BIOMOLECULES. I. (up to proteins)

Basic Molecules of Terrestrial Self-Replication (brief version)

Theory for origin of life by chemical evolution must explain following:

nuclei--->atoms--->molecules--->monomers--->polymers

It's the last step that is the problem: How to make molecules this complex?

First we'll list the basic monomers and polymers, and then discuss and show their properties in more detail.

Monomers are comparatively simple, and include:

amino acids (building blocks of **proteins**)

and

sugars, phosphates, and bases (building blocks of **nucleic acids** DNA,RNA)

These are not too difficult to make in the lab, and maybe in space (see the table of molecules found in the Murchison meteorite), which is what led to over-optimism about SETI.

Polymers, made from monomers, are much more complex, and *their origin is the basic problem in understanding the origin of life*. They include:

carbohydrates (used for food and structural materials)

fats (store and transport energy)

lipids (e.g. cell membranes) – these have a crucial "amphiphilic/amphiphobic" property due to their structure and that of water.

These are important, especially for life today and probably for the first cells (which may have predated proteins *or* nucleic acids). But in terms of getting complex organisms, the crucial polymers are:

proteins--made by combining 20 (out of thousands!) *specific* amino acid monomers. Nearly all have a type of symmetry called "left-handed" (levo-).

Many functions: enzymes, structure, contraction, gene-regulation, messengers, defense, transport.

and

nucleic acids--these typically contain anywhere from 10^5 -- 10^{10} atoms, so *very* long. Made up of 3 types of monomers:

a. sugars--nearly all the ones used in life have "right-handed" symmetry. (No one knows why.)

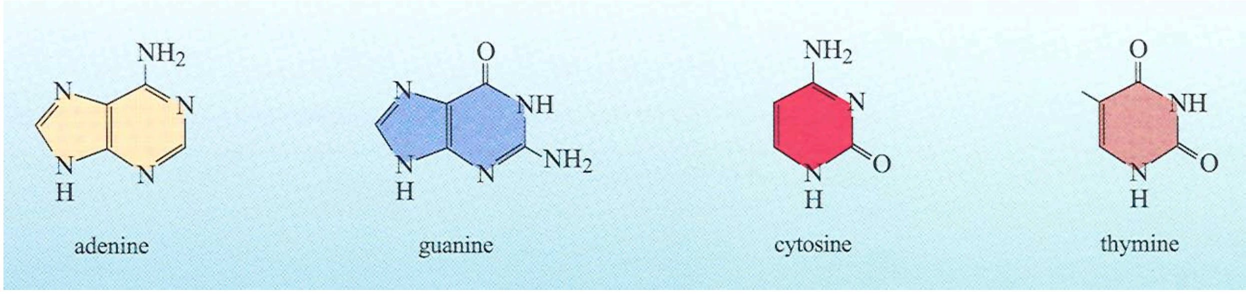
b. phosphates

Together, sugars and phosphates make up the sides of the "ladder" of nucleic acids: S--P--S--P

c. bases--these make up the "rungs" of the "ladder", and carry the code for reproduction. There are 4 types in DNA:

adenine (A), cytosine (C), guanine (G), and thymine (T) [replaced by uracil (U) in RNA] [You DON'T have to memorize the names. But *do* remember that these bases are used very much like letters that make words and larger structures of meaning.

Here is a nice picture of the 4 bases of DNA. Note that they come in two types, purines (single ring) and pyrimidines (double rings). Also notice that they all contain nitrogen (N)—they are called "nitrogenous bases." A fundamental question is: Why, out of all the possible bases that could be made, are these four used in living things (related to why there are 20 amino acids used.



Next we will look at these biomolecules in more detail, and how they function.

