Chapter 6 - Proteins: Three Dimensional Structure

Introduction:

The first x-ray structure for a protein was that for myoglobin in 1958 and indicated an apparent lack of regularity in the structure. Although such a lack of regularity is necessary considering the lack of regularity in function of proteins. Nevertheless there is a lot of regularity in protein structure, which is conceptualized as consisting of several levels.

Primary Structure is the (covalent) sequence

Higher order structure is noncovalent (secondary, tertiary, quaternary)

(Figure 6-1)
Secondary Structure

Pauling and Corey determined the X-ray structures of several dipeptides and found that the peptide bond is rigid and planar, which is due to its partial double bond character.

Most peptides adopt a trans, planar configuration (more stable than cis by about 8 kJ/mole). Cis proline is an exception - about 10% in cis configuration (Figure 6-3).

Note that since the peptide bond is rigid, no rotation about it is allowed, unlike the free rotation that occurs about single bonds. Thus, all the atoms in the green planes are fixed, and rotation can only occur about bonds which include the "carbon (Figure 6-4)
**N** (phi) and **R** (psi) are called torsion, or dihedral angles. The 180 degree of both angles is defined as that angle that corresponds to the fully extended conformation, as shown in Figure 6-2. Angles increase when rotation occurs in a clockwise fashion when viewed from Cα.

Not all values of **N** (phi) and **R** (psi) are allowed because of steric hindrance between groups on adjacent amino acid residues. A Ramachandran diagram indicates that actually very few values are allowed (Figure 6-6). Note that in repeating secondary structures, such as helices and sheets, the values of the dihedral angles **N** and **R** will be identical for all residues that form the helices or sheets.
If N, R values for a given residue are repeated, a helical structure will result. Pauling and Corey constructed models given the constraint of a rigid, planar peptide bond, and attempted to maximize H-bonding. Their efforts predicted the existence of the \"-helix (alpha helix) before its existence was established experimentally. Features of the \"-helix include:

- right handed, 3.6 residues/turn, 0.15 angstroms/residue, 5.4 angstroms/turn (pitch)

Intramolecular H-bonding between residues n and n+4 (C-O and N-H bonds parallel to helix axis, Figure 6-7).
Pauling and Corey also predicted another secondary structure, the β-sheet, (beta sheet) characterized by:

- H-bonds between neighboring chains (inter vs. intramolecular).
- Parallel and antiparallel. Parallel sheets containing fewer than 5 strands are rare, perhaps indicative of the fact that H-bonds are less stable for parallel than antiparallel sheets (Figure 6-9).
The polypeptide chain is more extended in a $\beta$-sheet than it is in a tightly-wound $\alpha$-helix. The chain is not fully extended, however, as shown above, but rather has a pleated appearance (Figure 6-10).
$\beta$-sheets typically exhibit a right-handed twist, as shown for the globular protein carboxypeptidase A (Figure 6-12).

![Diagram of a $\beta$-sheet in bovine carboxypeptidase A. (After a drawing by Jane Richardson, Duke University.)](image)

Copyright 1990 John Wiley and Sons, Inc. All rights reserved.

Although this twist distorts the H-bonds, it is found to be necessary, via conformational energy calculations, because of steric hindrance between amino acids in the sheet.

Note the topology (connectivity) in $\beta$-sheet when occurring in same strand (Figure 6-13).

![Connections between adjacent strands in $\beta$-sheets.](image)

Copyright 1990 John Wiley and Sons, Inc. All rights reserved.

Antiparallel strands can be connected by a small loop (a), whereas parallel strands
require a crossover connection that is out of plane of the $-$-sheet (b).

**Fibrous Proteins** - These proteins are found in connective tissue, and also play a role in protection (horns, nails, skin) and support (bone). Keratin, silk fibroin and collagen are examples of fibrous protein whose shapes are dominated by a single type of secondary structure, unlike the case for the other major category of protein, globular proteins (enzymes, immunoglobulins, etc.), whose shapes typically include a variety of secondary structures.

**Keratin**

Mechanically unreactive and chemically durable, keratin occurs in all higher vertebrates as principal component of hair, horn, nails, feathers. Classified as " (mammals) or $(birds, reptiles).

**Features of "keratin**

X-ray pattern similar to "-helix, except smaller pitch (5.1 vs. 5.4 angstroms). This is due to fact that two " keratin chains, each of which forms an "-helix, are wound around each other in a left handed fashion (Figure 6-14).

![Diagram of keratin structure](image)

The axes of each helix are inclined about $15^\circ$ relative to each other, and thus are inclined to the mutual axis of the duplex.

- Note the pseudorepeating sequence a-b-c-d-e-f-g (7 residues), where residues a and d are hydrophobic. Since there are 3.6 residues/turn and residues a
and d are 4 residues apart, residues a and d are on the same side of the helix. Hydrophobic residues on the partner helix line up against these residues, thereby creating a hydrophobic interaction on the interior of the dimer, or “coiled coil (Figure 6-14 b).”

