

2. FROM AMINO ACIDS TO THE PRIMARY STRUCTURE OF PROTEINS

As just about any biochemistry textbook will tell you, the word protein comes from the Greek word "πρωτειν" which means first. Well, that's not strictly true. So far as we know, RNA preceded protein on the evolutionary scene. Even today, the central dogma of molecular biology places protein last in the chain of information from DNA to RNA to protein. However, if you wanted to pick a class of molecules that were of the first importance in producing life as we know it today, proteins would certainly top the list. Later chapters will focus on how the structures of certain proteins contribute to their functions, but for now we're going to be content to take a bottom up approach in explaining how the structure of proteins can be tied to the structure of their building blocks, the amino acids. There are twenty genetically encoded amino acids¹ that are incorporated into growing protein chains on the **ribosome**, the cell's protein synthesis factory. The diversity of the amino acid building blocks, layered on top of some of their common structural features, is responsible for the diversity in protein structure itself, so that's where we'll start.

The Twenty Amino Acids

Basic Structure

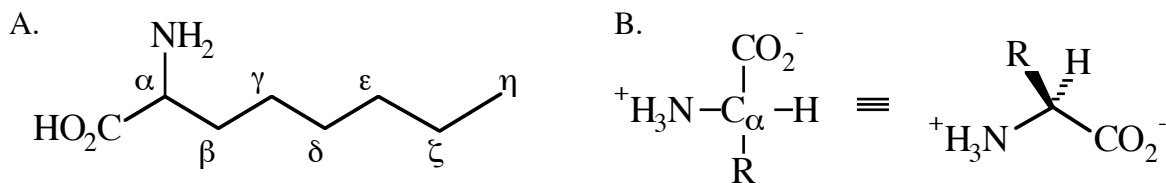


Figure 2.1. (A) The alphabetic (Greek) labeling of carbons in an alkanic acid, showing an amino group at the C_{α} . (B) A Fischer projection showing the L-configuration of the naturally occurring amino acids, converted to the standard 3D projection on a 2D surface.

The name "amino acid" fundamentally describes the chemical nature of these molecules; each contains an carboxylic acid function and an amino function (Figure 2.1A). Furthermore, each of the 20 is an α -amino acid, which refers to the position of substitution of the amino group with respect to the carboxylic acid functionality. As shown in Figure 2.1A, each of the carbons in an alkanic acid is given a label from the Greek alphabet, denoting that carbon's distance from the carboxyl carbon. The α -carbon (C_{α}) of an amino acid is directly adjacent to the carboxyl group, and is the position of attachment for the α -amino group. Among the twenty, there is an

¹Actually, there are 21 if you include selenocysteine, but that's a separate topic entirely.

additional "R" group, or side-chain, attached at C α that renders it a chiral center. So we add on another label, and specify the naturally occurring amino acids as α -L-amino acids. The "L" appellation for these amino acids refers to a specific chiral configuration according to Fischer's nomenclature, which is shown in Figure 2.1B.² The common chiral configuration of the twenty amino acids is essential for providing regular structural features as we'll see, but the reason for the choice of "L-" is a mystery, if one even exists. Amino acids isolated from carbonaceous meteorites (which are thought to be abiotic in origin) are racemic, so somewhere along the road, a (perhaps fortuitous) choice was made by the first common ancestor of all existing life forms to use this set of stereoisomers.

The structures of the twenty amino acids are shown in Figure 2.2. It would be hard to overemphasize the importance of these structures in understanding structural biochemistry. Just as the 26 letters of the alphabet are required to construct the variety of words in the English language, the 20 amino acids provide the fundamental alphabet for protein biochemistry. Included in Figure 2.2 are three and one letter codes for each amino acid, which along with the molecular structures, **must be committed to memory**. The categories used here to segregate the twenty by their chemical character is somewhat arbitrary, since in a few instances, one amino acid might be placed in any of a number of the boxes. Tyrosine, for example, which has a 4-hydroxybenzyl side chain is generally described as a non-polar aromatic amino acid - which is in part true. The phenyl ring contains a lot of non-polar surface area which is important to the role of Tyr in many proteins. However, it is also a polar non-charged amino acid, in that the hydroxyl group at C ζ is capable of acting as a hydrogen bond donor or acceptor. And finally, tyrosine might also be classified as an acidic amino acid. The phenolic hydroxyl group has a pK $_a$ of about 10, which means that the side chain could act as an acid under certain conditions. That's why the "classic" categories can be misleading. Often a given side chain may have several different characteristics, of which one or more may contribute to its function in a protein. In the following sections, we'll individually address each of the broad characteristics by which amino acids are judged.