Biomolecules and their Functions

The four main classes of biomolecules

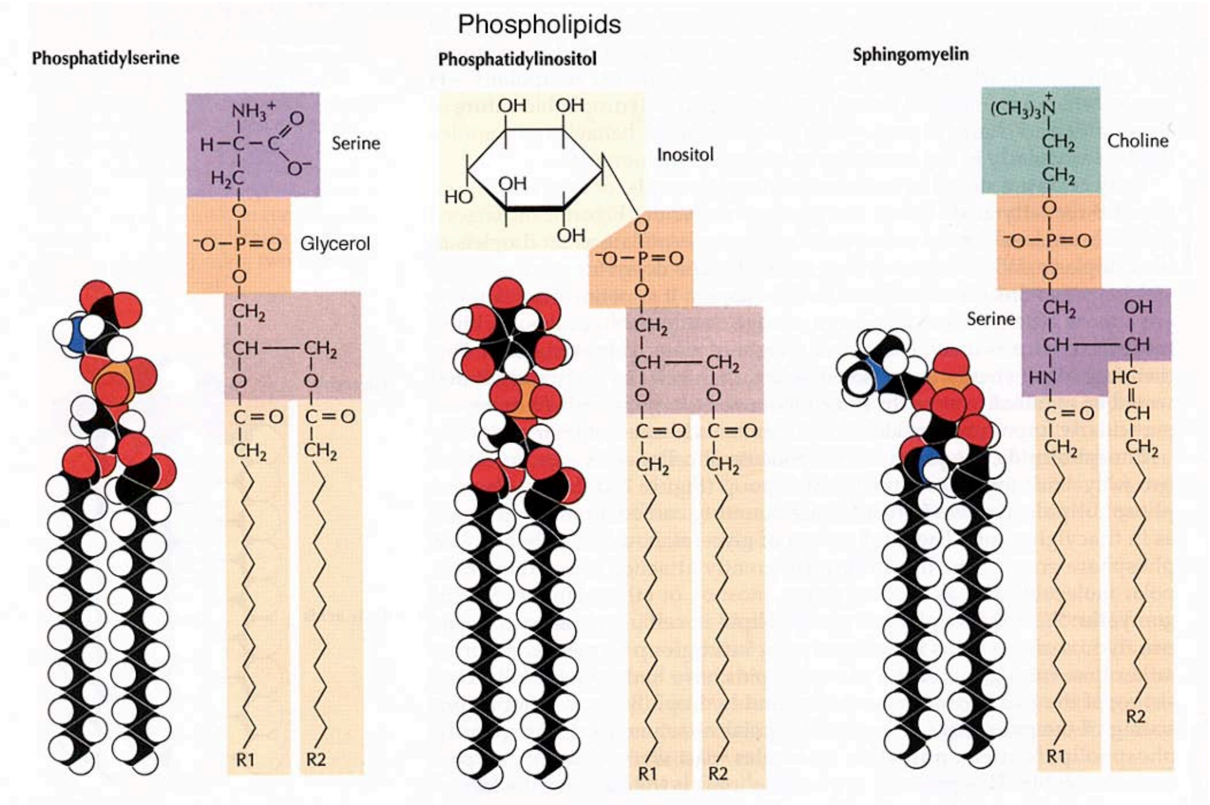
Lipids – nonpolar, do not dissolve in water (because of high proportion of C-H bonds). These include (I'll omit comment for some of them):

Fatty acids – simplest. Long hydrocarbon chain with carboxyl group at one end. Important for storing and transferring energy.

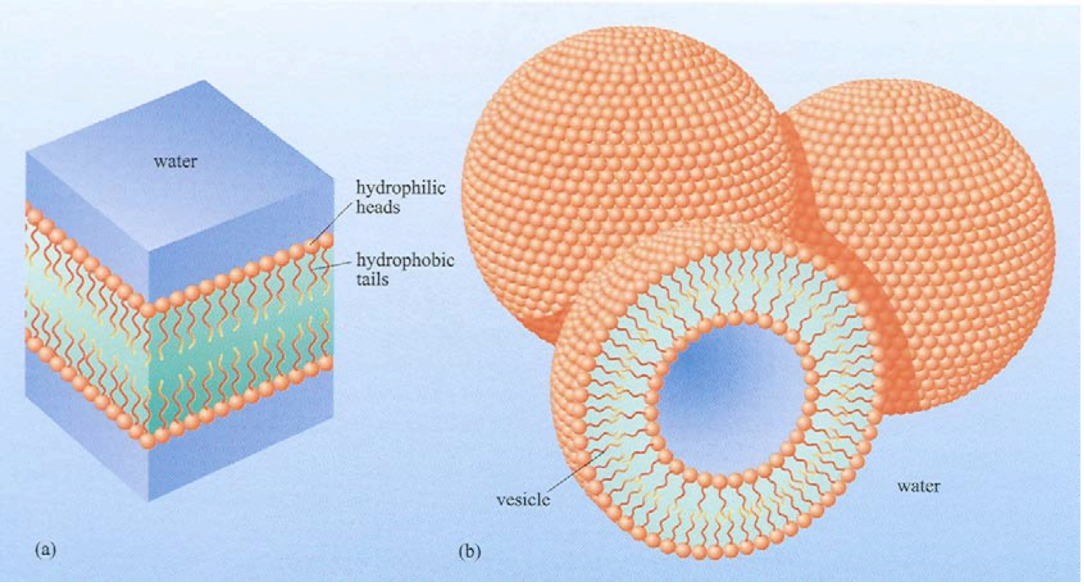
Waxes – fatty acids bonded to long-chain alcohol; e.g. waterproof covering on leaves and stems of plants

Triglycerides – 3 fatty acids bonded to glycerol; e.g. oil, fat. Used for energy storage in animals, some plants.

