

Last updated: September 14, 2010 02:43 AM with updated fig refs to Purves 7th edition
© Copyright 2010 Lawrence Chasin and Deborah Mowshowitz Department of Biological Sciences Columbia University New York, NY

Bio C2005/F2401x 2009 Lec.3 L. Chasin September 14, 2010

Props: tinker toys. molecular models for amino acids,
See rec. section assignments on website.
Recitations start this week. Quizzes start next week for C2005 students.

Anyone who needs to change or still sign up for a recitation section, email Jessie at jrk2159@columbia.edu

POLYSACCHARIDES.

Glucose ring formation
Starch and cellulose: function follows form
LIPIDS: Soluble in organic solvents
Steroids
Fats: Fatty acids and glycerol as monomers
Esters; fats vs. oils; saturated vs. unsaturated fats
Phospholipids, phosphoesters, phosphatidylcholine
Phospholipid bilayer; membranes
Proteins: monomers = amino acids
pK
Stereoisomers: L and D amino acids
Polypeptides
Peptide bonds
Primary structure

(Ring formation)

In 3-dimensional space, a hexose chain can easily curl up, such that the oxygen attached to carbon 5 can be juxtaposed next to carbon 1. A 6-membered **ring** forms preferentially in water, by attack of the hydroxyl of carbon-5 (C5) on the carbonyl double bond at C1. One bond of the carbonyl double bond opens up and forms a new bond between carbon-1 with the O of C5. The H leaves C5 and a new OH group is formed on carbon 1. Follow along with the diagram on the [glucose handout](#). So a 6-membered ring is formed, with O as one of its members (one of the vertices). One carbon (C6) is left sticking out away from the ring. Unlike most biochemical reactions, which require a catalyst to help them take place at a reasonable rate (more on this in a week or so), this intramolecular cyclization reaction takes place all on its own, as soon as a sugar is put into a water (aqueous) solution. This reaction is rapid because the players can't help but keep bumping into each other as the glucose chain dances in thermal motion. The ring structure can also open up, re-forming the straight chain. The 2 forms are in a dynamic equilibrium, but because the ring form is more stable, this species predominates in water.

{Q&A}

(Anomers: alpha and beta glucose)

Now, when the O attached to C5 approaches the carbon C1 which has the carbonyl double bond, it can do so from one side or from the other side. Depending on which side is attacked, the resulting ring comes out looking different in 3-dimensional-space, because the OH formed from the carbonyl oxygen is oriented distinctively in the 2 cases. That is, the resulting ring can be of [two different conformations](#) in space. The two conformations are formed at roughly equal frequencies. The 2 conformations are called alpha and beta:

Alpha, where the C1 OH that is formed ends up BELOW* the C1 hydrogen,
or

Beta, where the C1 OH that is formed ends up ABOVE* the C1 hydrogen (see [glucose handout](#), right side. See also a picture [\[Purves 3.11\]](#)).

This relative position information can be seen in the Haworth depictions in which the glucose ring is depicted as flat (right

sides of handout 2-7 and 2-8).

* "Above" and "below" refer only to the Haworth projection depiction and not real three-dimensional space, see blow.

[{Q&A}](#) [{Q&A}](#)

(Chair form)

The ring is actually not flat, but puckered into a reclining chair-like shape, but hard to draw: (see [flat vs. puckered](#) handout 2-8) -- in this chair-view the hydrogens and the hydroxyls can be seen to be not really up or down, but are rather either **axial** (vertical = sticking up OR down) or **equatorial** (horizontal = sticking out).

Note in glucose all the hydroxyls are equatorial except that of the #1 carbon in the alpha conformation. In beta-glucose this -OH is upper, relative to the hydrogen, and is in fact equatorial; but in alpha-glucose it is lower (relative to the hydrogen) and axial (and down).

You should study handouts 2-7 and 2-8 until you are satisfied that you can picture these chair conformations in 3-dimensional space.

Animation at <http://wetcbe.cmbi.ru.nl//organic/gluc/glucjmol.html> may or may not help.

The existence of these two seemingly very similar 3-dimensional structures for glucose can have important effects on the 3-dimensional structure of polysaccharides made from these glucose monomers, which in turn can determine the function of the polysaccharide, as we will see.

Glycosidic bonds)

As we consider a polymer built from glucose monomers, we can first consider a dimer. Two glucose monomers can be connected to form a DIMER. This connection, WHICH DOES NOT HAPPEN BY ITSELF (i.e., without some help from a catalyst), involves a **dehydration**, the removal of one molecule of water, from the 2 monomers:

2 monomers -----> dimer

R-OH + R-OH -----> R-O-R + HOH

This type of reaction is also referred to as a CONDENSATION, as it condenses two molecules into one.

The resulting -C-O-C- bond is called a **glycosidic** bond when it is connecting two sugars.

Conversely, the breakdown of polymers back to their constituent monomers involves the reversal of this chemistry, the addition of water, or **hydrolysis** (the products = a hydrolysate).

R-O-R + HOH -----> R-OH + R-OH

[{Q&A}](#) [{Q&A}](#)

Both of these reactions require different catalysts in the cell in order to occur, which is generally true for all the biochemical reactions we will discuss. See the [carbohydrates handout](#) (2-6), below the line, for a depiction of two dimers in the flat ring forms. Note the 1-4 linkage (C6 sticks out of the ring, so that is one way to figure out the numbering in the ring). Although the bonds are presented as bent at sharp angles, they are not really so, it is just a way of presenting both sugar monomers right side up and still connect them with a glycosidic bond.

