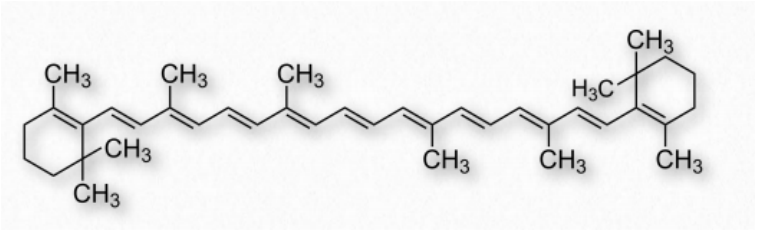


Figure 2.227 – β-Carotene

Good sources of vitamin A are liver and eggs, as well as many plants, including carrots, which have a precursor, β-carotene (Figure 2.227) from which retinol may be made by action of a dioxygenase.

Light sensitivity The conjugated double bond system in the side chain of vitamin A is sensitive to light and can flip between cis and trans forms on exposure to it. It is this response to light that makes it possible for retinal to have a role in vision in the rods and cones of the eyes. Here, the aldehyde form (retinal) is bound to the protein rhodopsin in the membranes of rod and cone cells.

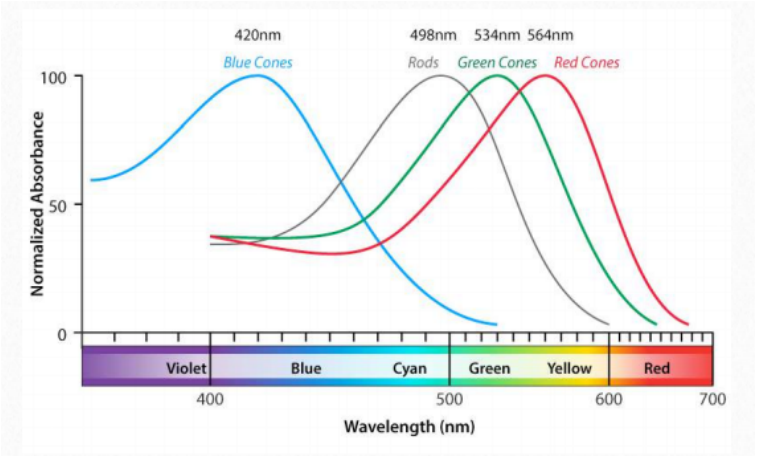


Figure 2.228 – Color sensitivity for cones and rods Image by Aleia Kim

When exposed to light of a particular wavelength, the "tail" of the retinal molecule will flip back and forth from cis to trans at the double bond at position 11 of the molecule. When this happens,

a nerve signal is generated that signals the brain of exposure to light. Slightly different forms of rhodopsin have different maximum absorption maxima allowing the brain to perceive red, green and blue specifically and to assemble those into the images we see (Figure 2.228). Cones are the cells responsible for color vision, whereas rods are mostly involved in light detection in low light circumstances.

Deficiency and surplus

Deficiency of vitamin A is common in developing countries and was inspiration for the design and synthesis of the genetically-modified golden rice, which is used as a source of vitamin A to help prevent blindness in children. Overdose of vitamin A, called hypervitaminosis A is dangerous and can be fatal. Excess vitamin A is also suspected to be linked to osteoporosis. In smokers, excess vitamin A is linked to an increased rate of lung cancer, but non-smokers have a reduced rate.

Vitamin D

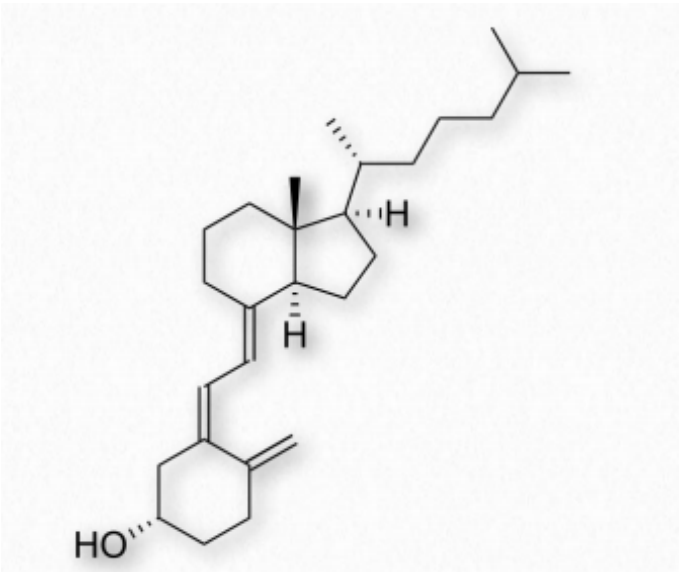


Figure 2.229 – Cholecalciferol – Vitamin D3

The active form of vitamin D plays important roles in the intestinal absorption of calcium and phosphate and thus in healthy

bones. Technically, vitamin D isn't even a vitamin, as it is a compound made by the body. Rather, it behaves more like a hormone.

Derived ultimately from cholesterol, vitamin D can be made in a reaction catalyzed by ultraviolet light. In the reaction, the intermediate 7-dehydrocholesterol is converted to cholecalciferol (vitamin D3) by the uv light (Figure 2.229). The reaction occurs most readily in the bottom two layers of the skin shown in Figure 2.230.

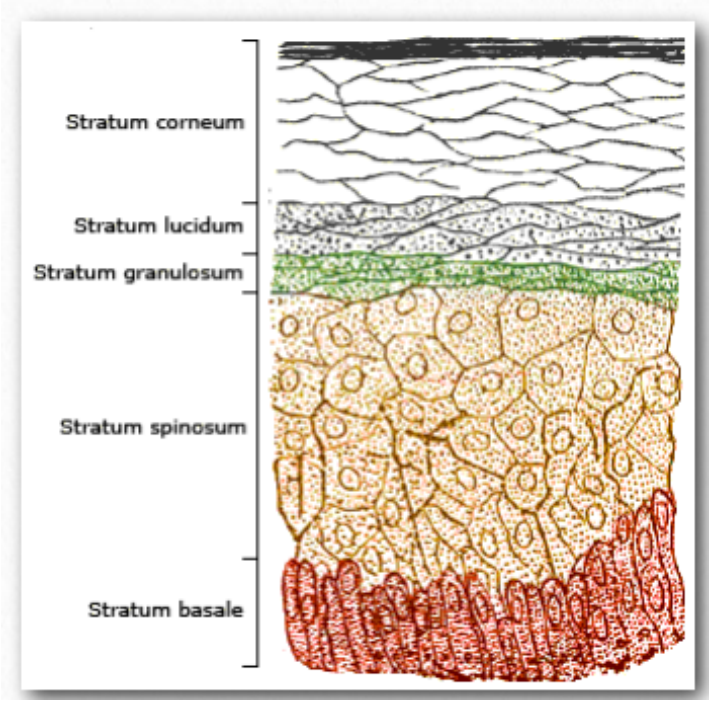


Figure 2.230 – Layers of the skin. Outside is at top.

Forms of vitamin D

Five different compounds are referred to as vitamin D. They are

Vitamin D1 – A mixture of ergocalciferol and lumisterol

Vitamin D2 – Ergocalciferol

Vitamin D3 – Cholecalciferol Vitamin

D4 – 22-Dihydroergocalciferol Vitamin

D5 – Sitocalciferol

Vitamin D3 is the most common form used in vitamin supplements and it and vitamin D2 are commonly obtained in the diet, as well. The active form of vitamin D, calcitriol (Figure 2.231), is made in the body in controlled amounts. This proceeds through two steps from cholecalciferol. First, a hydroxylation in the liver produces calcidiol and a second hydroxylation in the kidney produces calcitriol. Monocyte macrophages can also synthesize vitamin D and they use is as a cytokine to stimulate the innate immune system.

Mechanism of action

Calcitriol moves in the body bound to a vitamin D binding protein, which delivers it to target organs. Calcitriol inside of cells acts by binding a vitamin D receptor (VDR), which results in most of the vitamin's physiological effects. After binding calcitriol, the VDR migrates to the nucleus where it acts as a transcription factor to control levels of expression of calcium transport proteins (for example) in the intestine. Most tissues respond to VDR bound to calcitriol and the result is moderation of calcium and phosphate levels in cells.

Deficiency/excess

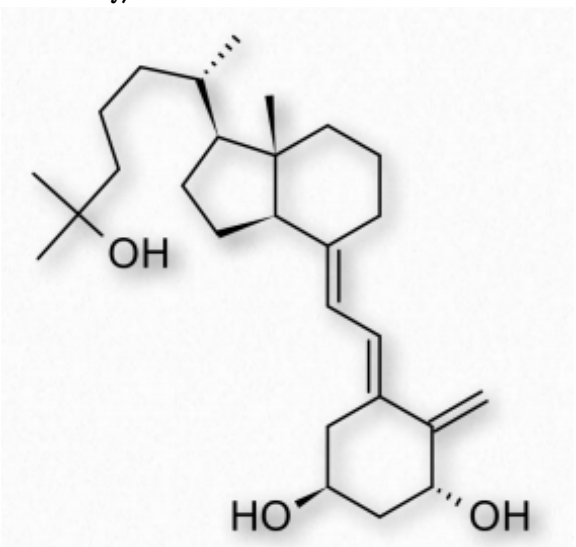


Figure 2.231 – Calcitriol – Active form of vitamin D

Deficiency of vitamin D is a cause of the disease known as rickets, which is characterized by soft, weak bones and most commonly is found in children. It is not common in the developed world, but elsewhere is of increasing concern.

Excess of vitamin D is rare, but has toxic effects, including hypercalcemia, which results in painful calcium deposits in major organs. Indications of vitamin D toxicity are increased urination and thirst. Vitamin D toxicity can lead to mental retardation and many other serious health problems.

Vitamin E

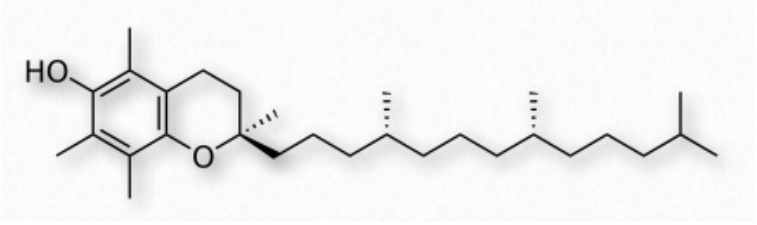
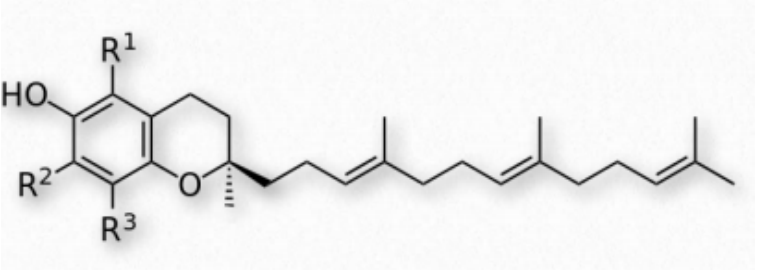


Figure 2.232 α-tocopherol – The most biologically active form of vitamin E Figure

Vitamin E comprises a group of two compounds (tocopherols and tocotrienols – Figure 2.232) and stereoisomers of each. It is commonly found in plant oils. The compounds act in cells as fat-soluble antioxidants. α-tocopherol (Figure 2.233), the most active form of the vitamin, works 1) through the glutathione peroxidase protective system and 2) in membranes to interrupt lipid peroxidation chain reactions. In both actions, vitamin E reduces levels of reactive oxygen species in cells.



2.233 – Structure of tocotrienols

Action

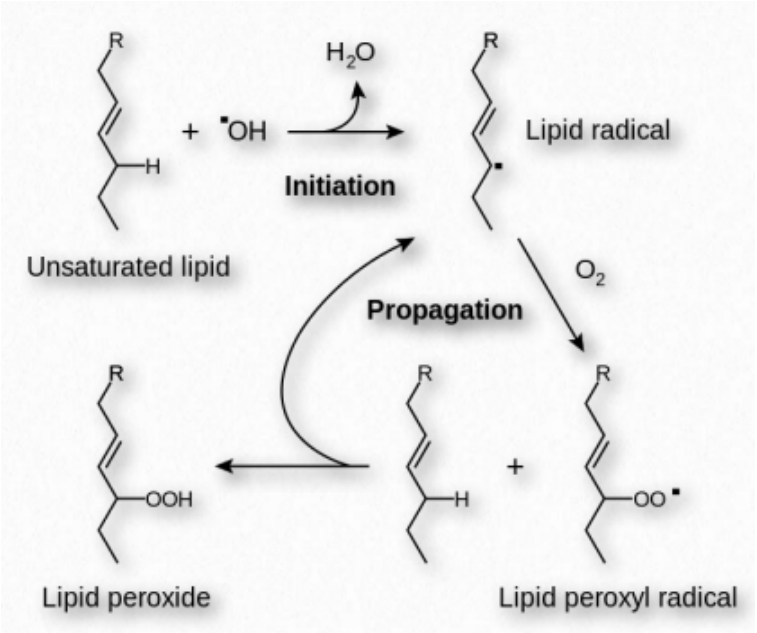


Figure 2.234 – Lipid peroxidation reactions

Vitamin E scavenges oxygen radicals (possessing unpaired electrons) by reacting with them to produce a tocopheryl radical. This vitamin E radical can be converted back to its original form by a hydrogen donor. Vitamin C is one such donor. Acting in this way, Vitamin E helps reduce oxidation of easily oxidized compounds in the lipid peroxidation reactions (Figure 2.234).

Vitamin E also can affect enzyme activity. The compound can inhibit action of protein kinase C in smooth muscle and simultaneously activate catalysis of protein phosphatase 2A to remove phosphates, stopping smooth muscle growth.

Deficiency/excess

Deficiency of vitamin E can lead to poor conduction of nerve signals and other issues arising from nerve problems. Low levels of the vitamin may be a factor in low birth weights and premature

deliveries. Deficiency, however, is rare, and not usually associated with diet.

Excess Vitamin E reduces vitamin K levels, thus reducing the ability to clot blood. Hypervitaminosis of vitamin E in conjunction with aspirin can be life threatening. At lower levels, vitamin E may serve a preventative role with respect to atherosclerosis by reducing oxidation of LDLs, a step in plaque formation.

Vitamin K

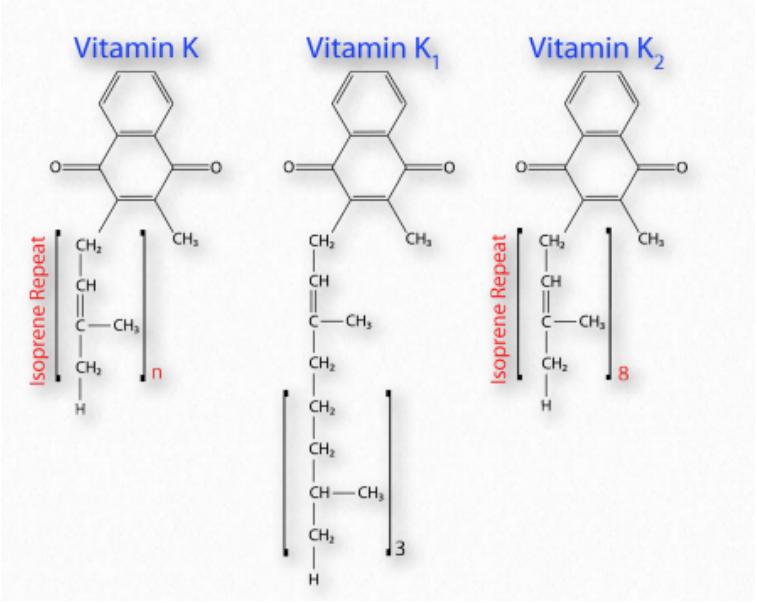


Figure 2.235 – Forms of vitamin K Image by Pehr Jacobson

Like the other fat-soluble vitamins, Vitamin K comes in multiple forms (Figure 2.235) and is stored in fat tissue in the body. There are two primary forms of the vitamin – K1 and K2 and the latter has multiple sub-forms . Vitamins K3, K4, and K5 are made synthetically, not biologically.

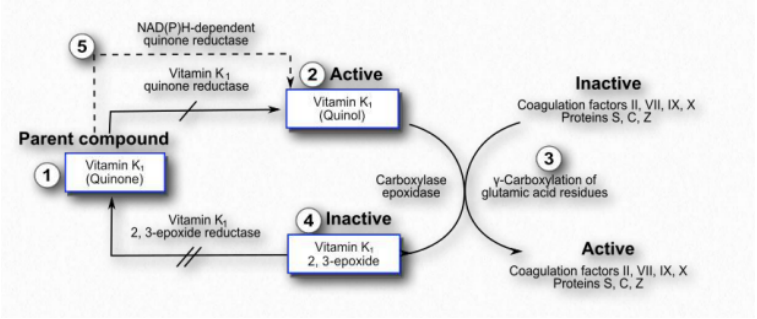


Figure 2.236 – Recycling of vitamin K WikipediaFigure 2.236 – Recycling of vitamin K Wikipedia

Action

Vitamin K is used as a co-factor for enzymes that add carboxyl groups to glutamate side chains of proteins to increase their affinity for calcium. Sixteen such proteins are known in humans. They include proteins involved in blood clotting (prothrombin (called Factor II), Factors VII, IX, and X), bone metabolism (osteocalcin, also called bone Gla protein (BGP), matrix Gla protein (MGP), and periostin) and others.

