

MACROMOLECULES, Part 2* -- Proteins

General: Proteins are long chain polymers of monomeric units called amino acids. Proteins perform an amazing number of tasks in cells. Generally speaking, these include:

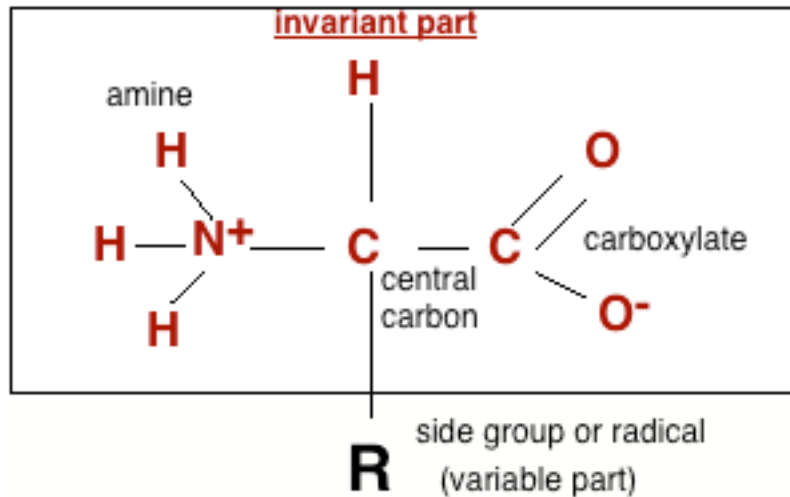
- **Structure** -- both within cells (examples: "intracellular skeletal proteins") and extracellular (connective tissue proteins such as collagen).
- Act as organic **catalysts** called **enzymes** -- these regulate the rate of chemical reactions that would proceed only slowly, if at all, under the typical conditions of temperature and pressure found in a cell.
- **carriers**: both across membranes (**membrane transport proteins** -- see next class) and as a means to increase the ability of the body to carry and/or store certain substances. A good example is hemoglobin which increases the amount of oxygen that can be carried in the blood.
- **hormones and other signal molecules** -- not all hormones are sterols. There are even more protein hormones. Examples include growth hormone, oxytocin (important in birth and nursing), insulin, glucagon (both important in glucose regulation), and tropic hormones such as thyrotropin or lutenizing hormone. There are many others
- **receptors**: proteins can recognize exact shapes and therefore are perfect to register the presence or absence of crucial molecules. Receptors can be found just about everywhere within the organism and are the most important means whereby the organism finds out about its immediate chemical environment.
- **immunity**: antibodies are proteins that recognize particular foreign substances and participate in their removal.
- **movement** -- contractile proteins in muscles and those responsible for cell division and cell movements.

Amino Acids and Proteins: The monomers that make up **proteins are called amino acids**. All amino acids have the same general **structure; it is shown on the next page and you should learn it**. The name "amino acid" comes from fact that each type of amino acid has both an amine (-NH₂) and carboxylic acid (-COOH) group.

All amino acids contain the same **invariant region**. This consists of the amino group, a central carbon atom and its hydrogen, and the carboxylic acid. Please note in the figure on the next page that, under conditions normally found in cells, the amine actually has gains a hydrogen and is thus -NH₃⁺ and the carboxylic acid has lost a proton and is -COO⁻. What gives different types of amino acids their own unique properties are the **side or "R" groups** that are

* © 2006 by K.N. Prestwich, Dept. Biology, College of the Holy Cross, Worcester, MA 01610

attached to the central carbon. Organisms use 20 different types of amino acids¹; thus there are 20 types of "side groups".



Of these 20, we must obtain eight in our diet (**essential amino acids**) and we can either synthesize the remaining 12 **non-essential** amino acids or get them from food.

Don't get the idea that non-essential amino acids are not needed to make proteins -- they certainly are. The reference for deciding if something is essential or not is whether or not the body can synthesize it.

The side groups can vary according to the following properties:

- **size**
- **shape**
- **polarity and charge arrangement** (*i.e.*, how polar (full or partial charge) and the positions of the polar areas). Polarity influences an amino acid's ability to react with other molecules (often an important step in catalysis). It is very important to remember that **some side groups are non-polar**.
- **acid/base characteristics** (some side groups contain amines or carboxylic acid)
- **ability to form cross linkages with other amino acids** -- the side group of one type of amino acid, **cysteine**, is able to react with other cysteines and form a covalent bond often called a "disulfide linkage".