Hydrophilicity/Hydrophobicity

In Chapter 1, we looked at electrostatic and entropic contributions to intermolecular complexes. In comparing the basic amino acid structures (Figure 2.2), one of the chief means of categorizing the twenty amino acids relies on their principal mode of interactions with other molecules. For example, the character of the side chains of amino acids like serine, asparagine, and aspartic acid is dominated by their polar functional groups: the hydroxyl, amide and carboxylate groups respectively. On the other hand, amino acids like valine, leucine and phenylalanine have non-polar aliphatic and aromatic side chains, which restricts their enthalpic

²An observant eye will note that Figure 2.1B shows the amino acid in a doubly ionized form, reflecting the predominant protonation states of the basic amine and acidic carboxylic groups at pH 7. This ionization state is referred to as a **zwitterion**, indicating that it's a neutral species, with paired and opposing charges within the molecule. As we'll discuss shortly, the acid-base chemistry of the amino acids is one of the keys to their structural and functional roles.

contribution in intermolecular contacts to relatively weak vdW forces, but provides opportunities for strong interactions through the hydrophobic effect. However, most amino acids' side chains have both hydrophobic and hydrophilic character - they are **amphiphilic**. For example, threonine has both a hydrophilic hydroxyl functionality and a hydrophobic methyl group, both in the γ position. Likewise, lysine, which is categorized as a basic amino acid, has four methylene groups in its side chain, which is equivalent in non-polar surface area to amino acids like leucine and methionine. Also note that two of the three aromatic amino acids, tyrosine and tryptophan, have polar functionalities combined with large non-polar side chains.

So, can one definitively categorize an amino acid side chain as hydrophilic or hydrophobic? No. Instead, the issue is typically addressed by assessing the *relative* solubilities of the amino acid in water and some non-polar solvent, like octanol or cyclohexane. Table 2.1 presents some statistics on the relative contributions of polar and non-polar groups to the total surface area of each amino acid as well as the change in free energy associated with transfer of a side chain analog (where the C_{α} atom is replaced by a hydrogen atom) from water to cyclohexane.

Table 2.1. The total surface area and polar surface area of the side chains of the twenty amino acids are compared. $\Delta G_{\text{transfer}}$ refers to the free energy of transfer from water to cyclohexane. Note that the greater the polar surface area of the side chain, the more positive the free energy of transfer.

Amino Acid	Total Surf. Area of side chain (\AA^2)	Polar Surf. Area in the Side chain (\AA^2)	$\Delta G_{\text{transfer}}$ of side chain (kcal/mol)
Alanine	113	-	-0.87
Arginine	241	107	15.93
Asparagine	158	69	5.22
Aspartate	151	58	9.71
Cysteine	140	69	-0.34
Glutamine	189	91	6.51
Glutamate	183	77	7.78
Glycine	85	-	0
Histidine	194	49	5.63
Isoleucine	182	-	-4.00
Leucine	180	-	-4.00
Lysine	211	48	6.52
Methionine	204	43	-1.42
Phenylalanine	218	-	-2.05
Proline	143	-	
Serine	122	36	4.36
Threonine	146	28	3.53
Tryptophan	259	27	-1.40
Tyrosine	229	43	1.09
Valine	160	-	-3.11