Modification of prothrombin is an important step in the process of blood clotting (see HERE). Reduced levels of vitamin K result in less blood clotting, a phenomenon sometimes referred to as blood thinning. Drugs that block recycling of vitamin K (Figure 2.236) by inhibiting the vitamin K epoxide reductase, produce lower levels of the vitamin and are employed in treatments for people prone to excessive clotting. Warfarin (coumadin) is one such compound that acts in this way and is used therapeutically. Individuals respond to the drug differentially, requiring them to periodically be tested for levels of clotting they possess, lest too much or too little occur.

Sources

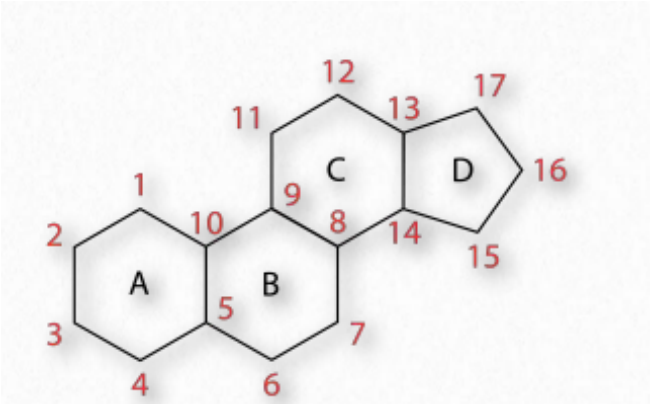


Figure 2.237 – Steroid numbering scheme Image by Pehr Jacobson

Vitamin K1 is a stereoisomer of the plant photosystem I electron receptor known as phylloquinone and is found abundantly in green, leafy vegetables. Phylloquinone is one source of vitamin K, but the compound binds tightly to thylakoid membranes and tends to have low bioavailability. Vitamin K2 is produced by microbes in the gut and is a primary source of the vitamin. Infants in the first few days before they establish their gut flora and people taking broad spectrum antibiotics may have reduced levels, as a result. Dietary deficiency is rare in the absence of damage to the small bowel. Others at risk of deficiency include people with chronic kidney disease and anyone suffering from a vitamin D deficiency. Deficiencies produce symptoms of easy bruising, heavy menstrual bleeding, anemia, and nosebleeds.

Steroids

Steroids such as cholesterol are found in membranes and act as signaling hormones in traveling through the body.

Steroid hormones are all made from cholesterol and are grouped into five categories – mineralocorticoids (21 carbons),

glucocorticoids (21 carbons), progestagens (21 carbons), androgens (19 carbons), and estrogens (18 carbons).

Mineralocorticoids

Mineralocorticoids are steroid hormones that influence water and electrolyte balances. Aldosterone (Figure 2.238) is the primary mineralocorticoid hormone, though other steroid hormones (including progesterone) have some functions like it. Aldosterone stimulates kidneys to reabsorb sodium, secrete potassium, and passively reabsorb water. These actions have the effect of increasing blood pressure and blood volume. Mineralocorticoids are produced by the zona glomerulosa of the cortex of the adrenal gland.

Glucocorticoids

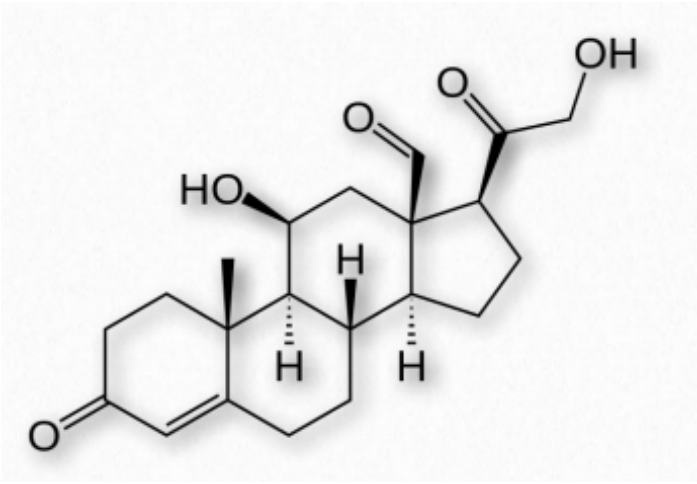


Figure 2.238 – Aldosterone – A mineralocorticoid

Glucocorticoids (GCs) bind to glucocorticoid receptors found in almost every vertebrate animal cell and act in a feedback mechanism in the immune system to reduce its activity. GCs are used to treat diseases associated with overactive immune systems. These include allergies, asthma, and autoimmune dis- Figure 2.237 – Steroid numbering scheme Image by Pehr Jacobson eases. Cortisol (Figure 2.239) is an important glucocorticoid with cardiovascular, metabolic, and immunologic functions. The synthetic glucocorticoid known as dexamethasone has medical applications

for treating rheumatoid arthritis, bronchospasms (in asthma), and inflammation due to its increased potency (25-fold) compared to cortisol. Glucocorticoids are produced primarily in the zona fasciculata of the adrenal cortex.

Progestagens

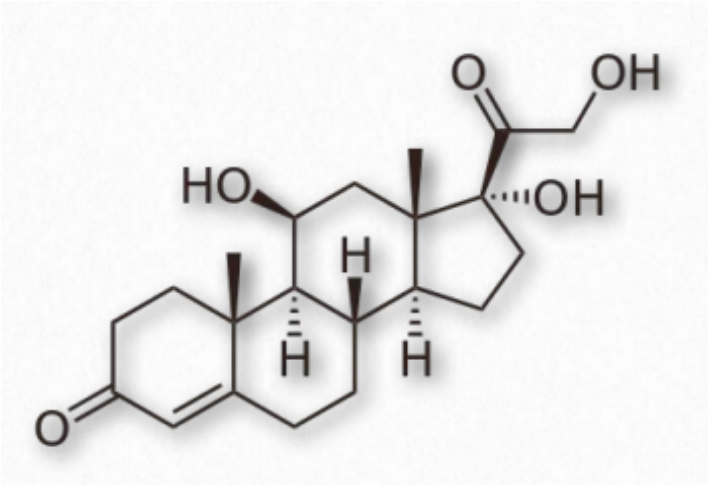


Figure 2.239 – Cortisol – A glucocorticoid

Progestagens (also called gestagens) are steroid hormones that work to activate the progesterone receptor upon binding to it. Synthetic progestagens are referred to as progestins. The most common progestagen is progesterone (also called P4 – Figure 2.240) and it has functions in maintaining pregnancy. Progesterone is produced primarily in the diestrus phase of the estrous cycle by the corpus luteum of mammalian ovaries. In pregnancy, the placenta takes over most progesterone production.

Androgens

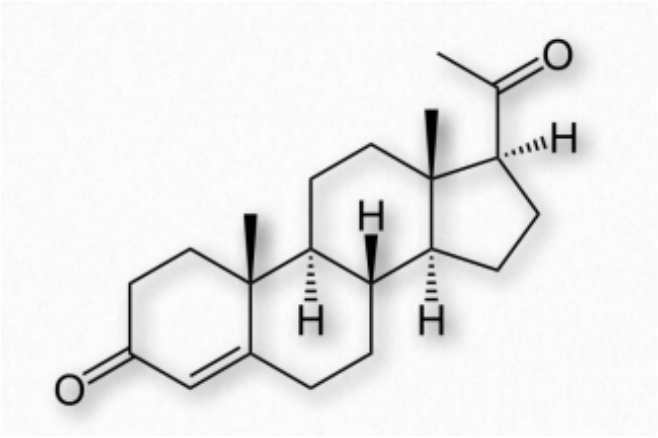


Figure 2.240 Progesterone – A progestagen

Androgens are steroid hormones that act by binding androgen receptors to stimulate development and maintenance of male characteristics in vertebrates. Androgens are precursors of estrogens (see below). The primary androgen is testosterone (Figure 2.241). Other important androgens include dihydrotestosterone (stimulates differentiation of penis, scrotum, and prostate in embryo) and androstenedione (common precursor of male and female hormones).

Estrogens

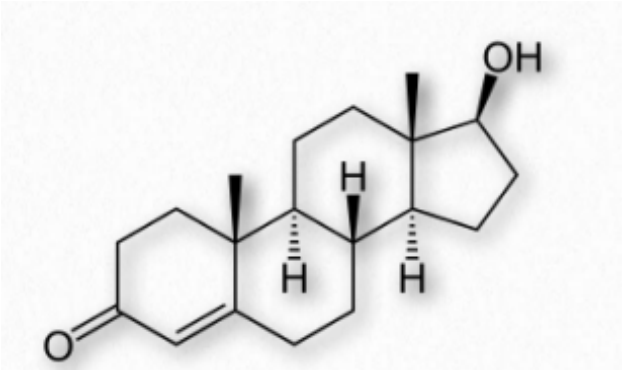


Figure 2.241 – Testosterone – An androgen

The estrogen steroid hormones are a class of compounds with important roles in menstrual and estrous cycles. They are the most

important female sex hormones. Estrogens act by activating estrogen receptors inside of cells. These receptors, in turn, affect expression of many genes. The major estrogens in women include estrone (E1), estradiol (E2 – Figure 2.242), and estriol (E3). In the reproductive years, estradiol predominates. During pregnancy, estriol predominates and during menopause, estrone is the major estrogen.

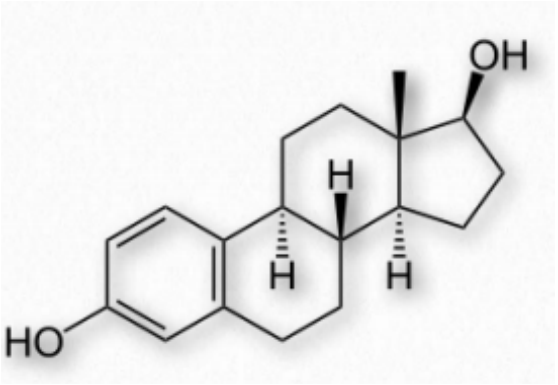


Figure 2.242 – Estradiol – An estrogen

Estrogens are made from the androgen hormones testosterone and androstenedione in a reaction catalyzed by the enzyme known as aromatase. Inhibition of this enzyme with aromatase inhibitors, such as exemestane, is a strategy for stopping estrogen production. This may be part of a chemotherapeutic treatment when estrogenresponsive tumors are present.

Other Lipids

Cannabinoids

Cannabinoids are a group of chemicals that bind to and have effects on brain receptors (cannabinoid receptors), repressing neurotransmitter release. Classes of these compounds include endocannabinoids (made in the body), phytocannabinoids (made in plants, such as marijuana), and synthetic cannabinoids (man-made).

Endocannabinoids are natural molecules derived from arachidonic acid. Cannabinoid receptors are very abundant, comprising the largest number of G-protein- 247 Figure 2.243 – Tetrahydrocannabinol – Active ingredient in marijuana coupled receptors found in brain. The best known phytocannabinoid is Δ-9-tetrahydrocannabinol (THC), the primary psychoactive ingredient (of the 85 cannabinoids) of marijuana (Figure 2.243).

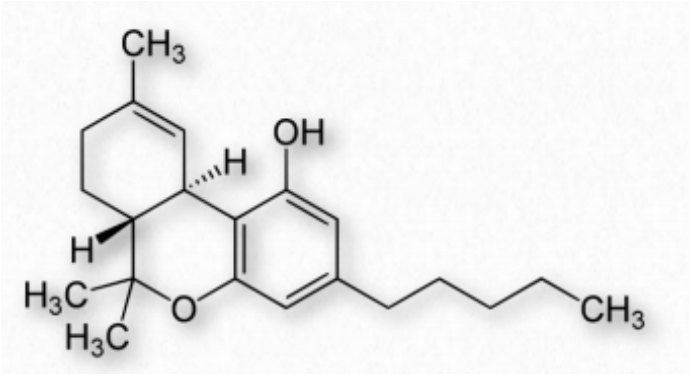


Figure 2.243 – Tetrahydrocannabinol – Active ingredient in marijuana

Anandamide

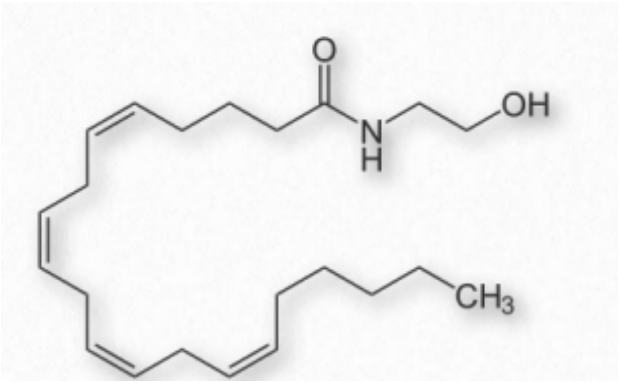


Figure 2.244 – Anandamide – An endocannabinoid

Anandamide (N-arachidonoylethanolamine – Figure 2.244) is an endocannabinoid neurotransmitter derived from arachidonic acid. It exerts its actions primarily through the CB1 and CB2 cannabinoid

receptors, the same ones bound by the active ingredient in marijuana, Δ9-tetrahydrocannabinol. Anandamide has roles in stimulating eating/appetite and affecting motivation and pleasure. It has been proposed to play a role in "runners high," an analgesic effect experienced from exertion, especially among runners. Anandamide appears to impair memory function in rats.

Anandamide has been found in chocolate and two compounds that mimic its effects (N-oleoylethanolamine and Nlinoleoylethanolamine) are present as well. The enzyme fatty acid amide hydrolase (FAAH) breaks down anandamide into free arachidonic acid and ethanolamine.

Lipoxins

Lipoxins (Figure 2.245) are eicosanoid compounds involved in modulating immune responses and they have anti-inflammatory effects. When lipoxins appear in inflammation it begins the end of the process. Lipoxins act to attract macrophages to apoptotic cells at the site of inflammation and they are engulfed. Lipoxins further act to start the resolution phase of the inflammation process.

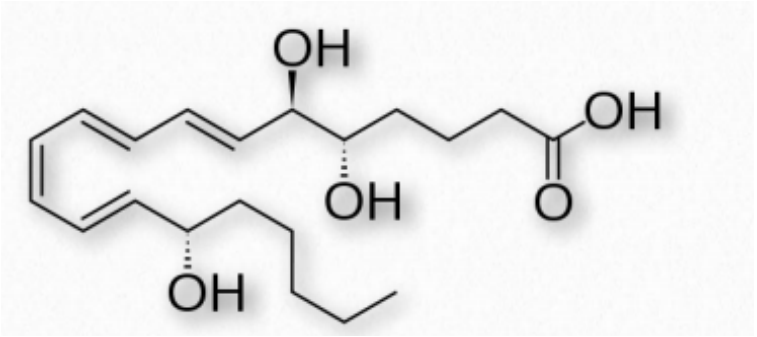


Figure 2.245 – Lipoxin A4

At least one lipoxin (aspirin-triggered LX4) has its synthesis stimulated by aspirin. This occurs as a byproduct of aspirin's acetylation of COX-2. When this occurs, the enzyme's catalytic activity is re-directed to synthesis of 15R-hydroxyeicosatetraenoic acid (HETE) instead of prostaglandins. 15R-HETE is a procursor of 15-epimer lipoxins, including aspirin-triggered LX4.

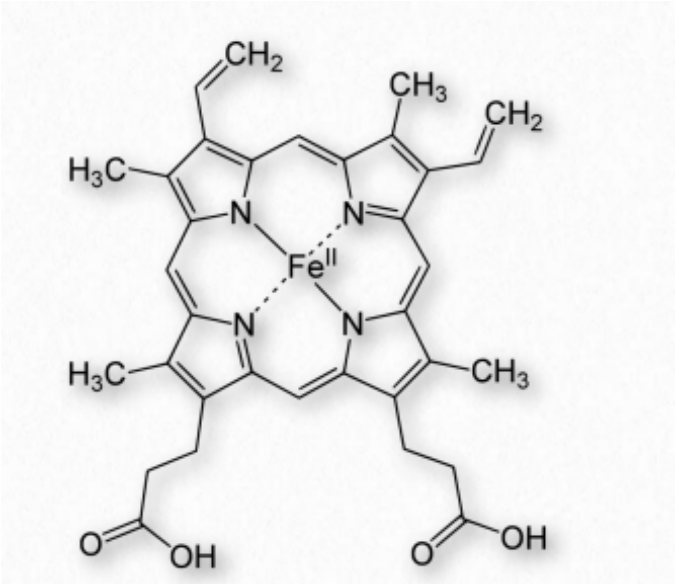


Figure 2.246 – Structure of heme B

Heme

Heme groups are a collection of protein/ enzyme cofactors containing a large heterocyclic aromatic ring known as a porphyrin ring with a ferrous (Fe++) ion in the middle. An example porphyrin ring with an iron (found in Heme B of hemoglobin), is shown in Figure 2.246. When contained in a protein, these are known collectively as hemoproteins (Figure 2.247).