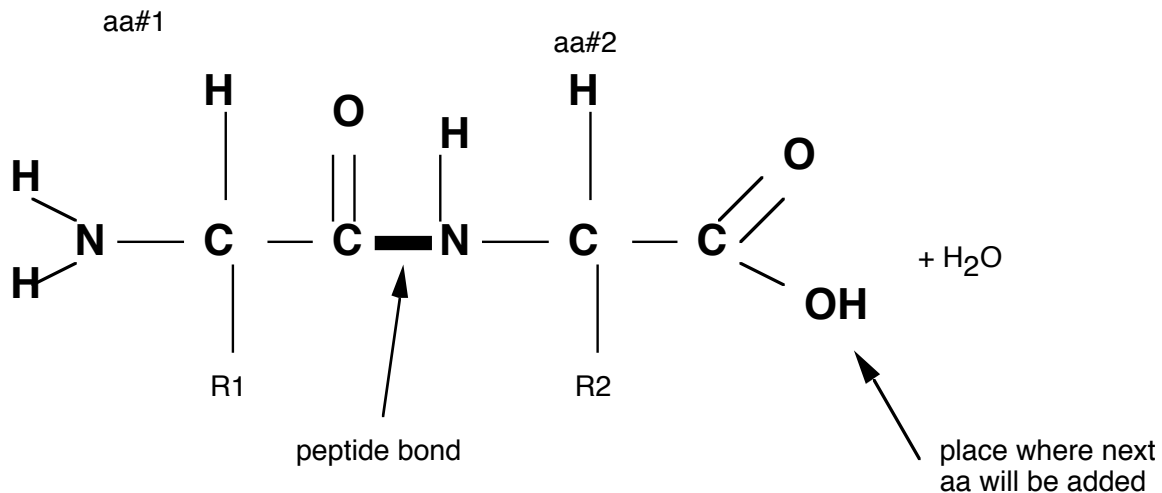
¹ There are actually a few more than 20 but most of these are rare and are slight modifications of one of the twenty amino acids. They are only rarely used in proteins and instead have other functions.

Protein Structure

At every level of the biological hierarchy, structure determines function. This is also true at the molecular level. The diversity in protein functions mentioned at the start of this handout related to an equal diversity in protein structures. In this section you will gain an appreciation for the complexity of protein structure and for the ways this complexity is achieved.

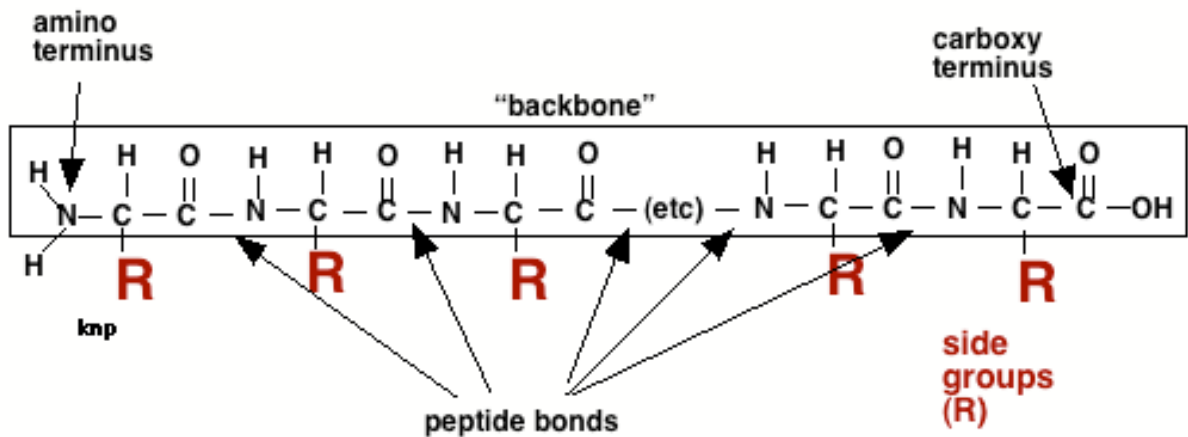
Primary Structure of Proteins: A chain of amino acids is called a peptide or polypeptide (depending on its overall length). Proteins consist of one or sometimes more polypeptides.