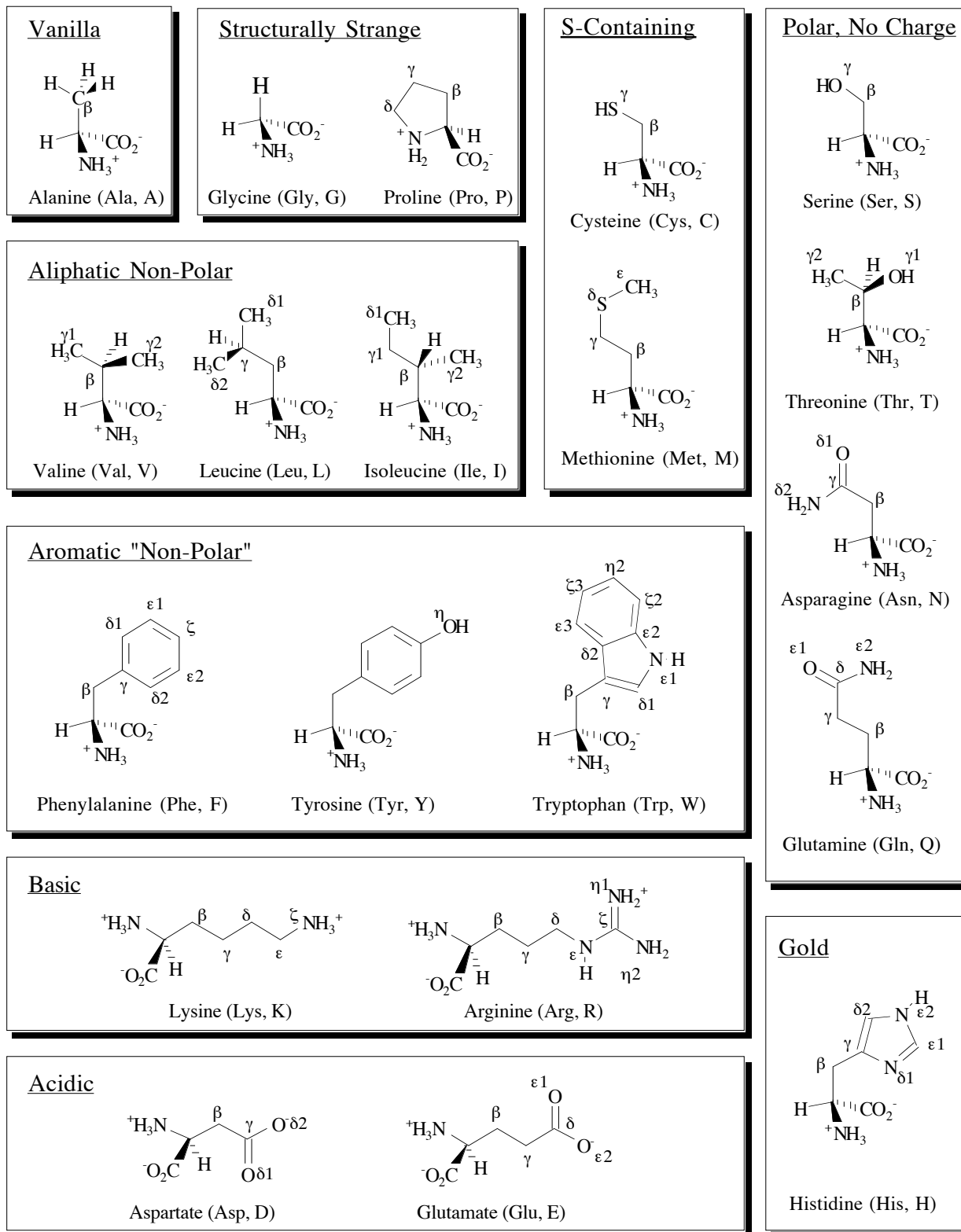


Figure 2.2. Structures of the 20 amino acids.

Acid-Base Chemistry

Another of the most fundamental aspects of amino acid chemistry relates to their acid-base chemistry. Equations 2.1-2.4 provide a brief review of some fundamental descriptors used in describing the properties of acids and bases.

$$K_a = \frac{[H^+][A^-]}{[HA]} \quad (\text{Eq. 2.1})$$

$$pK_a = -\log_{10} K_a \quad (\text{Eq. 2.2})$$

$$pH = -\log_{10}[H^+] \quad (\text{Eq. 2.3})$$

$$pH = pK_a + \log_{10} \frac{[A^-]}{[HA]} \quad (\text{Eq. 2.4})$$

Most importantly, the transfer of a proton between donors and acceptors is an equilibrium process (Eq. 2.1), and each acid has a defined acid dissociation constant (K_a) that corresponds to the transfer of a proton from the acid (HA) to water. Defining pH and pK_a in equations 2.2 and 2.3, one obtains the Henderson-Hasselbalch equation (Eq. 2.4), which provides a simple relationship between pH and pK_a . Consider the following situation: a weak acid of pK_a 4.0 is dissolved in a buffer at pH 7.0. Solving for the ratio of $[A^-]/[HA]$, one finds that there is a 1000 fold higher concentration of the conjugate base, A^- , in solution than the conjugate acid HA. The higher the pH of the solution, the less of the conjugate acid than will be found relative to the conjugate base. At neutral pH, acids with pK_a 's below 7 will be found predominantly in their conjugate base form, and acids with pK_a 's above 7 will largely be found in the protonated (conjugate acid) form.

Table 2.2 The pK_a 's of some functional groups relevant to the acid base chemistry of amino acids.

Molecule	Functionality	pK_a
Acetic acid	Carboxylic acid	4.7
Methylammonium ion	Ammonium group	10.6
Glycine	α -Ammonium group	9.6
Glycine	α -Carboxylic acid	2.3
Aspartic acid	γ -Carboxylic acid	4.0
Glutamic acid	δ -Carboxylic acid	4.4
Histidine	Imidazolium group	6.8
Cysteine	Sulfhydryl group	8.0
Tyrosine	Phenolic hydroxyl group	10.2
Lysine	ϵ -Ammonium group	10.7
Arginine	Guanidinium group	12.0

From Figure 2.1B, it should be apparent that amino acids readily react as acids and bases. In aqueous solution, only a miniscule fraction of the total number of dissolved molecules in fact exist as "amino acids". Perhaps a more appropriate name might be "ammonium carboxylates". The α -amino group is basic; its conjugate acid, the ammonium group, has a pK_a of 9.6 in free glycine, while the pK_a of the carboxylic acid is 2.3. From the Henderson-Hasselbalch equation (Eq. 2.4), it can be calculated that at pH 7, less than one percent of the amine group will be neutral, and only one part in 5000 will be in the acid form. Similarly, many of the other amino acid side chains are ionized at neutral pH (Table 2.2). In most instances, the pK_a 's of the side chains dictate that only one protonation state will predominate at pH 7. For example, the "acidic" amino acids, Asp and Glu, are typically found as their conjugate bases, whereas the basic amino acids, Arg and Lys, typically exist as their conjugate acids.

Histidine, on the other hand, is of interest because its pK_a (6.8) dictates that roughly equal fractions of the side chain will exist in the conjugate acid and base forms simultaneously. This makes histidine a particularly valuable amino acid (hence "gold") in that its conjugate base form³ represents the strongest base likely to be found in any abundance at neutral pH, while its conjugate acid is similarly the strongest acid to be found at high concentration at pH 7. Thus, while His turns out to be the least common amino acid found in proteins, it proves to be a frequent contributor to their function given its unusual properties. It may be used sparingly, but it contributes importantly to protein function in many instances.

