Contents I-III

Contents IV-V

Protein structure

The great complexity of proteins makes it unlikely that a complete structure determination for a protein will ever be made by X-ray methods alone.
Linus Pauling (1939) *J. Am. Chem. Soc.* **63** 1860

Molecular form defines function: huge variety in shapes and size

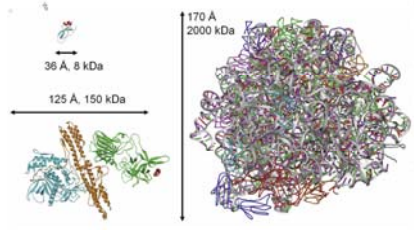
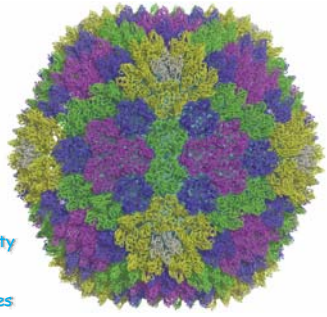


Figure 2-1 Protein structures determined by X-ray crystallography. The small protein (top left) is the bovine pancreatic trypsin inhibitor (58 residues, ~8 kDa), one of the first proteins whose structure was refined at atomic resolution.[4] Its two distinct secondary structure elements are shown in ribbon presentation (a red α-helix and two cyan anti-parallel β-strands).The larger molecule (bottom left, ~1300 residues, ~150 kDa) is the neurotoxin secreted by the bacterium Clostridium botulinum. Botulinum neurotoxins are the most toxic substances known to man. They consist of three functionally distinct domains: a Zn-protease (cyan ribbon), a translocation domain (orange), and the ganglioside-binding domain (green), which is shown bound to a recognition peptide from the neuronal cell surface (red helix).[5] The right panel shows the large (50S) ribosomal subunit of the bacterium *Haloarcula marismortui*. The huge, nearly 2 MDa large structure[6] contains 27 different proteins (~4000 residues) and the ribosomal 5S and 23S RNA, together having 2833 nucleotides. PDB entries 1bpi,[4] 3bta,[5] and 1ffk.[6]

X-ray crystallography provides these structures without size limitation and often at atomic detail level and relevant beyond solid state – with some caveats

Virus capsids are protein 'spheres'

Figure 2-2 The protein capsid of the 66 MDa head of the membrane-containing bacteriophage PRD1. The virion particle contains ~2000 subunits from 18 different proteins. Enclosed inside the head capsid are a host-derived lipid bilayer of ~12 500 lipid molecules and a core of double-stranded DNA. The crystals diffracted to 4 Å with unit cell constants of nearly 1000 Å, and the virion particle itself measures 640 Å across. Image courtesy of Nicola Abrescia, David Stewart, and Jonathan Grimes, Oxford University. PDB entry 1w8x.[7]



One of the intriguing aspects of X-ray crystallography is the ability to determine huge structures of viruses or multi-unit protein complexes

Hierarchy and organization of protein structure

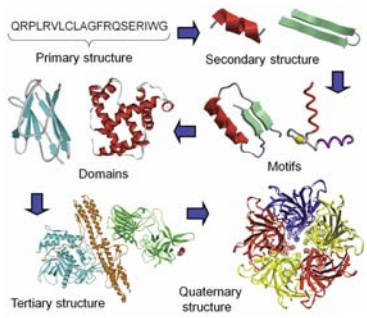


Figure 2-3 The hierarchy of protein structure organization. From a chain of defined amino acid sequence or primary structure, secondary structures form through distinct backbone interactions. Several secondary structure elements combine through side chain interactions into smaller structural motifs, which together with other secondary structure elements form distinct, independently folding structural domains. From the same protein chain, several domains with different functions can fold, assembling into the complete tertiary structure. Finally, several protein chains can form functionally important homo- or hetero-oligomeric assemblies, defining the quaternary structure.

Sequence-primary
Secondary
Motifs
Domains
Tertiary
Quaternary

Molecular geometry is determined by distances, angles, and torsions

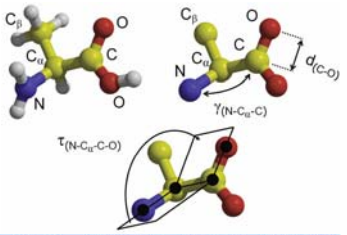


Figure 2-4 Definition of bond length, bond angle, and torsion angle. A ball and stick representation of the amino acid alanine is labeled with atom names and shows the definition of bond lengths, bond angles, and torsion angles. The color scheme applied to distinguish atoms is a common default scheme used by protein model display programs (carbon atoms yellow, oxygen red, nitrogen blue). Hydrogen atoms in riding positions are explicitly shown only in the top left panel (gray atoms) and are normally omitted. Because the torsion angle can be interpreted as the angle between two planes, torsion angles are also called dihedral angles.

