S. PROTEIN SECONDARY STRUCTURE

In the last section, the covalent structure of proteins was introduced. Amino acid residues are linked by peptide bonds in a particular sequence that defines the chemical identity of the protein. The amino acid sequence of a protein is its **primary structure**. Ultimately, the chemical structure of a protein leads to the formation of a particular three-dimensional conformation from which protein structure springs. At the next level of structure, **secondary structure**, the focus will be on local conformational choices made by the backbone of the polypeptide chain. As it turns out, the conformations adopted by the protein backbone cluster in certain categories, for reasons dictated by steric constraints and favorable intermolecular forces.

The Stability and Conformation of the Peptide Bond

The integrity of the protein depends upon the stability of the peptide bond. Otherwise, the information contained in the primary structure of the protein would be lost upon hydrolysis of the polypeptide to the thermodynamically more stable amino acids (the carboxylate and ammonium groups are preferred over the peptide bond). As it turns out, the peptide bond is extremely resistant to hydrolysis . It is thermodynamically unstable, but kinetically inert. The change in free energy for hydrolysis of the amide bond is -5 kcal/mol, reflecting the stability of the free carboxylate group in comparison to the amide derivative. Despite this instability, the half life of an amide bond is 7 years at pH 7. Typically, the cell will use catalysts called proteases to degrade a protein long before hydrolysis of the polypeptide has become a problem.

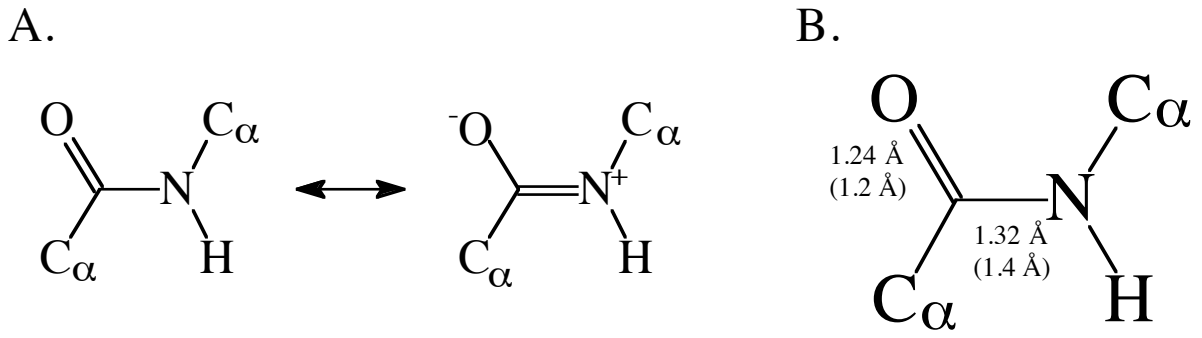


Figure S.1 (A) The resonance structures associated with the peptide bond. The C-N bond has about 40% double bond character. (B) The dimensions of the amide functionality. The carbon oxygen bond is somewhat longer than normal for a carbonyl group (1.24 Å vs. 1.20 Å) and the C-N bond is somewhat shorter than normal (1.32 Å vs. 1.4 Å).

Aside from its unusual resistance to hydrolysis, the amide bond is notable for possessing substantial double bond character between the carbonyl carbon and amide nitrogen. Meanwhile, the bond between the carbonyl carbon and oxygen is unusually long, reflecting weakened double bond character (Figure S.1). The simple explanation for this phenomenon is that the lone pair on the amide nitrogen can participate in resonance with the carbonyl group, leading to a three center conjugated π system. Two other significant results derive from this phenomenon.

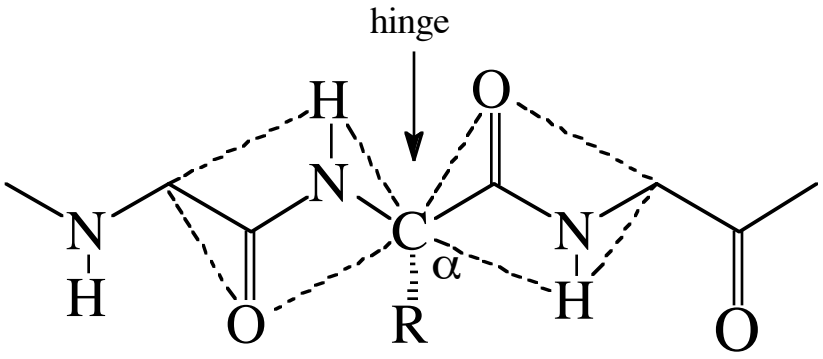


Figure S.2 The dashed lines connect two groups of six atoms flanking the central α-carbon which are held fixed in two "plates" that are hinged at the α-carbon.

The peptide bond is planar. The sp^2 hybridized amide nitrogen fixes its two substituents, the α-carbon of the second residue and a hydrogen, in the same plane as the substituents on the carbonyl carbon - namely the carbonyl oxygen and the α-carbon of the first residue (Figure S.2). This planarity reduces the conformational flexibility at each amino acid's α-carbon (see below), and has been described as creating a chain of plates on hinges. The partial double bond character between the amide nitrogen and carbonyl carbon provides an energetic barrier to the free rotation about the peptide bond of 20 kcal/mol (compared to 90 kcal/mol for a carbon-carbon double bond). Two conformations for the linkage are thus available for a dipeptide. The ***trans*** conformation (Figure S.3A) places the connected α-carbons opposite to each other, while the ***cis*** conformation (Figure S.3B) has the α–carbons placed in close proximity to one another. Because of the increased steric conflict between α-carbons in the *cis* conformation, it is rarely found in proteins, except in peptide linkages where proline, with its tertiary amide, is the C-terminal residue (Figure S.3C). For these peptide bonds, the *trans* isomer is favored by 2 kcal/mol, a ratio of thirty to one.

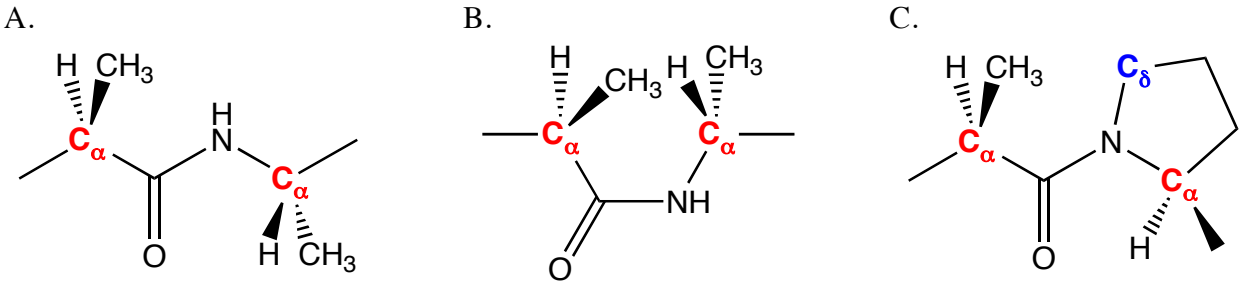


Figure S.3 (A) A *trans*peptide bond, placing the α-carbons 180° apart. (B) A *cis*peptide bond. Note that the C_α's are directly adjacent to one another. (C) A *trans*peptide bond involving proline. Since C_δ of proline is adjacent to the α-carbon of the N-terminal residue, instead of the usual N-H group, the relative stability of the *trans* conformation to the *cis* conformation is diminished with proline C-terminal to a peptide bond.

The peptide bond is polar. The second resonance structure of the amide bond (shown in Figure 2.6A) not only contributes to the planarity of the linkage, but also to its polarity. There are unusually large partial charges on the carbonyl oxygen and the hydrogen attached to the amide

nitrogen. The dipole associated with this charge difference is S.5 D (Figure 1.2D). The peptide bond functionalilty therefore makes a large electrostatic contribution to the energy of interactions with other hydrogen bond donors and acceptors.