Something that you may have noticed is that, despite a common structure, the carboxylic acid groups of glycine and acetic, aspartic and glutamic acids have a wide range of pK_a values, between 2.3 and 4.7. This is the result of differences in chemical environment. For example, the ammonium group of glycine acts as an electron withdrawing group which decreases the pK_a of the carboxylic acid, just as trifluoroacetic acid is a much stronger acid than plain old acetic acid. (In addition, the ammonium group provides enthalpic stabilization of the carboxylate form of the acid via an ion-ion interaction.) This effect shifts the equilibrium in favor of the carboxylate, yielding a relatively low pK_a of 2.3. Similar arguments can be made for the lower side chain pK_a of the α -ammonium group of glycine relative to methylammonium ion. It's also worth pointing out that some chemical environments favor the neutral species. For example, the pK_a of acetic acid changes dramatically depending on the solvent in which it is dissolved (Table 2.2). In solvents that are less polar than water, the conjugate base, acetate, receives less enthalpic stabilization and so is higher in free energy relative to the acid. Thus, the pK_a of acetic acid is raised in methanol relative to H_2O .

³Note that, like carboxylic acids, the imidazole side chain of histidine can exist in two tautomeric forms, in which the N-H bond can be on either the δ or ϵ nitrogen.

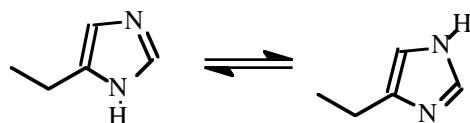


Table 2.3 The pK_a of acetic acid in various solvents.

Solvent	pK _a
Water	4.7
Methanol	9.6
Dimethylsulfoxide	12.6

All of this is merely a means of warning you that the pK_a of a given functional group is not an absolute. It varies substantially with its immediate chemical environment.

Redox Chemistry

While it is not a predominant feature of amino acid chemistry, redox behavior is associated with the two sulfur-containing amino acids, cysteine and methionine. From a biochemist's point of view, this possibility is more often an obstacle in working with a protein than a feature that reveals inherent protein function (though it is that too). The interior of the cell has a reducing environment, assuring that sulfur containing amino acids will remain in a reduced state. However, when a protein is isolated from cells it is exposed to the oxidizing environment of the atmosphere, which is 18% oxygen, unless precautions are taken. The oxidation of cysteine or methionine can be disastrous for a protein, since the resulting side chains are more polar than those of the starting materials (Figure 2.3). The cell is able to maintain strong reducing conditions by keeping high concentrations of glutathione, a molecule that contains a cysteine side chain. These free thiol groups can be used to scavenge oxygen, and even to reduce oxidized amino acids. In solution, β-mercaptoethanol and dithiothreitol, which each contain free sulfhydryl groups, are added for the same purpose.

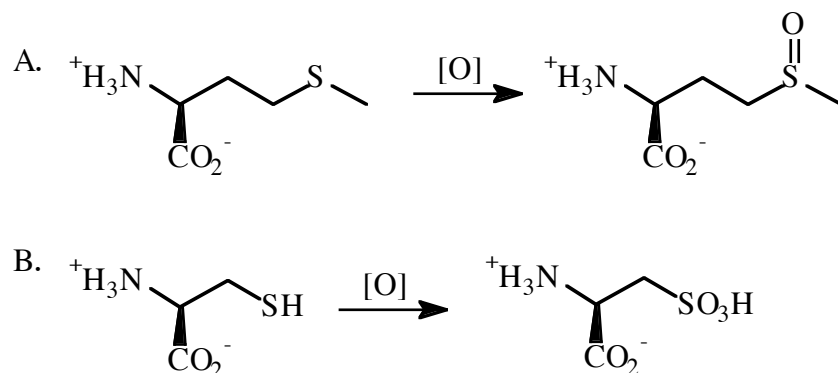


Figure 2.3 (A) The oxidation of methionine to its sulfoxide form. (B) The oxidation of the cysteine thiol to a sulfonic acid. Note that at pH 7, the sulfonate form would predominate.

On the other hand, the redox chemistry of cysteine does occasionally play a significant role in protein structure and function, usually through the oxidative formation of a disulfide bridge with another cysteine side chain. The resulting molecule is called **cystine**, which possesses a covalent linkage between the two sulfur atoms (Figure 2.4). Cystine turns out to be an important contributor to a number of extracellular proteins, since it can be used to cross-link different protein molecules together and even to provide crosslinking within a protein (as in the extracellular proteases trypsin and chymotrypsin). As with acid-base chemistry, redox chemistry is a reversible process with an equilibrium constant that varies depending on local conditions. The oxidizing half-reaction for the production of cystine from cysteine is sufficiently favorable that it can be coupled to the reduction of a variety of other molecules. For example, Hg^{2+} is reduced by cysteine side chains to produce $\text{Hg}_{(l)}$ and cystine.

