Chapter 12: An introduction of protein structure

12.1 Introduction

12.2 An overview of protein structural hierarchy and classification

  12.2.1 Sequence relationships between proteins: orthologs and paralogs
  12.2.2 Three-dimensional structural relationships between proteins: folds and domains
  12.2.3 Single-domain architecture (SDA) and multi-domain architecture (MDA) proteins:

12.3 Primary and secondary structure of proteins:

  12.3.1 Three torsional angles define the backbone configuration:
  12.3.2 Orbital overlap constrains the angle $\omega$:
  12.3.3 The peptide bond has a permanent dipole
  12.3.4 Most, but not all, peptide bonds are trans:
  12.3.5 The $\phi$ and $\psi$ angles are constrained by steric clashes- The Ramachandran Plot
  12.3.6 Secondary structures can be defined in terms of their $\phi, \psi$ angles.
  12.3.7 DSSP Code for secondary structural elements.
  12.3.8 Statistical preferences for amino acids to be found in $\alpha$-helix, $\beta$-sheet or turns
  12.3.9 Turns as secondary structural elements
  12.3.10 The $\alpha$-helix
  12.3.11 The $\alpha$-helix has a permanent dipole moment
  12.3.12 Amino acid preferences at the ends of the $\alpha$-helix
12.3.13 Helical distortions

12.3.14 Parallel and antiparallel $\beta$-sheet

12.4 Tertiary structure, protein folds and domains

12.4.1 Example: The immunoglobulin fold, an all-$\beta$ protein.

12.4.2 Example: $\beta$-barrels

12.4.3 Example: $\alpha$-helical coiled coils

12.4.4 Example: The helix-turn-helix homeodomain

12.4.5 Example: The chloride channel from Salmonella typhimurium

12.4.6 Comparing the helix-helix interface in soluble proteins and membrane proteins

12.4.7 Example: The $\beta\alpha\beta$ fold and the Rossmann fold

12.4.8 Example: The $\alpha/\beta$-barrel or TIM fold

12.5 The inside of the protein is tightly packed

12.6 Proteins are hydrated

12.7 Quaternary structure and protein-protein interactions
Chapter 12: An introduction of protein structure

12.1 Introduction

Biochemistry as a discipline is primarily concerned with molecular mechanisms of biological processes. These days, the starting point is often 3-dimensional structure, usually obtained from X-ray crystallography. Genomic sequencing and structural genomics projects continue to provide a wealth of information to biochemists. Many books are available which beautifully depict protein structures (e.g., see (2)), including introductory textbooks in Biochemistry. It is not our intention to review all that is known about protein structure, but to highlight what is necessary to appreciate the approaches taken to understand the issues of protein stability and protein folding, as well as providing the background needed to discuss the spectroscopic methods that are applied to probe protein structure and dynamics.

Protein stability is a thermodynamic concept. The stability of the fully folded, native form of a protein is a measurement of the difference in the standard state chemical potential of the fully folded form and the fully unfolded, or denatured, form of the protein. That is, stability is measured as an equilibrium constant, with which we should be quite familiar by now. Measuring protein stability is a common component of studies of the effects of mutations, quantifying the impact of the mutations on the behavior of the protein in solution. If one is interested to engineer a new protein or to redesign an existing enzyme to perform new tasks, protein stability is of crucial importance.

Protein folding, on the other hand, is a matter of kinetics, intermediates and pathways. Many proteins which have been isolated and then completely unfolded, can be induced to rapidly and spontaneously refold to form the proper 3-dimensional structure
with biological function (e.g., enzyme activity). These observations mean, at least in the
general case, that 1) the observed 3-dimensional structure of a protein represents the free
energy minimum; and 2) the information determining the 3-dimensional structure of a
polypeptide is encoded within the primary sequence of amino acids of that protein.

Understanding how this works and being able to predict the final folded structure
of a protein based on its amino acid sequence remain active fields of research.
Substantial progress has been made but there is much remaining to be revealed. In this
Chapter our goal is to define some concepts and tools useful in these endeavors which, at
the same time illustrate the applications of thermodynamics and kinetics which we have
already discussed. The next several sections are a survey of terms, concepts and
depictions of protein structures. In the following Chapter we will discuss protein stability
and the kinetics of protein folding.

12.2 An overview of protein structural hierarchy and classification

The traditional descriptive hierarchy of protein structure is

1) Primary structure: the amino acid sequence.

2) Secondary structure: mainly $\alpha$-helix and $\beta$-sheet.

3) Tertiary structure: the 3-dimensional structure of a single polypeptide chain.

4) Quaternary structure: the interaction between polypeptides or subunits in
oligomeric proteins.

The rapidly growing volume of sequence and 3-dimensional structural
information has required further classification in an effort to organize the large amount of
data. To a large extent the classification schemes are purely descriptive. Protein
structural diversity ultimately derives from the process of evolution and, conversely, this
information is used to deduce evolutionary relationships. Hence, there is an increasing
interest to take these evolutionary relationships into account in defining protein
classification schemes (3, 4).

12.2.1 Sequence relationships between proteins: orthologs and paralogs

In comparing two protein sequences, they are classified as

i) homologs: when the two proteins are phylogenetically related and are derived
from a common ancestral gene. As a general rule, protein sequences that have more than
25% sequence identity are classed as homologs. One compilation of related sequences is
PFAM (Protein FAMilies; http://pfam.sanger.ac.uk/)(5). One can identify sequence
homologies or sequence motifs that do not cover entire polypeptides but only portions of
each polypeptide. How one analyzes the data depends on the definition of the metric of
similarity and the boundaries of the sequences. In any event, “sequence profiles” can
define domains within polypeptides which have structural and functional significance (1).

ii) orthologs: when the two sequences are derived from a common ancestral gene and retain the same function.

iii) paralogs: when the two proteins are derived from a common ancestral

gene but have evolved to have different functions.

Primary sequence relationships form one basis of classifying proteins in families
(relating proteins with similar functions) and superfamilies (relating proteins with
different functions).