Heme, of course, is a primary component of hemoglobin, but it is also found in other proteins, such as myoglobin, cytochromes, and the enzymes catalase and succinate dehydrogenase. Hemoproteins function in oxygen transport, catalysis, and electron transport. Heme is synthesized in the liver and bone marrow in a pathway that is conserved across a wide range of biology.

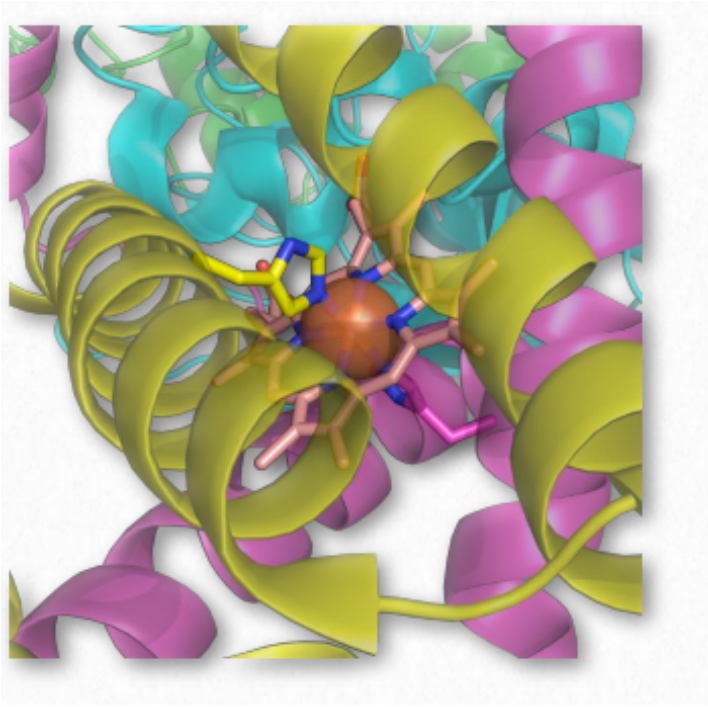


Figure 2.247 – Heme embedded in the succinate dehydrogenase hemoprotein Wikipedia

Porphobilinogen

Porphobilinogen (Figure 2.248) is a pyrrole molecule involved in porphyrin metabolism. It is produced from aminolevulinate by action of the enzyme known as ALA dehydratase. Porphobilinogen is acted upon by the enzyme porphobilinogen deaminase. Deficiency of the latter enzyme (and others in porphyrin metabolism) can result in a condition known as porphyria, which results in accumulation of porphobilinogen in the cytoplasm of cells.

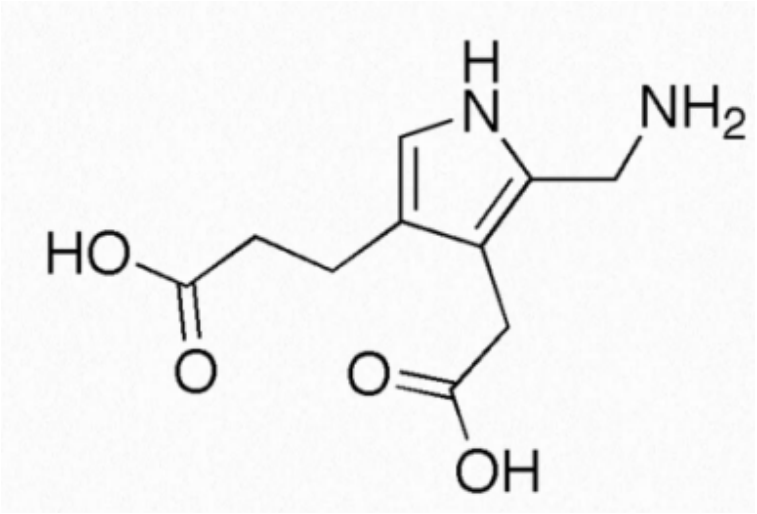


Figure 2.248 – Porphobilinogen

The disease can manifest itself with acute abdominal pain and numerous psychiatric issues. Both Vincent van Gogh and King ` George III are suspected to have suffered from porphyria, perhaps causing the "madness of King George III." Porphyria is also considered by some to be the impetus for the legend of vampires seeking blood from victims, since the color of the skin in non-acute forms of the disease can be miscolored, leading some to perceive that as a deficiency of hemoglobin and hence the "thirst" for blood imagined for vampires.

Dolichols

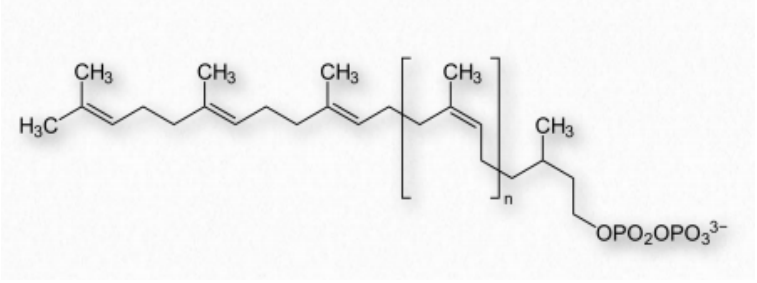


Figure 2.249 – Structure of dolichol pyrophosphate

Dolichol is a name for a group of non-polar molecules made

by combining isoprene units together. Phosphorylated forms of dolichols play central roles in the N-glycosylation of proteins. This process, which occurs in the endoplasmic reticulum of eukaryotic cells, begins with a membrane-embedded dolichol pyrophosphate (Figure 2.249) to which an oligosaccharide is attached (also see HERE). This oligosaccharide contains three molecules of glucose, nine molecules of mannose and two molecules of N-acetylglucosamine.

Interestingly, the sugars are attached to the dolichol pyrophosphate with the pyrophosphate pointing outwards (away from) the endoplasmic reticulum, but after attachment, the dolichol complex flips so that the sugar portion is situated on the inside of the endoplasmic reticulum. There, the entire sugar complex is transferred to the amide of an asparagine side chain of a target protein.

The only asparagine side chains to which the attachment can be made are in proteins where the sequences Asn-X-Ser or Asn-X-Thr occur. Sugars can be removed/added after the transfer to the protein. Levels of dolichol in the human brain increase with age, but in neurodegenerative diseases, they decrease.

Terpenes

Figure 2.250 – Pine tree resin – A source of terpenes Wikipedia

Terpenes are members of a class of nonpolar molecules made from isoprene units. Terpenes are produced primarily by plants and by some insects. Terpenoids are a related group of molecules that contain functional groups lacking in terpenes.

Terpenes have a variety of functions. In plants, they often play a defensive role protecting from insects. The name of terpene comes from turpentine, which has an odor like some of the terpenes. Terpenes are common components of plant resins (think pine) and they are widely used in medicines and as fragrances. Hops, for

example, gain some of their distinctive aroma and flavor from terpenes. Not all terpenes, however have significant odor.

Synthesis

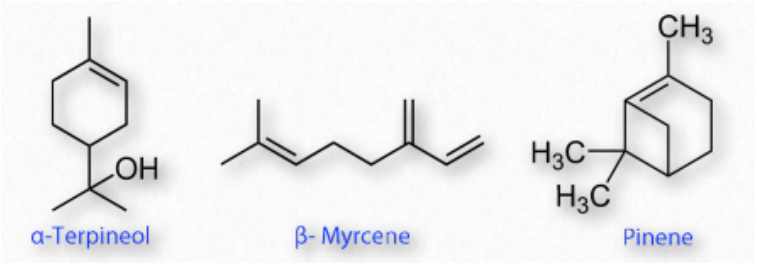


Figure 2.251 – Three monoterpenes

Terpenes, like steroids, are synthesized starting with simple building blocks known as isoprenes. There are two of them – dimethylallyl pyrophosphate and the related isopentenyl pyrophosphate and (Figures 2.252 and 2.253) which combine 1-2 units at a time to make higher order structures. Terpene synthesis overlaps and includes steroid synthesis.

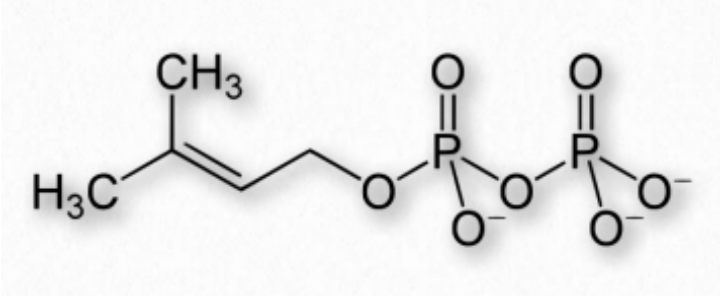


Figure 2.252 – Dimethylallyl pyrophosphate

Terpenes and terpenoids are classified according to how many isoprene units they contain. They include hemiterpenes (one unit), monoterpenes (two units), sesquiterpenes (three units), diterpenes (four units), sesterterpenes (five units), triterpenes (six units), sesquarterpenes (seven units), tetraterpenes (eight units), polyterpenes (many units). Another class of terpene-containing molecules, the norisoterpenoids arise from peroxidase-catalyzed reactions on terpene molecules.

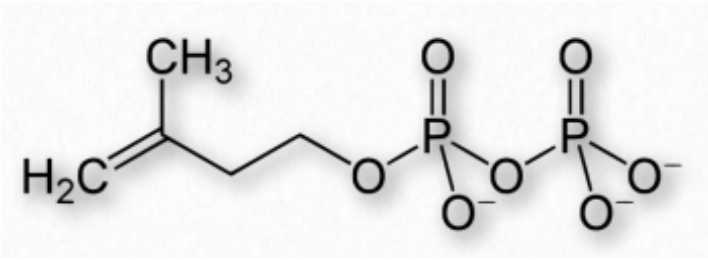


Figure 2.253 – Isopentenyl pyrophosphate

Examples

Common terpenes include monoterpenes of terpineol (lilacs), limonene (citrus), myrcene (hops), linalool (lavender), and pinene (pine). Higher order terpenes include taxadiene (diterpene precursor of taxol), lycopene (tetraterpenes), carotenes (tetraterpenes), and natural rubber (polyterpenes).

Steroid precursors geranyl pyrophosphate (monoterpene derivative), farnesyl pyrophosphate (sesquiterpene derivative), and squalene (triterpene) are all terpenes or derivatives of them. Vitamin A and phytol are derived from diterpenes.

Caffeine

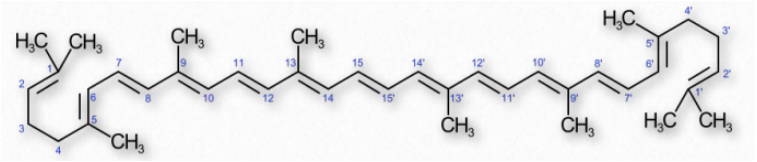


Figure 2.254 – Lycopene – A tetraterpene

Caffeine is the world's most actively consumed psychoactive drug (Figure 2.255). A methylxanthine alkaloid, caffeine is closely related to adenine and guanine and this is responsible for many effects on the body. Caffeine blocks the binding of adenosine on its receptor and consequently prevents the onset of drowsiness induced by adenosine. Caffeine readily crosses the blood-brain barrier and stimulates release of neurotransmitters. Caffeine stimulates portions of the autonomic nervous system and inhibits the activity of phosphodiesterase. The latter has the result of raising cAMP levels in cells, which activates protein kinase A and activates

glycogen breakdown, inhibits TNF-α and leukotriene synthesis, which results in reduction of inflammation and innate immunity.

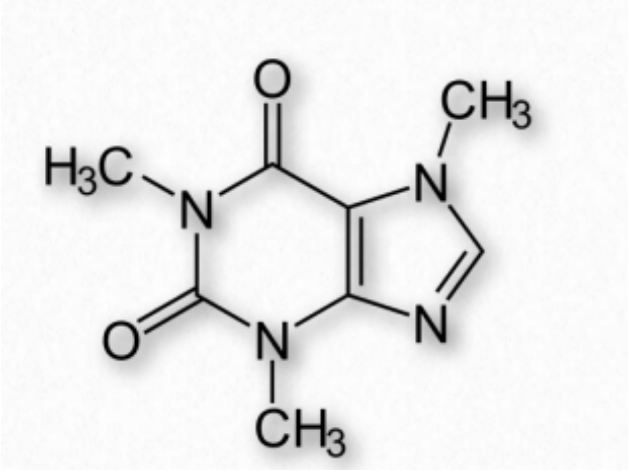


Figure 2.255 – Caffeine

Caffeine also has effects on the cholinergic system (acetylcholinesterase inhibitor), is an inositol triphosphate receptor 1 antagonist, and is a voltage independent activator of ryanodin receptors (a group of calcium channels found in skeletal muscle, smooth muscle, and heart muscle cells).

The half-life of caffeine in the body varies considerably. In healthy adults, it has a half-life of about 3-7 hours. Nicotine decreases the half-life and contraceptives and pregnancy can double it. The liver metabolizes caffeine, so the health of the liver is a factor in the halflife. CYP1A2 of the cytochrome P450 oxidase enzyme is primarily responsible. Caffeine is a natural pesticide in plants, paralyzing predator bugs.

7.4: Basic Concepts in Membranes

Lipid Bilayers

The protective membrane around cells contains many components, including cholesterol, proteins, glycolipids, glycerophospholipids, and sphingolipids. The last two of these will, when mixed vigorously with water, spontaneously form what is called a lipid bilayer (Figure 3.1), which serves as a protective boundary for the cell that is largely impermeable to the movement of most materials across it. With the notable exceptions of water, carbon dioxide, carbon monoxide, and oxygen, most polar/ionic require transport proteins to help them to efficiently navigate across the bilayer. The orderly movement of these compounds is critical for the cell to be able to 1) get food for energy; 2) export materials; 3) maintain osmotic balance; 4) create gradients for secondary transport; 5) provide electromotive force for nerve signaling; and 6) store energy in electrochemical gradients for ATP production (oxidative phosphorylation or photosynthesis).

Facilitated Diffusion

In some cases, energy is required to move substances across a membrane (active transport). In other cases, no external energy is required and they move by diffusion through specific cellular channels. This is referred to as facilitated diffusion.

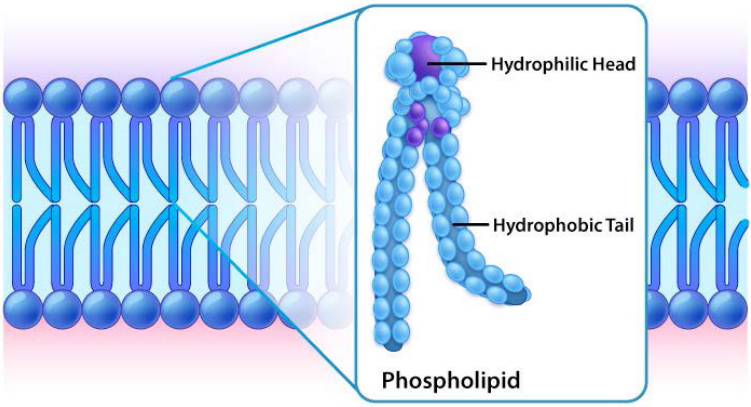


Figure 3.1: Lipid bilayer closeup Image by Aleia Kim

While all cells have cell membranes, the components of the membranes can vary. Figures 3.1 and 3.2 illustrate the structure and environments of plasma membranes. All plasma membranes contain a significant amount of amphiphilic substances linked to fatty acids. These include the glycerophospholipids and the sphingolipids. The fatty acid(s) are labeled as hydrophobic tails in the figures.

