

Lecture 9

Amino acids and protein structure

Today we're going to introduce some of the fundamentals of protein structure. This may seem far-removed from gene cloning, but it is the path to understanding the genes that we are cloning.

For this lecture, much of the material that will be presented is "off-site" and you will need to have the right type of browser plug-in to take full advantage of the site. Fortunately it is all free, though it may be more difficult to use if you are on a vintage computer or have the wrong flavor of browser.

- The plug-in module that you need is called "[Chime](#)", and it is a free program to show molecular structure in three dimensions.
- The browser that you need is Netscape version 4 (or later), but not Netscape 6. Internet Explorer will not work!

We've already had some opportunity to use Chime when we were looking at DNA structure, back in the first lecture, but Chime really becomes useful when we turn to study proteins. [Here's an example of a site using Chime](#). See if it will work for you.

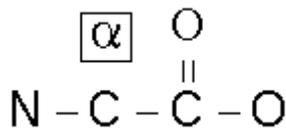
Of course I'm sure you remember the fundamentals of protein structure from your introductory biology course. We have such a thing as **primary structure**, which is simply the order in which you find the amino acids in a polypeptide sequence. When we talk about the primary structure of a protein, or as we sometimes say, the **primary sequence**, we are not talking about anything having to do with the shape of the protein in space. For that, we have other names.

We use the name **secondary structure** to refer to assembly of **alpha helices** and **beta sheets**, and these are regular shapes that depend on hydrogen bonding between elements of the peptide backbone. A protein may form a **tertiary structure** by bringing together amino acids from widely spaced segments of the primary sequence, and the types of interactions that are important in forming tertiary structure are **hydrophobic interactions**, **hydrogen bonding**, **ionic interactions** or "**salt bridges**", and **disulfide linkages**. Multiple polypeptide chains can be brought together to form a **quaternary structure**, as in the example of hemoglobin which has four polypeptide chains (alpha₂beta₂). So, does it matter if we use the word polypeptide or the word protein? It does. In increasing order of length, we would say amino acid, dipeptide, tripeptide, ..., polypeptide. Sometimes we use the word **oligopeptide** to talk about a short polypeptide, but the word does not imply a specific length. A **polypeptide** is a single polymer of amino acids. A **protein** on the other hand, may be an assemblage of one or more polypeptides.

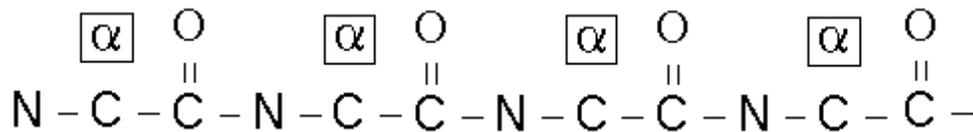
The peptide backbone So - let's review then.

Amino acids are the monomers that are linked through peptide bonds to form a polymer that we call a polypeptide. The process in vivo involves charged tRNAs and a peptidyl transferase activity in the ribosome, but the overall reaction is one of dehydration to produce an [amide linkage \(click here for an animation\)](#).

Amino acids differ in their side chains, which we refer to as **R groups**, and the R groups are attached to the alpha carbon of the backbone. Here's a really stripped down structure of an amino acid that doesn't show any of the hydrogen atoms. The amino group on the far left would usually be NH₃ with a positive charge. The carbon marked with the letter alpha would have a hydrogen atom and an R group bound to it, and the carboxyl carbon would be typically a dissociated carboxylic acid (COO⁻).