Polypeptides are formed when the amino end of one amino acid combines with the carboxyl end of another to form a union called the **peptide bond**. As in other macromolecules, this is a **dehydration reaction** -- water is released. On the other hand, the link between the amino acids (the peptide bond) is not a diester bond as in polysaccharides and lipids. Instead it is a C--N linkage.



As with other bonds formed by dehydration, the peptide bonds can be broken by the reverse "**hydration**" process, **hydrolysis** (as in carbs. and lipids). Note that there is nothing special about the "peptide bond". It is a typical covalent bond; peptide simply signifies that it joins amino acids in a peptide or protein.

The Primary Structure of Proteins: Thus, a polypeptide is a string of amino acids held together by peptide (C--N) bonds. It has a **repeating "backbone"** of N--C--C--N --C--C-- etc. that holds the molecule together covalently and a series of side groups that give the protein its unique properties. In this sense, it is very much like nucleic acids which have an invariant sugar phosphate backbone with nitrogenous bases that add unique characteristics to the molecule.



Note that every polypeptide has two ends -- for obvious reasons termed the **amino** and **carboxy terminuses**. All of the internal amine and carboxylic acid groups have reacted with each other. Thus, none of them act as acids or bases. As far as the "backbone" is concerned, a polypeptide has virtually no effect on pH . On the other hand, many polypeptides do affect the pH . The reason is that, as we learned earlier, the side groups of some amino acids are alkaline in character while others tend to be acidic. Thus, a protein that is rich in amino acids with acid side groups (example: aspartic acid) will tend to act as an acid while one with many amine groups on the side chain (examples: histidine and arginine) will tend to act as bases. So, if a particular protein tends to be somewhat acidic or alkaline, it is due to the chemistry of its side groups, not to the amino and carboxylic terminuses.

The sequence of amino acids on a polypeptide or protein is called the **primary (1°) structure** of the protein. The primary structure is determined by the structural gene for that protein. Each protein has its own structural gene – structural genes are nothing more than plans giving the primary structure of proteins.

A near infinity of primary structures is possible. This large number is due to

- the size of proteins (typically consisting of hundreds of linked amino acids),
- the fact that there are 20 different possible amino acids for each position on a protein, and,
- the fact that what is present at one position in theory does not affect what will be present at the next.

Let's look at the number of possibilities mathematically. If proteins were only one amino acid long (*i.e.*, not different than amino acids) there would be 20 possible "proteins" -- one for each amino acid. Now, what if we go to a two amino acid peptide. If what is present at the first (amino) position in no way influences what can be found at the second then there are $20 * 20 = 20^2 = 400$ possible combinations. Now, what if we look at three amino acid peptide. There are now

$20 \times 20 \times 20 = 20^3 = 8000$ combinations! A general formula for the number of possible combinations is the number of kinds amino acids (20) raised to the power of the number of amino acids in the primary structure. Thus:

The number of different sequences = $(20)^N$

Now, in fact organisms do not take advantage of all possible combinations -- many will not provide proteins with useful functions. But the fact remains that a staggering number of primary sequences can be constructed given the fact that proteins are usually of different lengths and that so many amino acid sequences are possible

Note -- a short protein that has a sequence identical to that found within a larger protein does not have the same structure or function. In fact, some enzymes are activated by cutting off a number amino acids. In the process the shape and function of an enzyme changes.

For all possible proteins with total lengths of exactly 200 amino acids, how many theoretically possible primary structures are there?

In DNA, how many different nitrogenous base sequences can be written on a section of a molecule that is 200 nucleotides long?

One final bit about primary structures. We are about to look at so-called higher order structure. We will see that higher order structure is determined by the interaction of the primary structure with the chemical and physical environment. Thus, different higher order structures and therefore different protein properties will be determined by features of the primary structure:

- a. the types of amino acids
- b. their proportions
- c. their order
- d. and the environment.

Higher Order Protein Structure: All proteins must have primary structure. However, rather than being a one-dimensional thread, peptides tend to interact with themselves and other molecules to produce complex shapes.