Conformational Flexibility of the Polypeptide

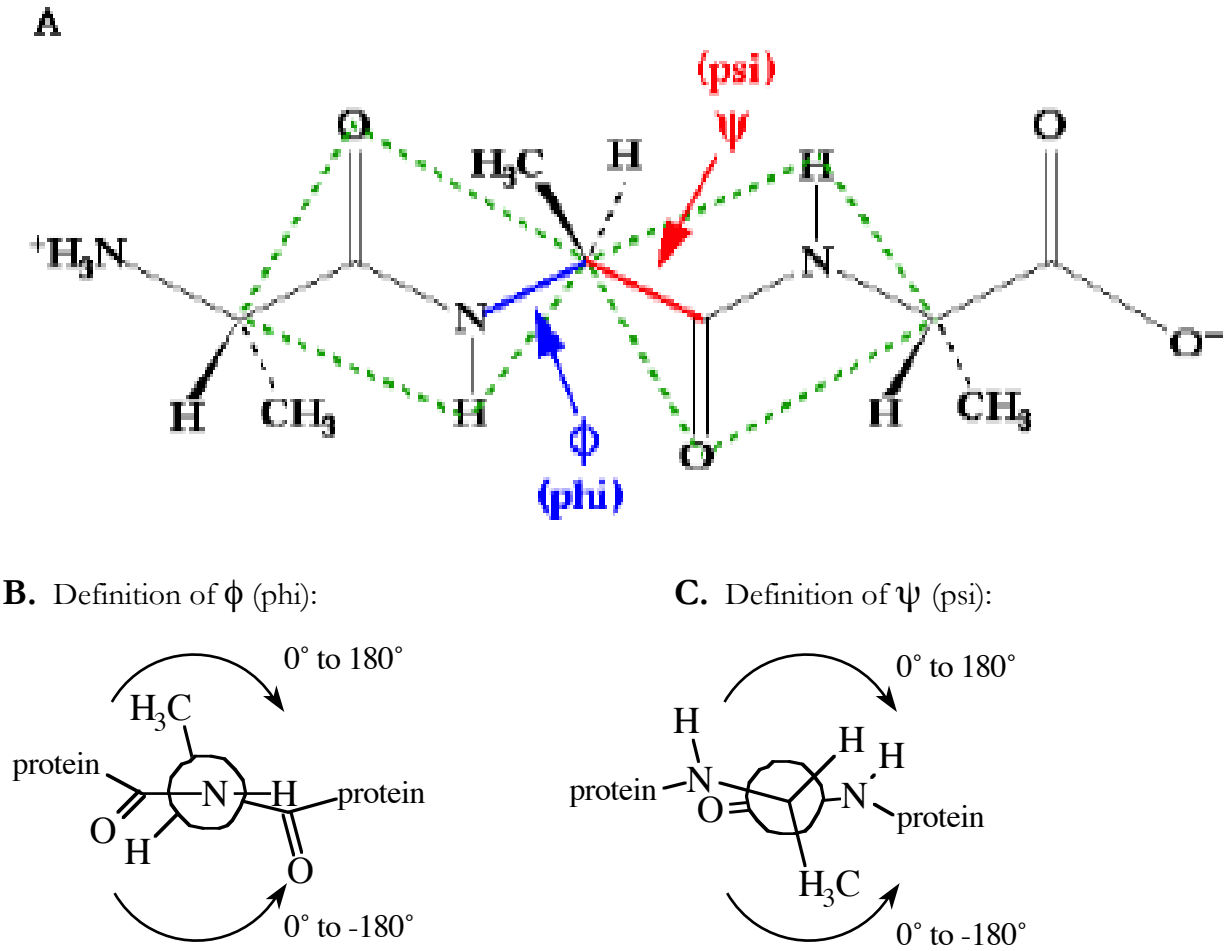


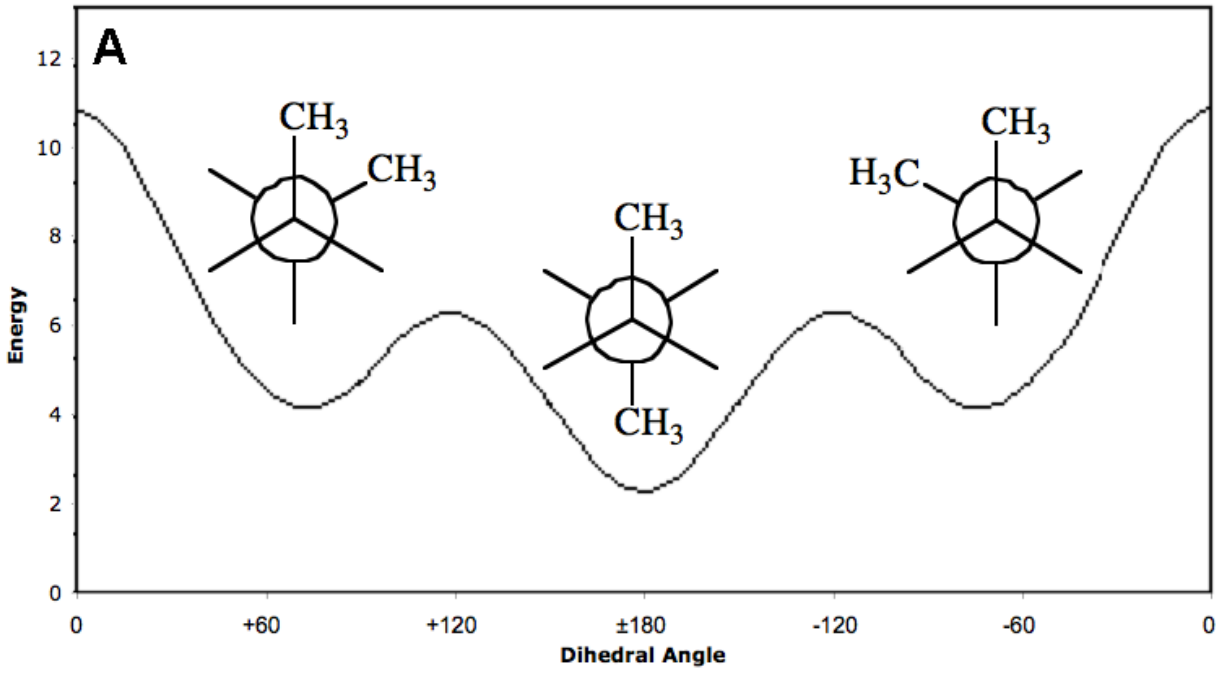
Figure S.4 Rotational freedom along the polypeptide backbone. (A) Structure of the tripeptide AlaAlaAla, showing ϕ and ψ for the central alanyl residue. (B) Newman projection of the dihedral angle ϕ. (C) Newman projection of the dihedral angle ψ.

Much of the structure of proteins can be related to the unique properties of a polymer of α-amino acids. If we ignore, for the moment, the side chains of amino acyl residues in a polypeptide, we see a simple repititious polyamide chain (Figure S.4A). The atoms of the polyamide chain are known as the **polypeptide backbone**, which includes the N, C_α, C and O atoms of each residue (ignoring hydrogens). The conformation of the backbone atoms can be described by three dihedral angles for each amino acid residue. A dihedral angle is defined by the offset of two atoms connected to a central bonded pair of atoms. The two most important angles, defined for each residue, are:

- The angle ϕ (phi) is defined by the offset of the carbonyl carbon atom of the n^{th} residue from the carbonyl carbon of the $(n+1)^{th}$ (Figure S.4B).
- The angle ψ (psi) is defined by the offset of the backbone nitrogen of the n^{th} residue from the nitrogen of the $(n+1)^{th}$ residue.

The third (called ω), related to the offset between C_α's associated with a single peptide linkage, is relatively uninteresting. As mentioned earlier, the peptide bond is restrained to a planar conformation and can only be found in either the *cis-* or *trans-* conformations, 180° apart from one another. Therefore, really only two dihedral angles, ϕ and ψ, are necessary to describe the local conformation of the polypeptide backbone at a given residue.