These dehydrations can continue in many cases in a repeated way to form chains that contains 1000's of monomers. E.g.,:

X--1,4--X--1,4--X--1,4--X--.....(where X represents a sugar ring)

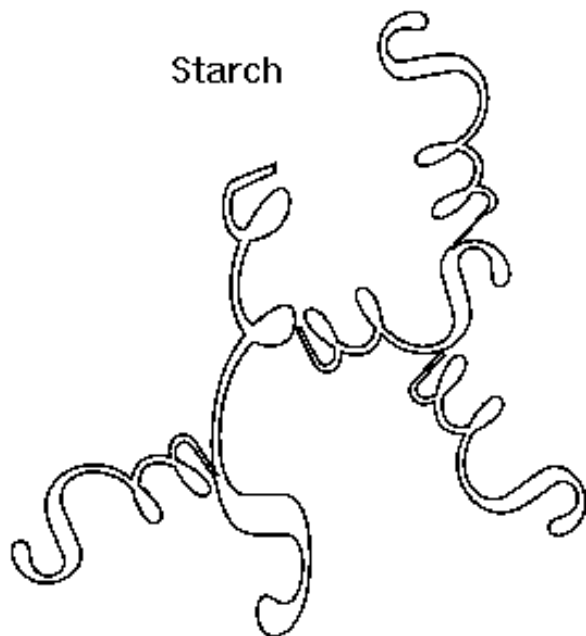
To be sure you understand disaccharides, try Problem 1-8C and 1-9 D & E.

(Starch and cellulose: function follows form)

If glucose molecules are put together with an alpha 1,4 link, then the 2 sugars in the dimer lie at an angle relative to each other. Here, the

C1 -OH is axial whereas the C4 -OH in glucose is (always is) equatorial. The disaccharide in this case is called maltose. The angle of this alpha 1,4 bond is such that the polymer bends at each glycosidic connecting bond, as can be seen in [handout 2-9, disaccharides in chair form](#). As a result, it takes on a helical shape that allows lots of hydrogen bonding between glucoses in each turn of the helix, thus stabilizing the polymer in this shape. [TINKER TOY demo]. Such is the case with **STARCH**, which consists of **alpha**-glucose molecules joined in 1,4 linkages. In addition, starch has branches [[Purves 6:3.14b, 7:3.16](#)] made by linking additional glucose molecules at the C6 OH of some of the glucose residues in the chain, via an alpha 1,6 bond). The branched chain continues with alpha 1,4 linkages (see Becker 7th: Fig. 3-24). The length and frequency of the side chains give rise to the different forms of starch (potatoes, corn) or of a starch-like polymer found in mammalian liver and muscles, **GLYCOGEN** (and see [[Purves 6:3.14a2, 7: 3.16](#)]). These polymers act as storage forms for glucose. When glucose is needed, they can be hydrolyzed (adding water back to the bond between the monomers) to regenerate the free monomer. Glycogen is more highly branched than starch, and its breakdown from the many ends so produced leads to rapid mobilization of the glucose moieties within it, a property more important in animals than plants. [{Q&A}](#).

When a catalyst (a different one) puts together glucose molecule that are in the beta conformation, polymer of a different shape results. Here the disaccharide is called cellobiose. A poly-glucose of this type is **CELLULOSE**, which contains exclusively glucose molecules in beta linkages. The beta linkage is a pretty straight connection between the C1 and C4 of adjoining carbon atoms, since they both are equatorial and so are sticking out, as can be seen on [handout 2-9, disaccharides in chair form](#). Thus a cellulose chain extends straight with its C6 OHs sticking out from the chain on either side. [[Purves 3.14a1](#)] Many cellulose molecules can then associate side by side (via hydrogen-bonds to each other) to form a fiber of great strength (e.g., in cotton, and it also contributes to rigidity of wood) [[Purves 6:3.14b, 7:3.16](#)] [TINKER TOY demo]. Although the hydrogen bonds that hold the cellulose polymers together in a bundle are individually weak, because there are hundreds or thousands of them, the bundle stays together. At any moment some will be broken but others will not, so there is strength in numbers here. Cellulose is the most abundant carbon compound in the biosphere, accounting for about half of all such carbon.



Here is our first good example of an important theme in biochemistry, the relationship between structure and function at the molecular level. The straight linear structure of cellulose made possible by the beta-linkages allows the assembly of thousands of aligned molecules to produce a cellulose fiber of great tensile strength. The alpha-linkage in starch produces a compact structure, not strong, which serves as a storehouse of glucose for energy when needed.

Your texts have additional examples of important polysaccharides. Some of the sugars have nitrogen-containing groups appended to the basic carbohydrate ring. The rigid bacterial cell wall is another example, like cellulose, of a polysaccharide used for structural support. So is the shell, or exoskeleton, of insects ([CHITIN](#)) [[Purves 6:3.15c, 7: 3.17](#)].

To go over the structure of polysaccharides, try problem 1-11. If you need more review, try 1-25.