Secondary Structure Secondary structure involves the backbone of the protein forming hydrogen bonds with itself. In both cases bonds form between the carbonyl (C=O) groups that are now present on the inside of the molecule just before each peptide bond (the carbonyl is all that is left of the carboxylic acid -- see earlier diagram in these notes) and the N-H of the former amine, also associated with the peptide bond. **However, the carbonyl's oxygen at one side of a given peptide bond DOES NOT H-bond with the amine's H on the other side of the same peptide bond.** This is because the bonds inside of the backbone rotate the adjacent C=O and N-H so that they are separated by

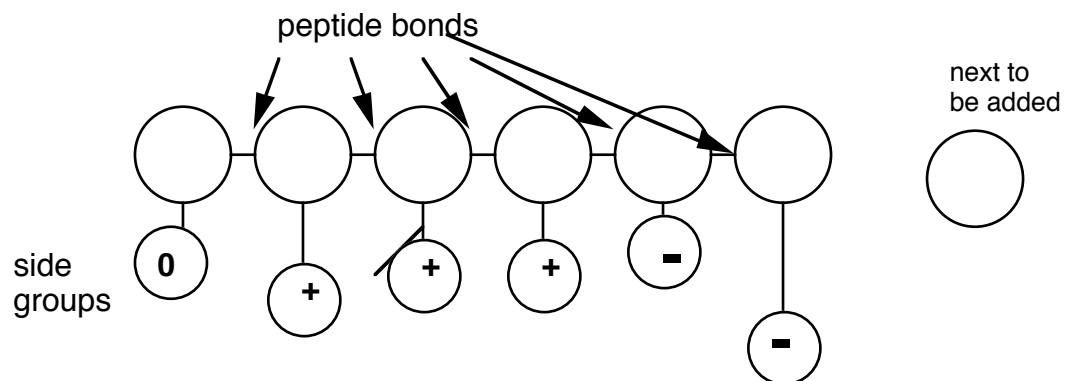
enough distance so that they cannot interact. However, eventually the molecule spirals around so that such bonding is possible and under the correct conditions it occurs.

There are two types of secondary structure and you should know them both

- **Alpha helix (α -helix)** -- in this structure, the C=O H-bonds with an N-H that is 3 amino acids away. This structure repeats over and over with the result being that the H-bonds between a whole series of amino acids and their near (3 positions away) neighbors twists the chain into a coil -- an alpha helix. Look in your textbook on page 76 and check your computer disk (CD) to see this structure.
- **Beta-pleated sheet (β -pleated sheet)** -- once again the link is between the oxygen of the main chain carbonyls and the hydrogen of the main-chain amines. However, this time, the H-bonded groups are located a long ways apart on the primary structure -- usually much further than 3 amino acid "residues". As the polypeptide chain folds back on forth on itself like a thread that has been bent, the result is a sheet like structure held together by a great many H bonds (again, between C=O and N-H on the backbone. Once again, check page 76 and/or your CD in your text to see a picture.

Most proteins show regions of secondary structure. However, please do not get the idea that it is mandatory and that every region of a polypeptide will show some secondary structure. That is simply not the case; typically there will be regions where there is no secondary structure and other areas of various sized α -helix and/or β -pleated sheet.

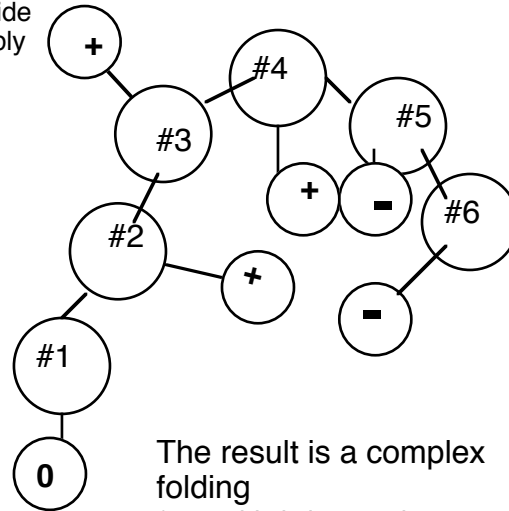
Tertiary (3°) Structure: Tertiary structure is caused by the interactions of the side (R) groups of amino acids with each other and with the environment. For example, imagine a polypeptide chain (a long-ish chain of amino acids) being formed with side groups with the following types of charge. Furthermore, imagine that is being formed in water (the environment in cells). Here is the growing chain:



What happens is that some amino acids repel each other, some attract each other, and others have no effect on each other. In addition, water may also get into the act. Here is a very rough idea of what results:

Aa#3 has a positive charge that is repelled outwards by other positive amino acid side groups. Water probably is attracted to it.