Even given the relatively fixed conformation of the peptide bond, if ϕ and ψ could each adopt any angle between -180° and +180° for each residue, then there would still be enormous conformational range allowed to the backbone with each added residue. Each residue could adopt 360 x 360 different conformations separated by 1° steps in each angle. However, it turns out that the nature of the substitution pattern at the alpha carbon restricts the number of likely pairs of ϕ and ψ. As one rotates about these two angles, the substituents on the central two atoms alternate between eclipsing and staggered conformations, leading to alternating maxima and minima in the potential energy surface. To understand this phenomena, consider the simpler case of butane, focusing on the dihedral angle about C2-C3 (Figure S.5A). In this system, there are local maxima in potential energy when eclipsed conformations result (at 0°, +120 and -120°), while local minima result at staggered conformations (+60°, ±180° and -60°.) The global maximum in potential energy arises when the two methyl groups eclipse each other at 0°, and the global minimum occurs when the methyl groups are *trans-* to one another at 180°. The preferred conformation of butane therefore is the *trans-* conformation, with other possible conformers sampled in relation to the energetic differences between them.



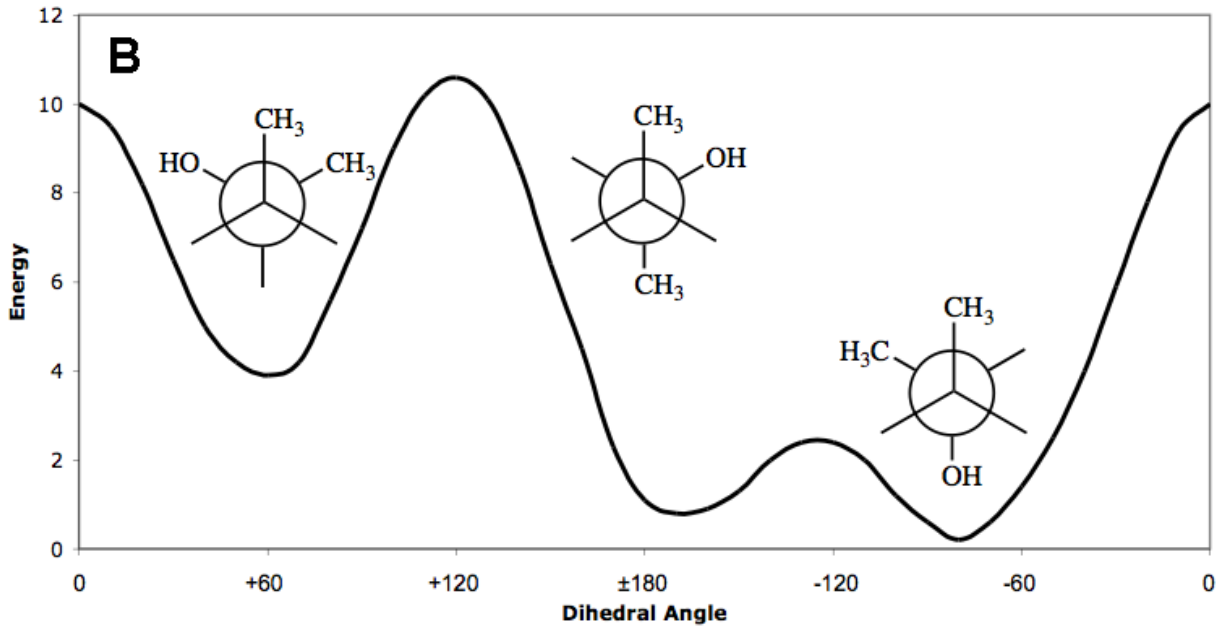


Figure S.5 Plots of energy vs. the central dihedral angle of (A) butane and (B) *S*-2-hydroxybutane. Note that maxima occur for eclipsed conformations, while minima accompany staggered conformations.

Moving to a slightly more complex case, consider an asymmetric molecule, such as *S*-2-hydroxybutane (Figure S.5B). A chiral compound such as hydroxybutane lacks the mirror plane of symmetry found in butane itself, and thus possesses a less symmetric energy profile with respect to the central dihedral angle. This asymmetry arises because there are three different staggered conformations in hydroxybutane, while two of the three in butane are energetically equivalent (at +60 and -60°). This asymmetry indicates that hydroxybutane is likely to reside in a set of conformations between -60° and -180°, perhaps occasionally sampling the less favorable conformation at +60°. Thus the conformational space of hydroxybutane is more constrained than it is for butane. That constraint within the polypeptide backbone as well.

Returning to the case of a polypeptide backbone, there will be global minima and maxima associated with rotation about ϕ and ψ. In Figure S.6, ϕ and ψ are alternately fixed (S.6A and S.6B) while the value of the other dihedral is varied. Unlike the plot of potential energy for butane, these plots are asymmetric, reflecting the asymmetry of the chiral center at the alpha carbon. Because two of the three rotating substituents on C_α are large (the R group and the opposite polypeptide chain) in comparison to the C_α's H atom, there is a distinct preference for conformations where the hydrogen atom is either staggered or eclipsed with respect to the large groups associated with the α-amino group (ϕ) or with the carboxyl group (ψ).

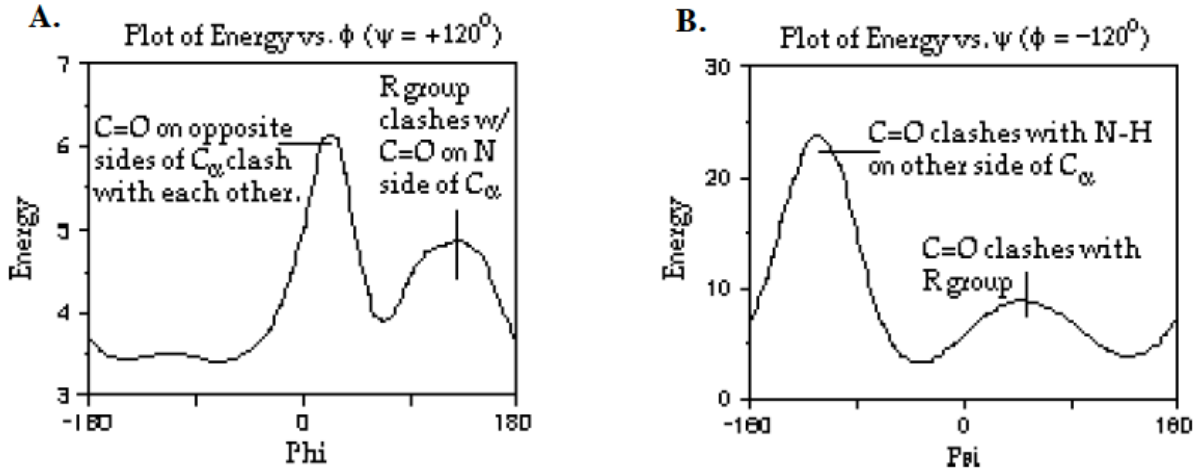


Figure S.6 (A) Plot of energy vs. ϕ while ψ is held at 120°. Note that local maxima occur when carbonyl groups are eclipsing (ca. +60°) and when the R group eclipses the carbonyl group N-terminal to the central C_α (+120°). (B) Plot of energy vs. ψ while ϕ is held at -120°. Local maxima occur when the carbonyl group clashes with the α-amino group (-120°) and when it clashes with the R group (+60°).

Since each of these two dihedral angles can rotate independently, each of the plots in Figure S.6 would have to be repeated for every value of ψ (in the case of S.6A) or ϕ (in the case of S.6B). Instead of displaying a large number of two-dimensional plots, we can simplify the accumulated information from these plots in a single three-dimensional plot (Figure S.7). This plot places ϕ and ψ on the x and y axes, and plots energy on the z-axis. Rather than trouble ourselves with a complicated 3D representation, we can get by with a contour plot similar to those used in topographical maps, where altitude is represented by contours at fixed altitude intervals. This representation is referred to as the **Ramachandran Plot** in honor of the Indian biochemist who first analyzed polypeptide conformation in this manner.[1]

[1]G. N. Ramachandran and V. Sasiskharan (1968) Adv. Protein Chem. 23, 283-437.