![Image of coiled coil](image)

Staggered rows of dimers, end to end, form a protofilament. 4 protofilaments constitute a microfibril, which associates with other microfibrils to form a macrofibril. A single mammalian hair consists of layers of dead cells, each of which is packed with parallel macrofibrils. (Figure 6-15).

"(alpha) keratin is rich in cys residues, which form disulfide linkages between adjacent polypeptide chains. Hair, horn and nails have lots of cys residues (hard), and are less pliable than in skin, with fewer cys (soft). A hair permanent consists of treating the hair with a reducing agent which reduces, or breaks the disulfide linkages, followed by curling the hair in the desired fashion, then applying an oxidizing agent of re-form the disulfides.
Note also that shrinking of wool represents a conformational change of the "keratin.

**Silk Fibroin**

Constitutes silks of insects and spiders

Consists of antiparallel $\alpha$ sheets whose chains are parallel to the fiber axis.

**Features of silk**

Sequence studies indicate the following 6-residue repeat: (-Gly-Ser-Gly-Ala-Gly-Ala)$_n$. Thus, all the ser and ala are on one surface, all the gly on the other.

Silk is very strong because of the $\alpha$-sheet structure, yet is flexible because neighboring sheets associate only weakly through van der Waals forces.

**Collagen**

The most abundant vertebrate protein, it occurs in all multicellular animals. It forms strong fibers (can support 10,000 times its own weight and is stronger than steel). Collagen comprises the stress-bearing components of connective
tissues such as bone, teeth, cartilage, tendon, and the fibrous matrices of skin and blood vessels.

**Features of Collagen**

About 30 genetically distinct forms in mammals.

Type I is common, has two $\alpha_1$ chains and one $\alpha_2$ chain, with a molar mass of about 285 kDa, a width of 14 Angstroms and a length of 3000 Angstroms.

Lots of Gly and 3- and 4- hydroxyproline. These non-standard amino acids are converted from proline, the required enzyme being prolyl hydroxylase. The OH groups are necessary for H-bonding, which in turn gives the collagen strength. Vitamin C is necessary to maintain the activity of prolyl hydroxylase. Scurvy, common several hundred years ago in sailors on long voyages, was alleviated by the addition of limes to British sailors diets by Captain James Cook (“limeys”).
A single molecule consists of three chains, each in a left-handed helix, wound around each other in a right handed manner. Recall that keratin had right handed helices wound around each other in a left handed manner. This is for strength.

![Collagen molecule](https://example.com/collagen.png)

Genetic, as well as nutritional diseases are associated with collagen. Mutations in Type I collagen, which constitutes the major structural protein in most human tissues, usually result in osteogenesis imperfecta (brittle bone disease). A Gly to Ala substitution can be lethal because the additional crowding can distort the collagen triple helix. In normal collagen there is a Gly residue at every third position. This is because the triple helix of collagen is so tight that only a Gly residue is small enough to fit in.

Collagen molecules assemble to form loose networks, or thick fibrils arranged in bundles or sheets. The fibrils are organized in staggered arrays that are stabilized by hydrophobic interactions, and also by cross links, which form from
lys and his residues (Figure 6-19). It is of interest to note that the only enzyme implicated in the cross linking process is lysyl oxidase, which forms the aldehyde of the lysine side chain.

**Globular Proteins**

Most proteins (enzymes, antibodies, binding and transport proteins, receptors, etc.) are globular, rather than fibrous.

The higher level structure of a fibrous protein consists of fibrils, or packets of secondary structure, hence is typically not referred to as tertiary.

Tertiary structure is typically confined to globular proteins.

Regular, secondary structures, such as $\alpha$-helix and $\beta$-sheet are found in globular as well as fibrous proteins. In order to attain compact, globular shape, there must be a mechanism to terminate secondary structure to avoid long, narrow structures as found in fibrous proteins. $\alpha$-helices and $\beta$-sheets must abruptly change direction in globular proteins. Such direction changes typically occur via reverse turns or $\beta$-bends (Figure 6-19).

![Diagram of Type I and Type II structures](image.png)

Note that in such a bend H-bonds occur between residue 1, initiating the bend and residue 4 ($n + 3$, rather than $n + 4$). Note also in Figure b), the peptide bond
between residues 2 and 3 is flipped by 180° relative to that in Figure a). This brings the carbonyl O of residue 2 in close proximity to the R group of residue 3, which is typically a glycine for this reason.

Proteins longer than 60 residues typically have a direction change via a so-called S (omega) loop (Figure 6-20) consisting of 6 - 16 residues and resembling a Greek omega. They are invariably located on the surface of the protein, thus may have a role in recognition.