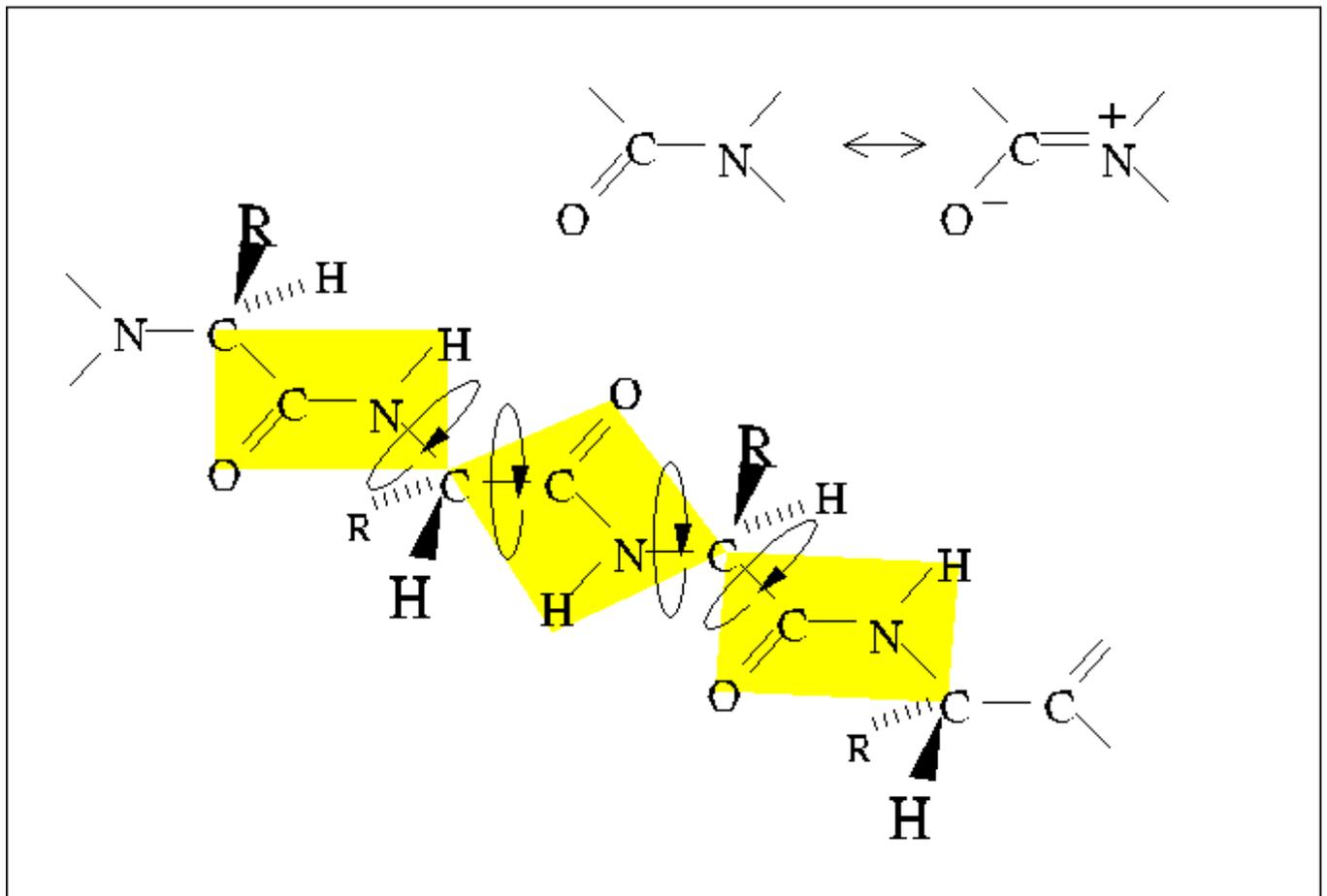
What do we mean by a peptide backbone? It is the regular structure that forms when we link these amino acids together through making an amide bond (which we call a peptide bond)



This is also shown as a projection below (from Tulane Univ.). The R groups are attached to the alpha carbon, and alternate being projected out of the screen and into the screen.

One important concept is that the carboxyl carbon is planar, due to the partial double-bond character of the amide group. That is, the three atoms bonded with the nitrogen are squashed into the same plane as the nitrogen, rather than forming a tetrahedral structure as in the case of the ammonia molecule. The planar regions in this picture below are shaded.

