

## Amino acids

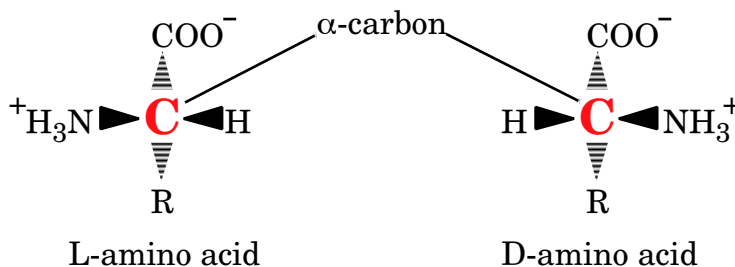
### Biological roles:

1. Amino acids can be metabolized to **produce energy**. This is especially important during fasting, when the breakdown of muscle protein is a major energy source.
2. Some amino acids act as **neurotransmitters**, and some act as starting materials for the **biosynthesis** of neurotransmitters, hormones, and other important biochemical compounds.
3. Amino acids are the primary building blocks for **proteins**.

**Definition:** an amino acid is an organic compound containing an amino group and an acidic group (usually, but not always, a carboxylic acid). This description fits a very large number of compounds, most of which are non-physiological. In addition, many physiologically important amino acids are not used in proteins. However, in biochemistry, the term “amino acid” generally refers to one of the 20 types of monomeric units most commonly used to construct proteins.

### General structure:

The amino acids that constitute proteins are all  $\alpha$ -amino acids. The “ $\alpha$ -carbon” is bonded to four different groups: an amino group, a carboxylic acid, a hydrogen, and the side-chain. Because these are four different chemical groups, the  $\alpha$ -carbon is chiral. Physiological amino acids are almost exclusively **L stereochemistry**.<sup>5</sup> The reason for the specifically L stereochemistry is not known; in principle D stereochemistry would also be possible. Experiments performed using a chemically synthesized protein made from D-amino acids showed the resulting protein to be fully functional, although folded into the mirror image of the normal protein.



In the above paragraph, it is stated that physiological amino acids are almost exclusively L stereochemistry. A few exceptions exist. Glycine is non-chiral, and therefore has no stereochemistry. In addition, some organisms use small amounts of D amino acids for specific purposes. However, the protein synthesis machinery cannot use D amino acids, and therefore protein synthesis results in proteins comprised of L amino acids and glycine. Finally, two of the standard amino acids,

<sup>5</sup>In general, biochemistry does not use the Cahn-Ingold-Prelog R/S nomenclature commonly used in organic chemistry. Although the older D/L nomenclature is difficult to assign to a novel molecule, the D/L nomenclature is actually somewhat simpler to use, especially for amino acids. While nearly all of the chiral amino acids used for protein synthesis exhibit S stereochemistry, cysteine and a few rarer amino acids have the R configuration. These amino acids are, however, still designated L due to the differing criteria for the nomenclature rules.

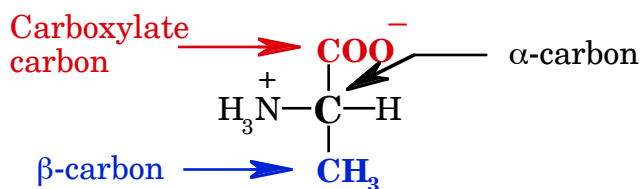
isoleucine and threonine, have chiral centers at both the  $\alpha$  and  $\beta$  carbons. Physiological forms of these two amino acids always have the L-stereochemistry at the  $\alpha$ -carbon, and have a set stereochemistry at the  $\beta$ -carbon as well; unless specifically stated otherwise, you should assume the correct stereochemistry at both carbons.

The  $\alpha$ -amino group and carboxylic acid are normally ionized at physiological pH, forming a **zwitterion**. Note that if both groups are fully ionized, the overall net charge of the amino acid (ignoring the side chain) is zero.

### Amino acid structure

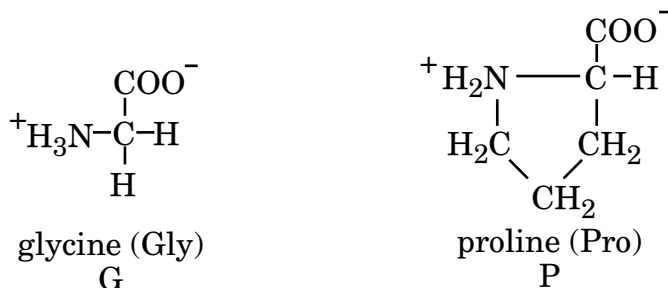
Structurally, the standard amino acids (with the exception of glycine) can be considered to be derivatives of alanine. (Note: these amino acids are not synthesized from alanine; instead, looking at the amino acids as alanine derivatives is merely an aid to understanding the structures.)