12.2.2 Three-dimensional structural relationships between proteins: folds and
domains
The comparison of the 3-dimensional structures of proteins show there are common folding patterns. These folding patterns involve secondary structural elements, connected in the same manner and which interact similarly. These are referred to as “supersecondary” structures, or “folds” or “domains”, somewhat vaguely defined and with overlapped definitions. There is no consensus on how to define the best unit of a three-dimensional structure (e.g., see [3, 6, 7]). Generally, the concept of a domain (typically 100 to 200 residues) implies a compact folded unit that has quasi-independence both structurally and functionally. There are a number of websites that provide protein classification schemes. For example, 1) The Structural Classification of Proteins (SCOP; http://scop.mrc-lmb.cam.ac.uk/scop/) classifies proteins in terms of folds, families and superfamilies; The Conserved Architecture Retrieval Tool (CDART; http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi) classifies sequences which have common domain folding architecture 3) The CATH site classifies proteins hierarchically in terms of Class, Architecture, Topology and Homogous Superfamily (http://www.cathdb.info/).

12.2.3 Single-domain architecture (SDA) and multi-domain architecture (MDA) proteins (I):

There is a vast and growing literature on the analysis of protein structural data. There is a large overlap of the “sequence profiles” that emerge from the consideration of sequence comparisons and from the consideration of regions of shared 3-dimensional structure (1, 8), but the correspondence is not perfect. There are a number of examples of very similar sequences having very different 3-dimensional structures (7). The extreme cases are called chameleon sequences, which can adopt different structures even within
the same protein (see (4, 7)). The malleability of the protein fold can, of course, be functionally important (4). There are also examples of proteins with very similar 3-dimensional structures having little or no sequence similarity (7). Although these cases are quite striking, these are the exceptions.

As sequence and structural information continues to grow rapidly, the number of unique families of proteins containing only one recognizable domain, as defined by the “sequence profile” approach (1), has been growing very slowly. About 25% of these families have at least one member whose structure is known. There continues to be a rapid growth of new families of multi-domain architecture (MDA) proteins. Most of the MDA proteins have at least one domain whose structure can be modeled based on the homology to a known structure of a protein sharing the sequence profile of the domain.

The data are consistent with the concept of protein evolution proceeding in large part by the generation of new proteins by new combinations of protein domains to form new multi-domain proteins. Not only do the domains themselves adopt new functions, but the interfaces between them, and between protein subunits, can also evolve functional sites.

We will now take a closer look at protein structural features.

12.3 Primary and secondary structure of proteins:

Proteins are linear polymers of amino acids linked by peptide bonds (Figure 12.1). Each amino acid contains a side chain, and the properties of the side chain, covalently attached to the Cα (α-carbon), determine the characteristics of the amino acid.
Figure 12.1: A peptide chain with two amino acids (side chains $R_1$ and $R_2$). From (2).

With the exception of glycine, which has a hydrogen atom as its side chain, all the other amino acids are chiral (Figure 12.2). The naturally occurring amino acids found in ribosome-encoded proteins are all L-amino acids.

Figure 12.2: Chirality or handedness of the amino acids

Figure 12.3 shows the structures of the 20 naturally occurring amino acids and Table 12.1 gives their molecular volumes and surface areas. The system of labeling the side chain carbons is shown for lysine in Figure 12.4. The side-chain carbon that is adjacent to the $\alpha$-carbon is called $C\beta$, the $\beta$-carbon, and all amino acids other than glycine have a
β-carbon. Typically, the amino acids are classified as either non-polar, polar (non-ionizable) and polar (ionizable), as indicated in Figure 12.4.

Figure 12.3: Structures, classification and designations of the 20 naturally occurring amino acids.
Table 12.1: Properties of amino acids and their designations. Surface and volume data are from (9, 10).

<table>
<thead>
<tr>
<th>Amino Acid Name</th>
<th>3-letter Code</th>
<th>I-letter Code</th>
<th>Surface [Å²]</th>
<th>Volume [Å³]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
<td>115</td>
<td>88.6</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
<td>225</td>
<td>173.4</td>
</tr>
<tr>
<td>Aspartic Acid</td>
<td>Asp</td>
<td>D</td>
<td>150</td>
<td>111.1</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
<td>160</td>
<td>114.1</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
<td>135</td>
<td>108.5</td>
</tr>
<tr>
<td>Glutamic Acid</td>
<td>Glu</td>
<td>E</td>
<td>190</td>
<td>138.4</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
<td>180</td>
<td>143.8</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
<td>75</td>
<td>60.1</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>195</td>
<td>153.2</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
<td>175</td>
<td>166.7</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
<td>170</td>
<td>166.7</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
<td>200</td>
<td>168.6</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
<td>185</td>
<td>162.9</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
<td>210</td>
<td>189.9</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
<td>145</td>
<td>112.7</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
<td>115</td>
<td>89.0</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
<td>140</td>
<td>116.1</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
<td>255</td>
<td>227.8</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
<td>230</td>
<td>193.6</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
<td>155</td>
<td>140.0</td>
</tr>
</tbody>
</table>

Figure 12.4: Structure of lysine, showing the system used for labeling the side chain carbon atoms.
12.3.1 Three torsional angles define the backbone configuration: The configuration of the polypeptide backbone is determined by only three torsional angles, \( \phi, \psi \) and \( \omega \), which are shown in Figure 12.5.

![Figure 12.5: Definition of the three torsional angles that determine the configuration of the polypeptide backbone. The Ca-CO-NH-Ca units are planar, and can swivel about the two bonds on either side of each Ca. After (2).](image)

12.3.2 Orbital overlap constrains the angle \( \omega \): The peptide bond has aromatic character, which can be also expressed in terms of resonance forms (Figure 12.6). One consequence is that the torsional angle \( \omega \) is constrained to be at or near either 0° or 180°, meaning that the Ca-C(=O)-NH units are planar. This maximizes the overlap of the \( \pi \) electrons, minimizing the energy. These planar units, shown in Figure 12.5, swivel about the bonds connecting on either side of Ca of each amino acid, which define the \( \phi \) and \( \psi \) torsional angles.

12.3.3 The peptide bond has a permanent dipole: A second consequence of the resonance forms (or electron delocalization) is that there is charge redistribution across the peptide bond, leading to charge separation and a net electric dipole of 3.5 Debye units. The partial charges on atoms are shown in Figure 12.6.
A Debye unit is the cgs unit for a dipole moment. The dipole moment of two charges of opposite charge with equal magnitudes of \(10^{-10}\) statcoulomb (or esu) separated by a distance of 1 Å is defined as one Debye. A statcoulomb is equal to \(3.35 \times 10^{-10}\) Coulomb. Since the dipole moment \(\mu_{el}\) is defined as the product of the separated charge times the distance between the charges \(\mu_{el} = z \cdot d\), 1 Debye is equal to \(10^{-10}\) statC-Å, or \(10^{-18}\) statC-cm. The Debye unit is convenient because it is in the range of the dipole moments of molecules. For example, KBr has a dipole moment of 10.5 D (Debye units).