There are several different hexoses in most cells. Fructose, galactose, and mannose are some common ones. Differences lie in the positions of the carbonyl along the chain and relative positions of the hydroxyls in space. Fructose has a ketone carbonyl at C2 which is its anomeric carbon; it cyclizes to form a 5-membered ring (still with one member oxygen, of course, so 2 CH₂OH groups stick out from the ring ([Carbohydrates handout](#)).

And there are several common disaccharides (as distinct from polysaccharides) found in nature (see Becker 7th Fig. 3-23):

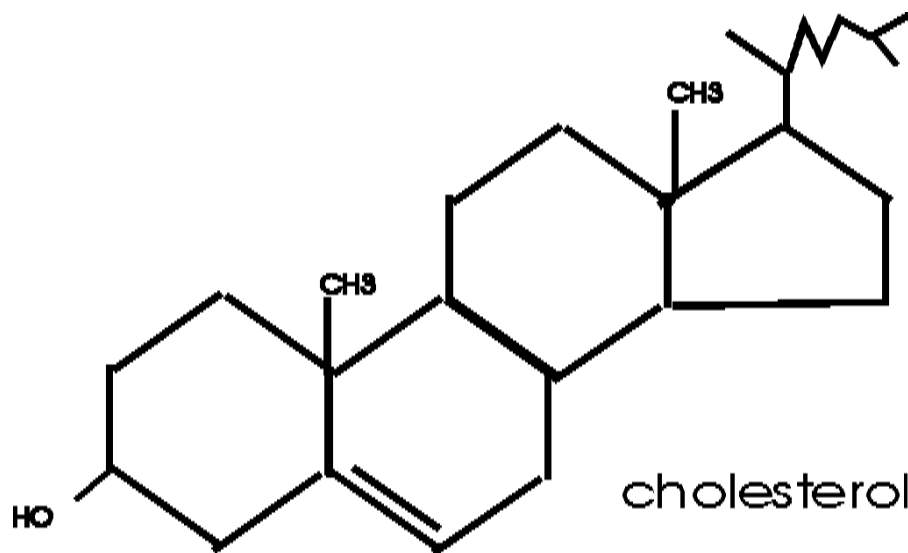
Glucose-glucose via a 1-4 alpha-link is **maltose**, where alpha refers to the state of the -OH in the monomer joined at its C1 carbon [[Purves 3.13a](#)]. Maltose is formed as you digest bread.

[{Q&A}](#)

Galactose + glucose [[Purves 6:3.12, 7:3.14](#)] via a 1-4 beta-link is **lactose** (in milk) ,

Glucose + fructose [[Purves 6:3.12, 7: 3.14](#)] via a 1-2 alpha-beta link = **sucrose** (table sugar).

(LIPIDS: Soluble in organic solvents)



LIPIDS: This is a more heterogeneous group, being defined as substances in a cell that are extractable in organic solvents. Non-polar compounds are not soluble in water, as they tend to coalesce. But they ARE soluble in non-polar solvents such as octane or benzene (a **hydrocarbon** - compounds made up of just hydrogen and carbon atoms, like the octane molecule we considered earlier). So lipids are molecules that have extensive non-polar regions.

(Steroids)

Steroids such as cholesterol and testosterone have multiple hydrocarbon rings, and are in this category [[Purves 6:3.24, 7: 3.23](#)],

(see Becker 7th: p.70). Note the drawing conventions, with further shorthand: the depiction (not) of C's and H's. C is assumed to always have 4 bonds; unless otherwise specified, C's are assumed to be present at the vertices on drawn bonds; bonds from such C's that are not shown are assumed to be to H, to make a total of 4. Almost no atoms are named, yet the structure is completely defined.

Steroids are small molecules that are **not** monomers, they do not become connected to form polymers. Cholesterol is a steroid that is a component of the cell membrane, which we discuss in a few minutes. Steroids such as testosterone, estrogen, cortisone, and vitamin D are hormones, compounds that circulate in the blood to send signals from cells in one part of the body into cells in another region. You will learn more about steroids in the physiology section in the second semester.

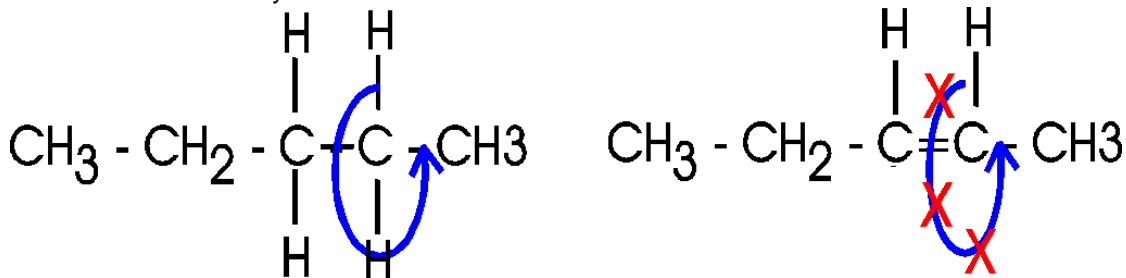
(Fats: Fatty acids and glycerol as monomers)

A major class of lipids are the fatty acids, long straight chain hydrocarbons with a carboxyl group (carboxylic acid) on one end. See [[Purves 6:3.19, 7:3.18](#)], and another [picture](#). [{Q&A}](#).