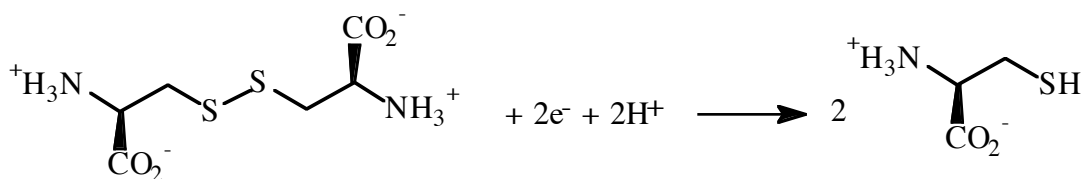


Figure 2.4. The half reaction corresponding to the reduction of cystine to cysteine. The standard reduction potential at pH 7 is -0.22 V.

Spectroscopic Properties of the Aromatic Amino Acids

Although the UV spectra of Phe, Tyr and Trp aren't of any significance to the function of amino acids in proteins, they are a useful diagnostic tool to be used in handling proteins. Some spectroscopic data for the three aromatic amino acids are given in Table 2.4. Of primary interest are the absorbance data. Tryptophan and tyrosine both absorb strongly near 280 nm, while phenylalanine absorbs more weakly with a maximum near 260 nm. Commonly, protein is identified by its absorbance at 280 nm thanks to the contributions of Trp and Tyr, and estimates of protein concentration can be made based on the absorptivity of protein solutions. As will be seen in Chapter 3, the fluorescence properties of Trp and Tyr are important in structural studies. Both absorbance and emission spectra can be used to learn about the chemical environment of an amino acid side chain.

Table 2.4. Spectroscopic properties of the aromatic amino acids.^a

Amino Acid	UV Absorbance		Fluorescence Emission	
	λ_{max}	ϵ ($\text{M}^{-1}\text{cm}^{-1}$)	λ_{max}	Quant. Yld.
Phe	257 nm	197	282	0.04
Tyr	275 nm	1420	303	0.21
Trp	280 nm	5600	348	0.20

Covalent Structure in Proteins

Peptide Linkage and Primary Structure

Amino acids are the building blocks of proteins. This simple analogy was used earlier in this chapter to rationalize the study of amino acid chemistry. Now we need to look at how proteins are constructed from amino acids. Proteins are linear polymers with amino acids acting as the monomers that combine to form the chain. The chemical linkage that holds the protein together occurs between the carbonyl carbon of one amino acid and the α -amino group of an adjacent amino acid - an amide bond. In proteins, the amide linkage is referred to as a **peptide bond** (Figure 2.5A). In organic chemistry, an amide is formed by the reaction of an amine and an activated carboxylic acid derivative, such as an ester or an acyl halide. In the cell, the peptide bond is synthesized by the ribosome by reacting an ester of one amino acid with the α -amino group of a second (Figure 2.5A). This occurs sequentially, until the full protein is synthesized as per the directions of the DNA sequence encoding the protein (much more on this later). The resulting polymer of amino acid **residues** (note that they are no longer amino acids, since both the amino and acid groups have been lost to other functionalities) is sometimes referred to as a **polypeptide chain**. And since this is a paragraph of jargon, let it be noted that a **peptide** (or oligopeptide) is a term that's used to describe chains of 50 (roughly) or fewer amino acid residues. In addition, a Greek numerical prefix can be used to specify exactly how many residues. For example, a tripeptide has three amino acid residues.