The “first” carbon of the amino acid is the carboxylate carbon. The next carbon is called the  $\alpha$ -carbon, because it is in the position  $\alpha$  to the first carbon. The side chain is then bonded to the  $\alpha$ -carbon. The side-chain carbons are given Greek letters in the order  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$  counting away from the  $\alpha$ -carbon. Because most amino acids are structurally based on alanine, the modifications to the alanine structure are varying compounds attached to the  $\beta$ -carbon.



### Amino acid classification

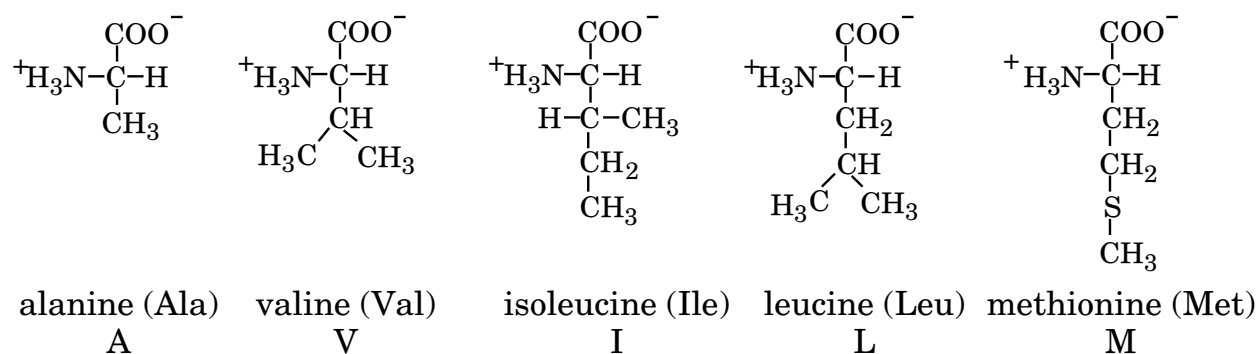
The standard scheme for amino acids classification is based on an assessment of the relative polarity of the side chain. **Glycine** and **proline**, however, do not fit into this standard classification scheme for amino acid side-chains. Glycine has the smallest side-chain (a hydrogen), and is the only non-chiral amino acid. Because it effectively lacks a side-chain, glycine has much less steric hindrance than the other amino acids. In addition, because it effectively lacks a side-chain, the glycine side-chain does not really have a polar or non-polar character.



Proline is often referred to as an **imino** acid; unique among the standard amino

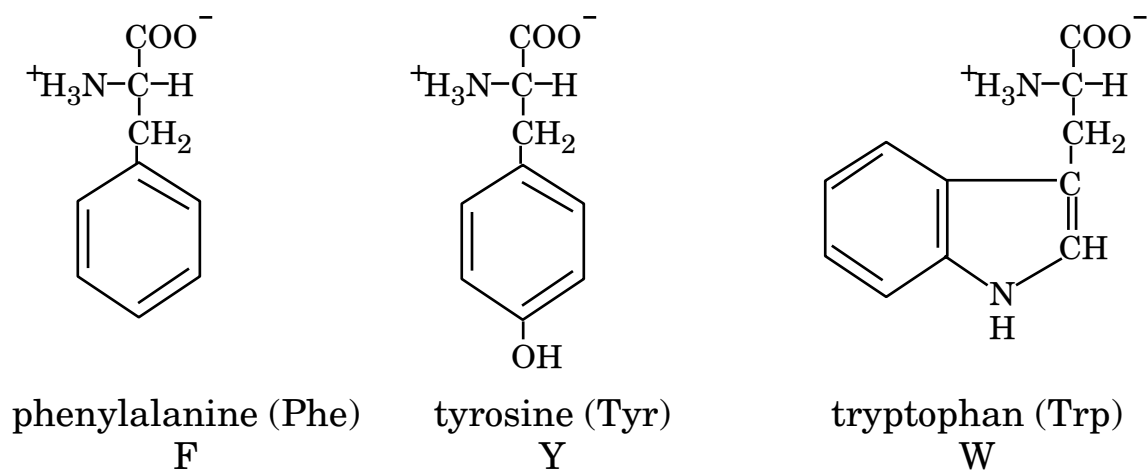
acids, the proline  $\alpha$ -amino group is actually a secondary amine. Proline has a conformationally restricted structure, with significantly fewer possible conformations of the  $\alpha$ -carbon bonds than the other amino acids. Although the proline side chain appears to be non-polar, proline has a fairly high tendency to appear on the solvent-exposed surface of proteins. Thus, although it appears to be non-polar, in a sense proline behaves more like a polar amino acid.