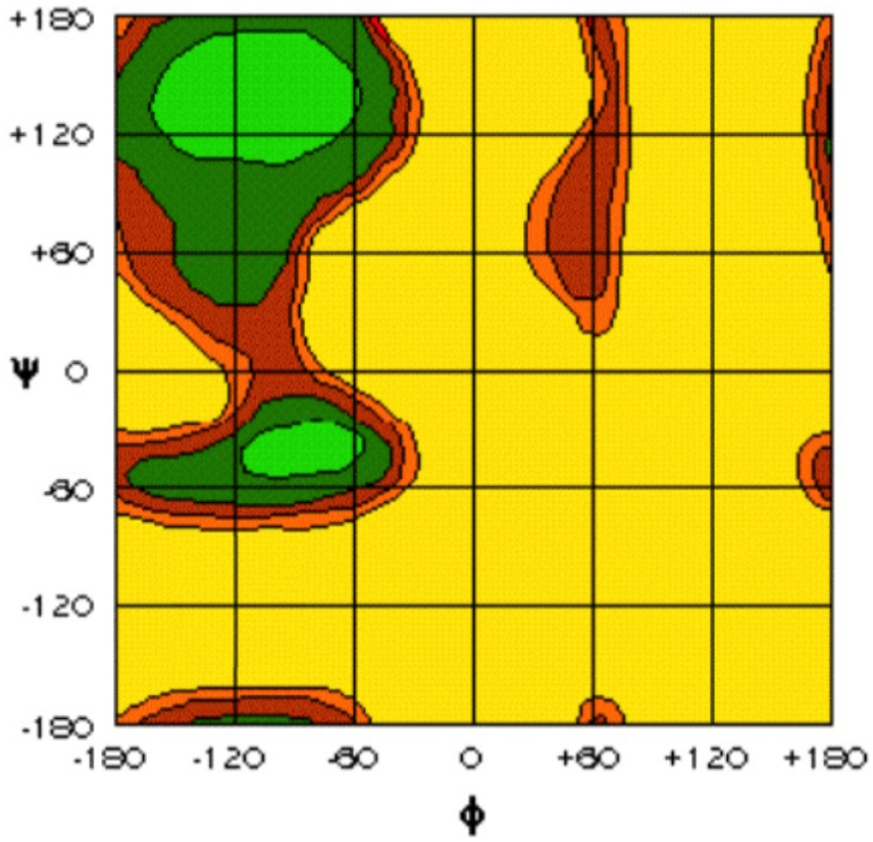


Figure S.7 The Ramachandran plot. Using $\phi = 0^\circ$, $\psi = 0^\circ$ as a reference point of high potential energy (unstable conformation), contours are drawn to demonstrate regions in which the energy drops below this high energy plateau. Local minima occur at about $\phi,\psi = (-120^\circ, +120^\circ)$ and $\phi,\psi = (-60^\circ,-60^\circ)$. A slightly less important minimum is found at $\phi,\psi = (+60^\circ, +60)$.

Inspection of the Ramachandran plot notes that there is a long, narrow valley of stable conformations associated with a given residue in a polypeptide chain. This occurs at $\phi = -60^\circ$ to about -120° (the narrow component), while ψ may vary between -60° to about $+180^\circ$ (the long component). Two deeper wells appear in this profile, labeled α and β. The importance of these wells will be described in the following sections. In the meantime, one should note that the polypeptide backbone is substantially limited in the conformations it may adopt. Only a very small region of the 2D space describes local minima, and a still quite-restricted space can be described as favorable. Thus, a potentially flexible polyamide chain can be seen to truly inhabit a sharply constrained set of conformations. Note, however, that the side chain group, R, plays a substantial role in restricting backbone conformation. Remove those groups, as would be the case for a tripeptide of glycine, and one would have a substantially more flexible backbone. In fact, the presence of glycine in a polypeptide substantially increases the chain's flexibility. On the other hand, a tripeptide of proline, which has a covalently constrained N-C_α dihedral (ϕ), has even more restricted conformation. In Figure 2.2, it was noted that proline and glycine are a structurally distinct pair of amino acids – they can have substantial impact on the conformation of the polypeptide, and therefore on the protein molecule as a whole.

Secondary Structure and Conformation

Secondary structure can be defined as the local conformation adopted by a polypeptide backbone, purposefully excluding more global issues of three-dimensional shape that can be found in a 300+ residue protein. As seen in Figure S.7, the polypeptide backbone is constrained to adopt only a limited range of conformations at the alpha carbon of each residue. It could be argued that these restraints force proteins to adopt the relatively narrow range of conformations that are observed. However, a different approach was taken historically. In 1948, sitting in a hotel room with the flu, Linus Pauling sought out structural explanations for the x-ray diffraction patterns observed from strands of keratin (wool/hair/etc.) Two structures are revealed from those early diffraction patterns, measured by William Astbury in the 1930's: α-keratin is present in unmodified fibers, but β-keratin forms from wet fibers that are stretched. Pauling made some simple models of amino acid residues out of cardboard and attempted to arrange these residues so that hydrogen bonding took place between nearby residues. He was looking for conformations of the peptide backbone that would be *favored* by stabilizing intermolecular forces rather than ones that would not be *disfavored* by steric clashes. In that effort, Pauling discovered the two most predominant conformations found in protein backbones. These common secondary structure elements are the **α-helix** and **β-strand**.[2] Each is generated by taking a particular pair of ϕ and ψ angles and repeating them for each residue along a stretch of the backbone. In the following sections, both of these forms of secondary structure will be discussed, along with **β-turns** (a common conformation in proteins that allows the peptide backbone to fold back on itself) and the **collagen triple helix** (a secondary structure peculiar to collagen, like the keratins, a fibrous structural protein).

The Alpha Helix[3]

There was ample evidence prior to Pauling's work that peptides can form helical structures. However, Pauling recognized both the relevance of the hydrogen bond in shaping molecular structure and appreciated the conformational restraints associated with the peptide bond. Guided by a desire to match peptide bonds in donating/accepting pairs, Pauling recognized a general class of structures in which the amide nitrogen of one residue can donate a hydrogen bond "ahead" to the amide carbonyl of a residue further along (towards the C-terminus) on the polypeptide. A helix forms when this donation is repeated sequentially with the same gap between donor and acceptor, creating a pattern in which the polypeptide loops around to bring peptide bonds that are some fixed number of residues apart into close physical proximity of one another.

---

[2] For a short history, with references, on these discoveries, see Eisenberg (2003) *Proc. Natl. Acad. Sci.* **100**, 11207-11210. I also strongly recommend Judson's great history of the birth of molecular biology (and all things structural), *The Eighth Day of Creation*. Buy a copy to read over winter break. You won't be sorry.

[3] In a tip of the hat towards the visual arts, let it be noted that a beautiful statue of the α-helix (created by Julian Voss-Andreae) is available for public viewing at 40th and SE Hawthorne, the childhood home of Linus Pauling. Check it out on your way to the Baghdad Theater sometime, where Pauling worked as an usher.