<u>Phospholipids</u> – principle component of cell membrane. They are <u>amphipathic</u>: they have **hydrophobic** tails (2 hydrocarbon chains) and **hydrophilic** "head groups" (phosphate + polar attachments). As a lipid <u>bilayer</u> (two layers of these lipids "back to back") they are able to perform amazing functions involving transport in and out of cells.



Here is an illustration of a lipid bilayer and the type of “cell-like” structure that it becomes in liquid water:

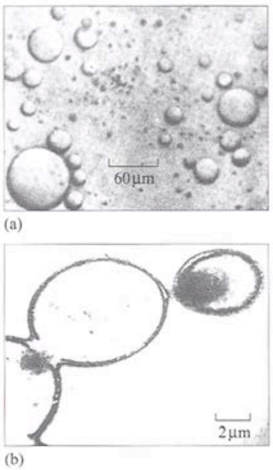


Steroids – hormone messengers consisting of 4 hydrocarbon rings + distinct functional groups.

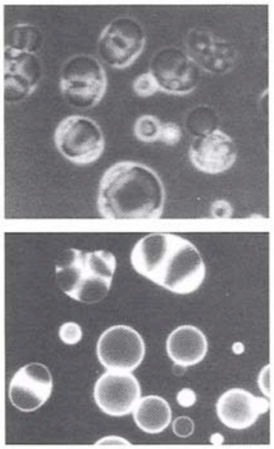
We shouldn’t underestimate the importance of lipids, since some people believe that life began with cell membranes and little else but random reactions going on inside. You could get reactions to produce other molecules going a lot faster if they can be contained in a compartment like a membrane, especially one that will let in (and out) some things but not others. Others think that RNA or DNA came before cell membranes, but there is no doubt that, whenever they occurred, they greatly increased the efficiency of life processes.

An optimistic fact is that lipid membrane structures are not very difficult to make in the laboratory--you basically have to drop a lot of hydrocarbons (seen in the atmospheres of, e.g. Titan) into water, and voila, “vesicles”! (Vesicles are small fatty bubbles.)

A few examples of how easy it is to form compartments are shown below.

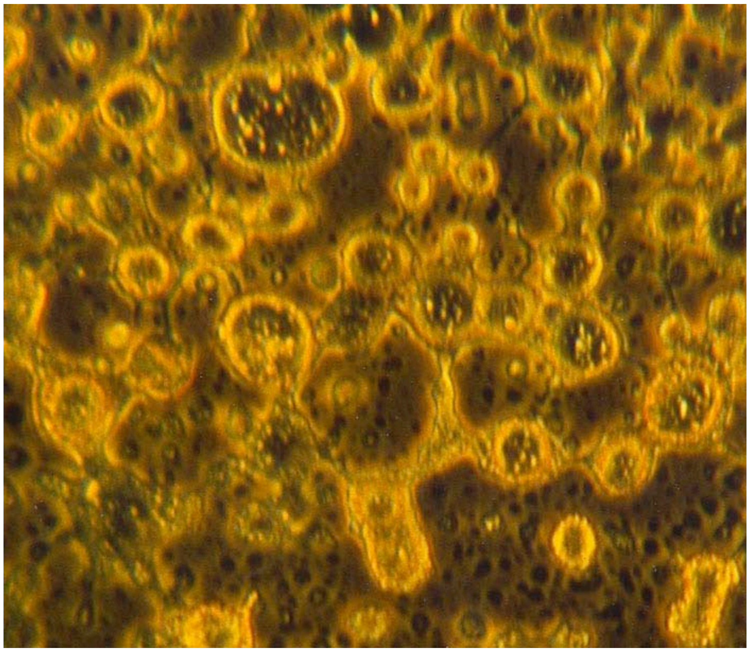


(a) Coacervates and (b) proteinoid microspheres.



Bilayers generated from the Murchison meteorite

Here is a picture, from a recent experiment, of vesicles (droplets) produced when organic residues from ice exposed to UV radiation (as would happen in space) were dropped in some water—they readily formed structures similar to cell membrances.