**Non-polar** amino acids contain hydrophobic side chains. These are the amino acids alanine, valine, isoleucine, leucine, and methionine. Note that in each case, the side chain consists of an aliphatic carbon chain. Although methionine has a thioether link prior to the terminal methyl group, this sulfur has generally similar properties to a methylene group.



An examination of the **valine** and **isoleucine** structures (above) and the structure of **threonine** (on the next page, below) reveals that these three compounds have **two** substituents on their  $\beta$ -carbon. This means that isoleucine and threonine residues contain two chiral centers (a minor point), and, more importantly, that these amino acids have greater steric hindrance, which allows these residues fewer possible conformations while forming part of a protein structure. These three amino acids are classed as  **$\beta$ -branched** to reflect their more restricted conformational flexibility.

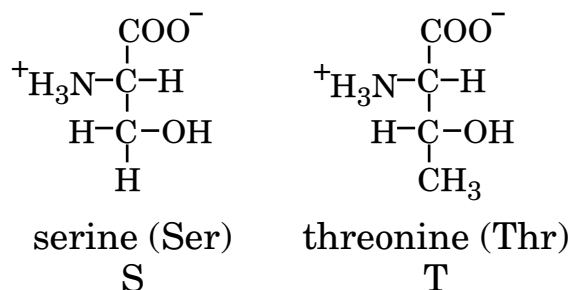
Three amino acids possess **aromatic** rings: phenylalanine, tyrosine, and tryptophan.



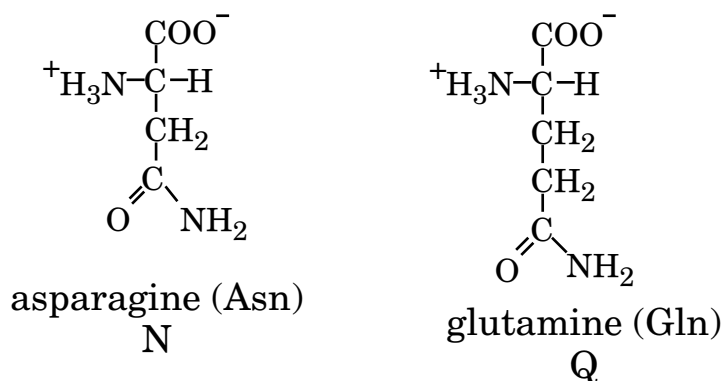
Phenylalanine is a hydrophobic amino acid; while tyrosine and tryptophan are borderline cases. Tyrosine is not very soluble in water, but contains a polar hydroxyl group in the *para* position of the phenyl ring (the phenolic proton has a  $pK_a = 10.5$ ). Because, in proteins, the tyrosine hydroxyl is often solvent-exposed, tyrosine is frequently listed as a polar amino acid. Tryptophan is also borderline hydrophobic/polar, because the proton attached to the indole nitrogen can act as a hydrogen bond donor to water or other hydrogen bond acceptors. However, the majority of the large tryptophan side-chain is non-polar.

The aromatic amino acids tryptophan and tyrosine absorb light at  $\sim 280$  nm, and account for nearly all of the ultraviolet absorbance characteristic of proteins that lack non-amino acid chromophores.

**Serine** and **threonine** contain hydroxyl groups. A comparison of their structures reveals that threonine is almost identical to serine except for the substitution of a methyl group for one of the serine  $\beta$ -carbon hydrogens. The hydroxyl groups on serine and threonine are essentially always protonated, and therefore uncharged.

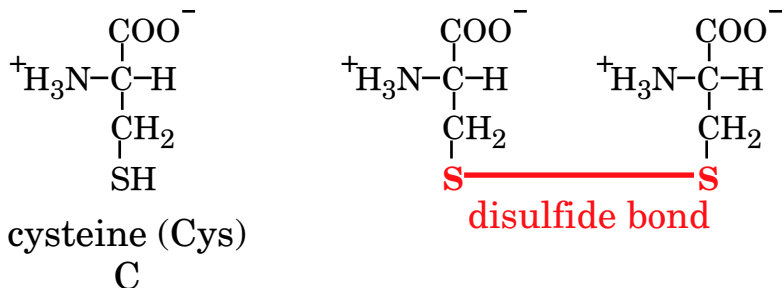


**Asparagine** and **glutamine** contain amide groups in their side chains. Amide groups are polar, and are capable of forming hydrogen bonds with water or with other compounds. However, **amide groups do not become charged** at biologically relevant pH values (*i.e.* in the pH range from 2-12).