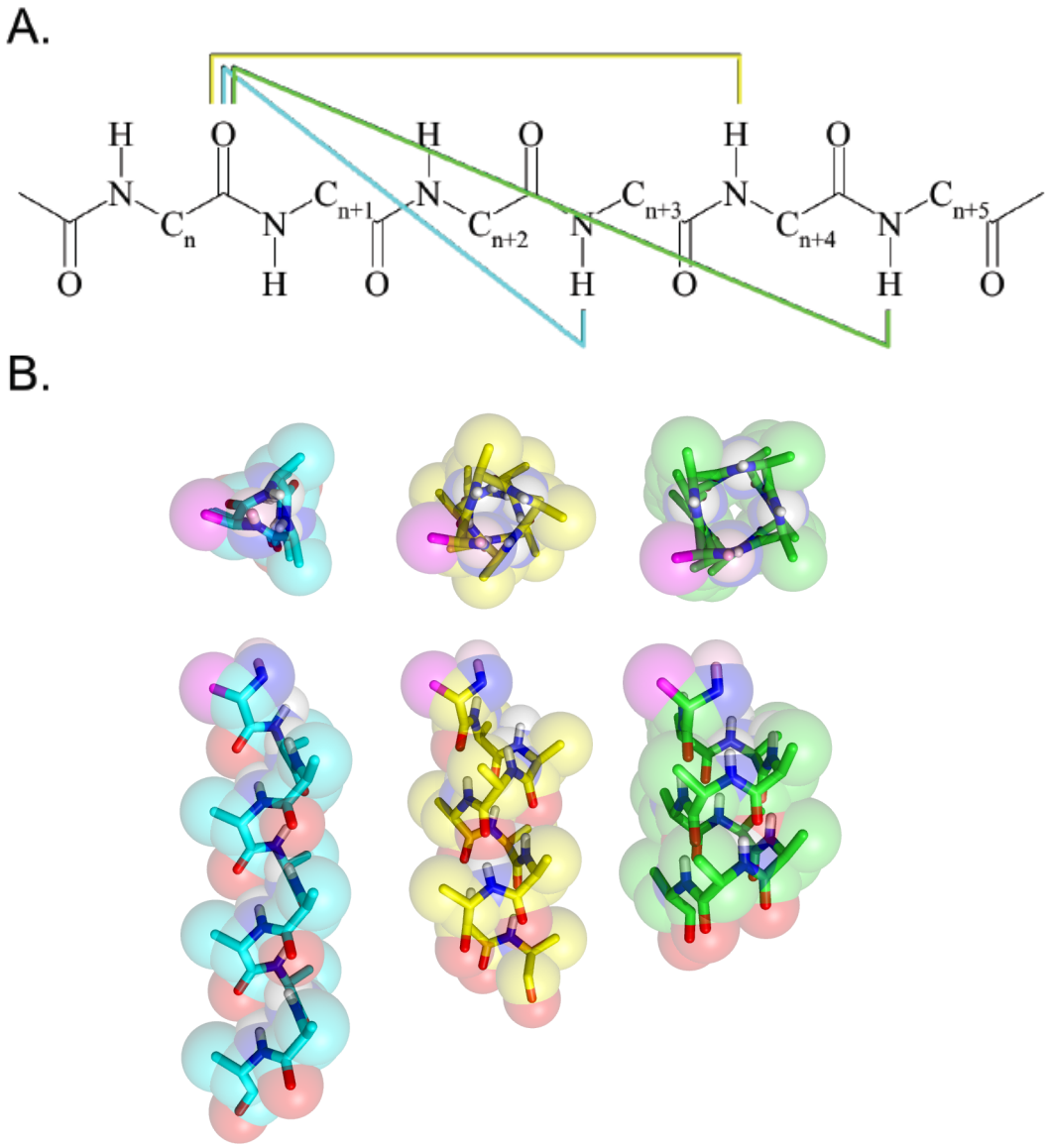


Figure S.8. (A) Hydrogen bonding scheme for 3_{10}, α and π helices in turquoise, yellow and green, respectively. Note that orientation of amide groups is arbitrary in this diagram. (B) Models, from left to right, of the 3_{10}, α and π helices, each 10 residues long, with the same color scheme as in A. On top, the projection is down the helix axis, looking from N- to C-terminus. On the bottom is a side view, with the N-terminus of each helix on top.

The α-helix forms when a section of the polypeptide backbone unites in having the amide carbonyl of the n^{th} residue accept an H-bond from the amide nitrogen of the $n+4^{th}$ residue (Figure S.8), while the 3_{10} helix is formed when the carbonyl of the n^{th} residue accepts an H-bond from the amide nitrogen of the $n+3^{rd}$ residue and the π helix is formed when the n^{th} residue accepts an H-bond from the $n+5^{th}$ residue (Figure S.8A, Table S.1). All three of these helices are **right-handed** helices. The handedness can be observed by directing the thumb of the right hand in the direction along the helix

axis from the N- to C-terminus and tracing the direction of the polypeptide backbone with the index finger of the right hand.

As it turns out, these three helices not only succeed in providing hydrogen bonding partners for each backbone amide group but also allow the backbone to adopt a favorable set of ϕ and ψ angles (Table S.1). Nevertheless, of the three, the α-helix is most commonly found in proteins; roughly 35% of all residues in globular proteins are found in the α-helical conformation.[4] Why? The α-helix is structurally superior to the 3_{10} and π helices from an energetic perspective. At the interior of the helix, the packing of backbone atoms in the core is perfectly attuned to the van der Waals radii of the atoms in the α-helix (Figure S.8B) and on the exterior, the side chains are staggered about the α-helical axis – thus reducing steric conflict. Furthermore, the backbone amide groups are particularly well oriented for hydrogen bonding in the α-helix. As will be discussed later in this chapter, there are other influences that can affect helical stability, but for now it should be sufficient to recognize that the α-helix is so readily formed thanks to its favorable backbone conformation, absence of steric conflict and availability of favorable hydrogen bonding opportunities along the backbone.

Table S.1 Parameters for three helical conformations.

Helix	Average ϕ, ψ	Residues/turn	Rise/residue (Å)	Helix diameter (Å)
α	-57°, -47°	3.7	1.5	12
3_{10}	-49°, -26°	3.0	2.0	10
π	-57°, -70°	4.3	1.2	14

Beta Strands/Sheets

Using opportunities for hydrogen bonding of backbone amides as a guide, Pauling identified a second secondary structure conformation that allows two different polypeptide segments hydrogen bond with each other. (Note that this is distinct from the case of the α helix, in which hydrogen bonds form *within* a polypeptide segment.) The so-called β-strand orients the amide groups in a plane such that hydrogen bonds may form on an axis perpendicular to the orientation of the side chains (Figure S.9). To achieve this orientation of the amide groups, each residue adopts ϕ and ψ angles of approximately -140° and +130°. Note that this conformation is likewise in a low energy region of the Ramachandran plot (Figure S.6).

[4] It has been estimated that roughly 15% of all helices in globular proteins are 3_{10} helices, but they are generally shorter (3-5 residues) than α-helices (10-12 residues), presumably due to the energetic costs of sustaining the 3_{10} heilx further. *See* Barlow & Thornton (1988) *J. Mol. Biol.* **201**, 601-619.

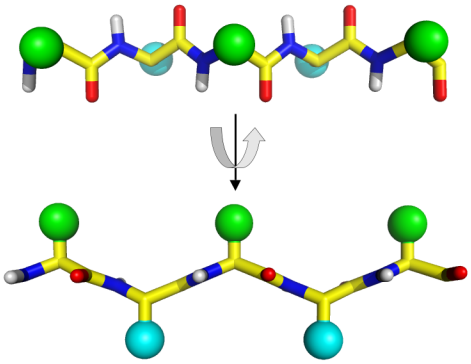


Figure S.9. Structure of the β strand. In the top orientation, the amide protons and carbonyl oxygens are oriented top and bottom. This is achieved, as seen in the lower orientation, by "pleating" the strand so that the side chains are oriented 180° from each other, top and bottom (in green and turquoise).