Carbohydrates -- sugars, their polymers, and related compounds

Monosaccharides – simple sugars, multiples of CH_2O. Classified by how many carbons, so 3 for triose, 4 for tetrose, 5 for pentose, 6 for hexose. So a pentose would have the formula $C_5H_{10}O_5$.

Disaccharides – two joined monosaccharides, e.g. sucrose, lactose.

Polysaccharides – polymers composed of 100s to 1000s of monosaccharides. The 3 most important are glycogen (animals store energy using this), starch (plants store energy with this), and cellulose (plant structural material, also the basis for plant photosynthesis).

We don't want to undermine the importance of carbohydrates, since they play such a crucial role in many functions such as energy production and cycling, but these functions probably arose later, after the origin of the first “living” molecules/organisms. Still, notice that sugars, especially the 5- carbon kind called “pentose” sugars, and particularly the pentose sugar **ribose**, plays a crucial role in making RNA and DNA.

Most people think that only the monosaccharides, in particular the pentose sugar ribose (and oxyrybose and dioxyribose), play a crucial role as part of the “nucleotides” (see below) that chain together to form the nucleic acids, the information component of all living organisms today (and maybe more functional than just information when life began—see “RNA world” below). Take a look at (but don’t worry about memorizing) the structure of some sugars shown below—notice for those with 5 or more carbons, they are basically rings of C with H and OH attached and a “sidegroup”:

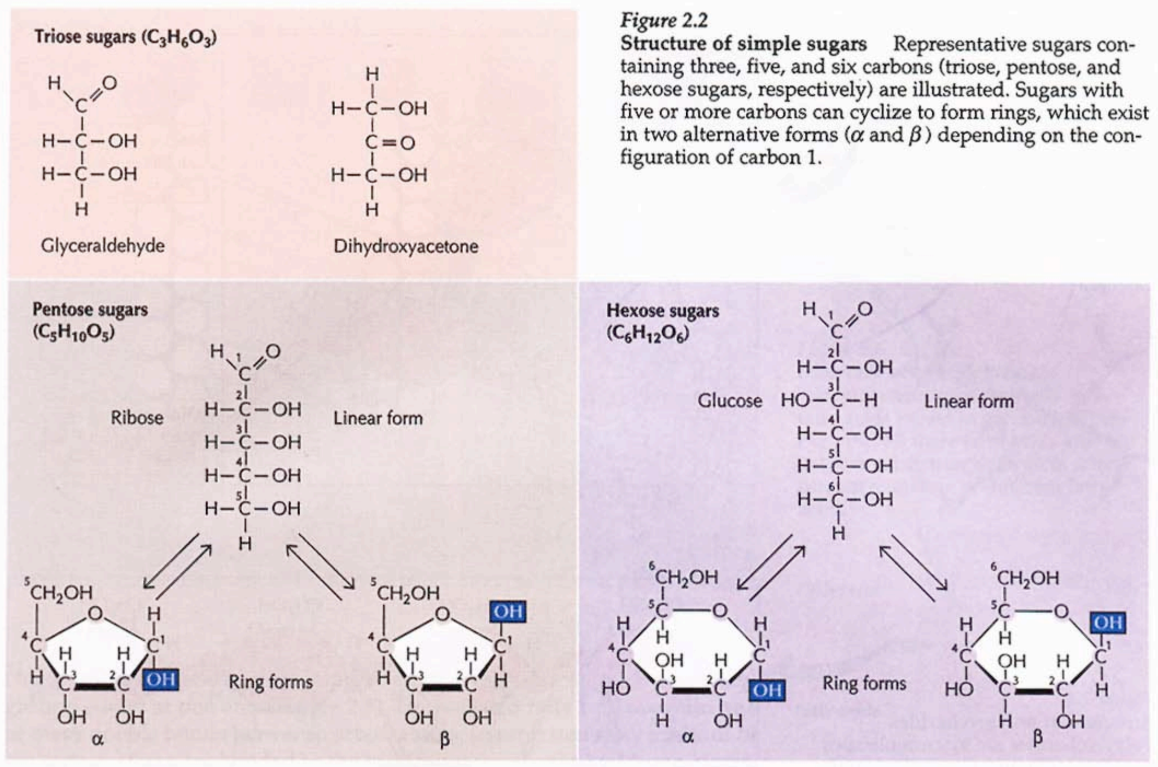


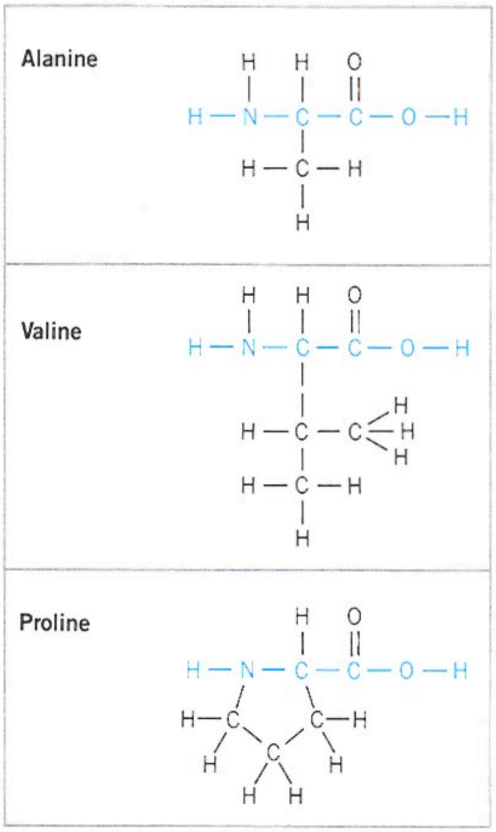
Figure 2.2
Structure of simple sugars Representative sugars containing three, five, and six carbons (triose, pentose, and hexose sugars, respectively) are illustrated. Sugars with five or more carbons can cyclize to form rings, which exist in two alternative forms (*α* and *β*) depending on the configuration of carbon 1.

Carbohydrates are not too difficult to make in the laboratory, in general, but simple ribose has proven especially problematic in the sense that it doesn't seem to want to bind with the rest of the nucleotide (base + phosphate) in the correct manner, at least in the laboratory. This is considered a serious problem for origin of life models.

You should note that almost without exception the sugars that are used biologically are exclusively right-handed in symmetry (they are "chiral" molecules, meaning a preference for one kind of symmetry over another—explained in the pictures shown earlier in the notes). No one knows why, but most people think they are important.

Proteins -- these are complex-shaped polymers made from the 20 different *amino acids* used by terrestrial organisms. Note: *there are many hundreds of amino acids that occur in nature, but for some reason living things use only a particular 20* (actually it is now known that a few organisms use a 21st and 22nd amino acid, but let's ignore this here); pictures shown in class.). Note that out of 20 amino acids, an enormous number (20 factorial=20x19x18x...x2) of proteins could be synthesized.

An amino acid = C atom ("α carbon") + carboxyl group COO^- + amino group (NH3 +) + H atom + distinctive side group. Here are a few examples.



The chemical structures of alanine, valine, and proline, three common amino acids found in life on the earth. Note that hydrogen, oxygen, carbon, and nitrogen—four of the most common elements in the cosmos—form the chemical basis of these compounds.

A very important property of *biological* amino acids is that they are all "left-handed." This is illustrated in the pictures below.