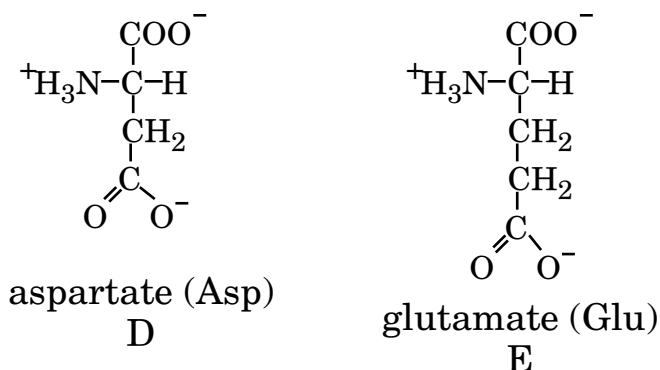


**Cysteine** is structurally similar to serine, with the substitution of a sulfur atom for the serine hydroxyl oxygen. The sulfhydryl group has a  $pK_a = \sim 8.4$ . This means that at pH 7.4, about 10% of free cysteine sulfhydryl groups are deprotonated, and therefore negatively charged. Defining the polarity of cysteine is difficult; the ionized cysteine is polar, while the protonated form is much less polar than serine.

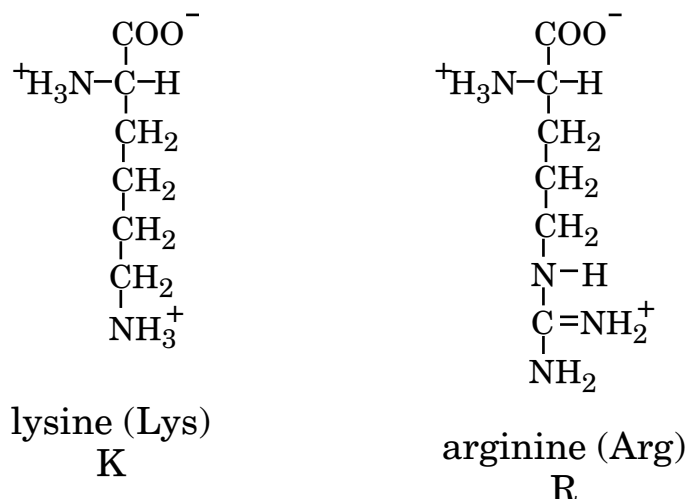
Cysteine also has the ability to form covalent bonds called disulfide bonds between cysteine side chains. This disulfide bond can be formed between cysteine molecules, or between cysteine side-chains from within proteins. Disulfide bonds are the most common form of covalent bond between amino acid side-chains in proteins. Note, however, that disulfide bonds are rarely observed in intracellular proteins.



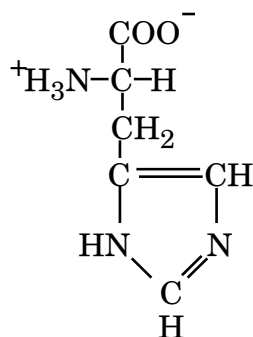
**Aspartic acid** and **glutamic acid** are the only standard amino acids with carboxylic acid groups in their side-chains. Although the side-chain  $\text{pK}_a$  depends on the local environment, these residues are generally negatively charged, because the  $\text{pK}_a$  value of the side-chain is  $\sim 4$  for both of these compounds.



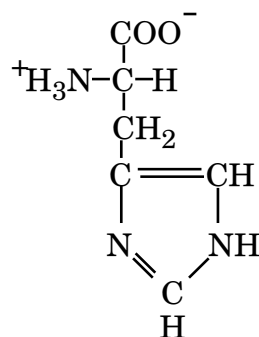
**Lysine** ( $\text{pK}_a = 10.5$ ) and **arginine** ( $\text{pK}_a = 12.5$ ) are basic amino acids that are positively charged at physiological pH. The side-chains of these compounds contain a primary amine (in lysine) or a guanidinyll group (in arginine) that can be readily protonated.