**12.3.4 Most, but not all, peptide bonds are trans:** Figure 12.7 shows that the peptide bond can be in either a cis or trans configuration. The vast majority of peptide bonds are trans. The majority of cis peptide bonds that are found in proteins are between the proline and the amino acid preceding it (on the N-terminal side), called the X-Pro peptide bond. In some cases, the rate of forming the cis X-Pro peptide bond can be the rate-limiting step to attain the native structure of a protein.
Although most peptide bonds are trans, some are cis. Most cis peptide bonds are X-Pro peptide bonds, as shown. The arrows on the peptide on the right indicate the direction of the peptide chain from the N-terminus to the C-terminus.

12.3.5 The $\phi$ and $\psi$ angles are constrained by steric clashes- The Ramachandran Plot: Because there are effectively only two torsional angles per amino acid that defines the backbone configuration, it is very convenient to summarize this information in the form of a two-dimensional Ramachandran Plot. Each amino acid in a protein is characterized by a $(\phi, \psi)$ pair, which is represented as a point on the Ramachandran Plot. Figure 12.8 shows experimental values for residues from about 500 high-resolution X-ray structures. The data are separated for i) glycine residues; ii) proline residues; iii) residues preceding proline (pre-proline); iv) all residues other than
glycine and proline (18 amino acids). In addition to the scatter plot showing the distribution of points, the contours show “favored” regions and regions that are “allowed”. The regions that are formally “not allowed” are ruled out due to steric clashes within each isolated amino acid. Within a protein, there are additional packing constraints which further severely limit the distribution of observed φ,ψ values (12).

Several points to note:

1. Glycine has the smallest side chain, a hydrogen atom, and has the largest allowable area in the Ramachandran Plot. Even for glycine, half of the area in the Ramachandran Plot is not allowed for the isolated amino acid. The φ,ψ angles for an “unstructured” peptide in water have been determined computationally for the pentapeptide series Gly-Gly-X-Gly-Gly (12). Even for the peptide with X = Gly, there were preferred conformations within the “allowed” regions, presumably due to solvation effects.

2. Proline, due to its 5-membered ring connecting the α-carbon and the amino group, is constrained to $\phi \approx -70^\circ$, though there is considerable experimental spread in observed φ values. There are three clusters or peaks centered at different ψ angles that are favored in proteins.

3. Non-Gly, non-Pro, residues are constrained by steric interactions between the backbone residues and Cβ (see Figure 12.4) which is, essentially, identical for all residues other than glycine (which has no Cβ) and proline (covalent constraint). Apart from this, there are clearly favored φ,ψ configurations found in proteins which do vary with each amino acid because of additional packing requirements.
4. Residues that precede proline in the sequence have a distinct distribution of \( \phi, \psi \) configurations.

5. The Ramachandran Plot is of value to summarize the backbone structure of a particular protein or region of a protein, but is also useful as a way to check the correctness of protein models from X-ray diffraction data. If it is found that the model has dihedral angles that are “not allowed”, then the model needs to be corrected.

Figure 12.8: Ramachandran plot showing the distribution of backbone configurations of residue in 500 high resolution structures. (A) All 18 amino acids other than proline and glycine; (B) Glycine residues; (C) Proline residues; (D) Residues preceding proline in the amino acid sequence. From (II).

12.3.6 Secondary structures can be defined in terms of their \( \phi, \psi \) angles.
Secondary structure describes repeating motifs that are defined by internal hydrogen bonding and/or a sequence of dihedral $\phi, \psi$ angles. The most prevalent secondary structural elements are the $\alpha$-helix (hydrogen bonding: $i_{c-o} \leftarrow (i + 4)_{NH}$, where the arrow indicates the direction from hydrogen bond donor to the acceptor), parallel $\beta$-sheet and antiparallel $\beta$-sheet. Each of these three structural motifs is also defined in terms of repeating $\phi, \psi$ dihedral angles in a contiguous stretch of amino acid residues. The next most prominent secondary structural elements are turns, in which the direction of the polypeptide reverses direction at the protein surface. There are a variety of turns with 3, 4 or 5 residues with defined internal hydrogen bonding, and with defined $\phi, \psi$ dihedral angles at each position in the turn. Other defined secondary structural elements, defined on the basis of hydrogen bonding patterns, are the $3_{10}$-helix (hydrogen bonding: $i_{c-o} \leftarrow (i + 3)_{NH}$), the $\pi$-helix (hydrogen bonding: $i_{c-o} \leftarrow (i + 5)_{NH}$) and the polyproline II helix. Table 12.2 summarizes the geometric parameters of some secondary structural elements.

### Table 12.2: Geometric characteristics of protein secondary structural elements

<table>
<thead>
<tr>
<th>Structure</th>
<th>$\Phi$</th>
<th>$\Psi$</th>
<th>Residues per turn</th>
<th>Axial displacement per residue (Å)</th>
<th>Pitch (Å) (displacement per turn)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$-helix</td>
<td>-60</td>
<td>-45</td>
<td>3.6</td>
<td>1.5</td>
<td>5.4</td>
</tr>
<tr>
<td>$3_{10}$ helix</td>
<td>-49</td>
<td>-26</td>
<td>3.0</td>
<td>2.0</td>
<td>6.0</td>
</tr>
<tr>
<td>$\pi$ helix</td>
<td>-55</td>
<td>-70</td>
<td>4.1</td>
<td>1.15</td>
<td>4.7</td>
</tr>
<tr>
<td>Polyproline II helix</td>
<td>-75</td>
<td>+160</td>
<td>3.0</td>
<td>3.1</td>
<td>9.3</td>
</tr>
<tr>
<td>Parallel $\beta$-sheet</td>
<td>-119</td>
<td>+113</td>
<td>2</td>
<td>3.2</td>
<td>6.4</td>
</tr>
<tr>
<td>Anti-parallel $\beta$-sheet</td>
<td>-139</td>
<td>+135</td>
<td>2</td>
<td>3.4</td>
<td>6.8</td>
</tr>
</tbody>
</table>
The $\phi, \psi$ pairs characteristic of the secondary structural elements define regions of the Ramachandran Plot, shown in Figure 12.9. Any amino acid in a protein might have $\phi, \psi$ angles that fall within these regions but not be part of the secondary structural element, since this requires the neighboring residues to also have the same $\phi, \psi$ angles (for $\alpha$ and $\beta$) or specified values (for turns). The polyproline II helix is a helical structure formed by contiguous prolines with trans peptide linkages (the polyproline I helix has cis linkages) and, although not found in globular proteins, many amino acids have $\phi, \psi$ angles that fall within this region of the Ramachandran Plot.