Peptide bond	Average length (Å)	Single bond	Average length (Å)	Hydrogen bond	Average length (Å)
Cα-C	1.525 ± 0.021	C-C	1.540 ± 0.027	O-H · · · O-H	2.8 ± 0.2
C-N	1.329 ± 0.014	C-N	1.489 ± 0.030	N-H · · · O=C	2.9 ± 0.2
N-C	1.458 ± 0.019	C-O	1.420 ± 0.020	O-H · · · O=C	2.8 ± 0.2

Table 2-1 Common bond lengths. Mean values and standard deviations of some covalent bond lengths and hydrogen bonds commonly occurring in protein structures. The average or mean values are derived from accurate small molecule X-ray structures and high resolution protein structures. Because they also serve as the basis for stereochemical restraints in macromolecular refinement (Chapter 12), the mean values are also known as restraint target values.[1]

Cα -chirality – the CO-R-N rule

Sidebar 2-2 Hydrogen atoms in protein structure models. In protein crystallographic work, the hydrogen atoms are generally only visible in electron density at ultra high resolution (about 1.2 Å or better). Because their so-called *riding* positions are known and can be calculated when needed, they are normally omitted in crystallographic models (the presence of hydrogen atoms is, however, implicitly and fully accounted for in crystallographic distance and energy restraints). The remaining panels of Figure 2-4 therefore show alanine in a representation typical for a crystallographic model, *sans* hydrogens. We will generally omit hydrogen atoms in riding positions in protein structure models, unless they are needed to emphasize a structural or stereochemical feature.

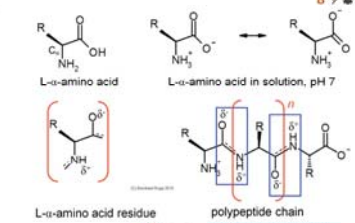
Biomolecular Crystallography B.Rupp (2010)

Figure 2-5 The CORN rule. The CORN rule is a practical aid for determining the configuration of chiral Cα centers of amino acids. When the central Cα-atom is viewed with the H atom pointing toward the observer (or out of the paper plane), the ligand sequence reads "CO-R-N" in clockwise rotation for a L-amino acid.

With exception of non-chiral glycine, all proteinogenic amino acids are chiral with 2S configuration – L- amino acids

Primary structure, sequence, peptide bond

Figure 2-19 Definition of L-α-amino acid, residue, and polypeptide chain. Hydrogen atoms at the chiral Cα atom are omitted. Top left: L-α-amino acid in uncharged state. In solution, the functional terminal groups are charged, and because of resonance the two oxygen atoms are equivalent. Bottom row: a residue of an L-α-amino acid. A residue lacks the second carboxyl oxygen and the second hydrogen atom on the amide group, which are lost during formation of the peptide bonds. The bottom right panel shows the formal composition of a polypeptide chain of n residues, with the blue boxes containing the planar peptide bonds.



99.9% non-Pro trans

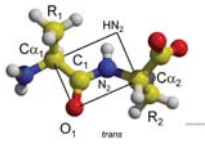


Figure 2-8 Trans- and cis-peptide bonds. The trans-peptide bond, recognizable by the two sequential Cα atoms opposed, has an ω-angle of 180° while the cis-peptide bond with an ω-angle of 0° can be easily recognized by the Cα atoms located on the same side of the peptide bond. In both cases, the peptide bond is planar and contains the same atoms, but in cis conformation (right panel) the torsional freedom is much more limited because of side chain collisions. The Cα-Cα distance in a cis-peptide is ~2.9 Å, while the Cα-Cα distance across a trans-peptide bond is ~3.8 Å.

Protein backbones prefer *trans* geometry

Figure 2-8 Trans- and cis-peptide bonds. The trans-peptide bond, recognizable by the two sequential Cα atoms opposed, has an ω-angle of 180° while the cis-peptide bond with an ω-angle of 0° can be easily recognized by the Cα atoms located on the same side of the peptide bond. In both cases, the peptide bond is planar and contains the same atoms, but in cis conformation (right panel) the torsional freedom is much more limited because of side chain collisions. The Cα-Cα distance in a cis-peptide is ~2.9 Å, while the Cα-Cα distance across a trans-peptide bond is ~3.8 Å.

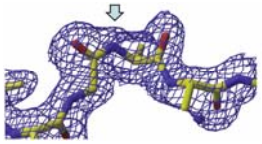
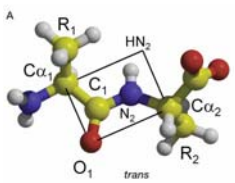
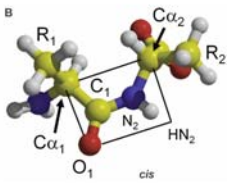


Figure 2-9 Cis-peptide bond and its electron density. A cis-peptide bond between Gly and Ala in a connecting loop is shown in experimental electron density (blue grid). Both the clear electron density and the fact that the cis conformation is structurally conserved in this particular fold family allow making this particular cis assignment with confidence. PDB entry 1ugs [1]





Secondary structure is defined by van der Waals repulsion and backbone torsion – not side chains!

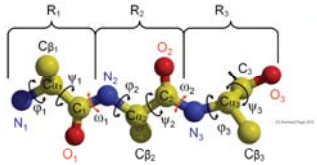
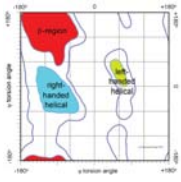


Figure 2-7 Backbone torsion angles. The N-terminal 3-residue stretch of a peptide Ala-Ala-Ala- containing three trans-peptide bonds is shown. Three torsion angles for each residue, φ (phi), ψ (psi), and ω (omega), define the conformation of the peptide backbone. While combinations of torsion angles φ and ψ are only restrained by van der Waals repulsion and fall into several allowed, energetically favored regions, the trans-omega-torsion around the partially delocalized, planar peptide bond (short red arrow) is highly restrained to 180°.



- Ramachandran energy surface
- outliers indicate points of high local energy
- ~2-5% high energy conformations are normal
- not restrained in refinement – will reliably reveal errors