**Histidine** can also be protonated, but the side-chain  $pK_a$  is 6.0 and therefore only a very small percentage of the side-chain is protonated at pH 7.4. The  $pK_a$  of the side-chain is heavily environment dependent, and varies considerably in proteins. (Note: histidine exists in two neutral forms,  $\delta$  and  $\epsilon$ , shown below. These are generally equivalent structures, with the only difference being the nitrogen that bears a proton, but they exhibit different geometric constraints on hydrogen bonding properties.)



$\delta$  histidine (His)  
H



$\epsilon$  histidine (His)  
H

Under most conditions, a total of six amino acid side-chains can be ionized: aspartate, glutamate, cysteine, histidine, lysine, and arginine. Aspartate, glutamate, and cysteine are ionized and **negatively charged** when they are **deprotonated**. Histidine, lysine, and arginine are ionized and **positively charged** when they are protonated. Of these, aspartate, glutamate, lysine, and arginine are predominantly ionized under physiological conditions, while histidine and cysteine are ionized only in some cases.

**All biochemists (and biochemistry students) need to know:**

1. The structures of all twenty standard amino acids.
2. The three and one letter codes for these amino acids.
3. The general properties of these compounds, such as size, charge, polar or non-polar character.

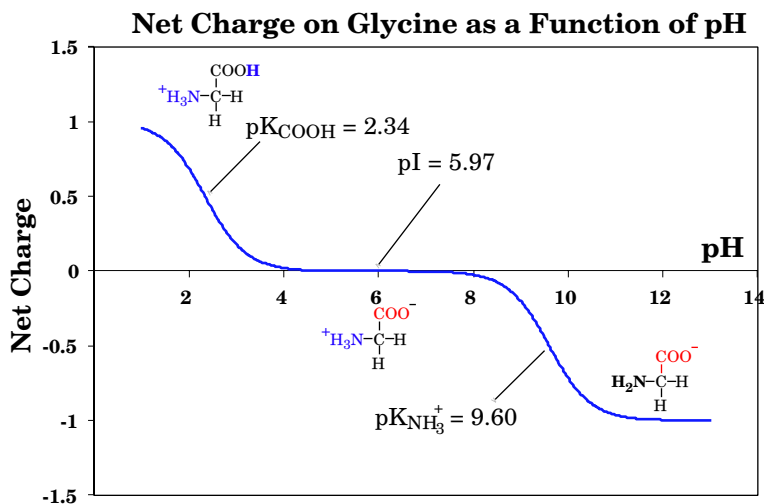
**Acid-base chemistry of amino acids**

At  $\text{pH} \gg pK_a$ , the ionizable group will be deprotonated. For carboxylic acids, this means that the group will be negatively charged; amines will be neutral.

At  $\text{pH} \ll pK_a$ , the ionizable group will be protonated. For carboxylic acids, this means that the group will be neutral; amines will be positively charged.

At  $\text{pH} = pK_a \pm 1$ ; the ionizable group will be a mixture of protonated and deprotonated forms (or, to phrase the concept slightly differently, the ionizable group will be protonated some fraction of the time, as predicted by the Henderson-Hasselbalch equation).

Let us look first at the least complex amino acid, glycine. A standard value<sup>6</sup> for the  $pK_a$  of the  $\alpha$ -amino group is 9.60 and of the carboxylate group is 2.34. Using these  $pK_a$  values, and a minor variation of the Henderson-Hasselbalch equation, it is possible to predict the charge on glycine as a function of pH. A graph of net molecular charge *versus* pH for glycine is shown below; note the glycine structures indicating the major species present in solution at various pH values.



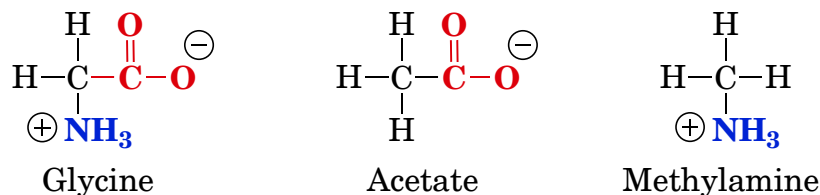
At pH values below 1, the glycine molecule has a +1 charge, because both the carboxylate and the amino group are charged. As the pH increases, the charge on the glycine becomes less positive (and thus more negative), reaching zero at 5.97. Note that the pI (the pH at which the molecule has zero net charge), 5.97, is the arithmetic mean of the  $pK_a$  values for the  $\alpha$ -carboxylate group and the  $\alpha$ -amino group, because these are the only ionizable groups, and if both are fully charged, the molecule has zero net charge. This zero net charge state is known as the **zwitterionic** state (a zwitterion is capable of both accepting and donating protons), and is characteristic of free amino acids with non-ionizable side-chains near pH 7.0. As the pH increases further, the charge on the glycine molecule becomes negative. This illustrates a general principle: **as the pH increases, the charge on the molecule becomes more negative**, and as the pH decreases, the charge on the molecule becomes more positive.