![Figure 12.9: Regions of the Ramachandran Plot defined in terms of secondary structural elements. From (13).](image)

12.3.7 DSSP Code for secondary structural elements.
Residues are assigned letters designating the secondary structural element in which they are located within a particular protein. These are designated by the “Dictionary of Protein Secondary Structure” or DSSP Code. These are

- G: $3_{10}$ helix
- H: $\alpha$-helix
- I: $\pi$-helix
- T: hydrogen bonded turn
- E: extended strand, either parallel or antiparallel $\beta$-strand
- B: isolated $\beta$-bridge (an isolated pair of residues hydrogen bonded as in $\beta$-strand)
- S: bend (not hydrogen bonded)

Residues that do not fit these categories are referred to as being “Coil” (C), meaning simply that they are not part of the regular secondary structures above.

### 12.3.8 Statistical preferences for amino acids to be found in $\alpha$-helix, $\beta$-sheet or turns

From the structures of proteins determined by either X-ray diffraction or by NMR, one can determine the statistical preference of each amino acid to be found in $\alpha$-helix, $\beta$-strand or in turns. This is a starting point for many algorithms to predict secondary structures in proteins for which there is no structure available. Table 12.3 shows the preferences obtained in one study of globular proteins. The preference $P_{ij}$ of amino acid $i$ for structure $j$ is defined in terms of

- $n_{ij}$, the number of residues of type $i$ in structure $j$
- $n_i$, the total number of residues of type $i$ in the data set
\( n_j \), the total number of residues in structure \( j \) in the data set

\( N \), the total number of residues in the data set

\[
P_{ij} = \frac{n_{ij}}{n_j/N} \tag{12.1}
\]

The statistical preference measures the frequency of finding an amino acid in a particular structural element \((n_{ij}/n_j)\) in relation to the frequency of occurrence of the amino acid in the protein \((n_j/N)\). Hence a value greater than 1 indicates a statistical preference of finding the amino acid in a particular structure.
Table 12.3: Conformational preferences of the amino acids. The highest and lowest preferences are highlighted in red and blue, respectively. From (15).

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>α-helix preference</th>
<th>β-strand preference</th>
<th>Turn preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>1.41</td>
<td>0.72</td>
<td>0.82</td>
</tr>
<tr>
<td>Arginine</td>
<td>1.21</td>
<td>0.84</td>
<td>0.90</td>
</tr>
<tr>
<td>Aspartic Acid</td>
<td>0.99</td>
<td>0.39</td>
<td>1.24</td>
</tr>
<tr>
<td>Asparagine</td>
<td>0.76</td>
<td>0.48</td>
<td>1.34</td>
</tr>
<tr>
<td>Cysteine</td>
<td>0.66</td>
<td>1.40</td>
<td>0.54</td>
</tr>
<tr>
<td>Glutamic Acid</td>
<td>1.59</td>
<td>0.52</td>
<td>1.01</td>
</tr>
<tr>
<td>Glutamine</td>
<td>1.27</td>
<td>0.98</td>
<td>0.84</td>
</tr>
<tr>
<td>Glycine</td>
<td>0.48</td>
<td>0.58</td>
<td>1.77</td>
</tr>
<tr>
<td>Histidine</td>
<td>1.05</td>
<td>0.80</td>
<td>0.81</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>1.09</td>
<td>1.67</td>
<td>0.47</td>
</tr>
<tr>
<td>Leucine</td>
<td>1.34</td>
<td>1.22</td>
<td>0.57</td>
</tr>
<tr>
<td>Lysine</td>
<td>1.23</td>
<td>0.69</td>
<td>1.07</td>
</tr>
<tr>
<td>Methionine</td>
<td>1.30</td>
<td>1.14</td>
<td>0.52</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>1.16</td>
<td>1.33</td>
<td>0.59</td>
</tr>
<tr>
<td>Proline</td>
<td>0.34</td>
<td>0.31</td>
<td>1.32</td>
</tr>
<tr>
<td>Serine</td>
<td>0.57</td>
<td>0.96</td>
<td>1.22</td>
</tr>
<tr>
<td>Threonine</td>
<td>0.76</td>
<td>1.17</td>
<td>0.90</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>1.02</td>
<td>1.35</td>
<td>0.65</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>0.74</td>
<td>1.45</td>
<td>0.76</td>
</tr>
<tr>
<td>Valine</td>
<td>0.98</td>
<td>1.87</td>
<td>0.41</td>
</tr>
</tbody>
</table>
Not surprisingly, there are clear preferences, which we will discuss in the following sections. For example, alanine is strongly favored to be found in helical segments, whereas glycine is much less likely to be within a helix. Residues with β-strands tend to have large, bulky side chains such as valine, isoleucine and tryptophan. Turns (next Section) have very specific conformational restrictions for which proline is well suited, as is glycine, which can attain the required $\phi, \psi$ dihedral angles to fit into a tight turn. The preferences of an individual amino acid by itself is not sufficient to predict with any accuracy whether it is within a particular secondary structural element within a protein structure. However, looking at the preferences in contiguous segments of 6 (α-helix), 5 (β-strand) or 2 (turn) allows one to predict with about 60% accuracy the secondary structure of a globular protein\(^{(15)}\). This tells us that local interactions within contiguous segments play a significant role in determining secondary structure, but additional factors are important.

### 12.3.9 Turns as secondary structural elements

Since globular proteins have defined sizes, there are limits to the length of contiguous stretches of α-helix and β-strand. At the protein surface, one finds turns where the polypeptide chain changes direction. The turn can be dramatic, making a sharp 180° turn to reverses direction, or may be less dramatic and be a “divergent turn”. There are a number of repeating motifs that are found among globular proteins, and turns are classified by the hydrogen bonds (donor→acceptor) between residues: α-turn ($i \rightarrow i \pm 4$), β-turn ($i \rightarrow i \pm 3$), γ-turn ($i \rightarrow i \pm 2$), and π-turn ($i \rightarrow i \pm 5$). The β-turn is also called a “reverse turn”, and is very common. There are a number of classes of reverse turns
defined by their $\phi, \psi$ angles. The Type I and Type II reverse turns are pictured in Figure 12.10. Note the orientation of the peptide bond in the middle is different in these two structures. Glycine is nearly always in position $(i + 2)$ in the Type II reverse turn because it is the only amino acid which can have the necessary $\phi, \psi$ angles.