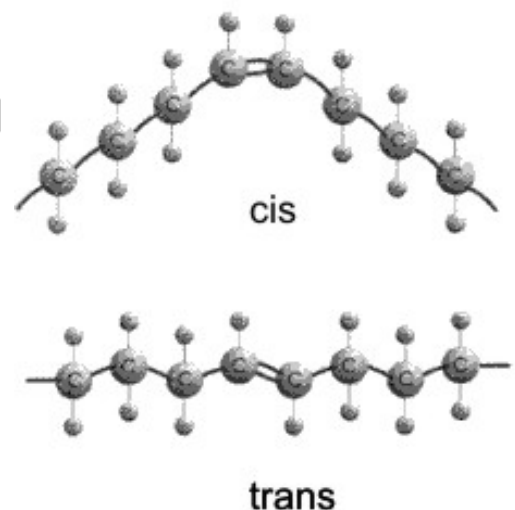
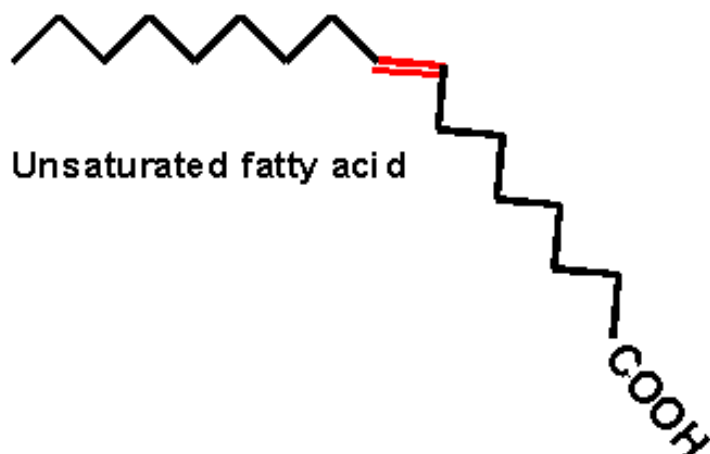
(Esters; fats vs. oils; saturated vs. unsaturated fats)

Inside cells, fatty acids (FA) are usually connected to a molecule of the tri-hydroxy (tri-alcohol) compound glycerol. Once again water is removed, this time producing an **ester** bond (acid + alcohol, see top right corner of [lipids handout 2-10](#)). If all 3 OH's on the glycerol are substituted with FA's, then we have a **triglyceride**. See [\[Purves 6:3.19, 7: 3.18\]](#), and another [picture](#). This is **fat**. Fats are very hydrophobic and are practically insoluble in water. You can also have mono- or di-substituted glycerol, but it is the triglyceride that is fat. Fats differ according to the exact nature of the FA's that are present. "Saturated" fats have $-CH_2-$ (methylene) groups, usually 18-20 of them, along the chain. They are **saturated** with hydrogens, compared to the **unsaturated** variety. The latter may have a double bond or two within the chain, and thus have less H's (**unsaturated** for H's). The presence of the double bond puts a crimp into the structure, since unlike single C-C bonds, there is no rotation about C=C double bonds) [\[Purves 6:3.20, 7:3.19\]](#). As a result, it is more difficult for the unsaturated fatty acid molecules to associate.



Actually, there are 2 ways a double bond can form in a fatty acid, called **cis** and **trans**. In the **cis** case, the two hydrogens are on the same side of the double bond (remember there is no rotation, so their position is fixed). Now the two bonds carrying the rest of the carbon atoms are also together on one side of the double bond, so the molecule is crimped, with a severe angle between the two hydrocarbon stretches. In the **trans** case, the two hydrocarbon stretches are on opposite sides of the double bond, and the overall chain is straighter. Most fatty acids in animals are saturated; with their relatively straight chains their hydrocarbon chains are free to associate with each other with no constraints and they aggregate into solid fat. Plants contain a lot of unsaturated fatty acids of the **cis** type. These unsaturated fats (WITH the double bonds) are usually liquids (oils), as their crooked fatty acid chains cannot approach each other so easily. Take vegetable oil (unsaturated), and add hydrogen across the double bonds and you get Crisco, or the creamy texture in peanut butter (read the label: hydrogenated).

Saturated fatty acid



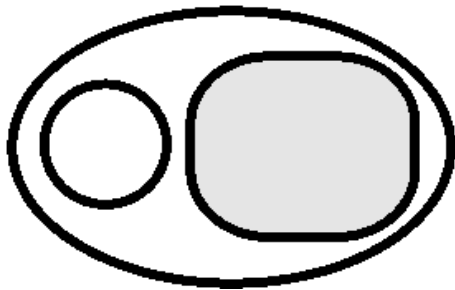
Trans fatty acids do occur made by bacteria in the stomachs of ruminants like cattle, so they end up in beef to some extent. A much greater source of **trans** unsaturated fatty acids comes from the chemical hydrogenation of oils, where they are formed somewhat ironically as a by-product of the hydrogenation process. The **trans** unsaturated fats resist

turning rancid, so are favored by the food industry. However, they are more equivalent to saturated fatty acids in their ability to form solid fat, which encourages the formation of atherosclerotic plaques. Thus margarine may be as bad for you as butter. Or worse, as trans fats show a stronger correlation with heart disease than saturated fat, possibly through more indirect effects (e.g., membrane structure).