The titration for glycine shown above illustrates a property common to all amino acids: these compounds are always ionized in aqueous solution. They may have a net positive charge, a net negative charge, or a net charge of zero, but these molecules always contain at least one ionized group.

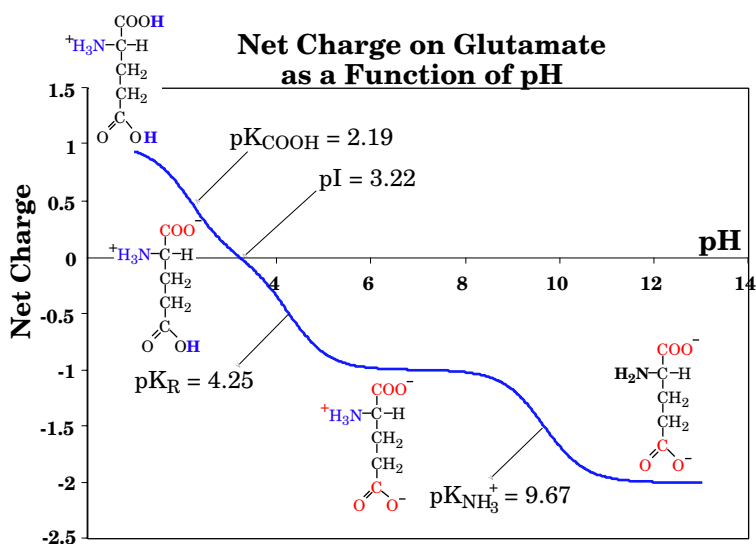
The observed  $pK_a$  values for glycine also illustrate a third important point: **the local environment alters the  $pK_a$  values for ionizable groups**. The only chemical difference between the amino acid glycine and acetic acid is the presence of the amino group attached to the  $\alpha$ -carbon. The presence of this amino group, however, has a profound effect. The  $pK_a$  for acetic acid is 4.76, while the  $pK_a$  of the

<sup>6</sup> Source: *Lehninger: Principles of Biochemistry*, 5<sup>th</sup> Ed, W.H. Freeman & Company, New York (2008), p.73.

glycine  $\alpha$ -carboxylate is 2.34, a value more than two pH units lower. This rather dramatic difference is caused by the  $\alpha$ -amino group proton, which repels the carboxylate proton, and stabilizes the negative charge on the carboxylate. Similarly, the  $pK_a$  of methylamine (*i.e.* glycine without the  $\alpha$ -carboxylate) is 10.6, which is one pH unit higher than the  $pK_a$  of the glycine  $\alpha$ -amino group. We will see further examples of  $pK_a$  alteration due to local environment as we look at protein structure and function.



As a second titration example, let us consider the amino acid glutamate. Glutamate contains three ionizable groups, an  $\alpha$ -amino group ( $pK_a = 9.67$ ), an  $\alpha$ -carboxylate group ( $pK_a = 2.19$ ), and a  $\gamma$ -carboxylate group ( $pK_a = 4.25$ ). As with glycine, at pH 1, the glutamate molecule has a +1 charge, because all of the ionizable groups are protonated. As the pH increases, the charge on the molecule becomes more negative. Note that the charge on the molecule reaches zero at 3.22. This value for the pI can be calculated by taking the average of the  $pK_a$  values for the two negatively charged ionizable groups. (Why does this work? What is the pI for lysine?)



As with the glycine titration, note that the charge on glutamate becomes more negative with increasing pH. In the case of glutamate, however, the amino acid is negatively charged at pH values close to 7.0, due to the side-chain carboxylate. Thus, glutamate typically has an overall negative charge in the usual physiological pH range.