![Figure 12.10: Type I and Type II reverse turns. The dotted line is a hydrogen bond.](image)

A special turn is found connecting hydrogen bonded, anti-parallel $\beta$-strands (Figure 12.11). For this, the $\beta$-hairpin is used. The $\beta$-hairpin has only two residues, and two types are shown in Figure 12.12. Again, the Ramachandran Plot shows that the second position in the Type I' $\beta$-hairpin can only be glycine.
Figure 12.11: A β-hairpin connects two anti-parallel β-strands

Figure 12.12: Two-residue β-hairpin turns. The white dots represent a hydrogen bond. Note the orientation of the peptide bond in the middle is different for the two classes of hairpin turns.

12.3.10 The α-helix

**Geometric features of the α-helix:** Approximately 35% of residues in globular proteins are in α-helix. The α-helix is defined both in terms of the $\phi, \psi$ angles of each residue and in terms of the hydrogen bonding pattern between the backbone donor NH at
position \((i+4)_{NH}\) to the carbonyl acceptor at position \(i_{C=O}\) (Figure 12.13). The \(\alpha\)-helix is a right handed helix, which means that when the helix is viewed down the helical axis (from either end), a clockwise progression corresponds to moving away from the viewer. The right handed helix is a mirror image of the left handed helix.

Because of the chirality of the amino acids, the right handed and left handed helices are not iso-energetic. The left handed helix, though “allowed” is disfavored energetically and is not observed.

Some of the geometric properties of the \(\alpha\)-helix are summarized in Table 12.2. The repeat unit of the \(\alpha\)-helix is 18 residues, meaning that the 19th residue occupies the same exact position of the first residue, but translated along the helix axis. The repeat unit contains 5 helical turns, so another designation of the \(\alpha\)-helix is the 18\(_5\)-helix. Each turn of the \(\alpha\)-helix requires \(18/5 = 3.6\) residues. Each turn of the helix (3.6 residues) represents a
displacement or rise of 5.4 Å along the helix axis (the “pitch” of the helix), which means that the rise per residue is $5.4/3.6 = 1.5$ Å. Hence, for example, an α-helix 30 Å long would contain 20 residues.

The side chains of the α-helix stick out, angled towards the amino-terminus of the helix (Figure 12.14). If the helix is viewed from the N-terminus down the helix axis, the side chains can be observed and displayed in the form of a helical wheel. An example of this is shown in Figure 12.14. Since it takes 3.6 residues to make a complete turn, each residue is rotated by 100° with respect to its nearest neighbors in the helix. If the residues have polar character on one side and non-polar character on the opposite side, as in the example shown in Figure 12.14, the helix is said to be amphipathic.

![Figure 12.14](http://cti.itc.Virginia.EDU/~cmg/Demo/wheel/wheelApp.html)

Figure 12.14: The side chains in an α-helix stick out to the side, as viewed in a ribbon diagram shown on the left. On the right is a helical wheel view of an α-helix with the 18 residues shown above. This example shows an amphipathic helix, with polar and non-polar residues on opposite sides. The helical wheel is taken from a Java Applet written by Edward K. O'Neil and Charles M. Grisham (University of Virginia in Charlottesville, Virginia): [http://cti.itc.Virginia.EDU/~cmg/Demo/wheel/wheelApp.html](http://cti.itc.Virginia.EDU/~cmg/Demo/wheel/wheelApp.html).
Amphipathic helices are typically found at the surface of globular proteins where one surface faces the solvent and has polar character, and the opposite helical surface faces the hydrophobic core of the protein. In these situations, it is often found that the α-helix is bent away from the solvent, allowing water greater access to hydrogen bond to the backbone C=O groups facing the solvent.

In addition, amphipathic helices are found in membrane proteins, located at the membrane surface and oriented with the helical axis parallel to the plane of the membrane. Hence, one helical surface faces the hydrophobic core of the membrane bilayer, and the opposite helical surface faces the aqueous solvent.

12.3.11 **The α-helix has a permanent dipole moment:** The dipoles of each amino acid (3.5 Debye; Figure 12.6) are all oriented in the same orientation with respect to the helical axis. The carbonyl groups are all essentially pointing towards the C-terminus of the helix (Figure 12.13). As a result, the dipole moments are additive, resulting in a net dipole moment of the α-helix with the amino-terminus being positive and the carboxyl-terminus being negative. The magnitude of the dipole is equivalent to a charge separation of about 0.5 to 0.7 charge units, negative at the carboxyl-terminus. It is frequently found that glutamic acid residues are found at the amino-terminus of an α-helix, stabilized by the positive end of the helical dipole. The α-helix dipole appears also to be functionally important. The amino-terminal region of α-helices is often engaged in binding to negatively charged ligands, such as phosphate.

12.3.12 **Amino acid preferences at the ends of the α-helix(16-18):** The preferences of amino acids to located at the ends of an α-helix are different from the
preferences to be found in the middle. Whereas in the middle of the helix, the backbone hydrogen bonding is fully satisfied (C=O●●●HN), this is not the case at the ends of the helix. For example, Figure 12.15 shows that for several residues at the N-terminus, the hydrogen bond donors (NH) are not partnered with the hydrogen bond acceptors, as they are in the middle of the helix. The opposite is true at the C-terminal region. The energetic cost of having an unsatisfied hydrogen bond is a strong driving force to adopt structures in which these hydrogen bonds are provided by other sources, such as amino acid side chains or water.

![Figure 12.15: At the N-terminal region of an α-helix, several residues lack hydrogen bond acceptors for the backbone NH hydrogen bond donors.](image)

The N-terminal and C-terminal sequence positions are designated as

...N''-N'-Ncap-N1-N2-N3......C3-C2-C1-Ccap-C'-C'...

The residues most often occupying the Ncap position are Gly, Ser, Asn and Asp. The N2 and N3 positions are most often occupied by Glu and Asp. We have already noted that negatively charged residues (Asp, Glu) are disproportionately found near the N-terminal region, rationalized by the positive dipolar helical charge. In addition, hydrogen bonding is satisfied, as shown by the kind of pattern shown in Figure 12.16. One favored
sequence at the N-terminus is Ser-X-X-Glu…where Ser is in the Ncap position. The presence of Gly in the Ncap position allows room for a water molecule to hydrogen bond to the NH group of the Ncap residue.