The bonds in the backbone on either side of the alpha carbon are able to rotate, and are called phi (on left) and psi (on right). Due to steric interference, they tend to adopt angles that lead to a nearly extended chain.



<http://www.tulane.edu/~biochem/med/second.htm>

These details are explained in a site from [Tulane University](http://www.tulane.edu/~biochem/med/second.htm).

The [Phi and Psi](#) angles are also represented in a Chime-based module that is available.

The R-group Let us discuss the structures of each amino acid R group, paying particular attention to the degree of hydrophobicity, the capacity for hydrogen bonding, and the capacity to carry a charge.

The twenty amino acids: [A Chime-based site](#)

On the matter of charge, [this site may explain some of the acid dissociations](#)

[Aliphatic amino acids](#)

First let's look at amino acids with aliphatic side groups (R groups), starting with the simplest of all: Glycine only has a hydrogen as a side chain. Alanine is the next largest,

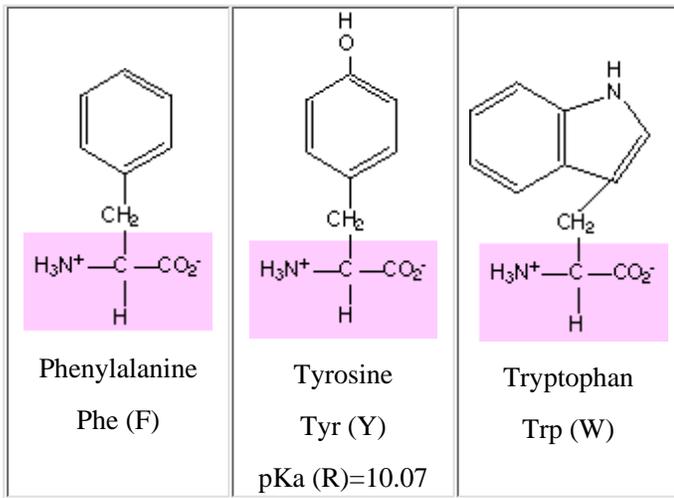
with a methyl group. Valine has a three-carbon side chain, while leucine and isoleucine have four-carbon side chains. Proline has a three-carbon side chain that reconnects with the amino nitrogen (making it an imino).

In the renderings shown below, the common peptide backbone is shaded in pink. Pay particular attention to the unshaded atoms.

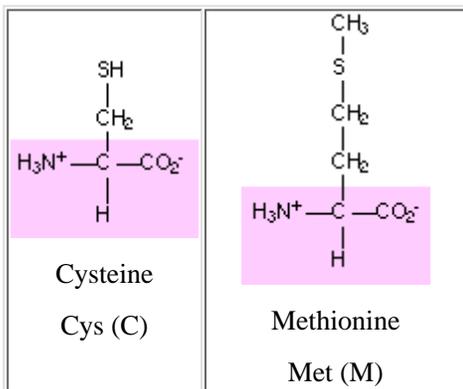
Art source:<http://www-mcb.ucdavis.edu/courses/bis102/>

$\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{CO}_2^- \\ \\ \text{H} \end{array}$ <p>Glycine Gly (G)</p>	$\begin{array}{c} \text{CH}_3 \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{CO}_2^- \\ \\ \text{H} \end{array}$ <p>Alanine Ala (A)</p>	$\begin{array}{c} \text{CH}_3 \quad \text{CH}_3 \\ \diagdown \quad / \\ \text{CH} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{CO}_2^- \\ \\ \text{H} \end{array}$ <p>Valine Val (V)</p>
$\begin{array}{c} \text{CH}_3 \quad \text{CH}_3 \\ \diagdown \quad / \\ \text{CH} \\ \\ \text{CH}_2 \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{CO}_2^- \\ \\ \text{H} \end{array}$ <p>Leucine Leu (L)</p>	$\begin{array}{c} \text{CH}_3 \\ \\ \text{CH}_2 \\ \\ \text{CH} - \text{CH}_3 \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{CO}_2^- \\ \\ \text{H} \end{array}$ <p>Isoleucine Ile (I)</p>	
$\begin{array}{c} \text{H}_2 \\ \\ \text{C} \\ / \quad \backslash \\ \text{H}_2\text{C} \quad \text{CH}_2 \\ \quad \\ \text{H}_2\text{N}^+ - \text{C} - \text{CO}_2^- \\ \\ \text{H} \end{array}$ <p>Proline Pro (P)</p>		