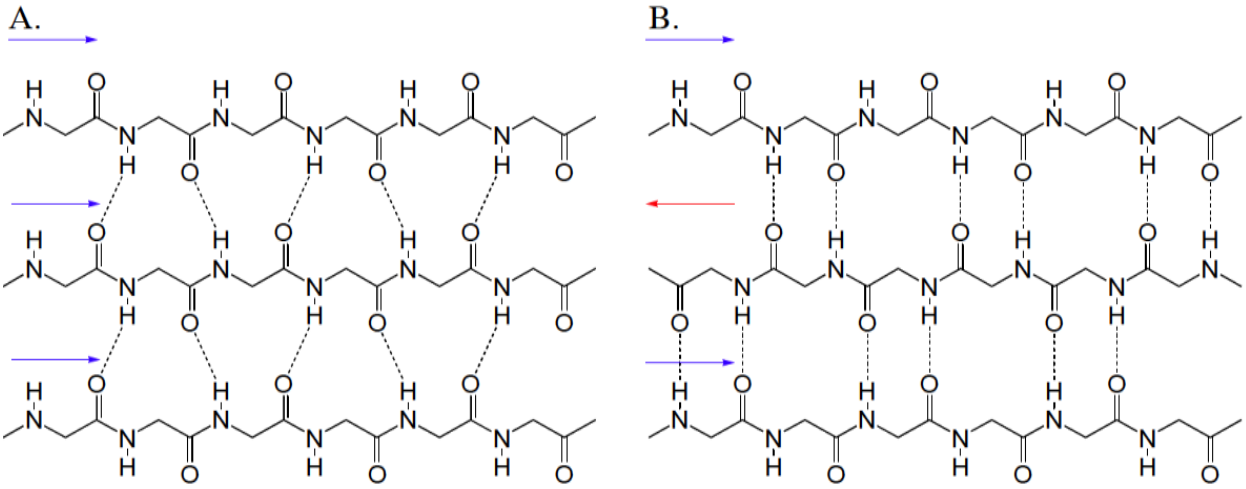


Figure S.10. (A) Parallel β sheet. All three strands are running in the same orientation from N to C-terminus, left to right. Note the pattern of H-bonding. (B) Antiparallel β sheet. The central strand is running in the reverse orientation of the top and bottom strands, yielding a different H-bonding pattern.

A peptide in a β strand conformation does not have any internal hydrogen bonds. To achieve that hydrogen bonding, additional β strands must be oriented side by side, in one of two orientations. The peptide orientation may be defined by the direction of travel from the N-terminus to the C-terminus. If two adjacent strands are oriented from N- to C-terminus in the same direction, then one obtains a **parallel sheet**. Additional strands may be added in parallel, creating extensive assemblies (Figure S.10A). It is also possible to set up intrastrand hydrogen bonding between strands running in opposite directions, as **anti-parallel sheet** (Figure S.10B). β sheet does not actually lie flat, and in fact each strand is somewhat twisted in a left-handed sense. In both instances, the side chains of the residues face above and below the plane of the sheet. Really, though, these strands don't lie perfectly in the same plane, and often there is some twisting of the strands to give any sheet some warping and texture.

A peptide in a β strand conformation does not have any internal hydrogen bonds. To achieve that hydrogen bonding, additional β strands must be oriented side by side, in one of two orientations. The peptide orientation may be defined by the direction of travel from the Nterminus to the C-terminus. If two adjacent strands are oriented from N- to C-terminus in the same direction, then one obtains a **parallel sheet**. Additional strands may be added in parallel, creating extensive assemblies (Figure S.10A). It is also possible to set up intrastrand hydrogen bonding between strands running in opposite directions, as **anti-parallel sheet** (Figure S.10B). β sheet does not actually lie flat, and in fact each strand is somewhat twisted in a left-handed sense. In both instance, the side chains of the residues face above and below the plane of the sheet. Because of the alternating up and down direction at each residue, β sheet is often referred to as β-pleated sheet (Figure S.11).

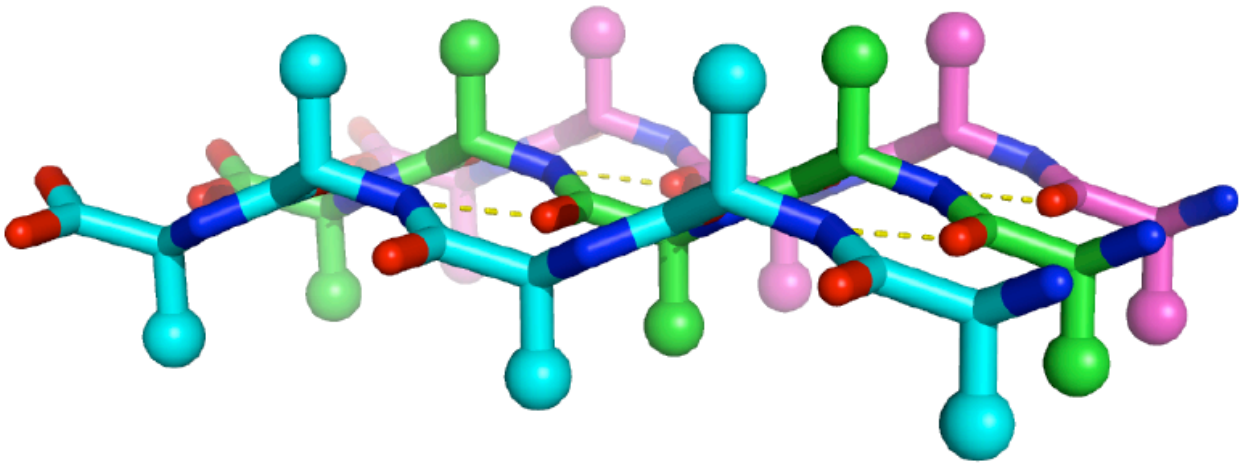


Figure S.11. Idealized cartoon of a 3-stranded parallel β sheet. Note the alternating orientations of side chains (small spheres). The alternating angles give rise to the name β-pleated sheet.

Reverse Turns

Reverse turns form a secondary structure class that is distinct from α helices and βstrands, in that there is no regularity of backbone conformation over a stretch of the polypeptide chain. However, these turns appear regularly in protein structures with a consistent internal structure and are thus worth considering as a fundamental element of protein structure. The reverse turn provides the quickest way for the polypeptide chain to change direction by 180°. Of these turns, so-called β **turns** are most common and typically connect two antiparallel β strands. They comprise four consecutive residues in the polypeptide chain and are bounded by a hydrogen bond from the carbonyl of the first (n_{th}) residue to the amide nitrogen of the fourth ($n+3^{rd}$) residue. There are six different common arrangements that permit this structure, but the principal forms are called Type I and Type II turns. Due to the tight nature of the turn, the residues involved must adopt strict ϕ and ψ angles in order to permit the reversal of the chain (Figure S.12; Table S.2).

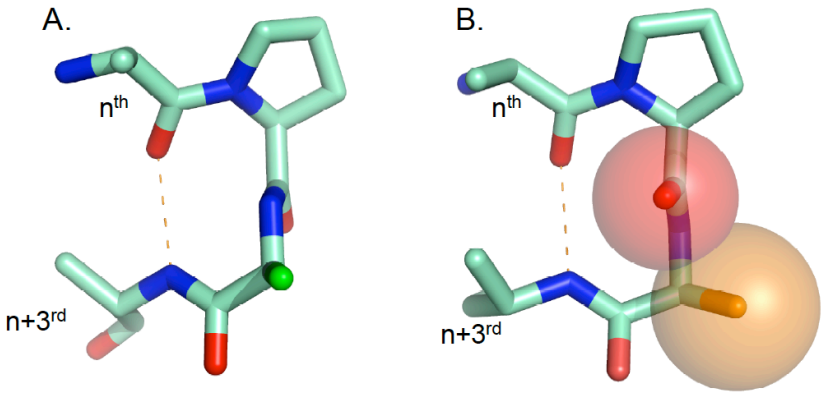


Figure S.12. (A) Type I β turn and (B) Type II β turn. Note that the chief difference between them is the orientation of the peptide bond connecting residues n+1 to n+2. In the Type II turn, the side chain of the n+2nd residue is in steric conflict with the carbonyl oxygen of residue n+1. Thus, the n+2nd residue in Type II turns is often glycine.