So here again as in the case of polysaccharides, the 3-dimensional structure of the molecule has a lot to do with its physical properties.

Fats are a good example of hydrophobic forces at work. Just think of a fatty chicken soup with those globules of fat floating on top, out of solution.

Fats serve as a storage form of energy. That is, like glycogen or starch, fats can be broken down and used for energy metabolism, as we will see later. Fats are stored in cells called adipocytes.



(Phospholipids, phosphoesters, phosphatidylcholine)

There is a special class of lipids that are related to the fats, but with a significant difference. These are the **phospholipids**, an example of which is shown in the middle of the [LIPIDS handout](#). Two of the glycerol hydroxyls are connected to long chain fatty acids, but the third is connected to quite a different group, a phosphate. Phosphoric acid (H_3PO_4) is an acid, The -OH groups attached to the phosphorous easily lose hydrogens at neutral pH. Phosphoric acid has 3 acidic hydrogens. [If you are shaky on pH, ask to review it in recitation section.]

Phosphoric acid is a strong acid, losing most of its hydrogen ions at pH7.

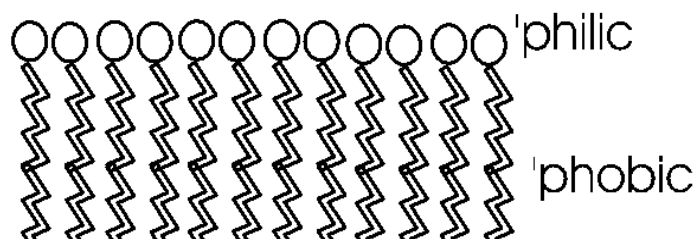
The ion that is formed is called phosphate, and we will treat the 2 names equivalently, considering them both acids (referring to their origin as the acid). Similarly we will use carboxylic acid and the carboxylate ion (the negatively charged unprotonated form) synonymously in most situations.

The phosphate group is connected to a glycerol hydroxyl, again by a dehydration that forms an ester (acid + alcohol). Whereas up until now we had a carboxylic acid ester linking the fatty acid to the glycerol, here we have a phospho-ester. The acid partner is the one named to specify a type of ester. In both cases the ester is formed by an alcohol linked to an acid. After linkage, the phosphate group is still charged, as shown. The rest of the phosphate may be free, as in a **phosphatidic acid**, or it may be esterified to yet another alcohol via another of its acidic groups; a common one is **ethanolamine**: $HO-CH_2-CH_2-NH_3^+$. The resulting phospholipid would be called phosphatidyl-ethanolamine, and it would be categorized as a phospho-di-ester (phosphodiester). Note that the presence of positively charged basic groups such as amines tends to neutralize the negative charge of the phosphate, but only adds to the hydrophilic character of the head of the phospholipid, by adding charged groups.

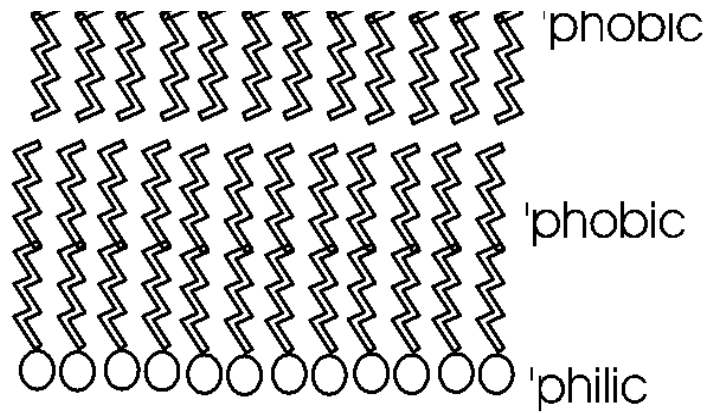
(Phospholipid bilayer: membranes)

Phosphatidyl-ethanolamine is a compound that is highly hydrophobic throughout most of the molecule, but then has a highly polar group at one end, with two complete, if opposite, charges. A further derivative has 3 methyl ($-CH_3$) groups

bonded to the nitrogen instead of H's. This moiety is choline (tri-methyl-ethanolamine; the nitrogen retains its positive charge. When esterified to a diglyceride one gets phosphatidyl choline, depicted in [\[Purves 6:3.21, 7:3.20\]](#). The polar end can interact strongly with water (it is hydrophilic), while the remainder of the molecule wants to come out of aqueous solution. This is a confused molecule. What happens is that the hydrophobic parts all line up with each other to minimize their interface with water (both side-to-side and end-to-end), while the charged ends remain in contact with water. See [\[Purves 6:3.22, 7:3.21\]](#), and [photo](#). It is in this way that biological **membranes** form, as a **phospholipid bilayer**, the charged ends of the double layer being on the outside in contact with water, with the cytoplasm on one side and the exterior of the cell on the other side: See [picture](#). And look in your textbooks for great diagrams of phospholipid bilayers



Such a bilayer presents a permeability barrier to water-soluble compounds, which cannot pass through the hydrophobic barrier. Special protein structures that are embedded in this membrane are then necessary to allow the passage of water soluble compounds in and out of the cell. These are the channels and pumps



mentioned earlier. See a diagram of a cell membrane at [\[Purves 6:5.1, 7:5.1\]](#) and in your texts. Yet again we see how the chemical properties of these molecules determine their structure, and how their structure provides a biological function. **To review phospholipid structure, try problems 1-12 & 1-13.**

Large amounts of cholesterol are embedded in the membranes of animal cells. The cholesterol is kept inside by hydrophobic forces. It acts to plug spaces that could cause leakiness, to impart more strength, and to prevent too much association of the saturated

fatty acids at low temperature (i.e., "freezing" of the membrane into fat).