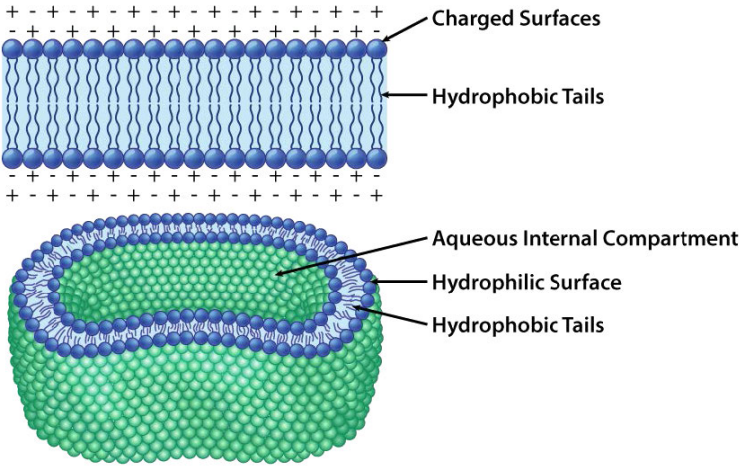


Figure 3.2: *Organization of the lipid bilayer Image by Aleia Kim*

Hydrophilic heads

The composition of the hydrophilic heads varies considerably. In glycerophospholipids, a phosphate is always present, of course, and it is often esterified to another substance to make a phosphatide (Figure 3.3). Common compounds linked to the phosphate (at the X position) include serine, ethanolamine, and choline. These vary in the their charges so in this way, the charge on the external or internal surface can be controlled. Cells tend to have more negative charges on the exterior half of the lipid bilayer (called the outer leaflet) and more positive charges on the interior half (inner leaflet).

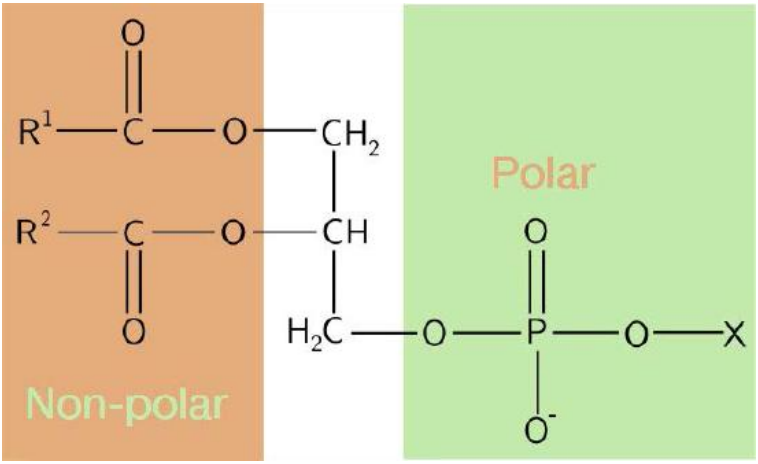


Figure 3.3 – Schematic diagram of a phosphatide

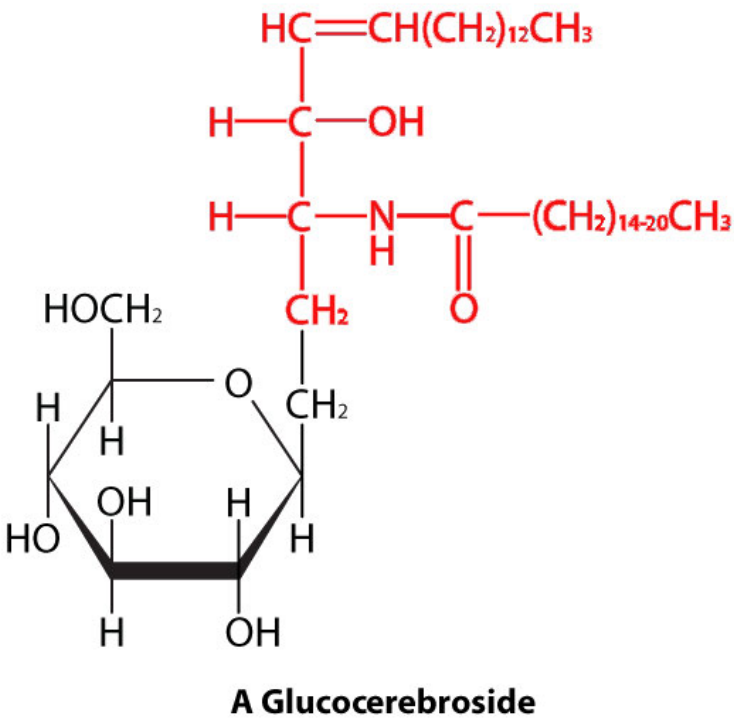


Figure 3.4: *A sphingolipid. Polar head in black. Nonpolar tail in red Image by Aleia Kim*

Sphingolipids

In sphingolipids (Figure 3.4), the hydrophilic head can contain a phosphate linked to ethanolamine or choline and this describes the structure of sphingomyelin, an important component of neural membranes. Most sphingolipids lack the phosphate and have instead a hydrophilic head of a single sugar (cerebrosides) or a complex oligosaccharide (gangliosides).

Water exclusion

If you recall how amino acids organize into folded proteins, the tendency is for hydrophobic components of a polypeptide chain to pack into the center of a globular protein. This leaves the hydrophilic components interacting with water. A similar phenomenon drives the non-polar portions of amphiphilic membrane molecules to arrange themselves so the hydrophobic portions avoid contact with water.

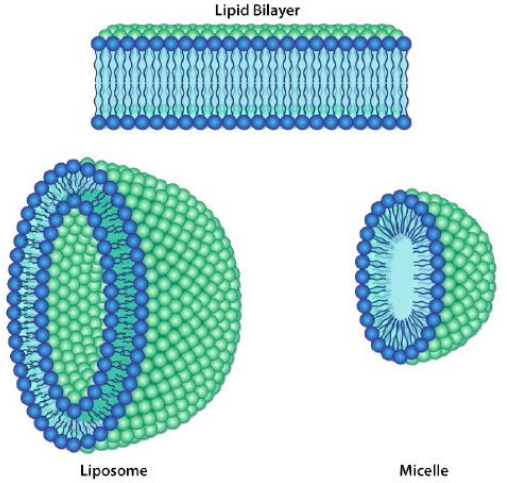


Figure 3.6 – Different lipid bilayer structures Image by Aleia Kim

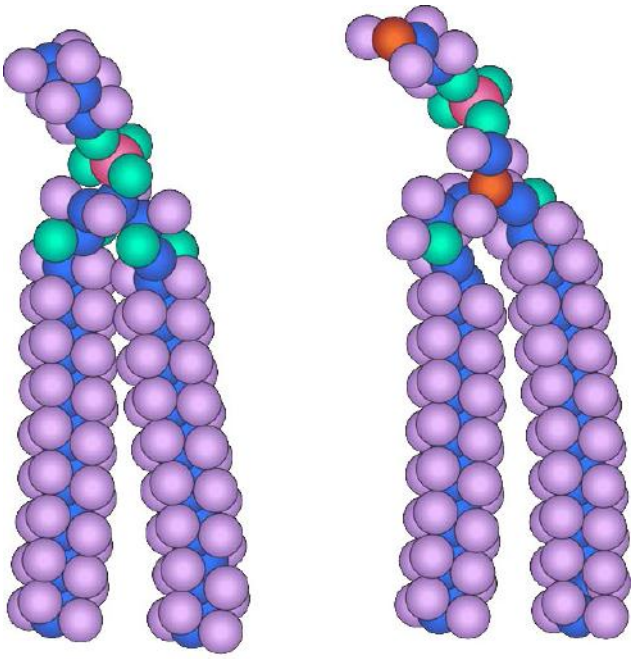


Figure 3.5 – Similarity of form between a phosphoglyceride and sphingomyelin Image by Aleia Kim

Lateral Diffusion

Movement of lipids within each leaflet of the lipid bilayer occurs readily and rapidly due to membrane fluidity. This type of movement is called lateral diffusion and can be measured by the technique called FRAP (Figure 3.10). In this method, a laser strikes and stains a section of the lipid bilayer of a cell, leaving a spot as shown in B. Over time, the stain diffuses out ultimately across the entire lipid bilayer, much like a drop of ink will diffuse throughout when added to a glass of water. A measurement of the rate of diffusion gives an indication of the fluidity of a membrane.

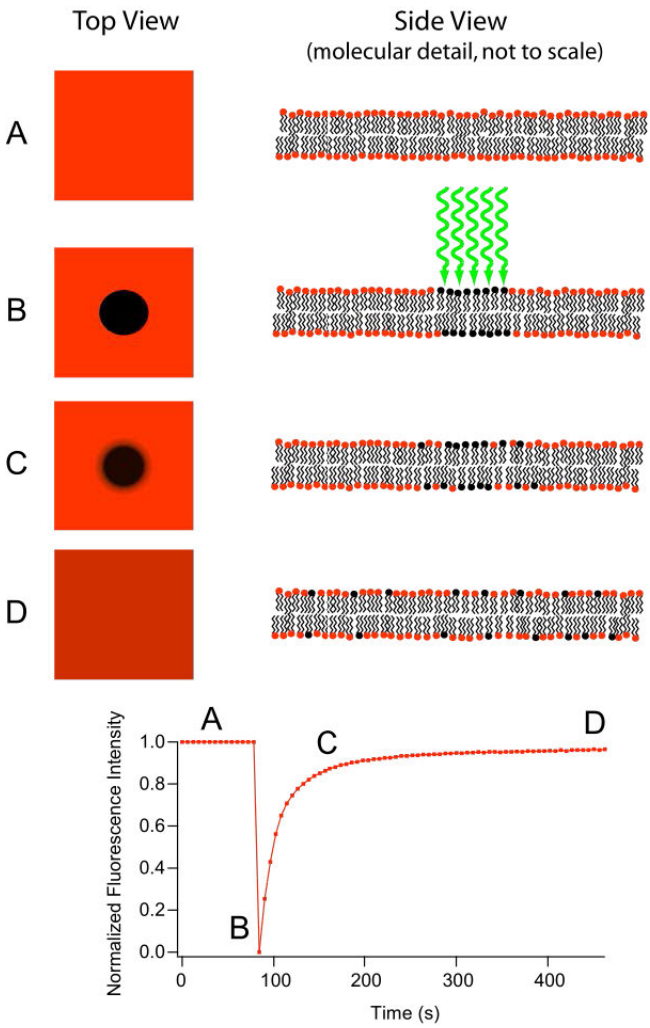


Figure 3.10 – Fluorescence recovery after photobleaching (FRAP)

Transverse Diffusion

While the movement in lateral diffusion occurs rapidly, movement of molecules from one leaflet over to the other leaflet occurs much more slowly. This type of molecular movement is called transverse diffusion and is almost nonexistent in the absence of enzyme action. Remember that there is a bias of distribution of molecules between leaflets of the membrane, which means that something must be moving them.

There are three enzymes that catalyze movement of compounds in transverse diffusion. Flippases move membrane glycerophospholipids/ sphingolipids from outer leaflet to inner leaflet (cytoplasmic side) of cell. Floppases move membrane lipids in the opposite direction. Scramblases move in either direction.

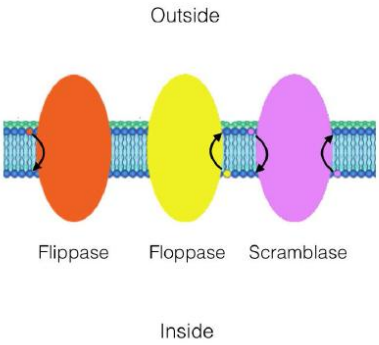


Figure 3.11 – Catalytic action of a flippase, a floppase, and a scramblase

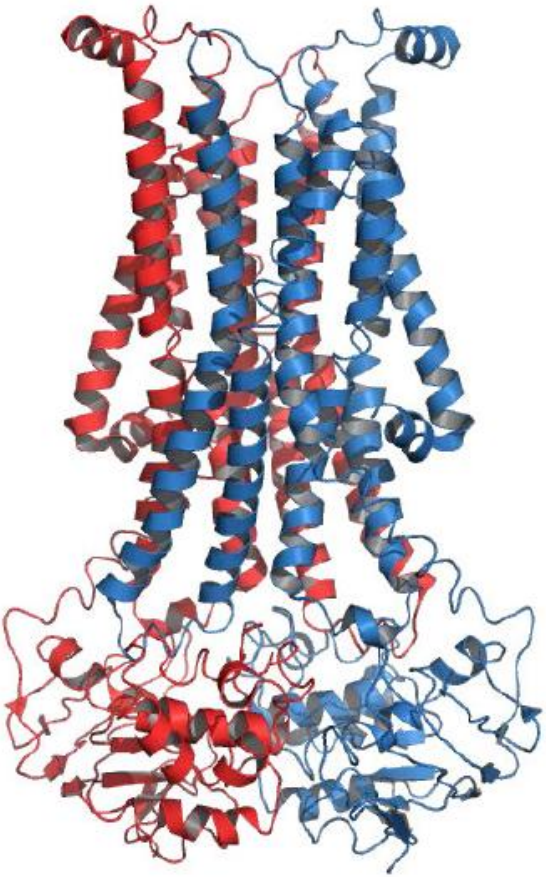


Figure 3.12 – Structure of a flippase Wikipedia

Other components of lipid bilayer

Besides glycerophospholipids and sphingolipids, there are other materials commonly found in lipid bilayers of cellular membranes. Two important prominent ones are cholesterol (Figure 3.13) and

proteins. Besides serving as a metabolic precursor of steroid hormones and the bile acids, cholesterol's main role in cells is in the membranes. The flatness and hydrophobicity of the sterol rings allow cholesterol to interact with the nonpolar portions of the lipid bilayer while the hydroxyl group on the end can interact with the hydrophilic part.

Cholesterol's function in the lipid bilayer is complex (Figure 3.13). It influences membrane fluidity.

Lipid Rafts

Cholesterol is also abundantly found in membrane structures called lipid rafts. Depicted in Figure 3.15, lipid rafts are organized structures within the membrane typically containing signaling molecules and other integral membrane proteins. Lipid rafts affect membrane fluidity, neurotransmission, and trafficking of receptors and membrane proteins. Lipid rafts may provide concentrating platforms after individual protein receptors bind to ligands in signaling.

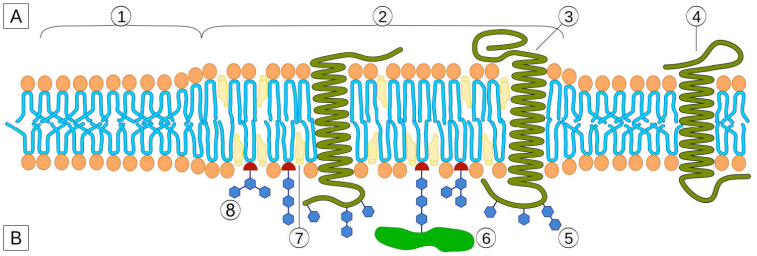


Figure 3.15 – A lipid raft – 1 = Non-raft membrane / 2 = Lipid raft / 3 = Lipid raft associated transmembrane protein / 4 = Non-raft membrane protein / 5 = Glycosylation modifications (on glycoproteins and glycolipids) / 6 = GPIanchored protein / 7 = Cholesterol / 8 = Glycolipid

Barrier

Transport of materials across membranes is essential for a cell to exist. The lipid bilayer is an effective barrier to the entry of most molecules and without a means of allowing food molecules to enter a cell, it would die. The primary molecules that move freely across the lipid bilayer are small, uncharged ones, such as CO2, CO, and O2, so larger molecules, like glucose, that the cell needs for energy, would be effectively excluded if there were not proteins to facilitate its movement across the membrane.

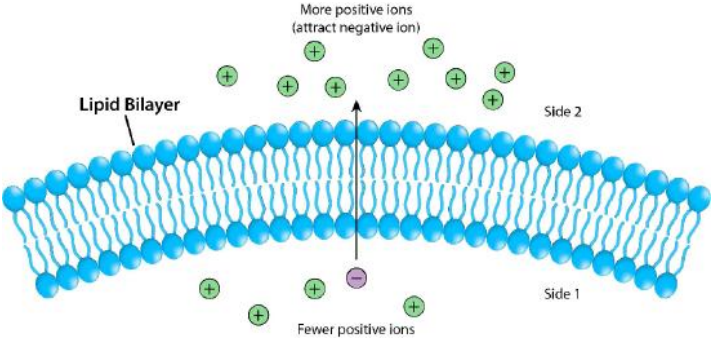


Figure 3.17 – The lipid bilayer as a barrier Image by Pehr Jacobson

Figure 3.17 depicts the barrier that the lipid bilayer provides to movement across it and the pressures (ionic attraction, in this case) that can affect movement. Potential energy from charge and concentration differences are harvested by cells for purposes that include synthesis of ATP, and moving materials against a concentration gradient in a process called active transport.

Membrane proteins

Proteins in a lipid bilayer can vary in quantity enormously, depending on the membrane. Protein content by weight of various membranes typically ranges between 30 and 75% by weight. Some

mitochondrial membranes can have up to 90% protein. Proteins linked to and associated with membranes come in several types.

Osmotic Pressure

Membranes provide barriers/boundaries for most molecules, but the permeability of water across a lipid bilayer creates a variable that must be considered. The variable here is osmotic pressure. Osmotic pressure (loosely) refers to the tendency of a solution to take in water by the process of osmosis. In Figure 3.22, one can see a visual representation of the concept of the pressure.

