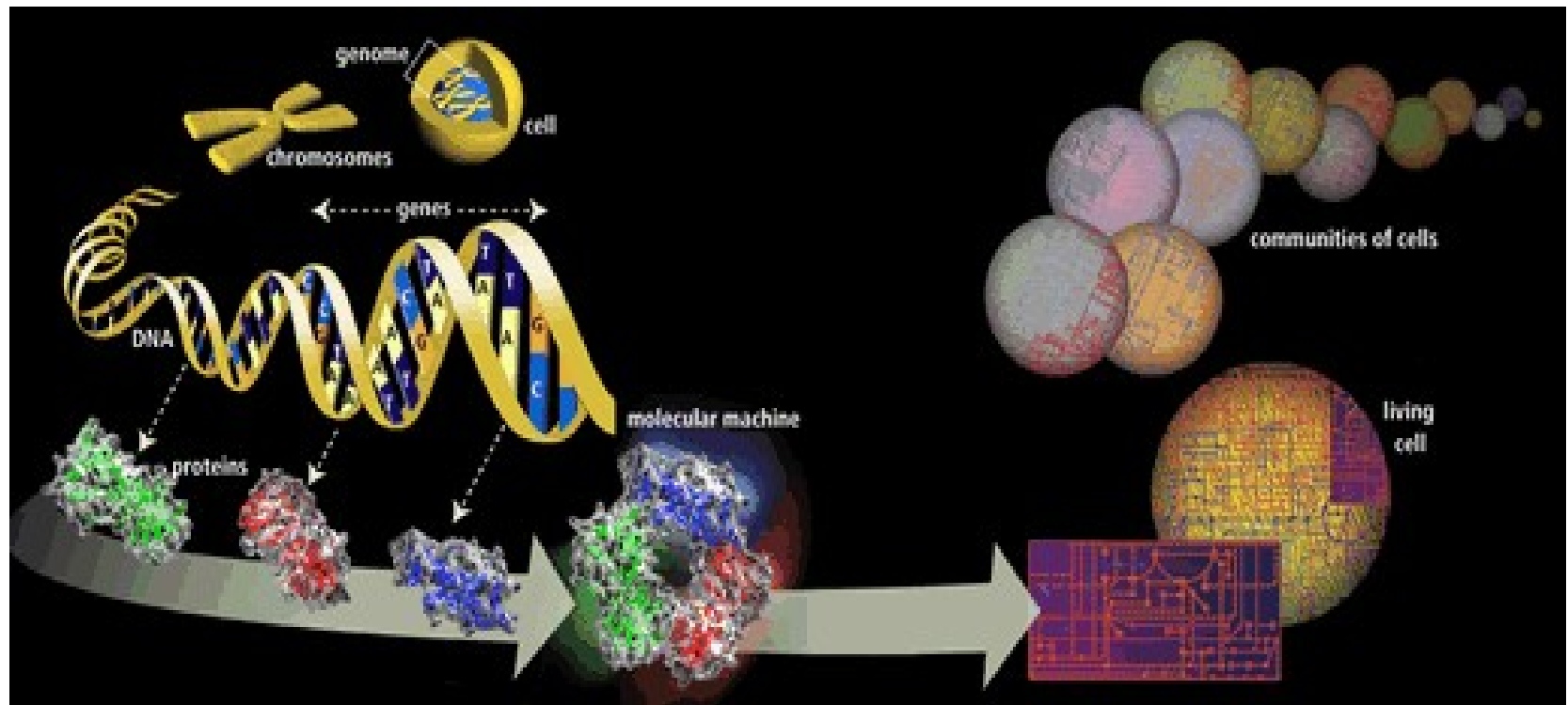
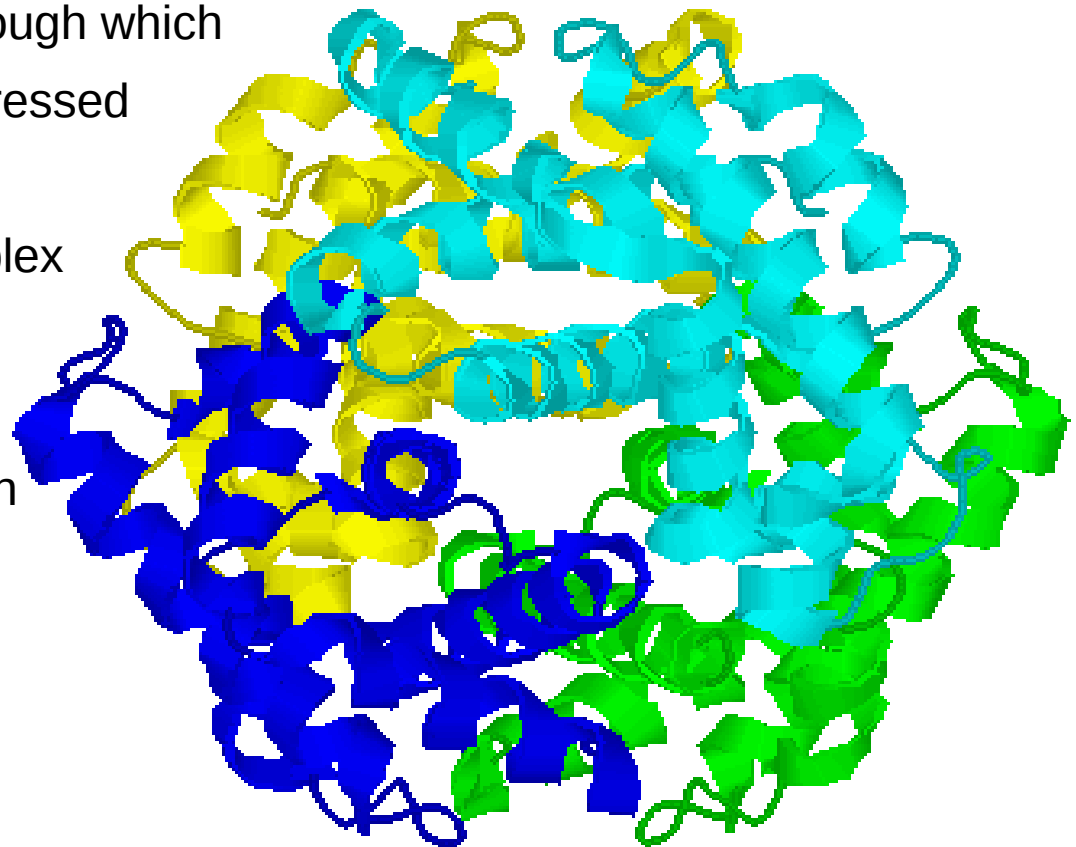