The unusual ϕ, ψ pairs associated with β turns lead to preferences in the residues that occupy positions with in the turn (Table S.2). In particular, Type II β turns require the n+2nd residue to adopt a conformation where (ϕ, ψ) is (+80°, 0°). This value creates a steric clash between the side chain of the n+2nd residue and the carbonyl oxygen of the n+1st residue (see Figure S.12B). That clash can only be removed by placing a glycine residue at the n+2nd position. Additionally, one often finds proline at the n+1st position of β turns. These are not required on the basis of Ramachandran angles, but their restricted conformational opportunities and their inability to donate H-bonds may make them ideal for constraining the conformation of the polypeptide backbone and disrupting any other secondary structure elements that might appear.

Table S.2. ϕ, ψ angles required for Type I and Type II β turns. Note the unusual conformation for the n+2nd residue in Type II turns.

	ϕ_{n+1}, ψ_{n+1}	ϕ_{n+2}, ψ_{n+2}
Type I	-60°, -30°	-90°, 0°
Type II	-60°, +120°	+80°, 0°

Stability of Isolated Alpha Helices

As noted above, alpha helices and beta sheets place their component residues in acceptable conformations and provide internal H-bonding opportunities for backbone amides that compensate for lost H-bonding opportunities to water. That enthalpic "satisfaction", however, is generally not enough to stabilize these conformations in isolation. The loss of entropy associated with adopting a single conformation overcomes the enthalpic issues, and most peptides fail to adopt stable

secondary structures absent other interactions. However, there are exceptions and they provide a useful view towards issues that can be exploited to stabilize conformational stability.

Amino Acid Preferences in Helices

Jane and David Richardson[5] explored the amino acid compositions of all alpha helices found in proteins and found distinct preferences for the positions of certain amino acid residues at particular positions within in the helix. They identified helix positions ranging from NCap, N1, N2, N3... to C3, C2, C1, CCap – with N1 being the first, N-terminal residue of the helix to have the appropriate phi and psi angles while C1 is the last residue to have helical backbone dihedrals. Ncap has the appropriate ψ value for a helix and Ccap has the appropriate ϕ value, but those are the positions at which helicity "breaks" and the peptide chain wanders away.

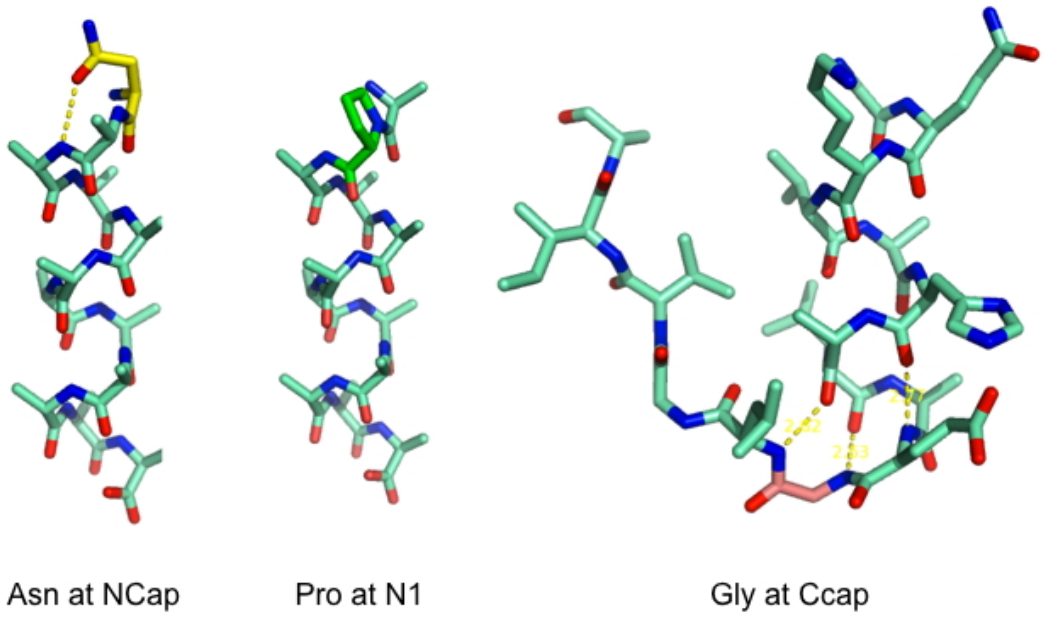


Figure S.13 Stabilizing substitutions at the ends of a helix.

Three residues stand out for typically being found at the ends of helices and much less frequently in the interior (sometimes much less frequently!). Asn and Pro are quite commonly found at the Ncap and N1 positions, respectively, while Gly is commonly found at the Ccap position (Figure S.14). Note that the ends of helices are distinct from the interior in that the N-H groups and the N-terminus and the C=O groups at the C-terminus. The interesting thing about each of these positional preferences is that they reduce unpaired H-bonding groups at the ends. The Asn at the Ncap has a side chain interaction with the amide N-H at N2. Proline at N1 simply doesn't have an N-H group, and the glycine at Ccap permits a ϕ, ψ combination, unattainable by other residues, that permits H-bond donation to the C=O of C2 and C3.

[5] Richardson & Richardson (1988) *Science* **240**, 1648-1652.

It is also worth noting that none of these three residues commonly inhabit the "interior" positions of the helix. The inability of proline to contribute an H-bond internally is obviously a problem, and the competition of the side chain of Asn for backbone interactions is likewise problematic. The case of glycine is a little more interesting. Glycine has greater conformational flexibility than other residues in the unfolded state. When it adopts a single conformation, it loses more entropy than other residues. Conceivably that contributes to the infrequent presence of glycine in helices.

It is also worth noting that alanine, methionine and leucine are particularly common residues in the internal positions of helices. An argument advanced by Brian Mathews at UO is that these residues lose little side chain entropy upon helix formation. β-branched residues such as valine and threonine have side chain collisions with the rest of the helix upon rotation (Figure S.14). Note that alanine loses the least side chain entropy in adopting a helical conformation, and accordingly it is generally a strongly helix stabilizing residue.

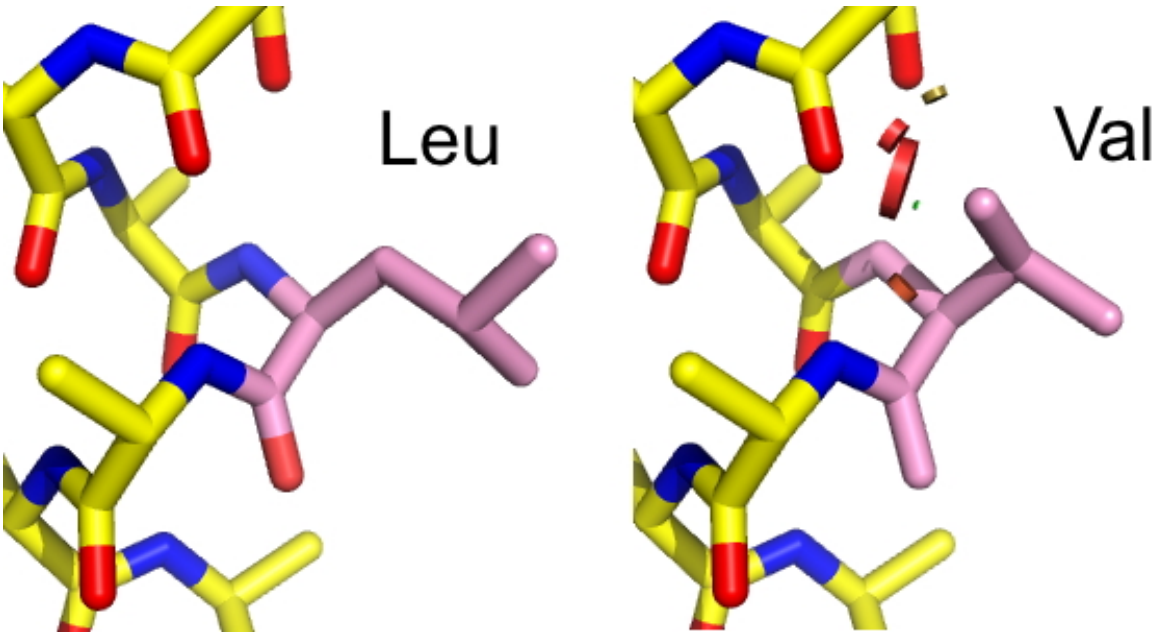


Figure S.14. Comparison of leucine and valine in an α-helix. Note that the branching atoms of leucine are well way from the helix backbone and less prone to attack than the methyl groups of valine, which will clash in two out of the three likely conformations. The red discs show points of steric conflict.