The texts have nice diagrams of all this.

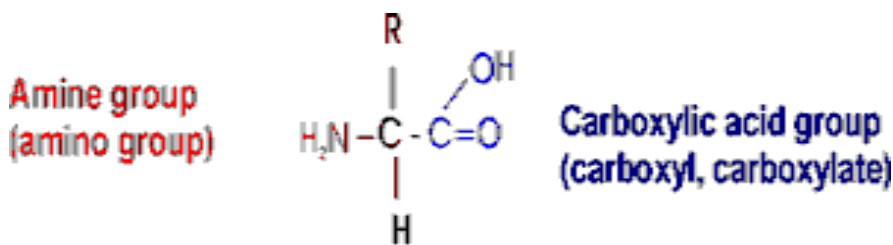
Lipids are impressive in their variety (see [picture](#)) and especially in membrane formation, but admittedly they are not really good examples of the linear biopolymers that we defined. But they have to go somewhere, and so they are stuck amongst the macromolecules.

NUCLEIC ACIDS: Unlike the catch-all category of LIPIDS, **NUCLEIC ACIDS** are biopolymers par excellence. There are 2 types, DNA and RNA, the monomers are nucleotides, that have nitrogen-containing rings, 5-carbon sugars, and phosphodiester linkages. There are four types of monomers in each polymer. We will discuss them in detail, but not for a few weeks yet.

(Proteins: Amino acids are the monomers (20))

PROTEINS. These are the most important class of macromolecules in the cell, and we will discuss them now in detail. The monomers that make up proteins are the amino acids, of which there are 20. The same 20 in E. coli, in elephants and in eggplants.

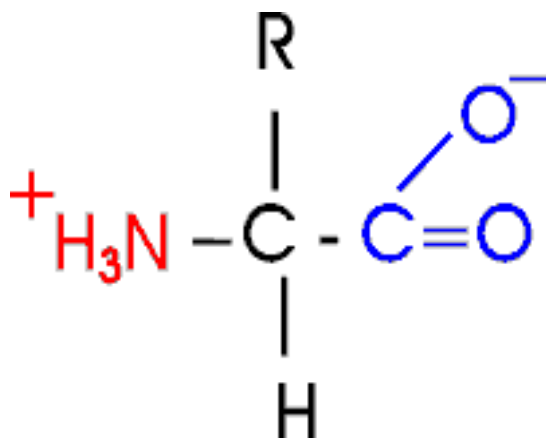
The general structure of an amino acid is:



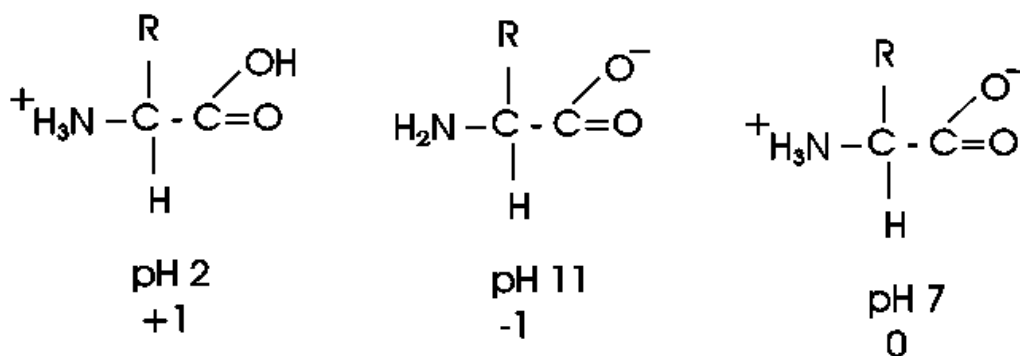
**R = ANY ORGANIC CHEMICAL GROUP
(SIDE CHAIN)**

Note the central carbon atom, to which 4 different groups are attached: an amino group (drawn by convention at the left), a carboxylic acid group (put at the right side), a hydrogen, and a side chain, or R-group. Only the R-group varies among the 20 different amino acids. This is the side chain, and so there are 20 different side chains. Look at the [amino acids and peptides handout](#) for some of the side chains. Your texts and hard copy handout show all 20, and you should examine all 20.