Protein Architecture



Foremost...

- Protein – derived from prótos = first, foremost
- Molecular instruments through which genetic information is expressed
- Workhorses of the body
Carry out all sorts of complex biological functions
- 20,000 to 25,000 protein-coding genes in the human body



First Some Definitions...

■ Protein

Polymer built up by amino acids linked into a peptide chain

■ Amino acid

Alpha-amino carboxylic acid

■ Residue

Amino acid within a peptide chain

(more generally, a building block of a protein molecule)

■ Peptide bond

amide bond between the carboxylic acid group and the alpha-amino group

links two amino acids together

■ Peptide

Two amino acids covalently linked by an amide bond

■ Polypeptide

Long chain of peptides

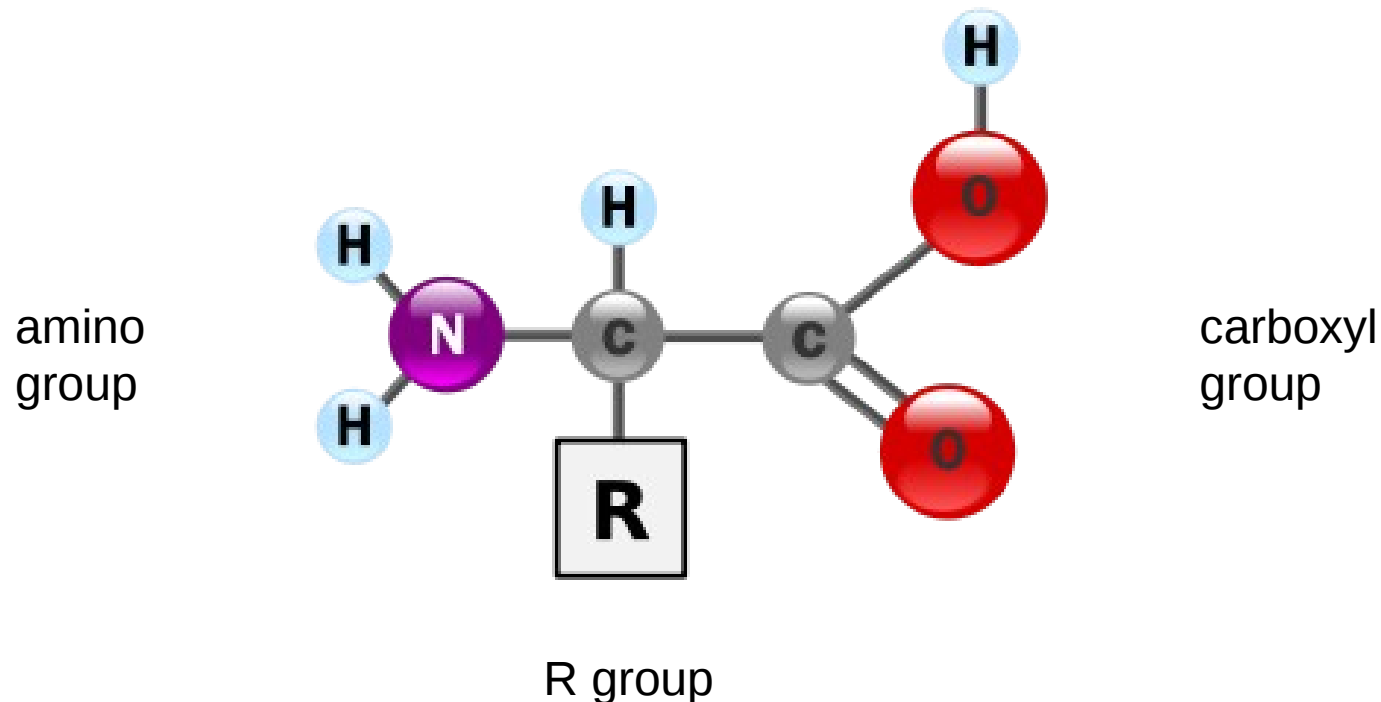
■ Protein

Macromolecule of polypeptides

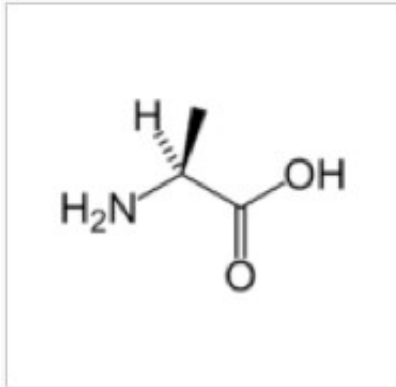
(one or more polypeptide chains)

Amino Acids

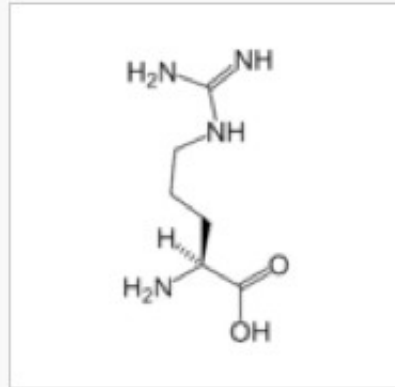
- Amino acids are the building blocks of proteins
- There are 20 common (classic) amino acids coded for in DNA
- All amino acids contain the same generic base (backbone atoms)
- The R-group varies between the different amino acid residues



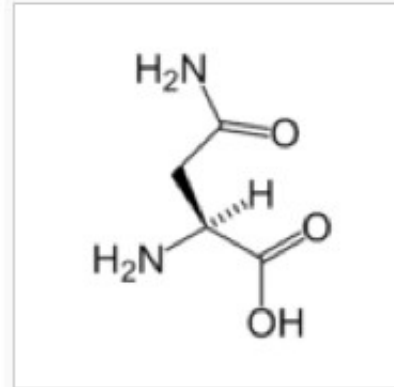
20 Common Amino Acids



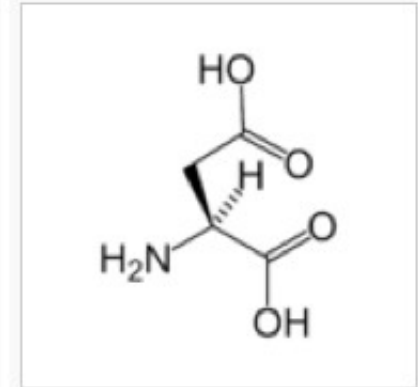
L-Alanine
(Ala / A)



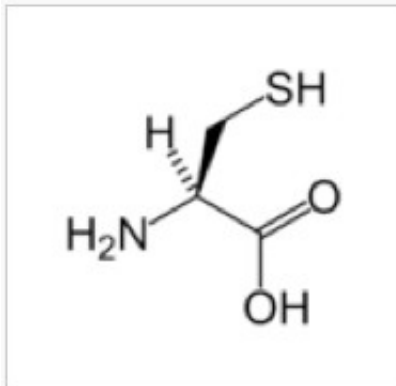
L-Arginine
(Arg / R)