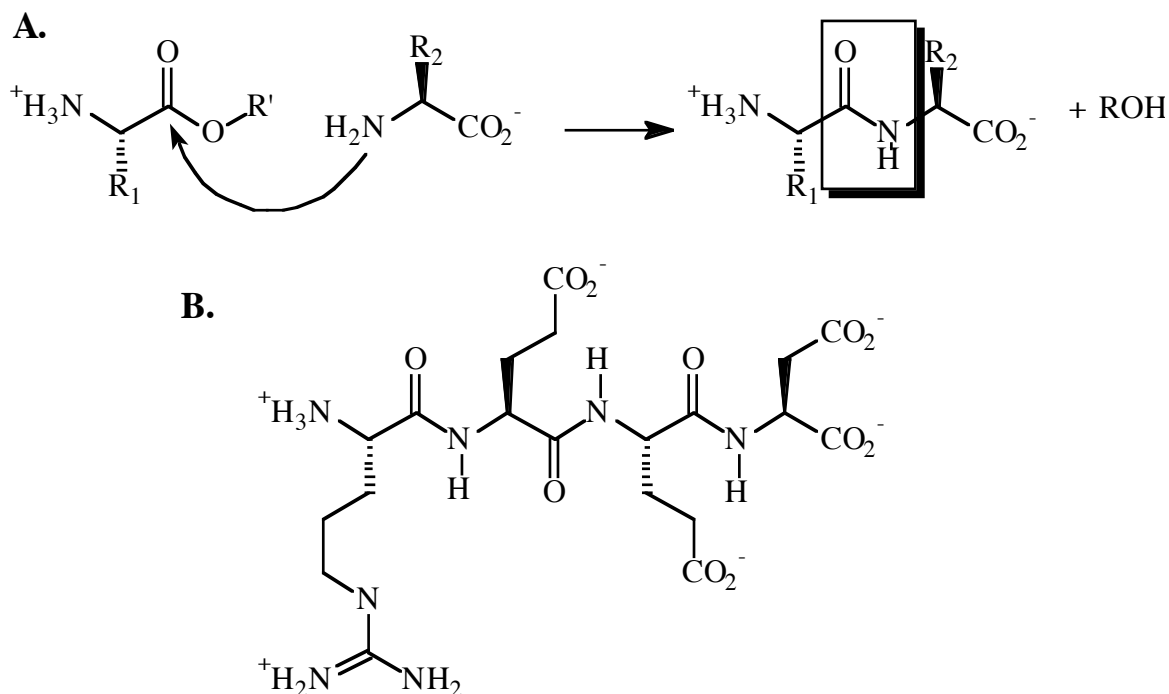


Figure 2.5 (A) The α -amino group of one amino acid condensing with an ester of the α -carboxylate of a second yields a peptide linkage (R' is the 3' hydroxyl of a tRNA molecule). (B) This tetrapeptide has the sequence REED.

The preceding definitions are general references to the peptide bond. We can, however, get more detailed and start to talk about the **sequence**, or **primary structure** of a specific peptide or protein. Any given protein, such as lysozyme or the α chain of hemoglobin, is a chemically defined compound that is most conveniently described by providing the names of the amino acid residues in order as they appear in the chain. By convention, the sequence is read from the amino acid residue with a free α -amino group towards the residue that has a free carboxylate - from the **N-terminus** to the **C-terminus**.⁴ For example, in Figure 2.5B the tetrapeptide's sequence is ArginylGlutamylGlutamylAspartic acid, abbreviated ArgGluGluAsp, or more succinctly, REED.

Disulfide Bonds

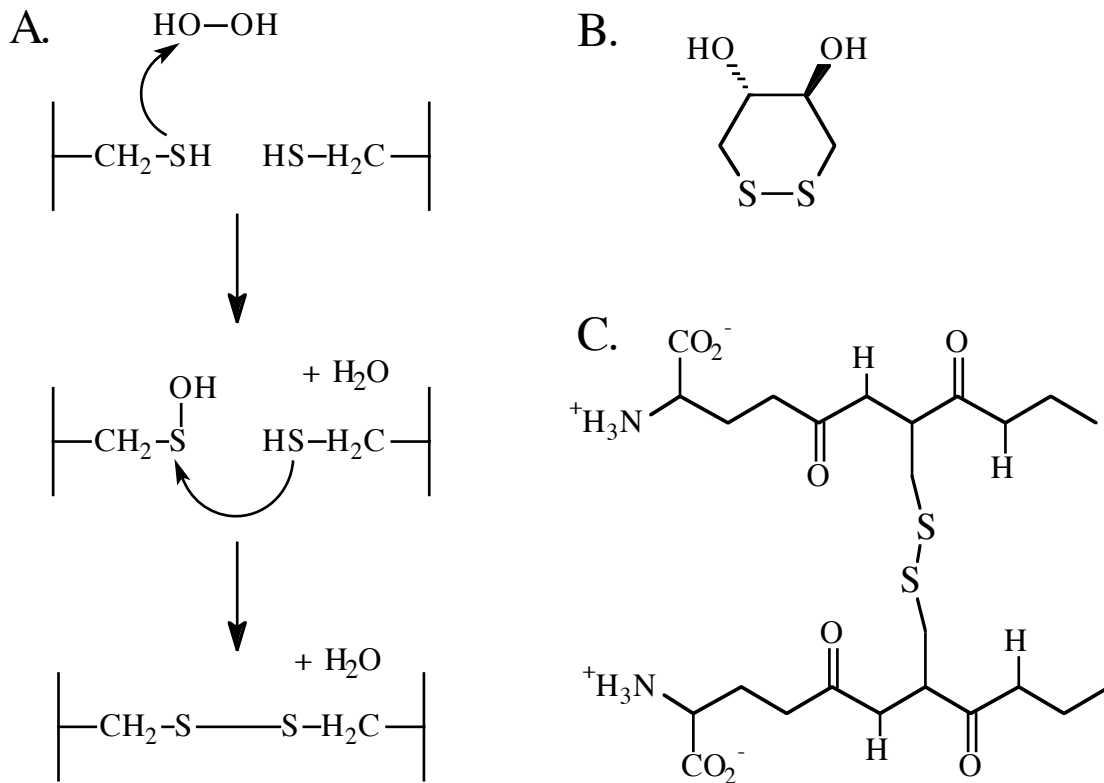


Figure 2.9 (A) The oxidation of two cysteines to a cystine disulfide bridge by hydrogen peroxide. (B) The oxidized form of dithiothreitol. (C) Oxidized glutathione. Both dithiothreitol and glutathione can oxidize other thiols to disulfide linkages.

Proteins are linear polymers of amino acids linked by amide bonds. There is, however, one additional source of covalent structure in polypeptides. Two cysteine residues may be

⁴This isn't totally arbitrary. The N-terminal amino acid is the first one to be put in place during protein synthesis on the ribosome and the C-terminal residue is the last to be added before the polypeptide chain leaves the ribosome.

oxidized to a cystine disulfide bridge, providing a covalent bond between residues that are distant from each other in sequence. The reducing environment of the cytosol (about -0.27 V in *E. coli*) means that most intracellular proteins will contain only reduced cysteine residues. However, disulfide bridges are particularly important in stabilizing the tertiary structures of small monomeric extracellular proteins, such as lysozyme and ribonuclease, which each contain four disulfides. The "hydrogen atom" of protein biochemistry - bovine pancreatic trypsin inhibitor, another secreted protein - contains three disulfides in a 60 residue polypeptide chain.