Figure 12.16: Amino acid side chains in the Ncap and N3 positions often hydrogen bond to the backbone NH groups.

Whereas proline is rarely found in the middle of the α-helix of globular proteins, it is found preferentially in the N1 position. Note that the backbone amino group in proline is unique insofar as it is not a hydrogen bond donor and, hence, suited to the N-terminal region of an α-helix.

At the C-terminal region, where the backbone carbonyl groups require hydrogen bonds, there is a very strong preference Gly specifically at the Ccap position. The presence of glycine allows the Ccap residue to accommodate unusual φ, ψ angles that are in the region of left handed 3_{10} and α-helices, which allow the C=O groups of two successive residues to hydrogen bond to backbone NH groups while, simultaneously changing the direction of the polypeptide chain, breaking the helix.

**12.3.13 Helical distortions:** Although it is unusual to find proline in the middle of an α-helix (Table 12.3), the frequency is higher in transmembrane helices in membrane proteins (19). When proline is within the middle of an α-helix, it generally results in a kink or bend in the helix up to 30°. Although classified as a separate element of secondary structure, the 3_{10} helix can also be viewed as a distorted helix since it is
often found at the ends of α-helices and is rarely longer than one turn long. The name
comes from the fact that there are 3 residues per turn and the ring formed by the
hydrogen bond contains 10 atoms, including the hydrogen. By this nomenclature, the α-
helix is a 3.613 helix. About 3% of residues are in 310 helices, which are clearly less
stable than α-helices.

12.3.14 Parallel and antiparallel β-sheet

The most prevalent secondary structure in globular proteins after the α-helix is β-
sheet. The basic unit of β-sheet is the β-strand (Figure 12.17), which is a fully extended
polypeptide chain. By itself, the β-strand is not a stable structure, but is stabilized when
hydrogen bonded to neighboring β-strands, thus forming a β-sheet. The β-sheet forms the
structural scaffolding in a number of globular proteins. In most cases, the β-sheet is not
planar, but has a right handed twist.
Figure 12.17: A $\beta$-strand is a fully extended polypeptide chain. From (2).

The backbone $\text{C}=\text{O}$ and $\text{NH}$ groups are fully hydrogen bonded as shown in Figure 12.18, and neighboring strands can be either parallel or anti-parallel. These are often mixed, so that a single sheet can contain strands oriented parallel and antiparallel to its neighbors. It is unusual to find parallel $\beta$-sheet containing fewer than 5 strands.

Figure 12.18 also shows the “pleated” nature of the $\beta$-sheet, with alternating residues having their side chains pointing towards to alternating side of the sheet. As with $\alpha$-helices, one can also have amphipathic $\beta$-sheet, with one side having polar character and the opposite side non-polar.
Because of the extended nature of the polypeptide chains in β-strands, the amino acids that are preferentially found in the middle of such structures have large aromatic or aliphatic branched side chains, Trp, Tyr, Ile and Val (Table 12.3). The distance between adjacent residues within each β-strand is about 3.3 Å, much more than the 1.5 Å in the α-helix. Hence, every “repeat” of 2 residues extends about 6.6 Å. The distance between neighboring strands is about 4.5 Å (see Table 12.2).

12.4 Tertiary structure, protein folds and domains

Secondary structural elements within individual polypeptide chains interact to define the tertiary structure of a protein. As discussed in Section 12.2, common folding
patterns, many with diagnostic sequence motifs, are observed. In this section we will show a few examples.

**12.4.1 Example: The immunoglobulin fold, an all-β protein.**

The immunoglobulin fold is found in many proteins and consists of a sandwich of two β-sheets with 7 antiparallel β-strands. Figure 12.19 shows the structure of the immunoglobulin domain that is part of the protein fibronectin,

![Figure 12.19: The immunoglobulin fold (fibronectin type III domain) from human tenascin (PDB accession code 1TEN), colored from blue (N-terminus) to red (C-terminus). From http://en.wikipedia.org/wiki/Immunoglobulin_fold](http://en.wikipedia.org/wiki/Immunoglobulin_fold)

Note that the strands cross from one sheet to the other so that a single layer is not made up of contiguous sequence. Typically, when β-sheets interact, as in this sandwich, the side chains do not intercalate because the distance between adjacent strands within a sheet is too small (4.5 Å). The crossing angle is either nearly orthogonal, as shown here, or aligned, with an angle of about 30°.
12.4.2 Example: β-barrels.

There are several variations of the β-barrel folding pattern. They all involve a β-sheet which wraps around to form a cylinder. Figure 12.20 shows the retinol binding protein, in which the bound ligand is held within the center of the barrel. The side chains facing outward are hydrophilic and those facing inward are hydrophobic, forming the inner core of the protein.

![Figure 12.20: The human retinol binding protein with retinol bound in the center. This is an 8-strand β-barrel. From http://en.wikipedia.org/wiki/File:Rbp_1brp.png](http://en.wikipedia.org/wiki/File:Rbp_1brp.png)

The β-barrel is also one of the designs for membrane proteins, particularly proteins that are found in the outer membrane of Gram-negative bacteria. Figure 12.21 shows the structures of two outer membrane proteins, OmpA and Porin, along with the structure of the toxin α-hemolysin, which is a soluble protein that inserts into the target membrane and creates a hole, allowing the inner contents to pour out. The β-barrel
membrane proteins are arranged with hydrophobic residues facing the lipid bilayer, on the outside, with hydrophilic residues facing the inside, forming, in many cases, a water-filled channel or funnel-shaped cavities for transport of water-soluble metabolites across the outer membrane.

![Figure 12.21: The structures of outer membrane proteins OmpA and porin from *Rhodobacter capsulatus* (20) and α-hemolysin from *Staphylococcus aureus* (21). In each case, there is a ring of aromatic residues that are located at the level of the head groups of the lipids in the bilayer. From (22).]

### 12.4.3 Example: α-helical coiled coils

Unlike the packing between β-sheets, the packing observed between α-helices usually involves the interdigitation of the side chains (Figure 12.22). One example that is observed in several groups of proteins is the coiled-coil motif. This is the basic structural motif of a number of α-coiled fibrous proteins such as tropomyosin and the α-keratins, which are principle components of skin, hair and nails.
Figure 12.22: Schematic of the two helix coiled-coil showing (left) the two helices with interdigitating hydrophobic residues at their interface and, on the right, the helical wheel view showing the pattern of the repeating heptad. Frequently, the residues labeled \( e \) and \( g \) have opposite charges on the opposing helices, resulting in strengthening the interactions. Modified from (23).