Globular proteins often contain regions of secondary structure characterized by irregular structures, in which residues have different N, R values. Note that this would result in points randomly distributed on a Ramachandran plot. These are sometimes called coils, distinct from random coils produced in denaturation. Coils may be considered to be a type of secondary structure.

The regular pattern of secondary structure in globular proteins is often distorted. For example an "α"-helix may vary from its ideal conformation in the initial and final turns of the helix. "α"-helices are often flanked by Asn or Gln residues whose side chains fold back and form additional H-bonds with one of the four terminal residues of the helix, a phenomenon known as helix capping. A strand in a $β$-sheet may contain an extra residue that is not H-bonded to a neighboring strand, producing a distortion known as a $β$-bulge.

Chou and Fasman studied many known protein secondary structures and compared them with their amino acid compositions. They found that different amino acid residues have different tendencies, or propensities, to form helices (Table 6-1).

**Tertiary Structure**

Determined by either X-ray crystallography or NMR

X-ray diffraction requires a crystal be formed to generate a repeating pattern
of electron density, which is responsible for diffraction of the X-rays (Figure 6-22).

Proteins tend to co-crystallize with solvent (up to 40 - 60% water). The resulting crystals tend to have a jelly-like consistency, which limits the resolution from about 1 to 2 - 3.5 Angstroms.

This results in an electron density contour map which gives the shape of the backbone (Figure 6-23). The orientation of individual side chains can usually be deduced, but knowledge of the protein sequence is typically required to fit this sequence to the electron density map. Mathematical techniques can then refine the structure to within 0.1 Angstroms in high-resolution structures.

NMR spectroscopy identifies positions of protons, rather than electron density. NMR is sensitive to interactions between protons, and interpretation of these interactions has led to structural determination of small proteins (<250 residues) has been possible since the mid-80's.

Nonpolar residues (Val, Leu, Ile, Met, Phe) are typically located in the interior of globular proteins, whereas charged, polar residues (Arg, Lys, Asp, Asp and Glu) are typically located on the exterior of globular proteins. Hydrophobic
interactions are responsible for this phenomenon. The overall structure of a globular protein is thus amphipathic, similar to that of a micelle in which the hydrophilic head groups of soap molecules, for example, are located on the exterior of the micelle, thus shielding the nonpolar aliphatic groups on the inside from water. The thermodynamic driving force here is, as it was described in chapter two, the internalization of nonpolar groups away from water to avoid the entropy loss in water that would occur near exposed nonpolar groups.

Uncharged polar residues (Ser, Thr, Asn, Gln and Tyr) are usually on the surface, but can be found in the interior. In this case, they are almost always H-bonded to other groups. Interior atoms of globular proteins tend to pack very tightly and exclude water.

Some globular proteins consist entirely of $\alpha$-helix (myoglobin and hemoglobin), some of $\beta$-sheet (concanavalin A, see panel b below), whereas others contain both (Triose phosphate isomerase, see panel c)

Certain groupings of secondary structure, called motifs, or supersecondary structures, are often found in unrelated globular proteins (Figure 6-29)
("", "hairpin", ",", and ") barrels). Sometimes motifs can have functional as well as structural significance: Two "," ") units can act as a nucleotide binding site (Figure 6-29 - each unit binds a nucleotide of NAD\(^+\)).

Large proteins can form domains (Figure 6-31).

Although there are thousands of unique sequences, it turns out that there are much fewer folding patterns. When folding patterns are compared without regard to amino acid sequence or the presence of surface loops, estimates place an upper
limit of about 1,000 proteins in nature, and only a few dozen account for half of all known protein structures.

The rapidly increasing database of known protein structures suggests the limited number of common protein structures may be due to the fact that there are a limited number of ways a polypeptide chain can fold up to form a functional protein. From an evolutionary point of view, only a limited number of folding patterns may be stable, or able to withstand inevitable amino acid deletions, insertions and substitutions. Also, only a limited number of patterns may have biological functionality.

It is of interest to draw comparisons with various prokaryotic c-type cytochromes (recall that we talked about eukaryotic cytochromes) with regard to both composition and structure). The functions of these various c-type cytochromes are similar, in that they are electron carriers. And their x-ray structures are similar. However, their sequences are often quite different. It can be concluded that from an evolutionary point of view, function, rather than sequence, is preserved during evolution, and also that very different sequences can have similar structures.

**Quaternary Structure**

Most proteins with >100 residues consist of more than one polypeptide chain, thus have quaternary structure (Figure 6-32 - hemoglobin). Advantages:

- Compare to pre-fab buildings
- Each subunit of a multi-subunit protein has an active site.
- Regulation of multi-subunit enzymes often occurs via their subunit assembly.
- Proteins with >1 subunit are typically symmetrically arranged.

Because of the chirality of individual amino acids, proteins can have only
rotational symmetry (Figure 6-34).