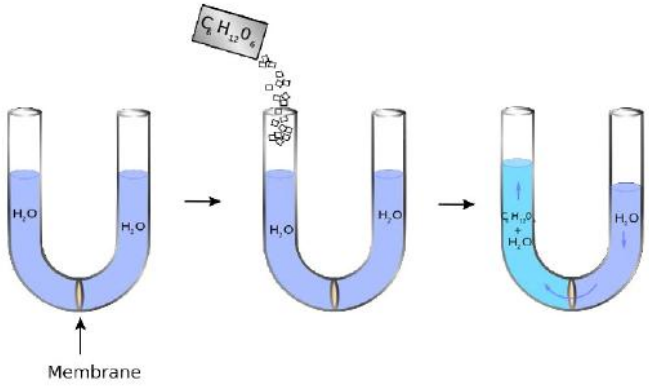


Figure 3.22

A U-shaped tube has at its bottom a semipermeable membrane. Water can pass through the membrane, but sugar molecules (C6H12O6) cannot. On the left side, sugar is added creating a concentration difference between the right and left chambers. Water diffuses across the membrane from right to left in an attempt to equalize the concentrations, causing the level of the right side to decrease and the left side to increase. The pressure resulting from the differences in height is felt at the membrane.

Equalizing concentrations

The liquid on the right does not completely move to the left,

though, as might be expected if the only force involved is equalizing the concentration of sugar across the membrane (no sugar on right = no water). Instead, an equilibrium of sorts of water levels is reached even though the concentrations don't equal out. The pressure existing at the membrane then from the differences in level corresponds to the osmotic pressure of the mixture. The osmotic pressure of a solution is the pressure difference needed to halt the flow of solvent across a semipermeable membrane. Osmotic pressure can also be thought of as the pressure required to counter osmosis. The osmotic pres- Figure 3.21 – Blood types arise from cell surface glycoproteins Figure 3.22 – Osmotic pressure. Water diffuses leftwards to try to equalize the solute concentration. The pressure realized at the membrane in the right figure is the osmotic pressure sure of a dilute solution mathematically behaves like the ideal gas law

$$P_{osmotic} = \frac{nRT}{V}$$

where n is the number of moles, R is the gas constant, T is the temperature in Kelvin, and V is the volume.

It is more convenient in solutions to work with molarity, so

$$P_{osmotic} = MR^*T$$

where M is the molarity of the dissolved molecules, R* is the gas constant expressed in (L atm)/(K mol), and T is the temperature. The Greek letter Π is used to refer to the $P_{osmotic}$term, so

$$\Pi = MR^*T$$

Remember when calculating the molarity to include the molarity of each particle. For example, when one dissolves sucrose in solution, it does not split into smaller particles, so

$$Molarity_{Particles} = Molarity_{Sucrose}$$

However, for salts, like KOH, which forms two ions in solution (K+ and OH-),

$Molarity_{Particles}$= 2* $Molarity_{KOH}$.

Hypotonic, hypertonic, isotonic

We consider three situations (Figure 3.23). First, if the

concentration of solutes is greater inside the cell than outside, water will tend to move into the cell, causing the cell to swell. This circumstance is called hypotonic. Conversely, if the solute concentration is greater outside the cell than inside of it, water will exit the cell and the cell with shrink. This is a hypertonic situation. Last, if the concentrations of solutes into and outside of the cell is equal, this is called an isotonic solution. Here, no movement of water occurs across the cell membrane and the cell retains its size.

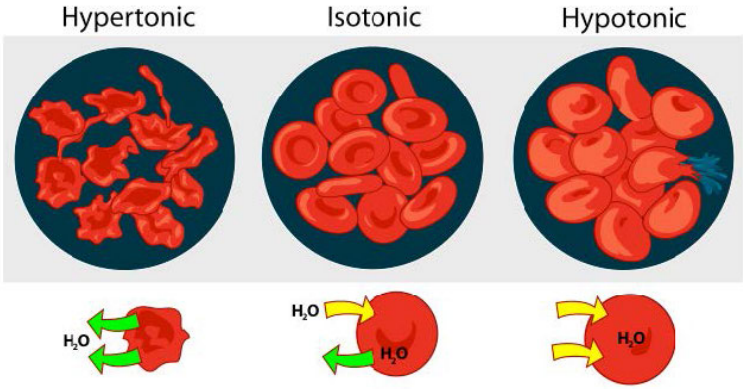
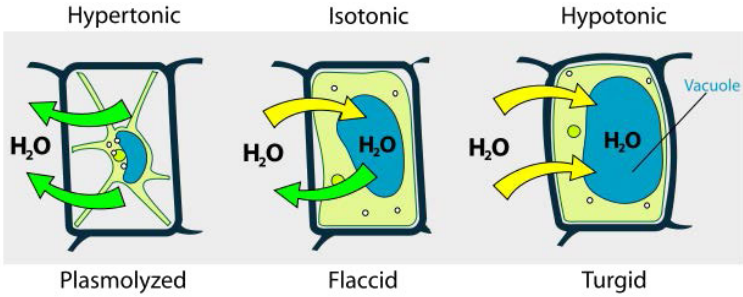


Figure 3.23 – The effect of three different osmotic conditions on red blood cells concentrations on either side of the membrane

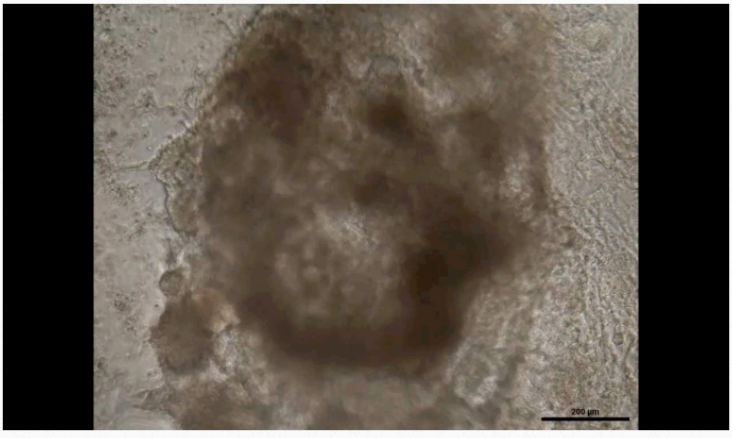
If the osmotic pressure is greater than the forces holding together a cellular membrane, the cell will rupture. Because of this, some cells have built in defenses to prevent problems. Plant cells, for example have a fairly rigid cell wall that resists expansion in hypotonic solutions (Figure 3.24). Bacteria also have a cell wall that provides protection.



CODA Protein Structure and Function motor proteins

To this point, the proteins we have discussed have not been catalysts (enzymes). The majority of proteins in cells, however, catalyze reactions. In this section we begin our discussion of a subclass of proteins that catalyze reactions releasing energy and convert it into mechanical force. These operate at the cellular and organismal level and are known as motor proteins. Motor proteins rely on globular structural proteins, so it is important that we describe how these cellular "railways" are assembled before discussing the motor proteins themselves. There are two relevant fibrous structures serving as rails for motor proteins. They are:

1. **microfilaments** (composed of an actin polymer) and
2. **microtubules** (composed of a polymer of tubulin.



Actin

The monomeric unit of actin is called G-actin (globular actin) and the polymer is known as F-actin (filamentous actin). Filaments of Factin comprise the smallest filaments of cells known as microfilaments (Figure 2.101). Actin is essential for muscular contraction and also has diverse roles in cellular signaling and maintenance of cell junctions. In conjunction with other proteins, actin has numerous interactions with the cell membrane. The β- and γ-forms of actin are components of the cytoskeleton and facilitate motility inside of cells. α-actin is important in muscle tissues, where it is used by myosin in the mechanical process of contraction (See HERE).

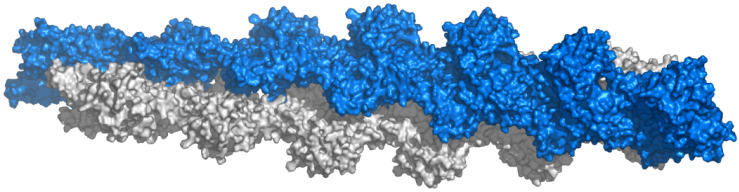


Figure 2.101 – Model of actin filaments. Image used with permission (CC BY-SA 3.0; Thomas Splettstoesser).

Monomeric and polymeric forms of actin play roles in cellular activities relating to motion. Two parallel F-actin strands can pair with each other and create a double helical structure with 2.17 subunits per turn of the helix. Helical F-actin in muscles contains tropomyosin, which covers the actin binding sites for myosin in resting muscles to prevent contraction. Other proteins bound to actin muscle filaments include the troponins (I, T, and C).

Actin Cellular Action

Examples of actin action at the cellular level include cell motility, cytokinesis, intracellular transport of vesicles and organelles, and cell shape. Each actin monomer is bound to a molecule of ATP or ADP and the presence of one of these is essential for proper G-actin functioning.

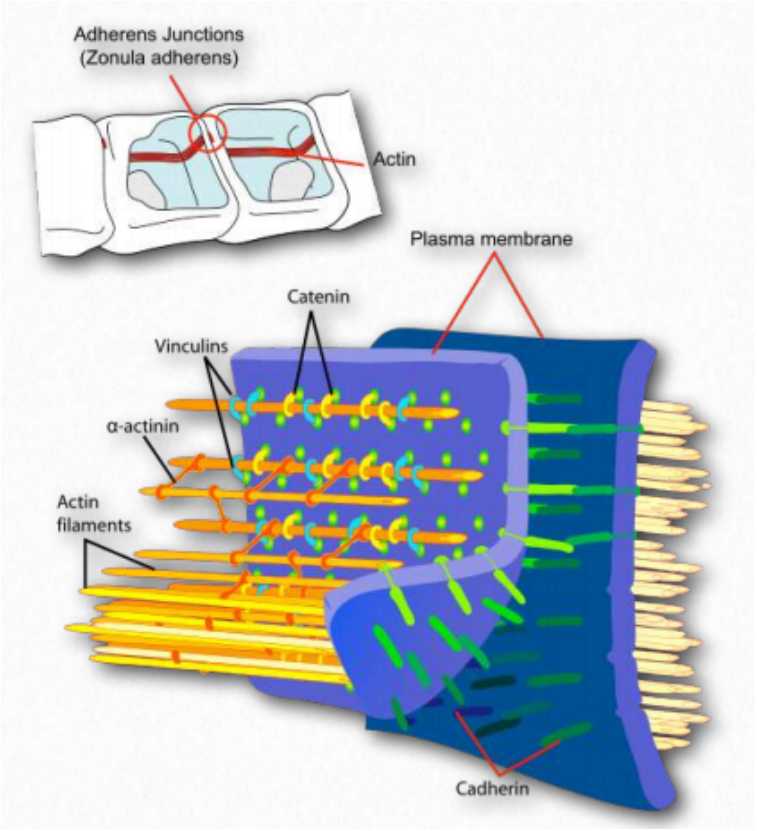


Figure 2.102 – Attachment of actin at the cell membrane complex known as the adherens junction Wikipedia

The role of ATP

In the monomer, actin is more commonly bound to ATP, whereas in the filaments, it is typically bound to ADP. Actin is an inefficient ATPase, breaking the molecule down slowly, but the catalysis speeds up as much as 40,000 fold when the monomer begins to polymerize. Actin also has a binding site for divalent cations – either calcium or magnesium. F-

Actin binds to structural proteins at the adherens junction (Figure 2.102). These include α-actinin, vinculin (provides a membrane connection and connections to the catenins and cadherin).

Polymerization

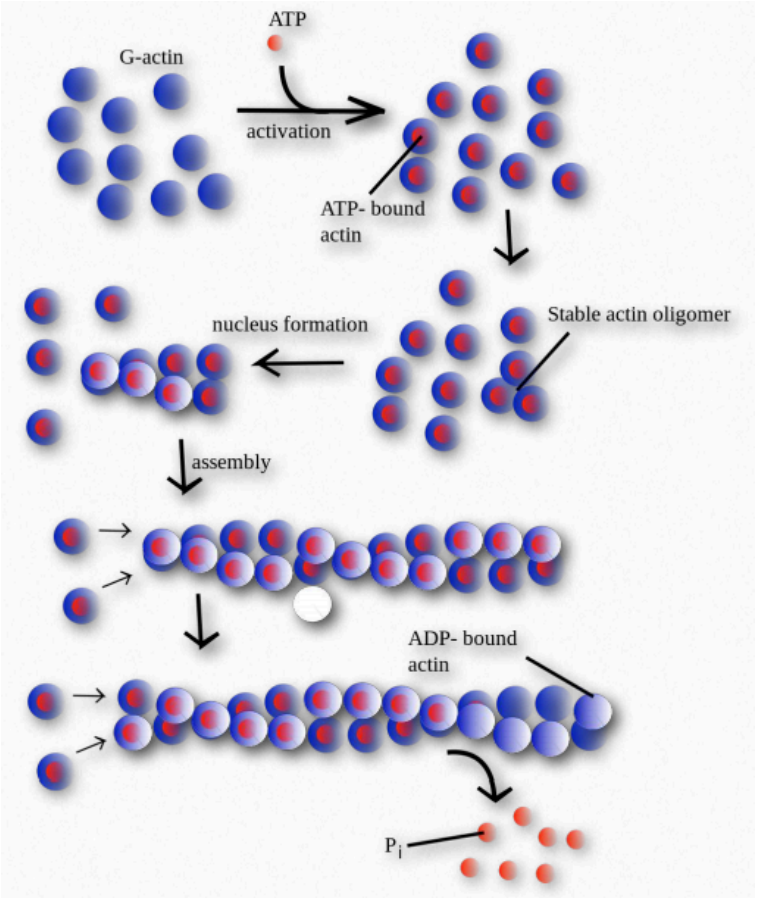


Figure 2.103 – Polymerization of G-actin monomers into Factin polymers

Polymerization of actin begins with a nucleating event (Figure 2.103). One factor known to affect the process is known as the Arp 2/3 complex. It does this by mimicking an actin dimer, starting an autocatalytic process of actin assembly. The Arp 2/3 complex plays roles both in the initiation of polymerization of new actin filaments as well as the formation of branches in the filaments.

Two proteins play roles in modulating polymer growth. Thymosin functions on the end of actin filaments to control growth. Profilin

works on G-actin monomers exchanging ADP for ATP, promoting addition of monomers to a growing chain.

F-actin filaments are held together by relatively weak bonds compared to the covalent bonds of the monomers of nucleic acids, thus allowing for easier disassembly when desired. Actin's amino acid sequence is optimized, having diverged only a relatively small amount (20%) between algae and humans. Mutations in the actin gene result in muscular diseases and/or deafness.

Tubulin

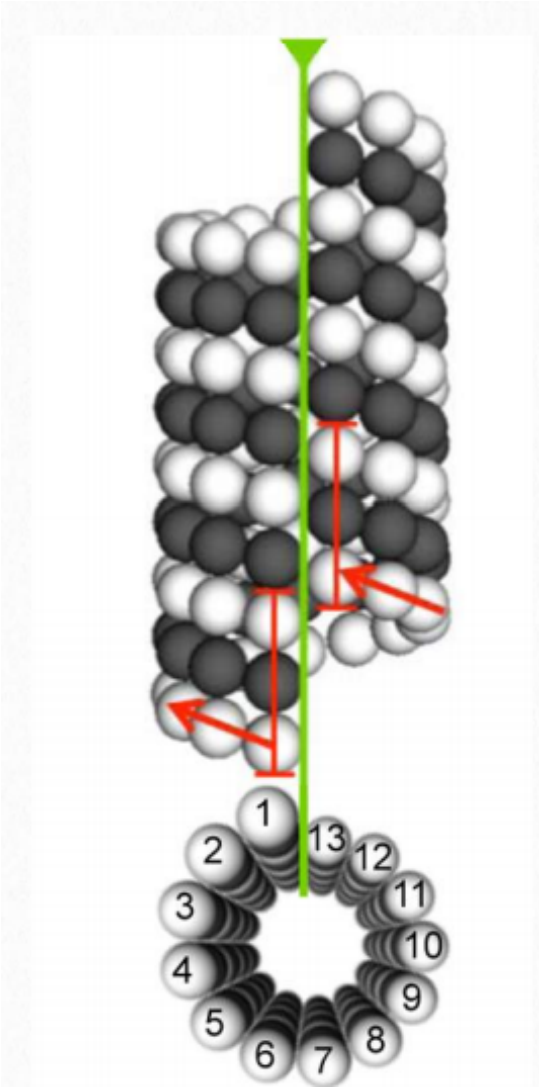


Figure 2.104 Microtubule structure Wikipedia

Tubulin proteins are the monomeric building blocks of eukaryotic microtubules (Figure 2.104 & 2.105). Bacterial (TubZ) and archaeon (FtsZ) equivalents are known. The α-tubulin and β-tubulin proteins polymerize to make microtubule structures in the cytoplasm of cells. Microtubules are major components of the cytoskeleton of

eukaryotic cells, providing structural support, transport within the cell, and functions necessary for segregation of DNAs during cell division.