AA#4 and #5 attract each other



AA#6 and 2 attract each other

The result is a complex folding
(note: this is just to give you an idea of the type of thing that happens).

Notice that we now have the beginning of a three-dimensional shape that is held together by many polar interactions. Notice also that this shape is **not absolutely rigid**. The many weak hydrogen bonds and ionic attractions tend to stabilize the shape but not absolutely.

There are other mechanisms associated with tertiary structure. The first is one you probably learned in high school -- the **disulfide bridge**. One type of amino acid, cysteine, contains on its end a "**sulfhydryl (S-H) group**". The side chains on different cysteines can be catalyzed to react with each other so that a "**disulfide (S-S) bond**" is produced². Unlike the ionic and hydrogen bonds mentioned above, this is a strong covalent bond. Typically only a few are found in a given protein, but they have very important effects on structure. Finally, tertiary structure also often owes something to the "hydrophobic interaction". Many amino acid side chains are hydrophobic. In a watery environment, these are pushed to the inside of the molecule. The chain may twist in a number of ways to "hide" these from the water (actually, the water forces the twisting). Once safely tucked away from the water, these groups bind weakly to each other with the VanderWaals attraction.

² Let's worry neither about what happens to the two hydrogens nor about the mechanism. Just realize it is a catalyzed reaction.

Parts of some proteins are often found surrounded by lipid. This happens in membranes for instance. Which types of side groups get forced into the center in this case and what sort of attractions are involved?

Further note the **if anything disrupts the interactions holding the molecule together, the protein's tertiary structure will change**. If the alterations are small, and if the factors causing them are removed (example -- a some hydrogen ions are added and then later removed) the molecule will go back to its original shape. This is what happens to proteins in your cells when lactic acid is produced. The protons tend to loosen or break up some of the H-bonds and the molecules shape change and function decreases. However, when the acid is removed later, the molecules go back to their normal shape and their full function is returned.

On the other hand, sometimes the changes in shape are so severe that the molecule cannot get back to its original tertiary structure. When that happens we say that the protein is **denatured**. A common way to denature a protein is to heat it – when egg “whites” turn from clear to white as a result of heating what has happened is that the protein's have taken on entirely new tertiary (not primary) structures. The old H bonds were broken and new ones, not possible under cellular conditions, formed and gave a new structure. When the egg cools it does not return to its original shape for the simple reason that it is energetically difficult to do so-- the present non-functional structure is stable.

Quaternary Structure: In a relatively small number of proteins, several separate polypeptide chains bind together to form one overall protein molecule. The binding of multiple polypeptides together, when it occurs, is called the **quaternary structure** of the protein. Proteins that shown quaternary structure often are very important in the regulation of processes or are themselves regulated. In hemoglobin, for instance, the addition of O₂ to one of the four polypeptides makes it easier to add O₂ to the remaining three.

? A molecule of hemoglobin consists of four polypeptides that are held to each other by hydrogen bonds. Each individual peptide is about 200 amino acids long and contains regions of α - helix and β -pleated sheet. Moreover, each of these peptides folds in a complex manner around a group called a heme that contains iron. It is to this iron that O₂ can bind. What levels of structure are shown in hemoglobin?

Myoglobin, on the other hand is an O₂ carrying protein found in muscle cells. For all intents, each myoglobin molecule is composed of what is essentially one of the polypeptide molecules of hemoglobin. What types of structure are shown by myoglobin? (read previous question before answering)

Looking Ahead: You should have the idea that protein structure is far from rigid. We will see in the next class that changes in the exact structure of proteins,

called **allosteric changes** are an important part of their function. In later classes, we will consider the relationship between a protein's structure, the genetic information that codes for the primary structure and the environment. This is one of the most exciting emerging topics in molecular biology and one that will engage the work of many biochemists over the next few decades.

QUESTIONS

1. Describe the four aspects of protein structure and explain the importance of primary and tertiary structures in terms of the correlation of structure and function.
2. Compare and contrast (1) proteins and nucleic acids and (2) DNA and RNA in terms of structure.
3. What is a dehydration reaction? Be able to draw out the formation of a dipeptide by the reaction of two amino acids. Know the general structure of an amino acid.
4. Learn the three main categories into which the 20 or so naturally occurring amino acids can be placed. If you see a structure, you should be able to decide which category the amino acid belongs in. You should not memorize the structures of any amino acids -- simply apply the rules we learned (reviewed) about polar covalent bonds earlier in the course.