When a 2-fold rotation axis is perpendicular to an n-fold axis, the protein is said to have dihedral symmetry. Higher symmetry is tetrahedral, octahedral, and icosahedral.

**Protein Stability**

Native proteins are only marginally stable under physiological conditions (only about 0.4 kJ/mole of residues is required to denature them). The stability of a fully-folded, 100-residue protein is only about 40 kJ/mole, to be compared with about 20kJ/mole to break a single H-bond.

As was mentioned above, the hydrophobic effect plays a large role in protein
stability. Note that although nonpolar residues are internalized, and charged residues externalized, there is typically an apparent random pattern of nonpolar, charged and neutral polar residues in the sequence. This is because the polypeptide chain follows a convoluted pattern, winding inside, then outside. A so-called hydropathy index (Table 6-2) provides a quantitative measure of hydrophobicity, amino acid residues with nonpolar side chains have positive hydropathies, whereas polar residues have negative hydropathies. A plot of hydropathies versus sequence is shown below for chymotrypsin (Figure 6-35). As expected, the pattern on nonpolar and polar residues is somewhat random (upper bars indicate residues in the protein’s interior, lower bars indicated residues located on the protein’s exterior.

As we’ll see in Chapter 10, a non-random pattern of hydropathies for a membrane protein may indicate a membrane-spanning series of residues.

Electrostatic interactions

H-bonds make only minor contributions (water can form), allthough they are important ("-helix (or $\beta$-sheet) must form in interior).

About 75% of charged residues form ion pairs or salt bridges

Cross links such as those formed by disulfide bonds are not essential stabilizing forces (Anfinsen). Disulfide bonds are rare in intracellular proteins because the cytoplasm is a reducing environment (Most are extracellular)
Metal ions can form cross links. Zinc fingers are motifs (about 10) in nucleic acid binding proteins that consist of 25-60 residues arranged around one or two zinc ions that are tetrahedrally coordinated by Cys, His and sometimes Asp or Glu side chains (Figure 6-37). The Zn$^{++}$ allows relatively short stretches of polypeptides to fold into stable units that can interact with nucleic acids. Note that the d electrons of zinc enable it to interact with a variety of ligands (S, N, O), and it has only one stable oxidation state.

**Protein Folding**

It is clear that proteins do not adopt their native conformations, or shapes, through a random search of possible conformations. Such a process would take a prohibitively long time.

Protein folding appears to begin with formation of local regions of secondary structure, which then serve as nuclei for a subsequent “hydrophobic collapse” into a so-called molten globule, which possesses much of the secondary structure of the native state, but lacks in a well-defined tertiary structure (see Figure 6-40).

Although the Anfinsen experiment indicated that sequence determines structure, there are cellular proteins which assist protein folding. Two examples are protein disulfide (PDI) and a family of proteins known as molecular chaperones.

PDI’s facilitate disulfide interchange in folding proteins that may have formed disulfide bonds not present in the native conformation via binding to a hydrophobic patch on an unfolded protein as depicted in Figure 6-42.

Molecular chaperones were originally termed heat shock proteins due to their increased rate of synthesis at elevated temperatures, presumably to recover heat-denatured proteins or to prevent misfolding. These proteins bind to exposed
hydrophobic surfaces of unfolded or aggregated proteins, thereby preventing improper folding as well as protein aggregation.

A number of typically fatal diseases are known that are involved with the formation of insoluble fibrous aggregates (amyloid) of normally soluble proteins. Examples include Alzheimer’s disease and mad cow disease (a member of a group of diseases known as transmissible spongiform encephalopathies, or TSE’s). Alzheimer’s is a neurodegenerative disease that mainly strikes the elderly, affection about 50% of those over 85. The amyloid plaques consist primarily of fibrils of a small (40 - 42 residue) protein (Aβ(beta)), derived from a larger, Aβ, precursor protein (betaPP). There is a genetic aspect to this disease. There are several rare variants in the Abeta region of the betaPP precursor gene that result in early onset of Alzheimer’s. Additionally, apo E, A serum apolipoprotein found in chylomicrons, VLDL and HDL, has been found in amyloid plaques and may function as a pathological chaperone protein, inducing beta sheet conformation in the Abeta proein. The frequency of apo E4, a variant of the predominant form of apo E is three times higher in patients with Alzheimer’s.

Mad cow disease (boving spongiform encephalitis, or BSE) is another neurological disorder, in this case the infectious element being a protein called a prion (proteinaceous infectious particle that lacks nucleic acid), originally thought to be a slow acting virus (viruses do not lack nucleic acids). In afflicted individuals, neurons develop large vacuoles that give brain tissue a sponglike microscopic appearance. One type of prion consists largely of hydrophobic residues leading to aggregation and formation of rodlike particles.

Problems: 2, 3, 8, 10, 12, 16