The same structural motif is found in a large group of transcription factors such as JUN, FOS and GCN4(24) (Figure 12.23). The sequences of coiled coils all contain a repeating heptad, illustrated in Figure 12.22. Starting at the N-terminus, the heptad has hydrophobic residues at positions \( a \) and \( d \). The \( \alpha \)-helix makes one complete turn every 3.6 residues, and these hydrophobic residues occur on average every 3.5 residues. Hence, the hydrophobic residues form a stripe along one side of the helix. Because the spacing is not exactly 3.6 residues, the stripe is at an angle to the helix axis (about 20°). In the case of the transcription factors, the most frequently found hydrophobic residue in positions \( a \) and \( d \) is leucine, and this motif is called the “leucine zipper”. The two helices shown in Figure 12.3 wrap around each other forming a left-handed coil (counter-clockwise rotation going away from the viewer) with leucines predominating.
along the contact stripe. In the case of the fibrous proteins such as \( \alpha \)-keratins, the coiled coil can be hundreds of Ångstroms long and other non-polar residues are also found at the helical contact regions.

**Figure 12.23**: (A) The structure of the leucine zipper from the GNC4 transcription factor; (B) A schematic of the role of the leucine zipper domain to form dimers of subunits that also contain a separate DNA binding domain. Part A is from (24), and Part B is from (25).

**Figure 12.24** illustrates another way of displaying the \( \alpha \)-helix which also shows the “knobs-into holes” packing of the side chains that is typical of helix-helix interactions. This diagram is made by wrapping a piece of paper around each helix, forming a cylinder, and marking the positions of the C\( \alpha \) atoms of each side chain. The cylinder is slit and the paper removed, and the representations of the two helices are placed together in a manner reflecting the dimer interface. In **Figure 12.24**, the side chains of one helix are represented by filled circles and those of the second helix by open circles. The heptad positions (a through g) are labeled. Two regions at the top of **Figure 12.24** which form part of the contact interface are enclosed by solid and dashed lines. Note that the side chains of one helix fit into “holes” between the side chains of the second helix.
Figure 12.24: A split cylinder view of the way in which two $\alpha$-helices interact in the GNC4 leucine zipper. Note how the side chains of one helix (for example, 2a) fits into a space in between residues of the second helix (1d', 1g', 2a' and 2d'). From (24).

12.4.4 Example: The helix-turn-helix homeodomain

One DNA binding domain that is found in eukarotic organisms is the homeobox, or helix-turn-helix domain. This domain contains three $\alpha$-helices, two antiparallel and one virtually perpendicular, as shown in Figure 12.25.
12.4.5 Example: The chloride channel from *Salmonella typhimurium*

With the exception of the outer membrane β-barrels discussed in Section 12.4.2, all other transmembrane proteins consist of bundles of α-helices. Most of these helices cross the entire membrane but there are exceptions. Another feature is that the transmembrane α-helices are frequently tilted with respect the the membrane normal. An example is the chloride channel from *S. typhimurium* (26). This structure, along with that of the potassium channel also illustrates the functional use of the α-helix dipole, discussed in Section 12.3.11. **Figure 12.26** shows the structure of the ClC Cl⁻ channel, which is a homodimer, with each subunit containing 18 α-helices.

![Figure 12.25: The helix-turn-helix antennapedia homeodomain protein from drosophila melanogaster bound to a fragment of DNA (PDB ID 1AHD). The recognition helix is the one binding to the major groove of the DNA. From http://en.wikipedia.org/wiki/File:Homeodomain-dna-1ahd.png](http://en.wikipedia.org/wiki/File:Homeodomain-dna-1ahd.png)
Figure 12.26: Structure of the ClC chloride channel from *S. typhimurium*. Each of the two identical subunits (red and blue) contains 18 α-helices, and each has a channel for chloride transport. (A) top view; (B) side view. The approximate width of the hydrophobic part of the lipid bilayer is indicated by the solid line. From (26).

Figure 12.27 schematically illustrates the chamber within this membrane protein in which negatively charged chloride ions are stabilized by the positively charged N-termini of helices. The opposite is observed in the structure of the potassium channel, in which the C-termini of α-helices contribute to the stabilization of positively charged potassium ions in an internal chamber.
Figure 12.27: Schematic showing the way in which helical dipoles are used to stabilize ions within (A) the chloride channel; and (B) the potassium channel. From (26).

12.4.6 Comparing the helix-helix interface in soluble proteins and membrane proteins (27).

In soluble proteins which consist of bundles of $\alpha$-helices, the pattern of hydrophobic residues packing at the interface between two helices is similar to that observed with the coiled-coils discussed in Section 12.4.3. The most prominent residues present are Leu, Ala, Ile and Val. More polar residue are found on the surface of the protein, accessible to solvent. Membrane proteins have two different motifs for helix interactions. The first is similar to that found for the helix-helix interactions with soluble proteins, with Leu, Ala, Ile and Val frequently located at the interface, and packed together in a knobs-into-holes manner, as diagramed in Figure 12.25. This is illustrated
in Figure 12.28, which shows the helix-helix interactions between the M2 and M6 transmembrane helices of the Ca\(^{2+}\)-ATPase. However, there is a second motif for helix-helix interactions found in membrane proteins in which small and polar residues form the interhelical contacts. The favored residues are Gly, Ala, Ser and Thr. The small residues allow the backbones of the two helices to approach more closely (\(\leq 6\) Å) with stronger van der Waals interactions.

![Figure 12.28: Examples of the helix packing motifs of membrane proteins.](image)

(A) A motif similar to that found in coiled-coil proteins and in soluble proteins with helical bundles. Large hydrophobic residues predominate and interdigitate. This is from the M2 and M6 transmembrane helices of the Ca\(^{2+}\)-ATPase (1EUL). (B) A motif often found in membrane proteins with small and polar residues at the contact region forming a flat interface with close interactions between the backbone residues of the two helices. This is from two transmembrane helices of the glycerol facilitator channel. Modified from (27).