Dimerization of the α-tubulin and β-tubulin proteins is necessary for polymerization and requires that the subunits bind to GTP. Microtubules only grow in one direction. β- tubulin is found on the plus end of the tubule (growth end = plus end) and α-tubulin is exposed on the other end (non-growth end = minus end). Dimers of α-tubulin/β-tubulin are incorporated into growing microtubules in this orientation. If a dimer is bound to GDP instead of GTP, it tends to be unstable and fall apart, whereas those bound to GTP stably assemble into microtubules.

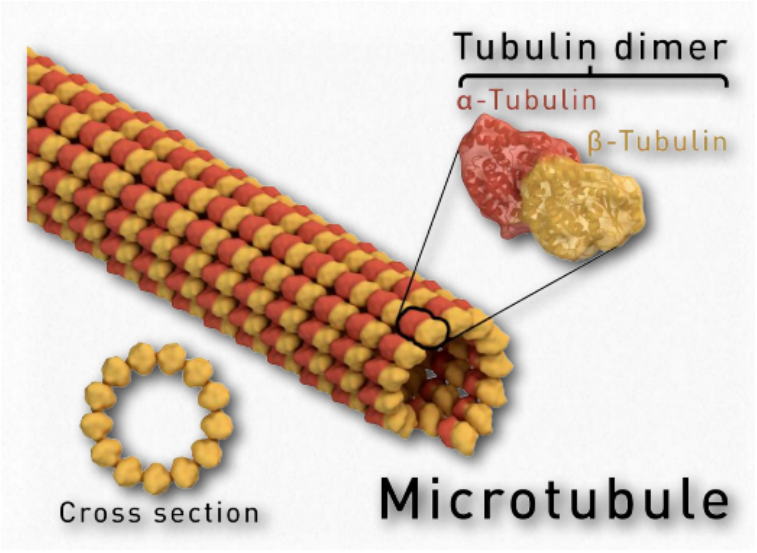


Figure 2.105 – Microtubule anatomy Wikipedia

Microtubules

Microtubules, along with microfilaments and intermediate filaments (see HERE) constitute the cytoskeleton of cells. Found in the cytoplasm, they are found in eukaryotic cells, as well as some bacteria. Microtubules help to give cells structure. They comprise the inner structure of flagella and cilia and provide rail-like surfaces for the transport of materials within cells.

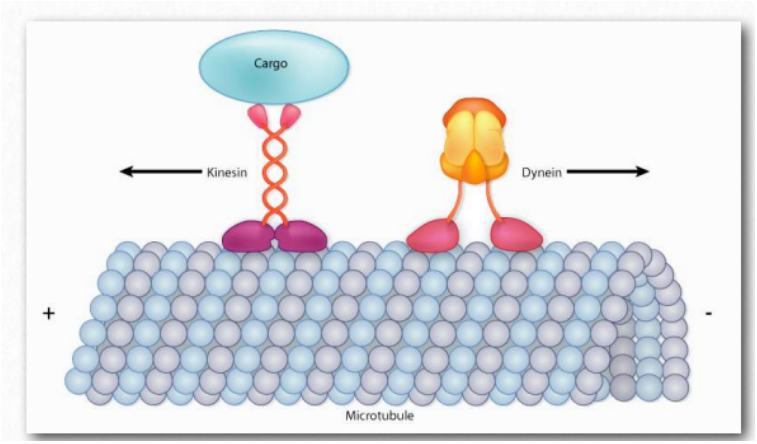


Figure 2.106 – Kinesin and dynein "walk" along microtubules, but move in opposite directions Image by Aleia Kim

Polymerization of α- tubulin and β-tubulin to form microtubules occurs after a nucleating event. Individual units get arranged in microtubule organizing centers (MTOCs), an example of which is the centrosome. Centrosomes are focal points of connection of microtubules. Basal bodies of cilia and flagella also help to organize microtubules.

Motor proteins

From the transport of materials within a cell to the process of cytokinesis where one cell splits into two in mitosis, a cell has needs for motion at the molecular level. Secretory vesicles and organelles must be transported. Chromosomes must be separated in mitosis and meiosis.

The proteins dynein and kinesin (Figure 2.106) are necessary for intracellular movement. These motor proteins facilitate the movement of materials inside of cells along microtubule "rails".

These motor proteins are able to move along a portion of the cytoskeleton by converting chemical energy into motion with the hydrolysis of ATP. An exception is flagellar rotation, which uses energy provided from a gradient created by a proton pump.

Kinesins and dyneins

As noted, kinesins and dyneins navigate in cells on microtubule tracks (Figure 2.108 & Movie 2.4). Most kinesins move in the direction of the synthesis of the microtubule (+ end movement), which is generally away from the cell center and the opposite direction of movement of dyneins, which are said to do retrograde transport toward the cell center. Both proteins provide movement functions necessary for the processes of mitosis and meiosis. These include spindle formation, chromosome separation, and shuttling of organelles, such as the mitochondria, Golgi apparatuses, and vesicles.

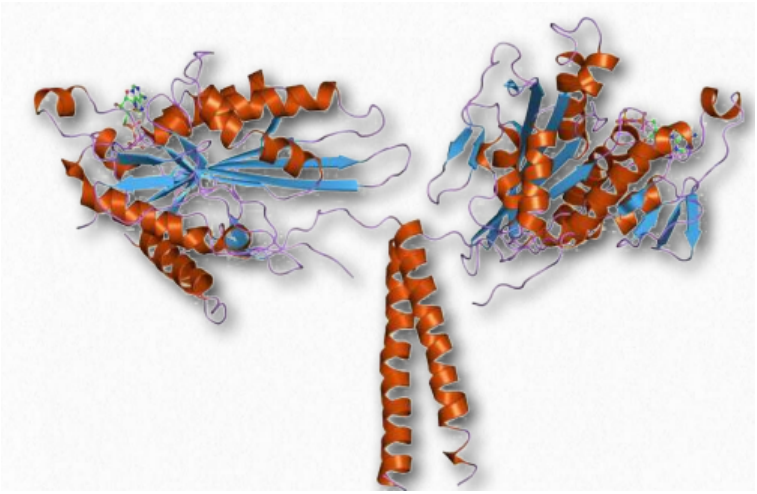


Figure 2.107 – Kinesin. "Feet" are at the top.

Kinesins are comprised of two heavy chains and two light chains. The head motor domains of heavy chains (in the feet) use energy of ATP hydrolysis to do mechanical work for the movement along the microtubules. There are at least fourteen distinct kinesin families and probably many related ones in addition.

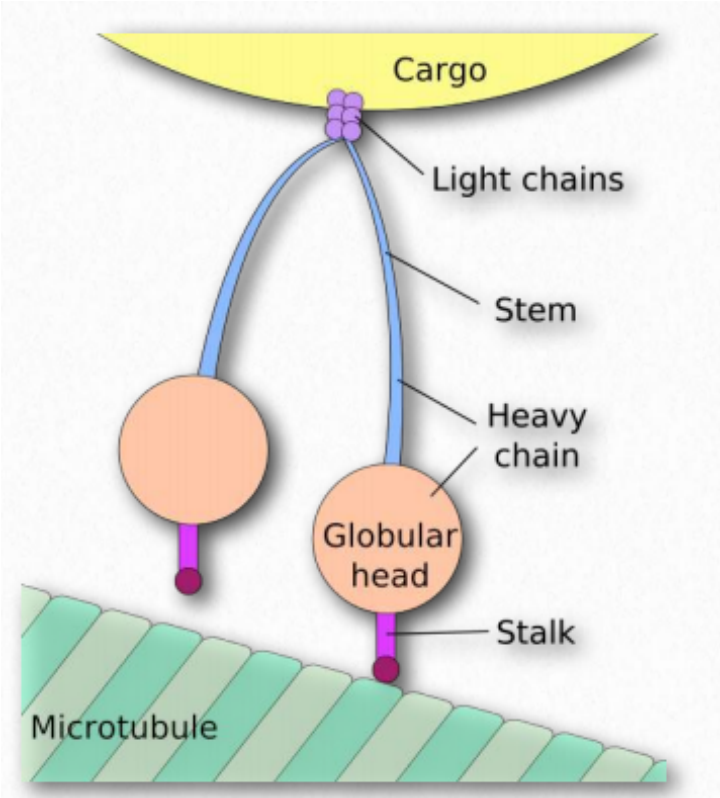
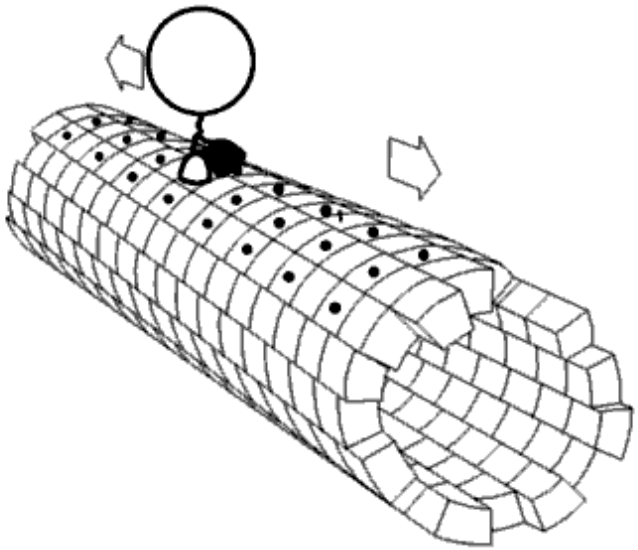


Figure 2.108 – Nomenclature of dynein. The "feet" of Figure 2.105 are the stalk and globular head of the structure here. Wikipedia

Dyneins are placed into two groups – cytoplasmic and axonemal (also called ciliary or flagellar dyneins – Figure 2.109). Dyneins are more complex in structure than kinesins with many small polypeptide units. Notably, plants do not have dynein motor proteins, but do contain kinesins.



Movie 2.4 The motor protein kinesin walking down a microtubule. Image used with permission (Public Domain; zp706).

Myosin

An important group of motor proteins in the cell is the myosins. Like kinesins and dyneins, myosins use energy from hydrolysis of ATP for movement. In this case, the movement is mostly not along microtubules, but rather along microfilaments comprised of a polymer of actin (F-actin). Movement of myosin on actin is best known as the driving force for muscular contraction. Myosins are a huge family of proteins, all of which bind to actin and all of which involve motion. Eighteen different classes of myosin proteins are known.

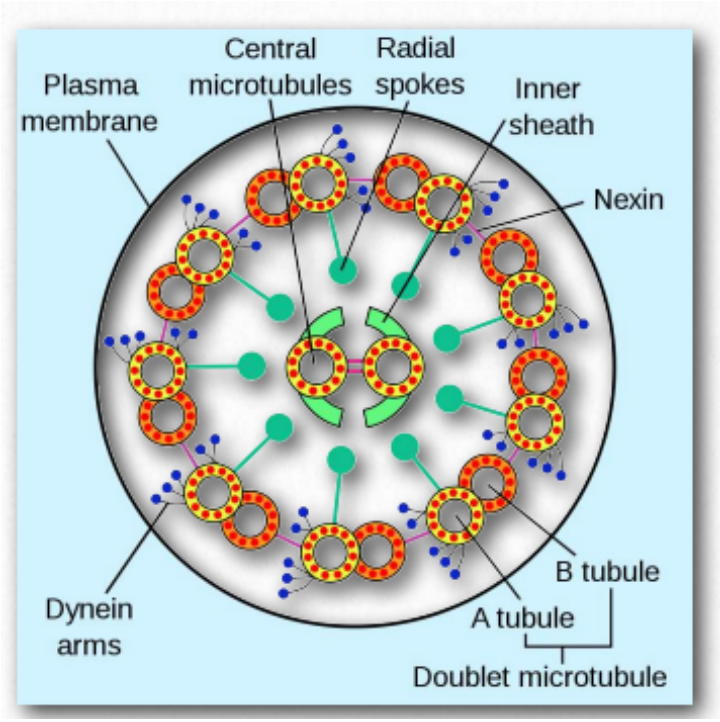


Figure 2.109 – Dynein in an axoneme Wikipedia

Myosin II is the form responsible for generating muscle contraction. It is an elongated protein formed from two heavy chains with motor heads and two light chains. Each myosin motor head binds actin and has an ATP binding site. The myosin heads bind and hydrolyze ATP. This hydrolysis produces the energy necessary for myosin to walk toward the plus end of an actin filament.

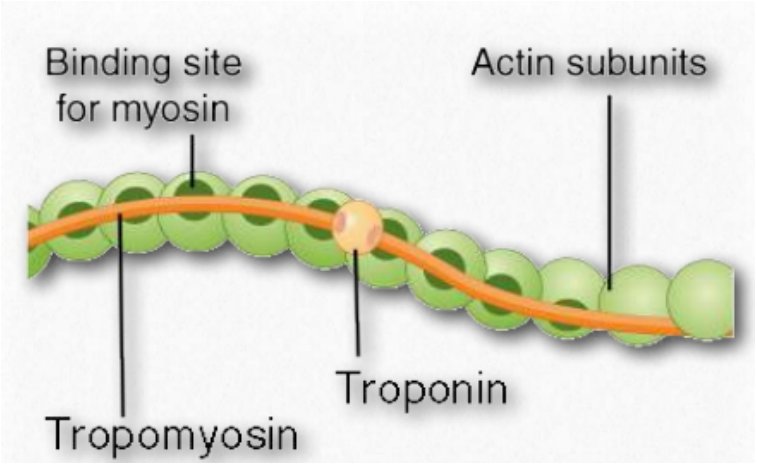


Figure 2.110 – Actin filament anatomy Wikipedia

Non-muscle myosin IIs provide contraction needed to power the action of cytokinesis. Other myosin proteins are involved in movement of non-muscle cells. Myosin I is involved in intracellular organization. Myosin V performs vesicle and organelle transport. Myosin XI provides movement along cellular microfilament networks to facilitate organelle and cytoplasmic streaming in a particular direction.

Structure

Myosins have six subunits, two heavy chains and four light chains. Myosin proteins have domains frequently described as a head and a tail (Figure 2.111). Some also describe an intermediate hinge region as a neck. The head portion of myosin is the part that binds to actin. It uses energy from ATP hydrolysis to move along the actin filaments. In muscles, myosin proteins form aggregated structures referred to as thick filaments. Movements are directional.

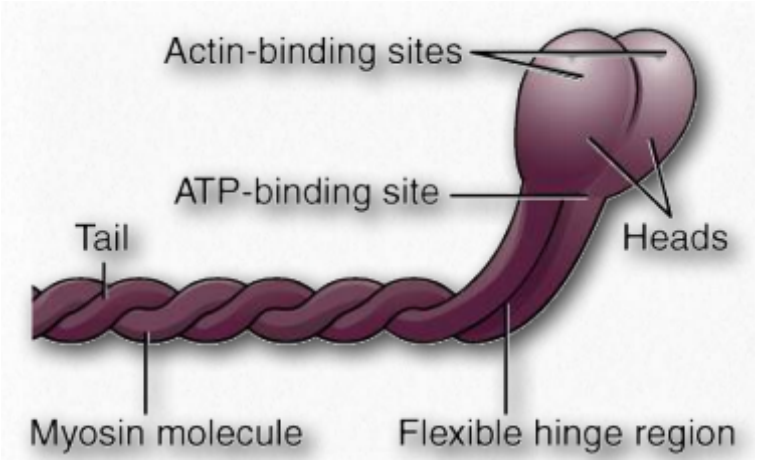


Figure 2.111 – Myosin protein anatomy Wikipedia

Structural considerations of muscular contraction

Before we discuss the steps in the process of muscular contraction, it is important to describe anatomical aspects of muscles and nomenclature.

There are three types of muscle tissue – skeletal (striated), smooth, and (in vertebrates) cardiac. We shall concern ourselves mostly here with skeletal muscle tissue. Muscles may be activated by the central nervous system or, in the case of smooth and cardiac muscles, may contract involuntarily. Skeletal muscles may be slow twitch or fast twitch.

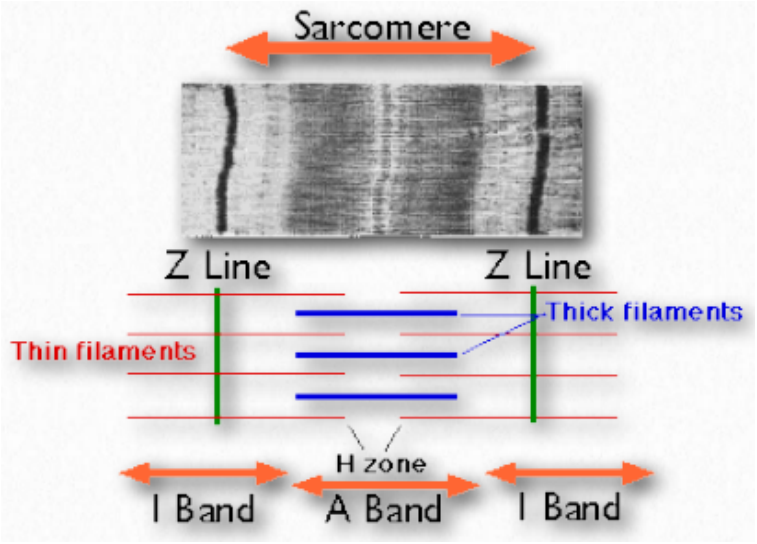


Figure 2.112 – Structural components of muscle. Wikipedia

Sarcomeres

Sarcomeres are described as the basic units comprising striated muscles and are comprised of thick (myosin) and thin (actin) filaments and a protein called titin. The filaments slide past each other in muscular contraction and then backwards in muscular relaxation. They are not found in smooth muscles.