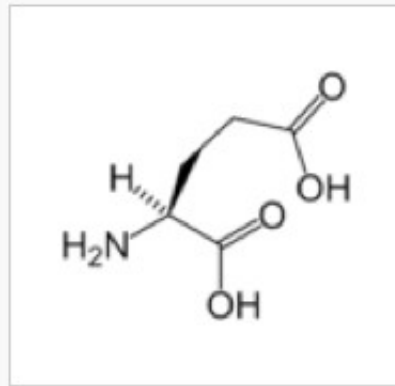
L-Asparagine
(Asn / N)



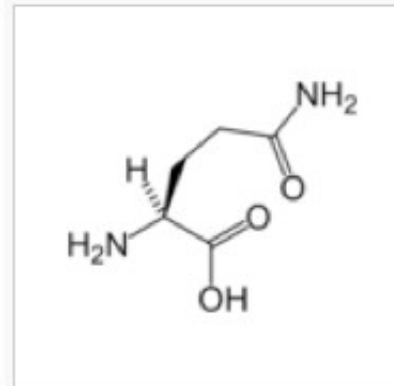
L-Aspartic acid
(Asp / D)



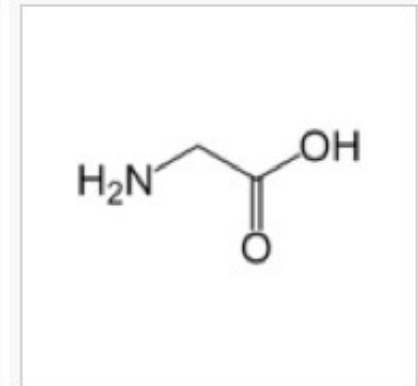
L-Cysteine
(Cys / C)



L-Glutamic acid
(Glu / E)

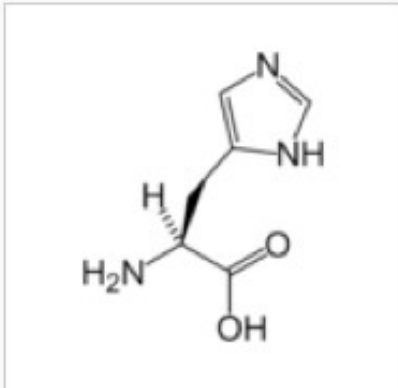


L-Glutamine
(Gln / Q)

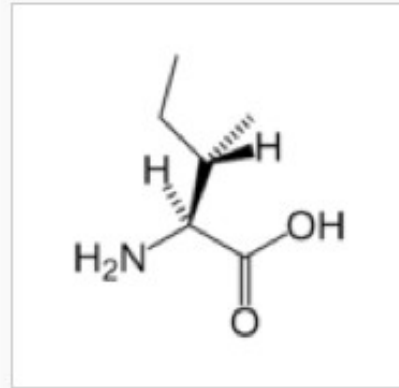


Glycine
(Gly / G)

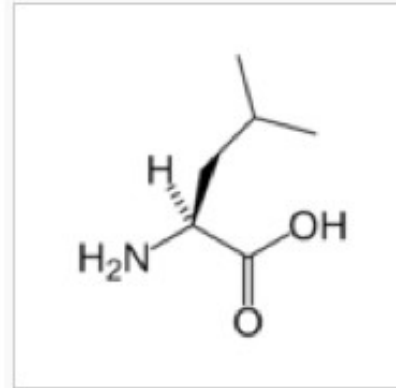
20 Common Amino Acids



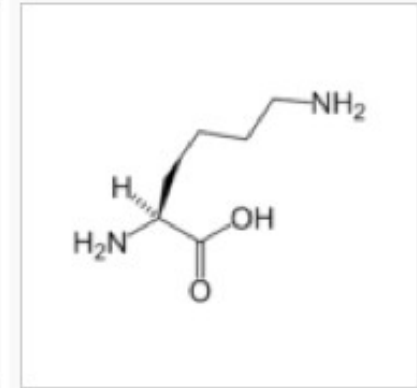
L-Histidine
(His / H)