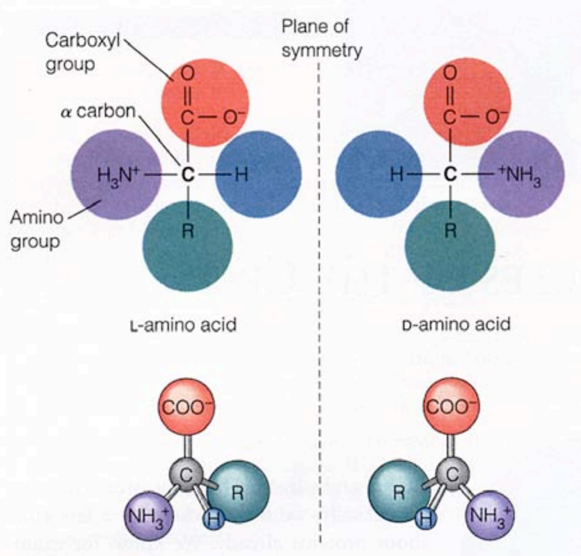
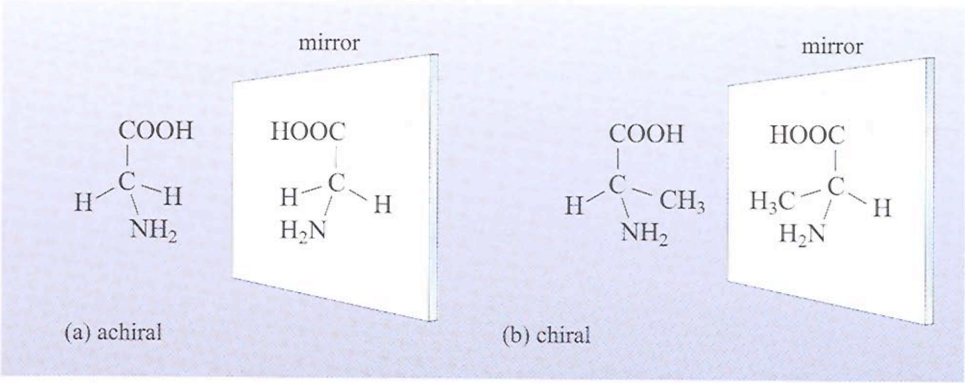
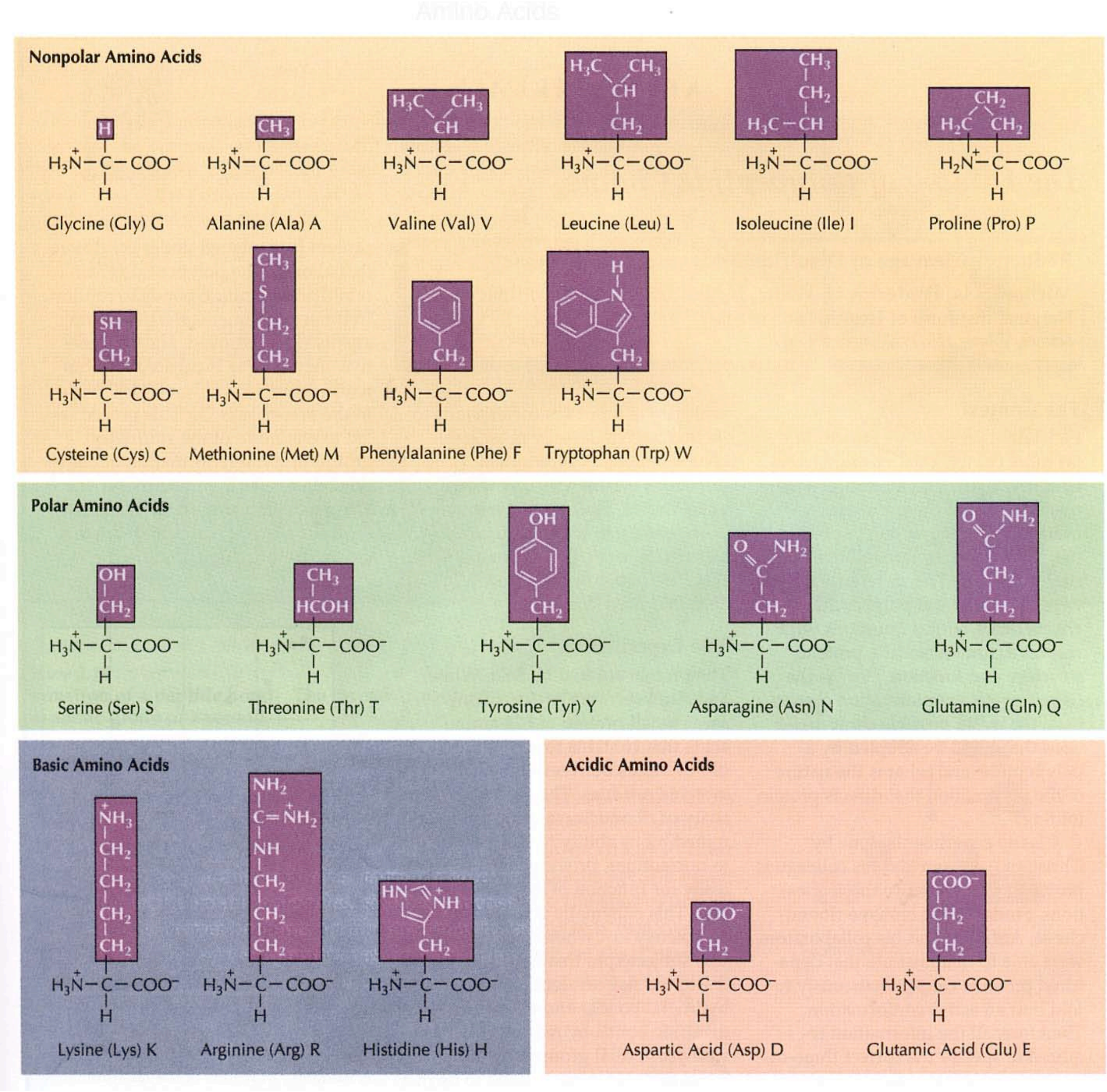


Figure 3-1 The Structure and Stereochemistry of an Amino Acid. Because the α carbon atom is asymmetric in all amino acids except glycine, most amino acids can exist in two isomeric forms, designated L and D and shown here as (top) conventional structural formulas and (bottom) ball-and-stick models. The L and D forms are stereoisomers, with the vertical dashed line as the plane of symmetry. Of the two forms, only L-amino acids are present in proteins.