Under the microscope, a sarcomere is the region between two Z-lines of striated muscle tissue (Figure 2.112). The Z-line is the distinct, narrow, dark region in the middle of an I-band. Within the sarcomere is an entire Aband with its central H-zone. Within the Hzone are located tails of myosin fibers, with the head pointed outwards from there projecting all the way to the I-band. The outside of the Aband is the darkest and it gets lighter moving towards the center.

Within the Iband are located thin filaments that are not occupied with thick myosin filaments. The Aband contains intact thick filaments overlaying thin filaments except in the central H zone, which contains only thick filaments. In the center of the H-zone is a line, known as the M-line. It contains connecting elements of the cellular cytoskeleton. In muscular contraction, myosin heads walk along pulling their tails over the actin thin filaments, using energy from hydrolysis of ATP and pulling them towards the center of the sarcomere.

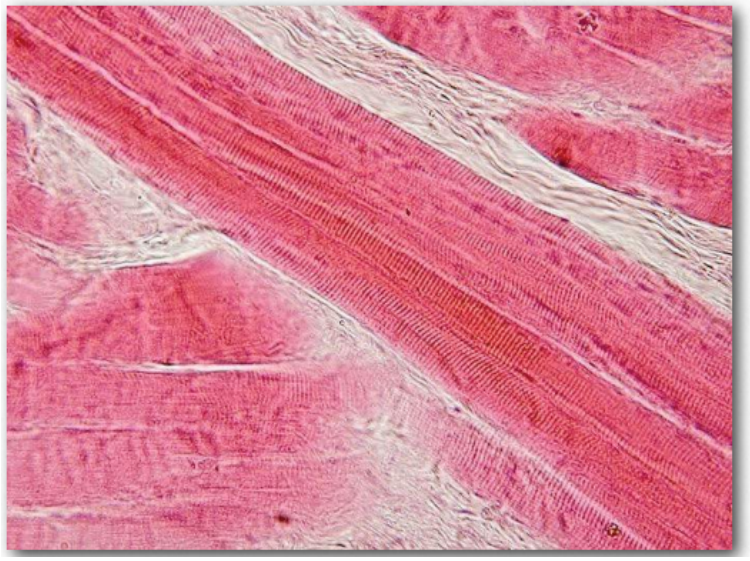


Figure 2.113 – Skeletal muscle longitudinal section. Wikipedia

Sarcolemma

The sarcolemma (also known as the myolemma) is to muscle cells what the plasma membrane is to other eukaryotic cells – a barrier between inside and outside. It contains a lipid bilayer and a glycocalyx on the outside of it. The glyocalyx contains polysaccharides and connects with the basement membrane. The basement membrane serves a s scaffolding to connect muscle fibers to. This connection is made by transmembrane proteins bridging the actin cytoskeleton on the inside of the cell with the basement membrane on the outside. On the ends of the muscle fibers, each

sarcolemma fuses with a tendon fiber and these, in turn, adhere to bones.

Sarcoplasmic reticulum

The sarcoplasmic reticulum (Figure 2.114) is a name for the structure found within muscle cells that is similar to the smooth endoplasmic reticulum found in other cells. It contains a specialized set of proteins to meet needs unique to muscle cells. The organelle largely serves as a calcium "battery," releasing stored calcium to initiate muscular contraction when stimulated and taking up calcium when signaled at the end of the contraction cycle. It accomplishes these tasks using calcium ion channels for release of the ion and specific calcium ion pumps to take it up.

Movement direction

All myosins but myosin VI move towards the + end (the growing end) of the microfilament. The neck portion serves to link the head and the tail. It also a binding site for myosin light chain proteins that form part of a macromolecular complex with regulatory functions. The tail is the point of attachment of molecules or other "cargo" being moved. It can also connect with other myosin subunits and may have a role to play in controlling movement.

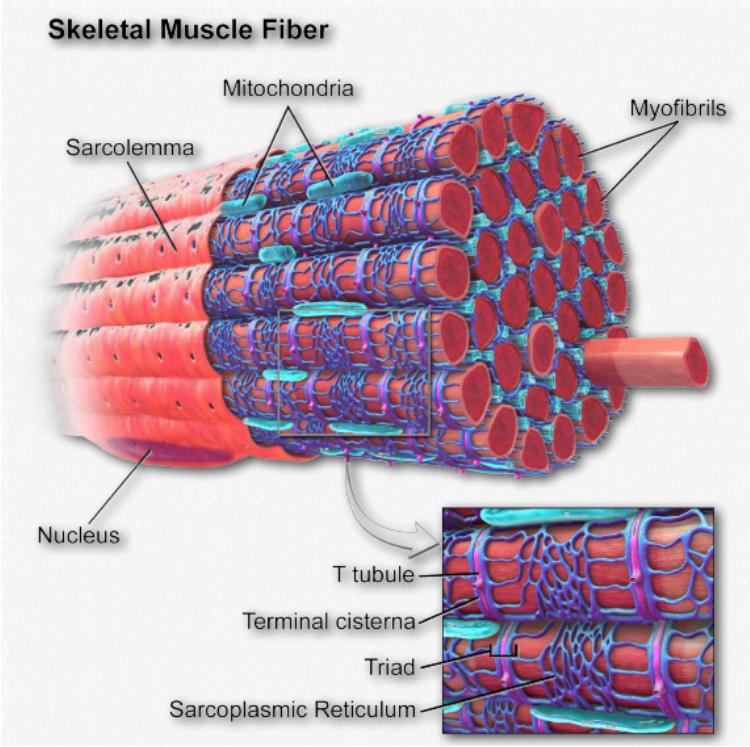


Figure 2.114 – Anatomy of a muscle fiber Wikipedia

Muscular contraction

The sliding filament model has been proposed to describe the process of muscular tension/contraction. In this process a repeating set of actions slide a thin actin filament over a thick myosin filament as a means of creating tension/ shortening of the muscle fiber.

Steps in the process occur as follows:

A. A signal from the central nervous system (action potential) arrives at a motor neuron, which it transmits towards the neuromuscular junction (see more on the neurotransmission part of the process HERE)

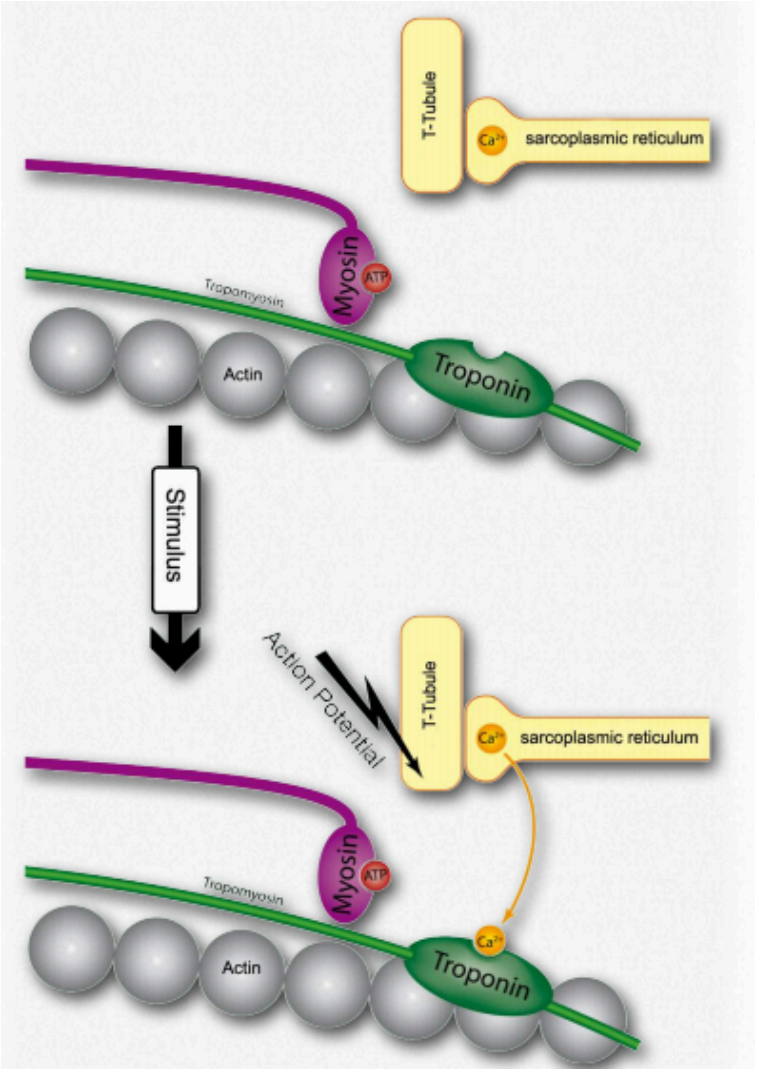


Figure 2.115 – 1. Activation of a muscle cell by release of calcium (step H) Wikipedia

B. At the end of the axon, the nerve signal stimulates the opening of calcium channels at the axon terminus causing calcium to flow into the terminal.

C. Movement of calcium into the axon of the nerve causes

acetylcholine (a neurotransmitter) in synaptic vesicles to fuse with the plasma membrane. This causes the acetylcholine to be expelled into the synaptic cleft between the axon and the adjacent skeletal muscle fiber.

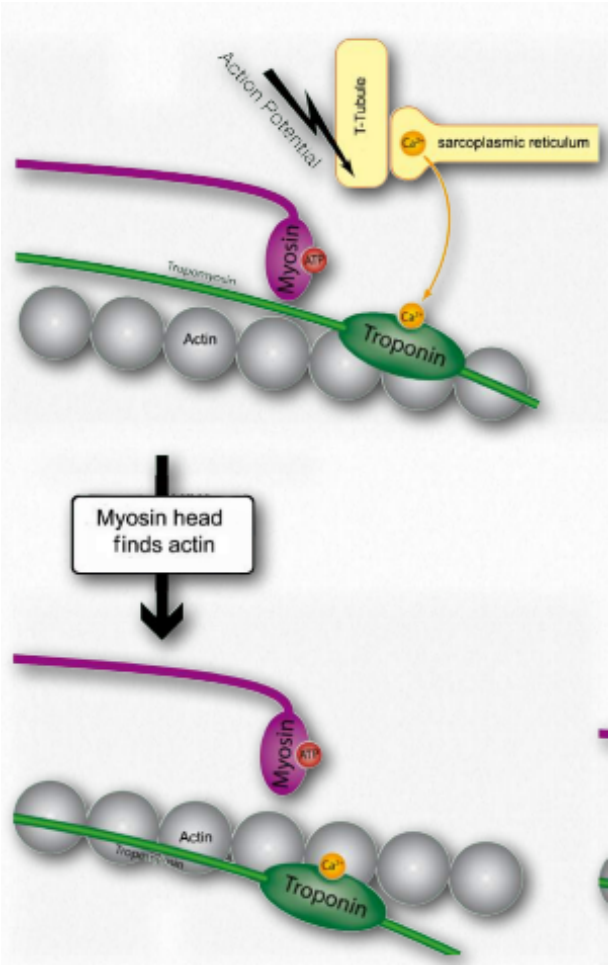


Figure 2.116 – 2. Calcium binding by troponin allows myosin to access actin sites (I). Wikipedia

D. Acetylcholine diffuses across the synapse and then binds to nicotinic acetylcholine receptors on the neuromuscular junction, activating them.

E. Activation of the receptor stimulates opening gates of sodium and potassium channels, allowing sodium to move into the cell and potassium to exit. The polarity of the membrane of the muscle cell (called a sarcolemma – Figure 2.111) changes rapidly (called the end plate potential).

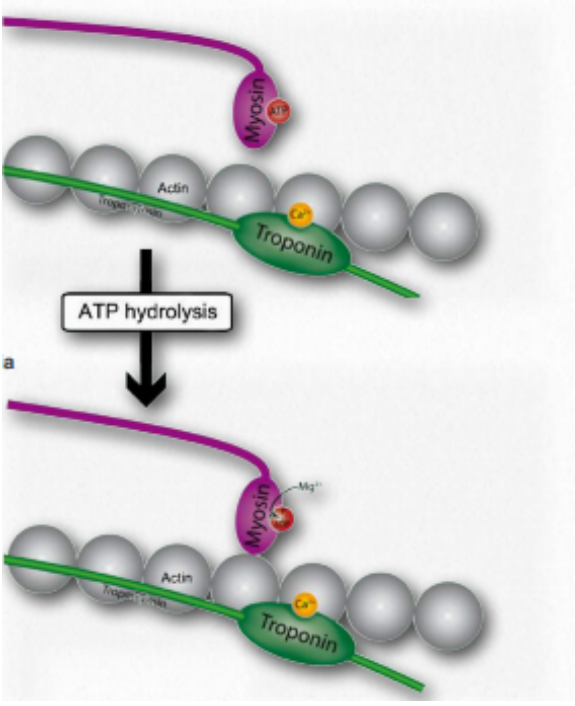


Figure 2.117 – 3. ATP cleavage by myosin allows actin attachment (J) Wikipedia

F. Change in the end plate potential results in opening of voltage sensitive ion channels specific for sodium or potassium only to Figure 2.117 – 3. ATP cleavage by myosin allows actin attachment (J) Wikipediaopen, creating an action potential (voltage change) that spreads throughout the cell in all directions.

G. The spreading action potential depolarizes the inner muscle fiber and opens calcium channels on the sarcoplasmic reticulum (Figure 2.115).

H. Calcium released from the sarcoplasmic reticulum binds to troponin on the actin filaments (Figure 2.115).

I. Troponin alters the structure of the tropomyosin to which is it bound. This causes tropomyosin to move slightly, allowing access to myosin binding sites on the microfilament (also called thin filament) that it was covering (Figure 2.116).

J. Myosin (bound to ATP) cleaves the ATP to ADP and Pi, which it holds onto in its head region and then attaches itself to the exposed binding sites on the thin filaments causing inorganic phosphate to be released from the myosin followed by ADP (Figure 2.117).

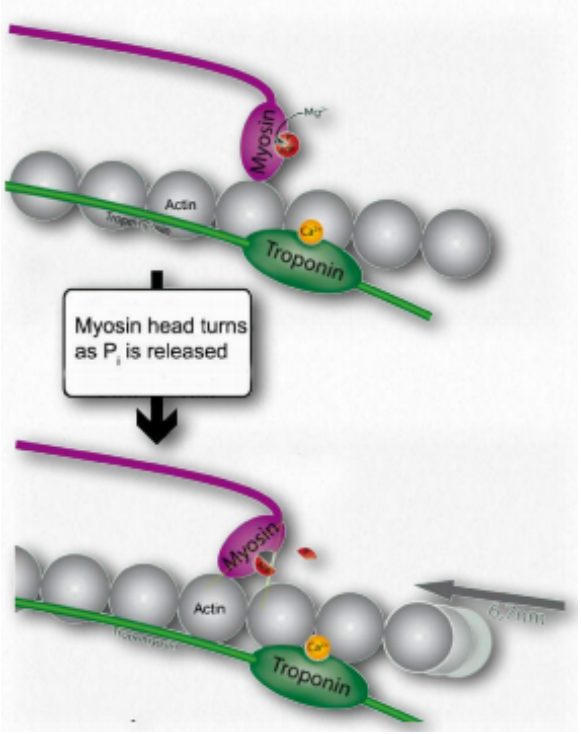


Figure 2.118 – 4. Release of Pi causes myosin hinge to bend. Thin filament pulled left (K). Wikipedia

K. Release of ADP and Pi is tightly coupled to a bending of the myosin hinge, resulting in what is called the power stroke. This

causes the thin filament to move relative to the thick fibers of myosin (Figures 2.118 & 2.119).