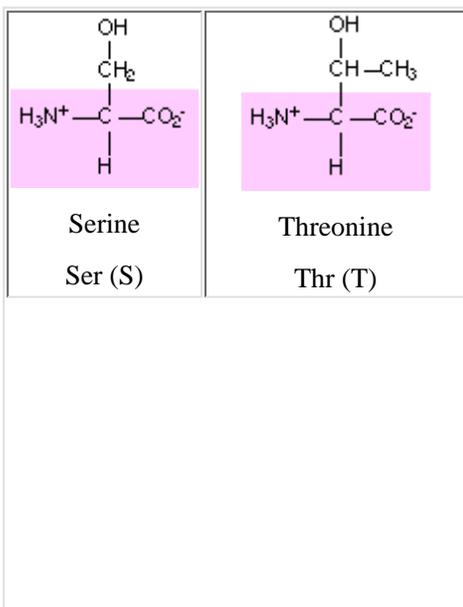
[Aromatic amino acids](#)

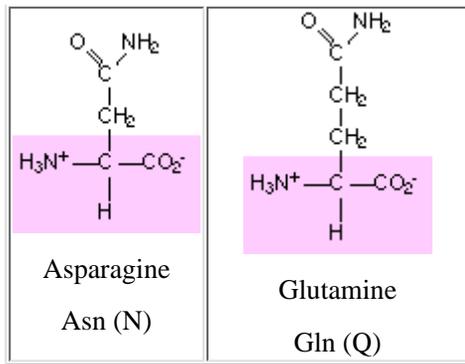


Sulfur-containing amino acids



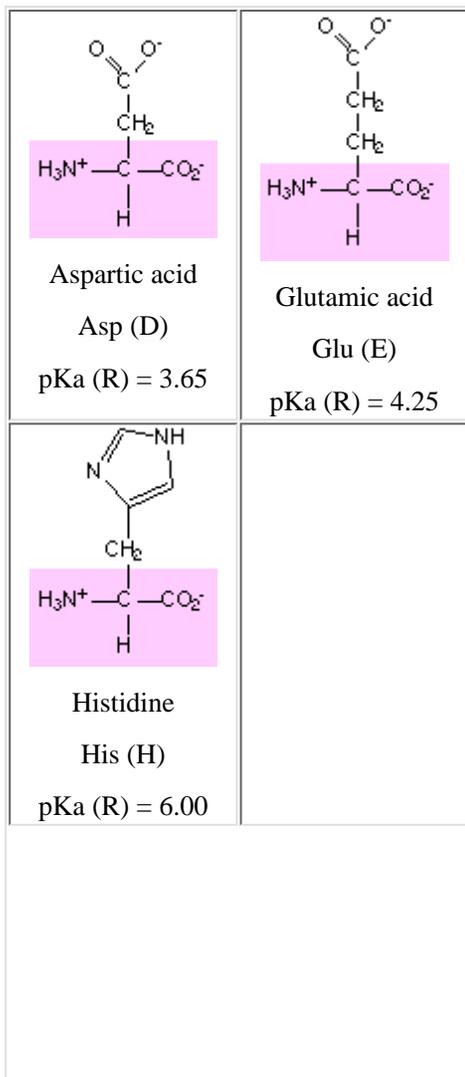
Polar, uncharged amino acids

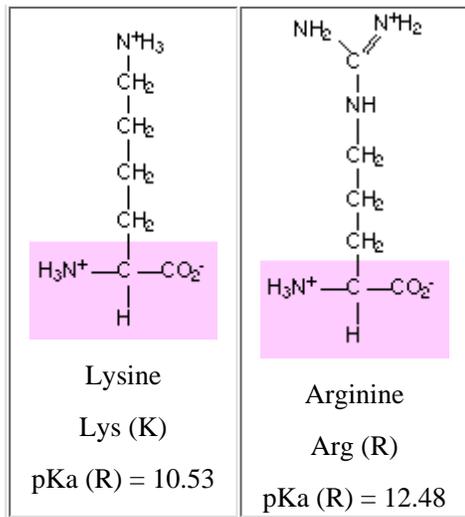




Polar, charged amino acids

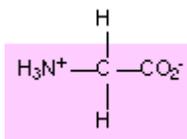
Look closely at the pKa's of the R groups. Aspartic and glutamic are carboxylic acidic side chains, tending to be dissociated (negatively charged) at neutral pH. Histidine has a pKa of 6, and lysine and arginine are basic with pKa's of about 10.5 and 12.5 respectively. These would be protonated (positively charged) at neutral pH.





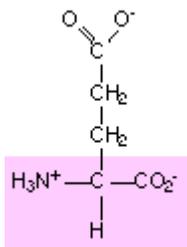
Meet the amino acids: [A chime-based site showing electrostatic and lipophilic surfaces](#)

Here's a question for you Now that you've learned a bit about amino acid side chains, and the peptide backbone, what would be the net charge on an amino acid such as glycine, at a pH of 7?

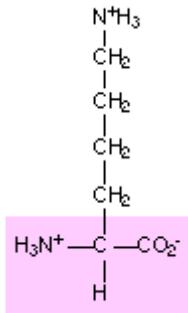


Well, you have a dissociated carboxylic acid (with a pKa of about 2) and a protonated amine (with a pKa of about 9). The net charge is zero, but of course that is because it has one negative and one positive charge. We call that a **zwitterion**.

Suppose I asked the same question about an amino acid like glutamic acid (pKa (R) = 4.25)?



Or suppose we were trying to figure the net charge on lysine (pKa (R) = 10.53)?



Now a more complicated question. Is there an easy way to figure the net charge of the following protein at pH 7.0?

ILIKECALSTATE