The formation of disulfides between reduced cysteines requires the participation of an oxidizing agent. For example, the OxyR regulatory protein of *E. coli*⁵ contains a redox active pair of cysteine residues that are selectively oxidized to cystine by hydrogen peroxide (Figure 2.9A). However, the oxidizing agent can be any of a number of species, though other disulfides, such as oxidized dithiothreitol (Figure 2.9 B) or oxidized glutathione (Figure 2.9C) are most commonly used *in vivo* and *in vitro*, respectively.

Primary Structure of Proteins

Recall that the primary structure of proteins is defined by the covalent linkages between aminoacyl residues and can be read as a sequence of residues stretching from the N-terminus to the C-terminus. A simple listing of residues in one or three letter code defines the chemical structure of a given protein, and is an important beginning step in understanding the function of a protein.

Genetic methods

The simplest method for obtaining protein sequence is to sequence the gene that encodes the protein. The "central dogma of molecular biology", a somewhat derogatory term coined by Francis Crick, notes that the information in the sequence of DNA (a linear polymer) is **transcribed** to a related linear polymer, RNA, which is then **translated** to an unrelated polymer, the protein. In the 1960's the combined work of several driven individuals and their hard-working graduate students yielded the "Genetic Code", a decoding device that allows any yokel to convert a given DNA sequence into protein sequence. Since DNA sequencing is now extraordinarily simple and inexpensive to perform, it is by far the most common approach to obtaining sequence information for a new protein, since in many instances the gene (a stretch of DNA) is discovered before the protein. We'll discuss the mechanism of protein synthesis later this semester, but this is enough to know for now.

Chemical methods

Automated sequencing of polypeptides is possible and tedious/expensive (depending on whether you do it, or pay someone else to). The principal method is Edman degradation, which

⁵Zheng, M., Åslund, F. &Storz, G. (1998). Activation of the OxyR Transcription Factor by Reversible Disulfide Bond Formation. *Science* 279, 1718-1721.

pursues peptide sequences by sequential degradation of a peptide from the N-terminal end (Figure 2.10). As each amino acyl residue is cleaved from the N-terminus, a new cycle begins. However, because each step possesses limited efficiency, a sample being treated by digestion gradually becomes heterogeneous as any given peptide molecule in the sample may or may not have reacted fully in the last round, leading to a variety of different N-terminal residues as the sequencing progresses. As a result, this technique is restricted to the determination of peptides less than 30 residues in length. By digesting a large protein selectively into a series of peptide fragment, either chemically or enzymatically, one can sequence the pieces and by comparing overlapping fragments, reconstruct the full sequence of the protein.

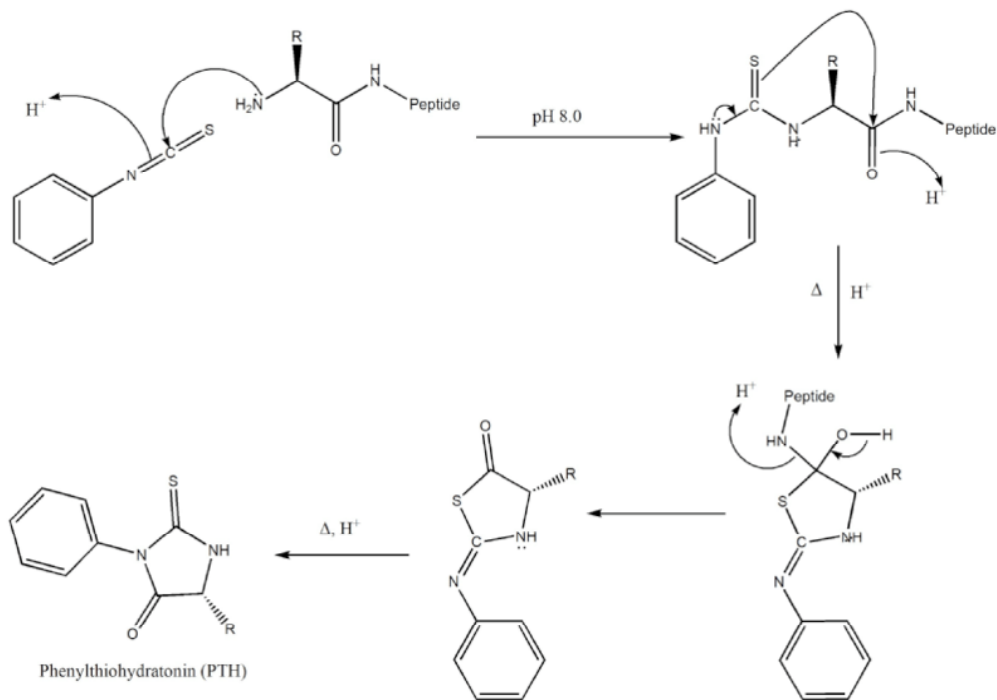


Figure 2.10. One cycle of Edman degradation using phenylisothiocyanate as a derivatizing agent. The N-terminal amine reacts with the electrophilic carbon, ultimately forming a cyclic product that self-cleaves from the 2nd residue in the peptide chain, freeing a new N-terminus to react in subsequent rounds. (Image taken from Wikipedia – thanks open source!).