(a) The achiral amino acid glycine. Its mirror image can be superimposed on the original. (b) The chiral amino acid alanine. The mirror image cannot be superimposed on the original.

The illustration below shows the 20 amino acids.



So amino acids are already pretty complex organic molecules, but they are only the *monomers*, or basic units of the proteins, which are extremely long chains of amino acids that fold into extremely complex and precise three-dimensional structures that control their functions.

Amino acids are joined by “peptide bonds” between an amino group of one amino acid and a carboxyl group of another ⇒ **polypeptides.** So a protein is a very complex polypetide chain. Look at the peptide bond shown below. Notice that a water molecule is *released* during this chaining together, or *polymerization*, of amino acids. This is symptomatic of a big problem: Polymerization of biomolecules generally *releases* water; the *presence* of water would break the bonds apart! So how could life begin in water?

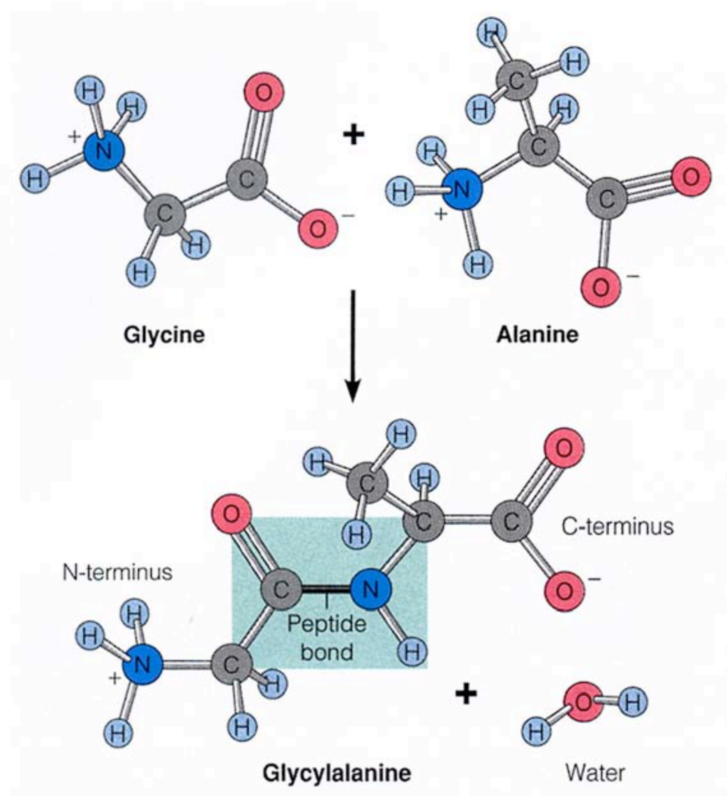
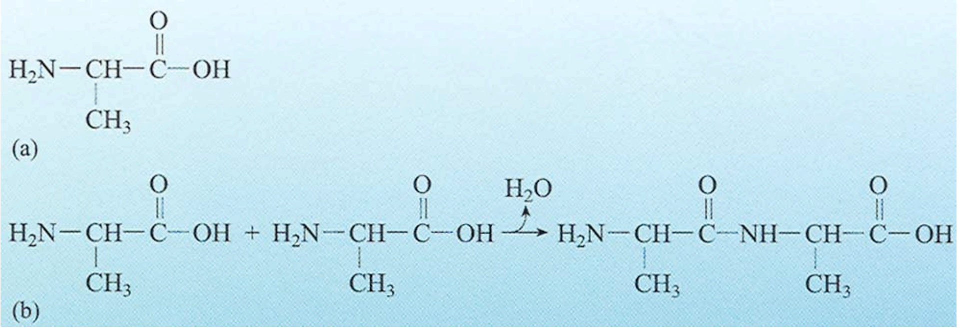


Figure 3-3 Peptide Bond Formation. Successive amino acids in a polypeptide are linked to one another by peptide bonds between the carboxyl group of one amino acid and the amino group of the next. Shown here is the formation of a peptide bond between the amino acids glycine and alanine.