Out of laziness, I drew the general amino acid incorrectly: Actually at neutral pH, the molecule is charged, because the carboxylic acid group is an acid, and the amine group is a base, so more accurately: (also see [3-D picture](#))



Let's take this opportunity to discuss the charge on organic molecules a bit more. In living systems, the carboxylic acid group is mostly charged and the amine is mostly charged, but that is at pH7, the cellular pH under most circumstances. Is an acid always charged in aqueous solution? No. It depends on the pH of the environment. In the laboratory we do not have to keep things at pH 7, as it is in the cell. We can vary the environment at will, adding strong acids such as hydrochloric acid as a source of hydrogen ions (lowering the pH), or a strong base such as sodium hydroxide (raising the pH). The strength of an acid is a measure of how readily it gives up a proton. Carboxylic acids are always in equilibrium with the hydrogen ions (protons) in the solution, so if the hydrogen ion concentration is high (acidic) then the equilibrium will shift toward the protonated (uncharged) species. At pH 2.5 an amino acid carboxyl group is protonated about half the time; for each pH unit this proportion of protonated species will drop by a factor of 10, so very little of the carboxyl group is protonated at the neutral pH of 7 found in most cells. A similar situation pertains to the amine base end: at a very low H⁺ ion concentration (e.g., 10⁻¹¹ M H⁺, a high pH of 11), it will tend to lose its extra proton, but at pH 7 it will mostly remain protonated, with a positive charge. There will always be some intermediate pH at which we find the the group half charged/ half uncharged. This pH is called the pK of the group, and it can be influenced by the remainder of the molecule. The pK is an indication of the acidic or basic strength of the group (the lower the pK is the stronger the acid, the higher the pK the stronger the base).



So at pH7, most amino acids are **neutral** (no net charge), but they are **highly charged** nonetheless. [{Q&A}](#)

A molecule that is charged but electrically neutral is called a zwitterion.

Now, what are some of these 20 different side groups?

Here are 2 charged side group, e.g.:

asp: R= -CH₂-COO⁻ , there is a second carboxyl group on this amino acid)

lys: R= -CH₂-CH₂-CH₂-CH₂-NH₃⁺ , there's a second amine on lysine, so lysine will have 3 charged groups, and a net charge of +1 (two +'s and one -) at pH7.

There is a convention for *numbering* amino acid carbons; actually it's a lettering. It starts from the central carbon, called alpha: so lys has (count with me) an alpha, beta, gamma, delta, EPSILON-amino group as well as an alpha-amino group (and an alpha-carboxyl).

The average molecular weight of an amino acid is ~120, but the range is from 75 to 203.

The smallest amino acid (a.a.) is glycine (gly), MW = 75. Here the side chain is merely hydrogen.

The largest is tryptophan (trp), MW = 203 [-CH₂- bridge to a 5-membered ring containing a N plus a fused 6-membered ring] and is fairly hydrophobic.

Look over the structures of the 20 amino acids in the textbook. It is the properties of the **functional groups** on the 20 different **side chains** of the 20 different amino acids that determine the function of a protein, so they are all-important. The handout shows all 20 aa's, but without indicating the ionization of the acidic and basic groups. We will discuss many of the side chains within the context of the discussion as we go along.

There are two that deserve special mention: arginine contains a functional group that will not be found elsewhere in this course; it is -NH-C⁺(NH₂)NH₂, called the guanido group. The guanido group is a strong base, even stronger than an ordinary amine, so it is positively charged at pH7 (like lysine). You can consider the positive charge to be distributed over all 3 N atoms. Proline has a side chain that folds back and forms a covalent bond to the amine nitrogen of the amino acid, thus producing a ring structure.

You should be able to recognize the properties of the side chains as polar or non-polar, charged or not charged. You will not be responsible for recalling a specific amino acid structure from the English name or vice versa, but given the structure you should know how it behaves. [{Q&A}](#).

To be sure you understand amino acid structure, try problem 1-15 except C. For additional review of the effects of pH, try problem 1-16. For the effects of different R groups, try 1-20.

(Stereoisomers)

Now let's consider the structure of an amino acid in **3** dimensions:

When carbon forms 4 single bonds, it makes them spaced equally apart from each other in space, in the form of a tetrahedron as in this representation of glycine [a model with 2 white groups (the H's) is shown].

Now consider this other molecule of an amino acid [again with 2 white groups], with 2 H's of glycine, e.g. Are these the same molecule, that is, are they distinguishable or are they indistinguishable?

They are indistinguishable, since I can rotate them and superimpose their atoms.

But now suppose I make this alanine instead of glycine. I replace one white H with a -CH₃ group [orange] on each molecule [I am being sure to make them stereoisomers].

I can no longer superimpose them. They are both alanine, as they have the same four groups attached to the central carbon. But in three dimensions they are actually mirror images of each other. See [\[Purves6ed+7th: 2.21a\]](#). We call one D-ala and one L-ala. See [\[Purves6ed+7th: 2.21b\]](#).

This one is D, or is it this one... ? I can't remember .. it's not too important here.

What is important is that in general, you have this situation, the possibility of two **stereoisomers**, whenever there is what is called an **ASYMMETRIC** carbon atom in a molecule, that is, a carbon with four different groups attached. No matter how you might divide an asymmetric molecule in two by placing a plane through it, the two sides look different.

These stereoisomers are sometimes called optical isomers, since the two forms, in solution, will bend a beam of polarized light one way or the other. Thus the D designation originally meant dextro, for to the right, whereas L stood for levo, to the left.

All amino acids except glycine have an asymmetric carbon, which is the alpha-carbon. So we can draw 19 of the amino acids in 2 stereoisomeric forms.

So do we really have 39 a.a.'s? No. All the stereoisomeric forms of the amino acids in proteins are **L**-amino acids, so we only have to worry about 20.