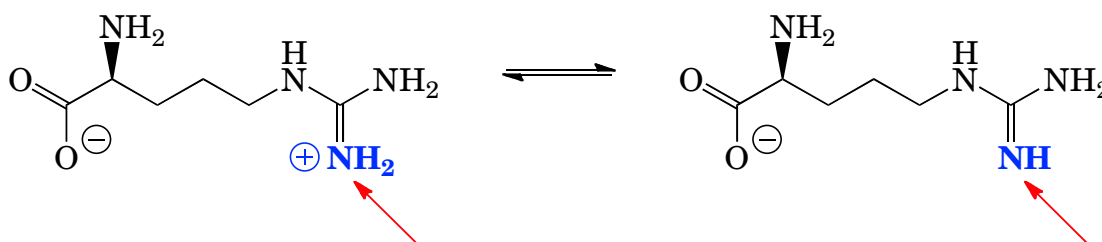
### Side-chain acid-base chemistry

As we will see, in peptides and proteins, the  $\alpha$ -amino and carboxylate groups become part of the backbone amide linkages that hold the peptide together. As a

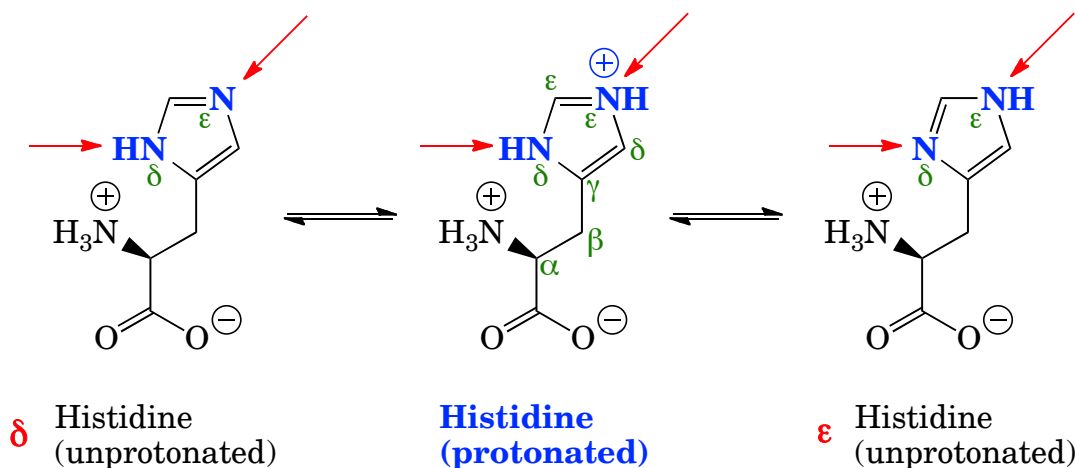
result, only the first amino acid residue has a free  $\alpha$ -amino group, and only the last amino acid residue has a free carboxylate. Other than these two groups (which are sometimes chemically modified to form uncharged derivatives, the only ionizable groups in most peptides and proteins are the amino acid side chains.

**Glutamate** and **aspartate** are generally deprotonated and therefore negatively charged at physiological pH, because the  $pK_a$  of these groups is usually  $\sim 4$ . However, these side-chains may become partially or fully protonated in some conditions (*e.g.*, in lysosomes, which are subcellular organelles with an internal pH of  $\sim 4$ , or in the mammalian stomach lumen, which typically has a pH of  $\sim 2$ ).

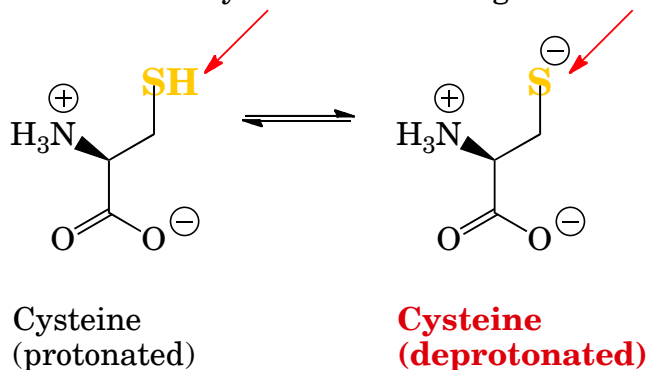
**Arginine** (side-chain  $pK_a = 12.5$ ) and **lysine** (side-chain  $pK_a = 10.5$ ) essentially always have protonated (and therefore positively charged) side-chains at physiological pH (it is extremely rare for the pH of living organisms to rise high enough to deprotonate these residues). (Note that the  $\alpha$ -amino group is not protonated in the diagram below. Why was this group drawn this way?)