Mass Spectrometry

Increasingly, mass spectrometry is used as a tool in identifying the primary structure of proteins. It is an excellent technique for measuring the molecular mass of any given protein, and can additionally with some clever analysis, be used to analyze the sequence of smaller peptides obtained by limited and specific hydrolysis of a large protein. If one knows the sequence of a variety of constituent peptides, it is often possible to reconstruct the full sequence of the protein.

Mass spectrometry measures the mass to charge ratio of an ion. The basic principle is to accelerate an ion through a potential field. The force applied to each ion is proportional to its

charge, and the acceleration will be equal to that force divided by the mass (thanks, Newton). Thus, acceleration is proportional to the mass to charge ratio (m/z). In the simplest kinds of mass spectrometers, known as time of flight (TOF) instruments, one simply measures the length of time it takes for an ion to transit a potential field across a fixed distance, and correlating that time to a reference mass to give a mass to charge ratio for each ion detected. Other mass analyzing techniques, such as sector field and quadrupole mass analyzers, use other geometries of force to sort masses, but the idea is generally the same – to distinguish mass to charge ratio.

The other important technical distinction among mass spectrometers is the ion source technology. All mass spec techniques require ions in the gas phase to operate. Getting proteins into the gas phase is not much different than teaching elephants to fly. The most common techniques used in protein MS are matrix-assisted laser desorption and ionization (MALDI) and electron spray ionization (ESI)⁶. In MALDI, the protein is embedded in a UV-absorbent plastic resin. When struck by a laser, the matrix evaporates suddenly and the protein finds itself in the gas phase – remarkably with a single positive charge overall. The physical reasons for the single charge are unknown, but it makes for simple analysis since the m/z ratio is simply the mass of the protein, plus a hydrogen ion (the source of the single positive charge). ESI is a little more complicated in analysis, because ions of multiple positive charges are produced. A protein solution in acidic organic solvent (often trifluoroacetic acid and methanol) is sprayed into a vacuum. As the drop evaporates, positive charge accumulates on the protein through the addition of H^+ ions. Ultimately a “manifold” of peaks arise in the mass spectrum, each reflecting differences in the m/z ratio due to differences in z (the ionic charge). Additional complexity in the spectrum is also introduced by the isotopic contribution of the protein molecule which becomes significant given the number of atoms in a typical protein molecule.

Protein sequencing by mass spectrometry is typically done by a couple MS/MS technique. Prior to MS, the protein is digested by an enzyme that catalyzes the hydrolysis of a few specific locations in the protein (for example, trypsin is an enzyme that will only cleave peptide bonds directly C-terminal to lysine and arginine). The peptides that result are then separated by liquid chromatography and injected onto a tandem mass spectrometer. In the first mass analysis step, a single mass ion is allowed to pass into a bombardment chamber, where collisions with an inert gas lead to the production of a variety of fragments. Those fragments then pass into a second mass analyzer (hence the name MS/MS). By examining the masses of all fragments that result, which will each vary in mass by the presence or absence of other residues in the peptide sequence, one can by logic reconstruct the sequence of residues in the peptide.⁷

⁶ For a good, if somewhat dated, introduction, see Fenn et al. (1989) Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **246**, 64-71. For a salacious example of a really bad bit of personnel management, the inventor of ESI (John Fenn, Nobel Prize in Chemistry 2002) left Yale at the age of 70 because of required downsizing of his lab. His Nobel Prize winning work took place later at Virginia Commonwealth University. Not all your good ideas come before you turn 25.

⁷ For more details on this technique, see K. G. Standing (2003) Peptide and protein *de novo* sequencing by mass spectrometry. *Curr. Op. Struct. Biol.* **13**, 595-601. For more on the logic of figuring out the sequence, see <http://www.ionsource.com/tutorial/DeNovo/DeNovoTOC.htm>

Summary

- It is important to memorize the 20 amino acids!
- Amino acids can be classified on several grounds. In particular, we looked at the differing intermolecular contacts that could be made by various side chains, acid-base chemistry and redox chemistry.
- The covalent structure of proteins is dominated by the planar peptide bond. As a linear polymer, the polypeptide has a specified sequence of amino acid residues that is read from the amino terminus to the carboxy terminus.
- Disulfide bridges between cysteines provide covalent crosslinks between regions of the polypeptide that are distant from each other in sequence.
- Due to the uniform stereochemical substitution pattern at C_{α} and the planar peptide bond, the polypeptide chain is constrained to certain conformations, severely restricting the flexibility of the backbone.

Further Reading

Thomas E. Creighton, Proteins: Structure and Molecular Properties, 2nd Ed. W. H. Freeman, New York, 1993.

G. E. Schulz and R. H. Schirmer, Principles of Protein Structure, Springer-Verlag, New York, 1979