Here is another illustration of amino acids linking up:



$$\mathrm{H_2N-\underset{CH_3}{\underset{|}{CH}}-\overset{O}{\overset{\|}{C}}-OH}$$

$$\mathrm{H_2N-\underset{CH_3}{\underset{|}{CH}}-\overset{O}{\overset{\|}{C}}-OH + H_2N-\underset{CH_3}{\underset{|}{CH}}-\overset{O}{\overset{\|}{C}}-OH \xrightarrow{H_2O} H_2N-\underset{CH_3}{\underset{|}{CH}}-\overset{O}{\overset{\|}{C}}-NH-\underset{CH_3}{\underset{|}{CH}}-\overset{O}{\overset{\|}{C}}-OH}$$

Proteins are *really* amazing molecules—by twisting and folding their length into complicated three-dimensional shapes (their “conformation”), they transform the purely informational content of nucleic acids into mechanically functional shapes: every turn and fold of a protein molecule is a geometrical key that carries out a huge number of functions.

Look at this illustration of the structure of a protein, to appreciate the complexity of these molecules.

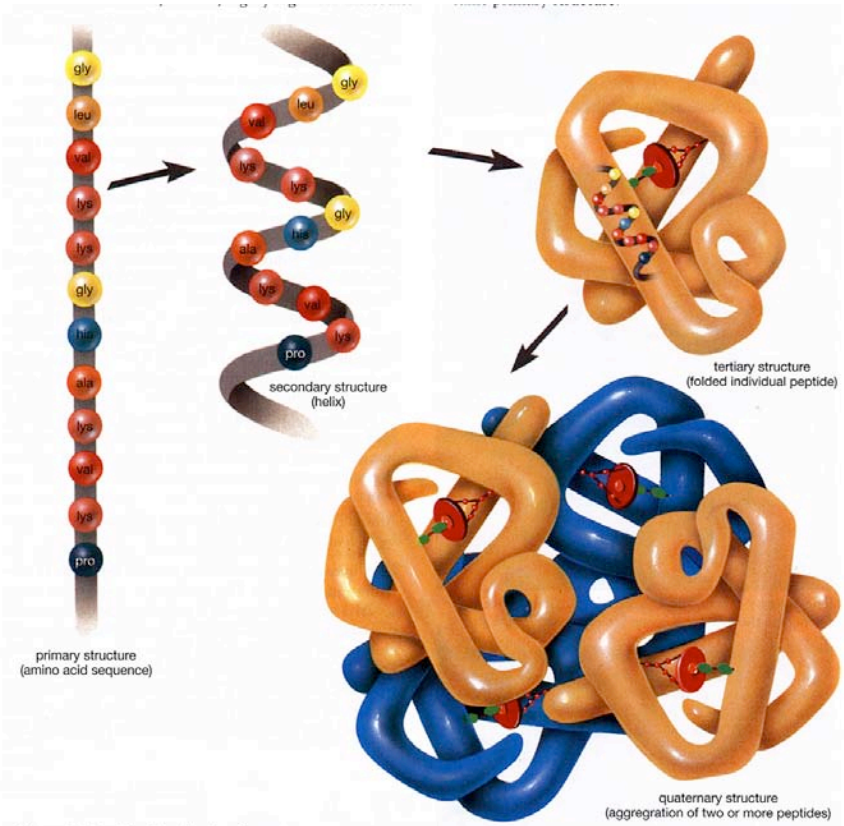


Figure 3-14 The four levels of protein structure

Levels of protein structure are represented here by hemoglobin, the oxygen-carrying protein in red blood cells. All levels of protein structure are determined by the amino acid sequence of the protein, interactions among the R groups of the amino acids (primarily hydrogen bonds and disulfide bridges between cysteines), and interactions between the R groups and their surroundings (usually water or lipids).

So proteins, whose structure is encoded in nucleic acid base sequences (below), are able to transform “talk” into “action”

Information (DNA) ➙ conformation (3D shape of protein) ➙ functionality (replication, damage repair, … endless list)

Later we’ll see than RNA might be able to carry out all three of these processes on its own, which is why it is a favorite for the first “living” thing.

The point here is that proteins carry out so many diverse functions for all organisms, and in particular for the functioning of the genetic code, that it is difficult to see how you could have any form of life without them. A central question in astrobiology became: What came first, the genetic code (or some more primitive replication system; “gene-first” theories), or the proteins (“protein-first” theories)? How could a genetic code work without proteins? How could you have proteins that weren't coded for by genes? So there is a big chicken-and-the-egg problem here. We'll see (and you'll read) that the conundrum was partially broken by an amazing discovery involving “self-catalytic” RNA, leading to a model for early life called the “RNA World” scenario, but we'll get to that shortly