The N and C termini are going to be of equal and opposite charge, and you also have positive charges from each lysine (K) and arginine (R). If the pH is below the pKa of histidine (H), you would also add positive charges for those as well. You will find negative charges on each aspartic (D) and glutamic (E) side chain. So, the net charge is $+1(\text{N term}) + +1(\text{K}) + -1(\text{E}) + -1(\text{E}) + -1(\text{C term}) = -1$

Here's a slightly different question. What would be the net charge of ILIKECALSTATE at pH 3.0? Would it be different, considering that the R groups of the glutamic acids (E) would be protonated?

Protein structure primer There are several types of protein structural elements that we are going to study:

- alpha helices
- beta sheets (parallel and antiparallel)
- beta turns

The key to understanding these structures is to look at the hydrogen bonding pattern. In an alpha helix, for example, H bonds are forming between a donor (amide nitrogen) at one residue (n) and acceptor (amide oxygen) three residues away (n+3). The helix is typically right-handed, with 3.6 residues per turn.

We will spend a considerable amount of time working with these links in lecture:

- The U.Mass [hemoglobin site](#), cited earlier, illustrates these structural elements nicely.
- Also, McClure's [Protein G primer](#) is an outstanding site.
- [Protein Explorer](#) is a program you can use to download and inspect proteins from the Brookhaven crystallographic database.

Test your knowledge Take the [amino acid quiz](#)

Take the [protein structure quiz](#)

Stan Metzenberg
Department of Biology
California State University Northridge
Northridge CA 91330-8303
stan.metzenberg@csun.edu

© 1996, 1997, 1998, 1999, 2000, 2001, 2002