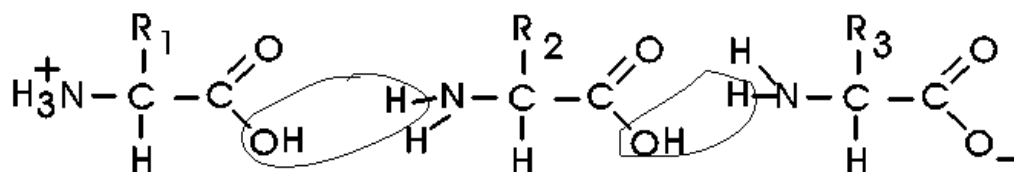
Note that the sugars we discussed, like glucose, have several asymmetric carbon atoms. Aside from L and D designations, the sugar stereoisomers are given different English names (e.g., D-glucose, D-mannose, L-rhamnose, etc.).

To review the various chemical groups discussed so far, try problem 1-19.

(Polypeptides, peptide bond)

Polymerization of aa's

OK, now let's now string these L-amino acids together, polymerize them. The bond that connects two amino acids is an **AMIDE** bond (-CO-NH-) between the carboxyl of one amino acid and the amino group of the next. Once again, a molecule of water is removed in the formation of the connecting bond:

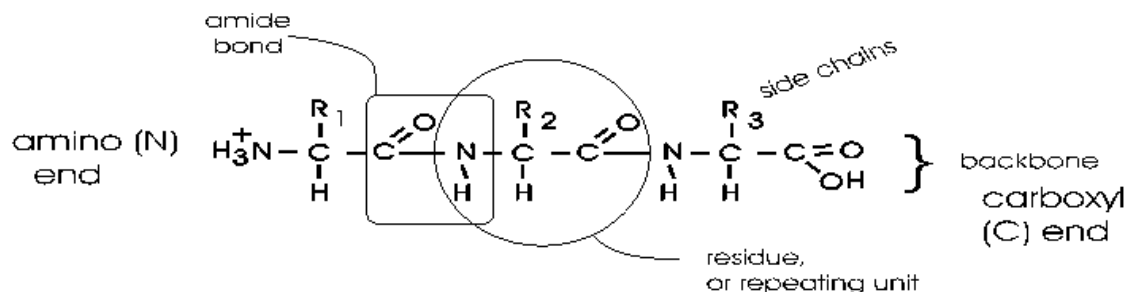


In the special case of proteins, this amide bond is called a **PEPTIDE BOND**, and the resulting product a **PEPTIDE**, a dipeptide (or we could go on to a tri-peptide, oligo-peptide, or finally, **POLY-PEPTIDE**). (See also [polypeptide handout](#)). Also see another [picture](#). {Q&A}

By convention, the amino group is written on the left for an amino acid and also for a peptide.

In the tripeptide in the diagram, note the peptide bond (boxed), and the repeating unit, or aa "residue" [circled]. Residue refers to what's left of the amino acid monomer after it has been incorporated into a polypeptide, which is most of it: it just lacks one H at what was the amino end and one OH at what used to be the carboxyl end. Note also that the charged amine and carboxyl groups no longer exist inside the polypeptide, having been replaced by the amide, an uncharged (but polar) functional group.

Almost all polypeptides have 2 ends, the **amino end** and the **carboxyl end**, which **do** remain charged at pH7.



The "**backbone**" of the polypeptide is defined as all of the atoms except the side chains.

The only **free** amino and carboxyl atoms of the backbone are at the 2 ends.

The side chains, then, stick out of this backbone (also see [polypeptide handout](#)).

Nomenclature: e.g., alanine-methionine-alanine, or ala-met-ala, or alanyl-methionyl-alanine, or AMA

To review peptide structure, try problem 1-15, C, and then try 2-3 part A.

The length of polypeptides is commonly 100-1000 amino acids, but smaller and larger ones also can be found.

Each and every protein molecule in the cell has an identity defined by its particular sequence of amino acids. Each E. coli cell contains about 3 million polypeptides molecules, but only about 3000 different ones. Each of these individual protein types has a name to go along with its chemical identity.

Are there enough combination to specify 3000 different polypeptide sequences? Well, if the average polypeptide is 500 amino acids long, then the possible combinations are 20^{500} or 10^{650} , which is a number of inconceivable magnitude. So evolution has settled on about 100,000 of these combinations to do biological jobs.

Some examples of polypeptides, taken not from E. coli, but from more familiar organisms include:

Carriers: hemoglobin, which carries oxygen in red blood cells;

Nutrients: egg albumin, a nutrient in the white of a hen's egg;

Structural: keratin, providing toughness in skin, fingernails, and wool;

collagen, providing a strong connection between cells in tendons;

Signal reception: estrogen receptor (intracellular)
epidermal growth factor receptor (spanning the cell membrane)

Recognition of foreign substances: immunoglobulins (antibodies)

Enzyme catalysts: beta-galactosidase, which helps digest the milk sugar lactose.

We will discuss enzymes in some detail as an important category of proteins.

(Primary (1°) structure = linear sequences of AAs)

Each of these proteins contains a polypeptide with a particular sequence of amino acids, usually all 20 are represented, although not at all equally. Unlike polysaccharides, this sequence usually exhibits no obvious simple regularity, or repeating subsequence:

This **linear** sequence of amino acids is called the **primary (1°) structure** of a protein. [{Q&A}](#)

It might be, for instance: met-ala-leu-leu-arg-glu-leu-val How is this sequence determined?