**Histidine** has a  $pK_a$  that varies over a relatively wide range, because its  $pK_a$  depends heavily on the local environment. When  $pK_a = 6$ , very little is charged at pH 7.4 (a pH value of 7.4 is observed in serum and most cellular compartments). However, when the local environment raises the histidine side-chain  $pK_a$  to  $\sim 7$ , the side-chain will have a significant partial positive charge at pH 7.4. In addition, histidine side-chains are frequently protonated when the protein is exposed to the lower pH of lysosomes or the stomach lumen. Histidine can act as a pH-dependent switch in proteins by inducing changes in the protein structure in response to the additional charge introduced by the proton.



Cysteine has a  $pK_a$  of  $\sim 8.4$ . Free cysteine side-chains will therefore be slightly negatively charged at some physiological pH ( $\sim 10\%$  at pH 7.4). Cysteine is much less likely than histidine to act as a pH-dependent switch, because reductions in pH will only protonate the  $\sim 10\%$  of the cysteine side-chains that are deprotonated, while pH values above  $\sim 8$  are rarely observed in living cells.



### Isoelectric point

As discussed above, amino acids and the proteins they form are always ionized molecules in aqueous solution. However, it is possible for the number of positive charges and negative charges to be exactly equal, resulting in no **net** charge. This state of zero net charge is pH dependent. The pH at which the molecule has no net charge is the isoelectric point, and is usually termed the **pI**.

Molecules with **single** types of ionizable groups can only be either charged or not charged. If, on the other hand, a molecule contains at least one group capable of forming a positive charge *and* at least one capable of forming a negative charge, that molecule will have an isoelectric point. For free amino acids with charged side chains, the pI is between the  $pK_a$  values of the side chain group and the  $pK_a$  of the  $\alpha$ -carbon group with the same charge (*e.g.* for lysine, the pI is 9.8, the average the  $pK_a$  of the side-chain, 10.54, and the  $pK_a$  of the  $\alpha$ -amino group, 9.06).

For **proteins**, the pI is the result of the relative charge of all of the ionizable side chains, and varies considerably, depending on the number of each type of ionizable amino acid and on structural features. The pI can be predicted based on the amino acid composition, but the predictions are not entirely accurate due to local environment-induced changes in  $pK_a$  of the side chains.

The Chemistry 330 course website has a downloadable Excel spreadsheet that estimates the molecular charge (based on peptide composition) as pH varies, and can generate a plot of charge *versus* pH for any amino acid or protein. As downloaded, it includes  $pK_a$  values based on the average values for proteins, which differ somewhat from those of free amino acids. If you wish to examine the plot for a free amino acid, enter the relevant  $pK_a$  values from your Biochemistry textbook.

### Spectroscopic properties

Proteins composed solely from amino acids are white solids (or clear, colorless aqueous solutions). This means that amino acids do not absorb light in the visible

range. However, **tryptophan** and **tyrosine** absorb in the near ultraviolet (at ~280 nm). Tryptophan (and to small extent tyrosine) exhibits fluorescence. Because fluorescence spectra are very sensitive to the local environment, fluorescence spectroscopy can be used to perform real-time monitoring of changes in the structure of the protein. It is especially useful for observing structural changes in the close proximity of tryptophan residues.

Other than tryptophan and tyrosine, none of the standard amino acids have significant spectral properties.

### Summary

Physiological amino acids have a variety of side-chains with a range of properties. These property differences are important biologically.

The **protonation state** of a functional group, such as a side-chain carboxylate or an  $\alpha$ -amino group, can be determined by comparing the  $pK_a$  of the functional group to the pH of the solution. If the pH is below the  $pK_a$ , the group will be largely protonated; if the pH is above the  $pK_a$ , the group will be largely deprotonated, and if the pH is near the  $pK_a$ , the group will exist as a mixture of the two forms.

The **charge** on a functional group cannot be directly predicted from the protonation state alone, because the charge is dependent on the type of functional group. However, the charge on a molecule will always become more positive as pH decreases.

Amino acids are always ionized in aqueous solutions. However, depending on the pH, the net charge on the molecule may be zero. The pH at which the net charge on a molecule is zero is called the **isoelectric point** for the molecule or the pI.

The  $pK_a$  value for a particular type of functional group can vary, depending on the local environment. This is somewhat evident when comparing  $pK_a$  values for carboxylate groups in different free amino acids. It is very important in considering protein structure and function.

Four amino acid side-chains are nearly always ionized under physiological conditions, with two being negatively charged (glutamate and aspartate), and two being positively charged (lysine, and arginine). Depending on its local environment, a histidine side-chain may be protonated (and therefore positively charged) under physiological conditions. Somewhat less commonly, cysteine side-chains may become deprotonated (and therefore negatively charged), under physiological conditions.