In addition, the Ser and Thr residues can form interhelical hydrogen bonds as well as intrahelical hydrogen bonds. An example is shown in Figure 12.28, showing portions of the M2 and M6 transmembrane helices from the glycerol facilitator channel. Note that
the contact region in the more polar interhelical motif appears flat and smooth, in contrast to the rough surface of the “knobs” in the traditional helical interface. In membrane proteins, the surface of the protein faces the lipid bilayer and the residues tend to be hydrophobic. Hence, the side chains tend to be more polar on the inside compared to the exterior surface.

12.4.7 Example: The $\beta\alpha\beta$ fold and the Rossmann fold

A very common fold pattern that combines both $\alpha$ and $\beta$ structures is the $\beta\alpha\beta$ fold. The helix lies over the plane formed by the $\beta$-sheet (Figure 12.29). Extending this to $\beta\alpha\beta\alpha\beta$ gives a fold that is often present in nucleotide binding proteins, called the Rossmann fold. Figure 12.29 shows the topology of the Rossmann fold as well as an example from the structure of the domain binding to the mononucleotide FMN. Enzymes that bind to dinucleotides such as NADH and FAD typically have two Rossmann folds.
12.4.8 Example: The α/β-barrel or TIM fold

Extending the βαβ fold even further to include 8 α-helices and 8 β-strands, and hydrogen bonding the first and last of the β-strands gives another common fold known as the TIM barrel or α/β-barrel. This was first observed in, and named after, triose phosphate isomerase (TIM). A very large superfamily of enzymes utilize this fold (28). The active sites of the enzymes are always formed by the loops connecting to the carboxyl ends of the β-strands. Two views of the TIM barrel is shown in Figure 12.30.

12.5 The inside of the protein is tightly packed:

As a rule, regardless of the domain structure, globular proteins are very tightly packed. Two points are worth emphasizing.
1. The packing of atoms in the protein interior is similar to that found in crystals of small organic molecules. If each atom is considered to a hard sphere (like a billiard ball) with a radius equal to the van der Waal’s radius, then the amount of space occupied by the atoms inside a protein corresponds to about 0.75 of the total internal volume. This is about equal to the fraction of volume occupied by the closest packing of hard spheres (71%). When packing billiard balls into a box, about 25% of the volume is the space between the balls. For comparison, in a typical liquid, the atoms occupy about 45% of the available space. The inside of a protein is not like a liquid. The tight packing means that the van der Waal’s forces between neighboring groups will be maximized and serve as a contribution to stabilize the folded protein.

2. Nearly every hydrogen bond donor and hydrogen bond acceptor inside a protein has a hydrogen bond partner. More than 90% of groups that can form hydrogen bonds, do form hydrogen bonds inside the protein. In the predominant secondary structural elements, \(\alpha\)-helix and \(\beta\)-sheet, the backbone hydrogen bonds are satisfied internally. Recall also that the capping motifs at the end of the \(\alpha\)-helix also involves hydrogen bonding. Clearly, proteins are organized to avoid leaving any hydrogen bond donor or acceptor without a binding partner.

12.6 Proteins are hydrated:

At sufficiently high resolution, the X-ray structures of proteins include a number of water molecules that are bound to the protein. Most of these water molecules are located at the surface, hydrogen bonded to either polar side chain, or to the water-accessible backbone amide group or to other water molecules anchored to the protein(29). Figure 12.31 shows the structure of plastocyanin, with water molecules
observed on the surface. The organized water that is observed does not extend beyond the first shell. One layer of water around a globular protein amounts to a mass of about 0.2 to 0.3 g of water per gram of protein, and when a protein moves in solution, as in a centrifuge or by diffusion, this amount of water moves with it. The fact that waters of hydration are apparent in the X-ray structures indicates that these water molecules are ordered, i.e., the average position of each water oxygen atom is relatively well defined and the fractional occupancy in the observed position is high. However, an important point is that the surface water molecules are highly dynamic and exchange with the solvent water on a fast (nsec) time-scale(30).

Figure 12.31: The structure of plastocyanin (left) and a surface representation of showing the waters of hydration.

Water is often found inside proteins, where it can serve both structural and functional roles. Figure 12.32 shows an example of a structural water molecule that is conserved in three subfamilies of ribonuclease.
12.7 Quaternary structure and protein-protein interactions (32, 33)

When proteins form a complex, a considerable surface area is buried at the interface, reducing the water-accessible surface area. It is useful to divide the protein complexes into two groups: transient vs obligate complexes. The transient complexes are generally those in which each protein partner has an independent biochemical function and they exist as independent, non-complexed species. Redox partners which briefly associate and then dissociate after electron transfer would be in this set, for example, as are antibody-antigen complexes. These complexes tend to be relatively weak and to have rapid dissociation rates. The obligate complexes are those where the affinity is very strong with very slow dissociation rates.

Figure 12.33 shows the distribution of the amount of buried surface area calculated from known structures. The high affinity complexes (obligate) have considerably more buried surface area than do the transient complexes. For transient complexes, the buried surface has the same composition and quality as does the overall surface composition. For obligate complexes, the buried surface tends to be have a higher
proportion of non-polar side chains, though these are not clustered. In both cases, there are hydrogen bonds that form between residues in the two protein partners, and in a number of instances, water molecules are intermediaries in these hydrogen bonds. Hence, water plays an important role in protein-protein interactions and is not totally squeezed out of the interface.

**Figure 12.33**: The composition of the buried interface in protein-protein complexes for (A) transient, weak complexes; and (B) obligate, strong complexes. The white regions of the bars on the right are from weak homodimers, and the grey and black portions are from different compilation of structures. Modified from (32).

The interface regions can be further subdivided into a core and rim based on structural criteria. The core residues are those which are totally buried and inaccessible to water. The rim residues are partially accessible to water. **Figure 12.34** shows one example, showing the surface of elastase which binds to the ovomucoid inhibitor. The “dry” core region is in red, surrounded by the rim (blue). Panel B shows the extent of residue conservation in related proteins, showing that residues in the core tend to be more conserved than those in the rim.
Analysis by site-directed mutagenesis shows that many residues at the protein-interface can be mutated without much cost to the interaction free energy. Some, but not all protein-protein interfaces have “hotspots” where mutations have a much larger perturbing effect. These tend to be in the core region of the interface.

Figure 12.34: The surface of elastin that is buried upon complex formation with the ovomucoid inhibitor. (A) Core residues totally inaccessible to water are in red, and rim residues, partially in contact with water, are in blue. (B) Red residues are those which are more highly conserved in sequence alignments with related proteins. Blue residues are less conserved. From (32).