Another contribution to residue preference at positions in the α-helix comes from charge stabilization at the end. In 1968, Wim Hol noted that all of the backbone amide groups of the helix are oriented in the same direction (Figure S.15). He argued that the sum of these individual dipoles are summed up over the length of the helix to create a **helix macrodipole** that is destabilizing to the helix since it implies a separation of charge. The Richardsons noticed that glutamate and aspartate, two anionic residues, predominate at the N-terminus of the helix (where the positive end of the dipole lies), while lysine and arginine, two cationic residues, predominate at the C-terminus (where the negative end of the dipole resides).

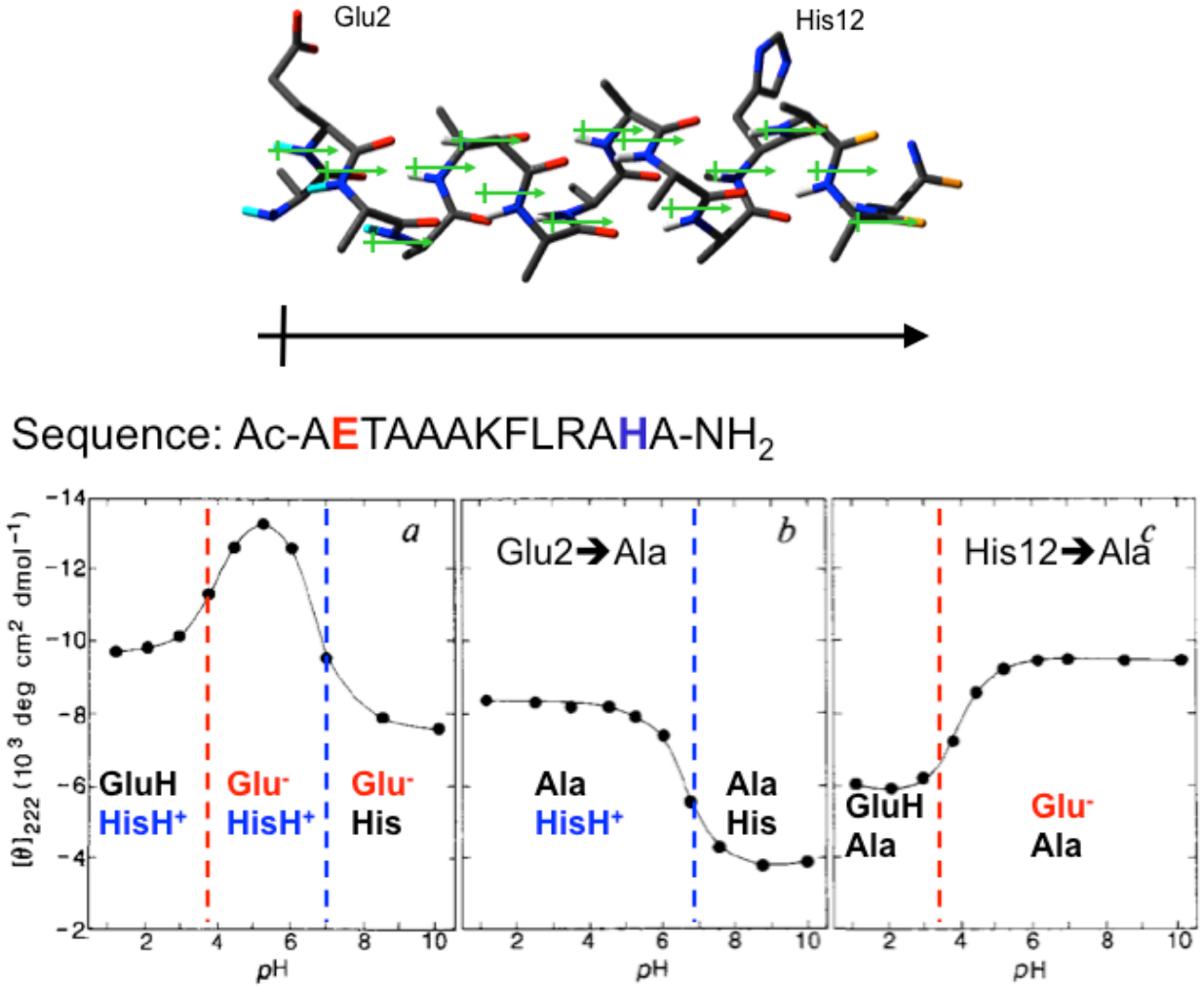


Figure S.15. (Above) Depiction of the helix macrodipole. Note that each peptide bond has a dipole (in green) locally associated with the orientation of the amide functionality. They are all oriented in the same direction and could be added to create the larger (black) macrodipole. (Below) CD spectra of three different peptides varying with pH. Note that maximum helicity is achieve in (a) at pH 6 when there is both an anionic Glu2 and cationic His12 to counterbalance the macrodipole. Adapted from Shoemaker et al. (1987) *Nature* **326**, 563.

The Baldwin lab at Stanford[6] tested the importance of charge at the termini by measuring helix stability using circular dichroism in response to pH. Studying a 14-residue peptide with Glu at position 2 and His at position 12, they noted that maximum helicity was obtained near pH 6, where both residues are expected to be charged, and stabilizing to the macrodipole. Helicity decreased at lower pH, presumably due to protonation of Glu2, and at higher pH, presumably due to deprotonation of His. When either the Glu or His residue was substituted with alanine, the relevant

[6] Shoemaker et al. (1987) *Nature* **326**, 563.

limb of the pH plot was lost, confirming the contribution of ionized residues to the stability of the helix (Figure S.15).

Predicting Secondary Structure

A lot of effort has been made to use the above trends in helix stability to identify secondary structure elements in proteins from sequence alone. While these have become quite advanced, it is worth noting early efforts made in this area.. Chou and Fasman, working in the mid-1970's, developed a scheme by which one could predict the secondary structure of a polypeptide given only the sequence. While many advances have taken place since then, the results are still meaningful because they highlight the capacity of certain amino acid residues to influence conformation.

The key feature of the Chou-Fasman method are calculated **propensities (P**) for each residue to adopt a given secondary structure. The propensity draws on a data base of many protein structures, and notes the frequency with which a given residue appears in a particular conformation relative to the frequency with which all residues appear in that conformation. For example, the P_α of alanine is determined by equation S.1, below:

$$P_\alpha = \left(\left[\frac{\#\text{ Ala residues in helices}}{\#\text{ Ala residues total}} \right] \Big/ \left[\frac{\#\text{ all residues in helices}}{\#\text{ all residues total}} \right] \right) \qquad \text{(Eq. S.1)}$$

Where P_α is large (as it is for alanine, with a value of 142), the residue is considered helix forming. Where P_α is low, the residue is considered a helix breaker (consider proline, with a value of 57). Comparable evaluations can be made with respect to β strands and reverse turns. Propensities for alpha helix and beta sheet formation are given in Table S.3.

Table S.3. Propensities of the amino acid residues to promote either α helices (P_α)or β strands (P_β).

Amino Acid	P_α	P_β
Alanine	142	83
Arginine	98	93
Aspartic Acid	101	54
Asparagine	67	89
Cysteine	70	119
Glutamic Acid	151	137
Glutamine	111	110
Glycine	57	75
Histidine	100	87
Isoleucine	108	160
Leucine	121	130
Lysine	114	74
Methionine	145	105
Phenylalanine	113	138
Proline	57	55
Serine	77	75
Threonine	83	119
Tryptophan	108	137
Tyrosine	69	147
Valine	106	170

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