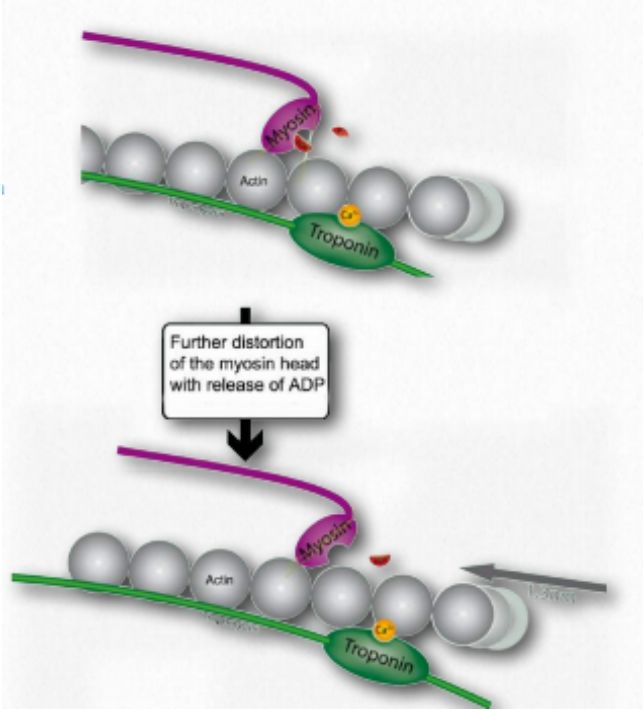


Figure 2.119 – 5. Release of ADP favors further bending of hinge and movement of thin filament leftward (K). Wikipedia

L. Such movement of the thin filaments causes the Z lines to be pulled closer to each other. This results in shortening of the sarcomere as a whole (Figure 2.122) and narrowing of the I band and the H zones (Figure 2.123). M. If ATP is available, it binds to myosin, allowing it to let go of the actin (Figures 2.120 & 2.121). If ATP is not available, the muscle will remain locked in this state. This is the cause of rigor mortis in death – contraction without release of muscles

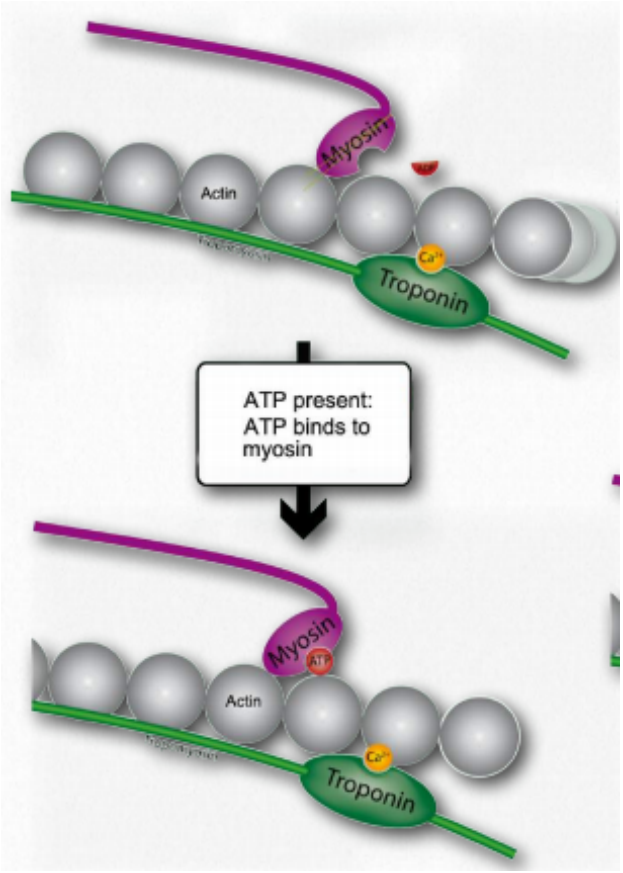


Figure 2.120 – When ATP is present, it binds to myosin (M). Wikipedia

N. After myosin has bound the ATP, it hydrolyzes it, producing ADP and Pi, which are held by the head. Hydrolysis of ATP resets the hinge region to its original state, unbending it. This unbent state is also referred to as the cocked position.

O.If tropomyosin is still permitting access to binding sites on actin, the process repeats so long as ATP is available and calcium remains at a high enough concentration to permit it to bond to troponin.

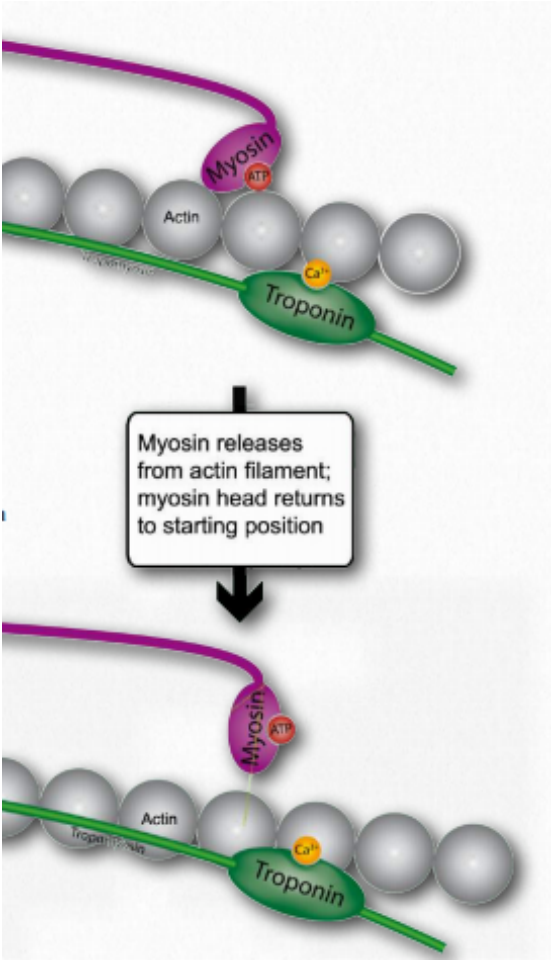


Figure 2.121 – Binding of ATP favors release of myosin from actin site (N) Wikipedia

Relaxation of the muscle tension occurs as the action potential in the muscle cell dissipates. This happens because all of the following things happen 1) the nerve signal stops; 2) the neurotransmitter is degraded by the enzyme acetylcholinesterase; and 3) the calcium concentration declines because it is taken up by the sarcoplasmic reticulum.

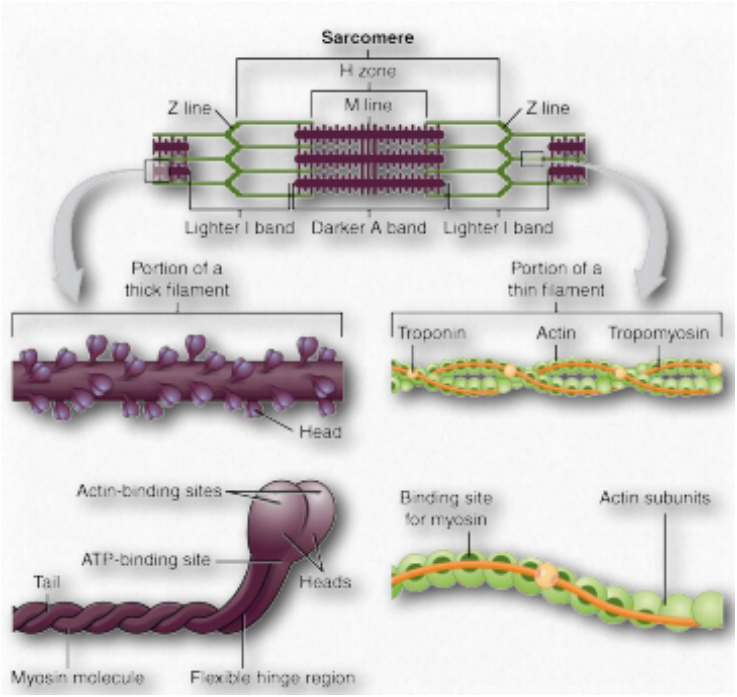


Figure 2.122 – Sarcomere Anatomy Wikipedia

It should be noted that the sarcoplasmic reticulum is always taking up calcium. Only when its calcium gates are opened by the action potential is it unable to reduce cellular calcium concentration. As the action potential decreases, then the calcium gates close and the sarcoplasmic reticulum "catches up" and cellular calcium concentrations fall. At that point troponin releases calcium, tropomyosin goes back to covering myosin binding sites on actin, myosin loses its attachment to actin and the thin filaments slide back to their original positions relative to the myosin thick filaments.

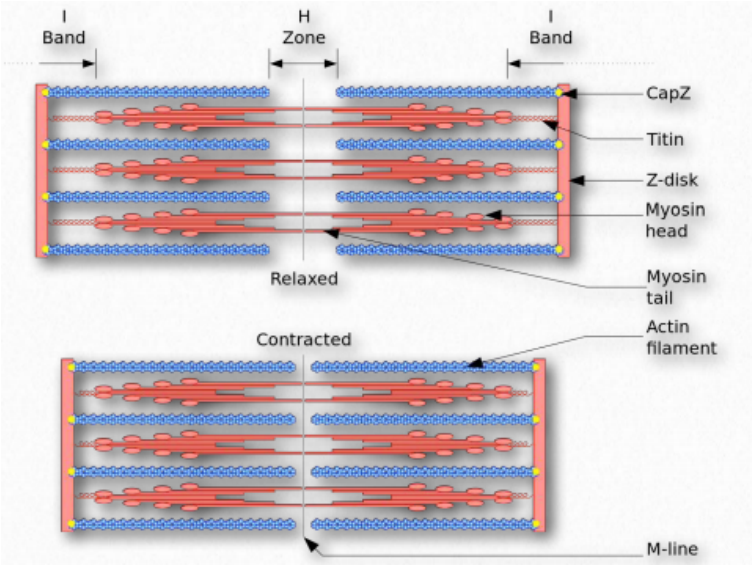


Figure 2.123 – The Sliding Filament Model of Muscular Contraction Wikipedia

Tropomyosin

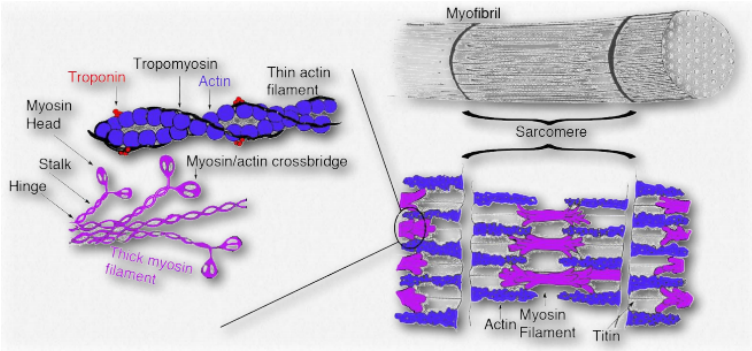


Figure 2.124 – Tropomyosin and troponin in muscle anatomy Wikipedia

Tropomyosins are proteins that interact with actin thin filaments to help regulate their roles in movement, both in muscle cells and non-muscle cells (Figure 2.124). Tropomyosins interact to form head-to-toe dimers and perch along the α-helical groove of an actin filament. The isoforms of tropomyosin that are in muscle cells

control interactions between myosin and the actin filament within the sarcomere and help to regulate contraction of the muscle. In other cells, nonmuscle tropomyosins help to regulate the cytoskeleton's functions.

The interactions of tropomyosin with the cytoskeleton are considerably more complicated than what occurs in muscle cells. Muscle cells have five tropomyosin isoforms, but in the cytoskeleton of non-muscle cells, there are over 40 tropomyosins.

Troponin

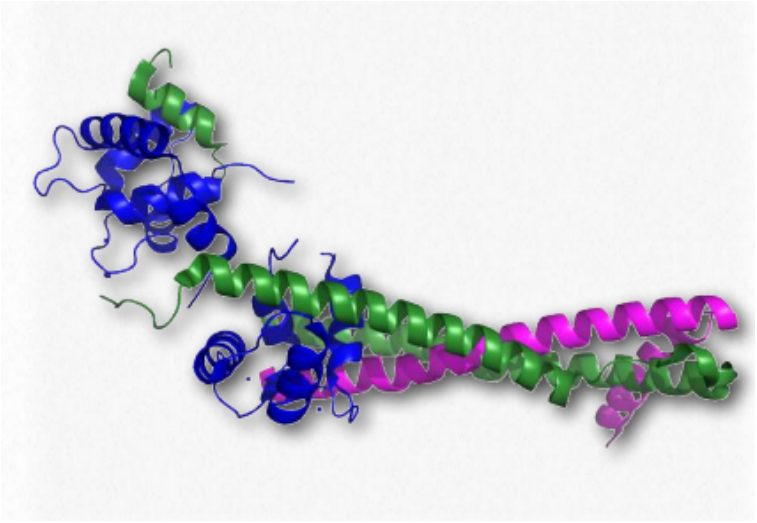


Figure 2.125 – Troponin complex of muscle. Blue = troponin C, magenta = troponin T , green = troponin I

The troponins involved in muscular contraction are actually a complex of three proteins known as troponin I, troponin C, and troponin T (Figure 2.125). They associate with each other and with tropomyosin on actin filaments to help regulate the process of muscular contraction. Troponin I prevents binding of myosin's head to actin and thus prevents the most important step in contraction.

Troponin C is a unit that binds to calcium ions. Troponin T is responsible for binding all three proteins to tropomyosin. Troponins in the bloodstream are indicative of heart disorders. Elevation of

troponins in the blood occurs after a myocardial infarction and can remain high for up to two weeks.

Actinin

Actinin is a skeletal muscle protein that attaches filaments of actin to Z-lines of skeletal muscle cells. In smooth muscle cells, it also connects actin to dense bodies.

Titin

Titin (also known as connectin) is the molecular equivalent of a spring that provides striated muscle cells with elasticity. It is the third most abundant protein in muscle cells. The protein is enormous, with 244 folded individual protein domains spread across 363 exons (largest known number), with the largest known exon (17,106 base pairs long), and it is the largest protein known (27,000 to 33,000 amino acids, depending on splicing).

Unstructured sequences

The folded protein domains are linked together by unstructured sequences. The unstructured regions of the protein allow for unfolding when stretching occurs and refolding upon relaxation. Titin connects the M and Z lines in the sarcomere (Figure 2.123). Tension created in titin serves to limit the range of motion of the sarcomere, giving rise to what is called passive stiffness.

Skeletal and cardiac muscles have slight amino acid sequence variations in their ti tin proteins and these appear to relate to differences in the mechanical characteristics of each muscle.

Energy backup for muscle energy

Myoglobin was described as a molecular batter for oxygen. Muscle cells have a better of their own for ATP. The is important for animals, but not for plants because a plant's need for energy is different than an animal's. Plants do not need to access energy sources as rapidly as animals do, nor do they have to maintain a constant internal temperature. Plants can neither flee predators, nor chase prey. These needs of animals are much more immediate and require that energy stores be accessible on demand. Muscles, of course, enable the motion of animals and the energy required for muscle contraction is ATP. To have stores of energy readily available,

muscles have, in addition to ATP, creatine phosphate for energy and glycogen for quick release of glucose to make more energy. The synthesis of creatine phosphate is a prime example of the effects of concentration on the synthesis of high energy molecules. For example, creatine phosphate has an energy of hydrolysis of -43.1 kJ/mol whereas ATP has an energy of hydrolysis of -30.5 kJ/mol. Creatine phosphate, however, is made from creatine and ATP in the reaction shown in Figure 2.126. How is this possible?

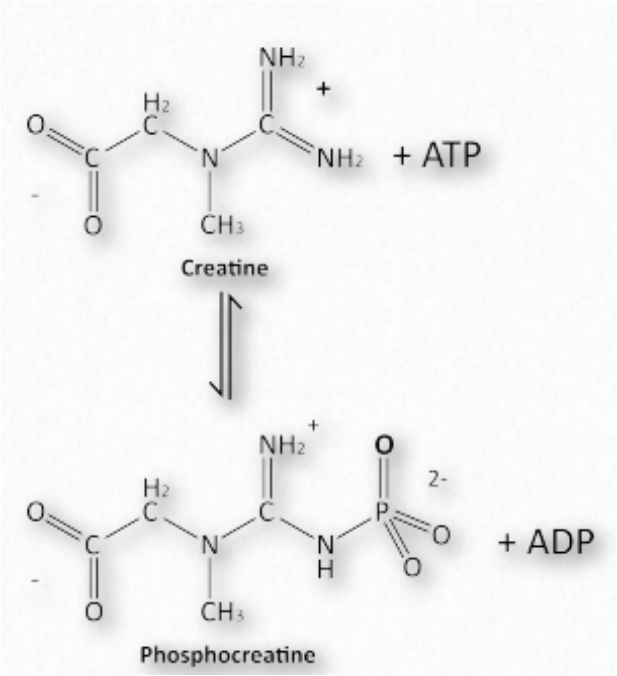


Figure 2.126 – Phosphorylation of creatine (phosphocreatine) – making of a creatine phosphate battery Image by Aleia Kim

The $\Delta G^{\circ\prime}$ of this reaction is +12.6 kJ/mol, reflecting the energies noted above. In a resting muscle cell, ATP is abundant and ADP is low, driving the reaction downward, creating creatine phosphate. When muscular contraction commences, ATP levels fall and ADP levels climb. The above reaction then reverses and proceeds to synthesize ATP immediately. Thus, creatine phosphate acts like a

battery, storing energy when ATP levels are high and releasing it almost instantaneously to create ATP when its levels fall.

This is where you can add appendices or other back matter.