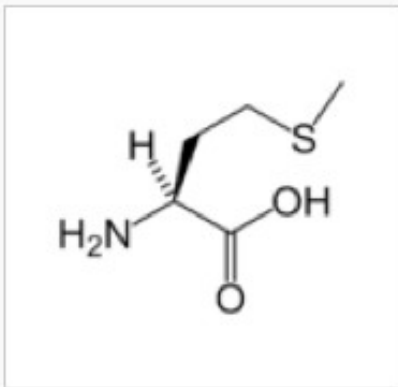
L-Isoleucine
(Ile / I)



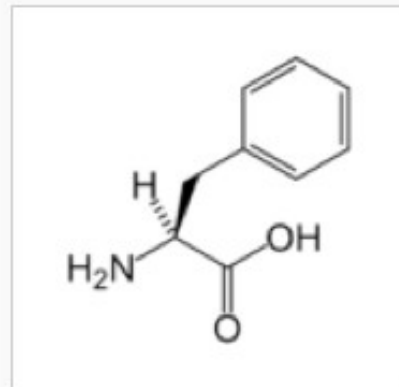
L-Leucine
(Leu / L)



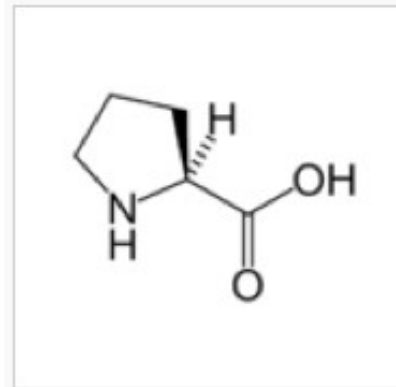
L-Lysine
(Lys / K)



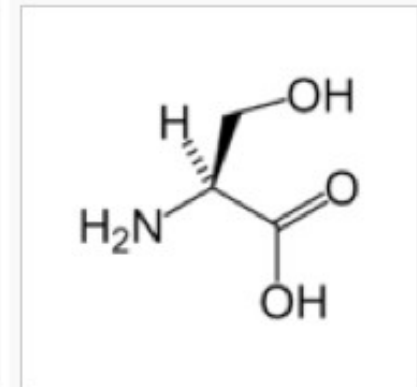
L-Methionine
(Met / M)



L-Phenylalanine
(Phe / F)

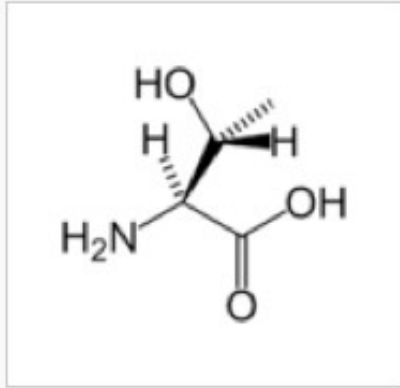


L-Proline
(Pro / P)

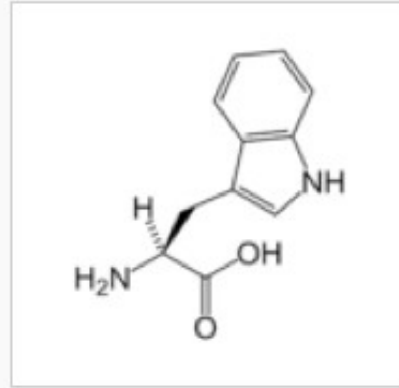


L-Serine
(Ser / S)

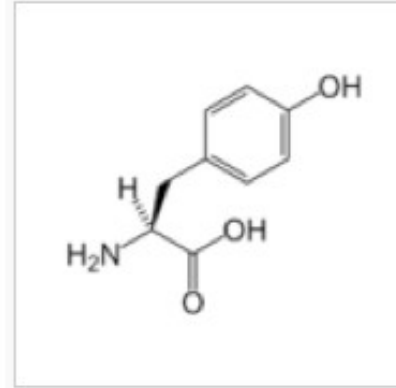
20 Common Amino Acids



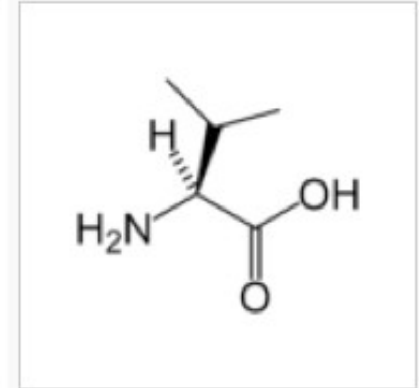
L-Threonine
(Thr / T)



L-Tryptophan
(Trp / W)

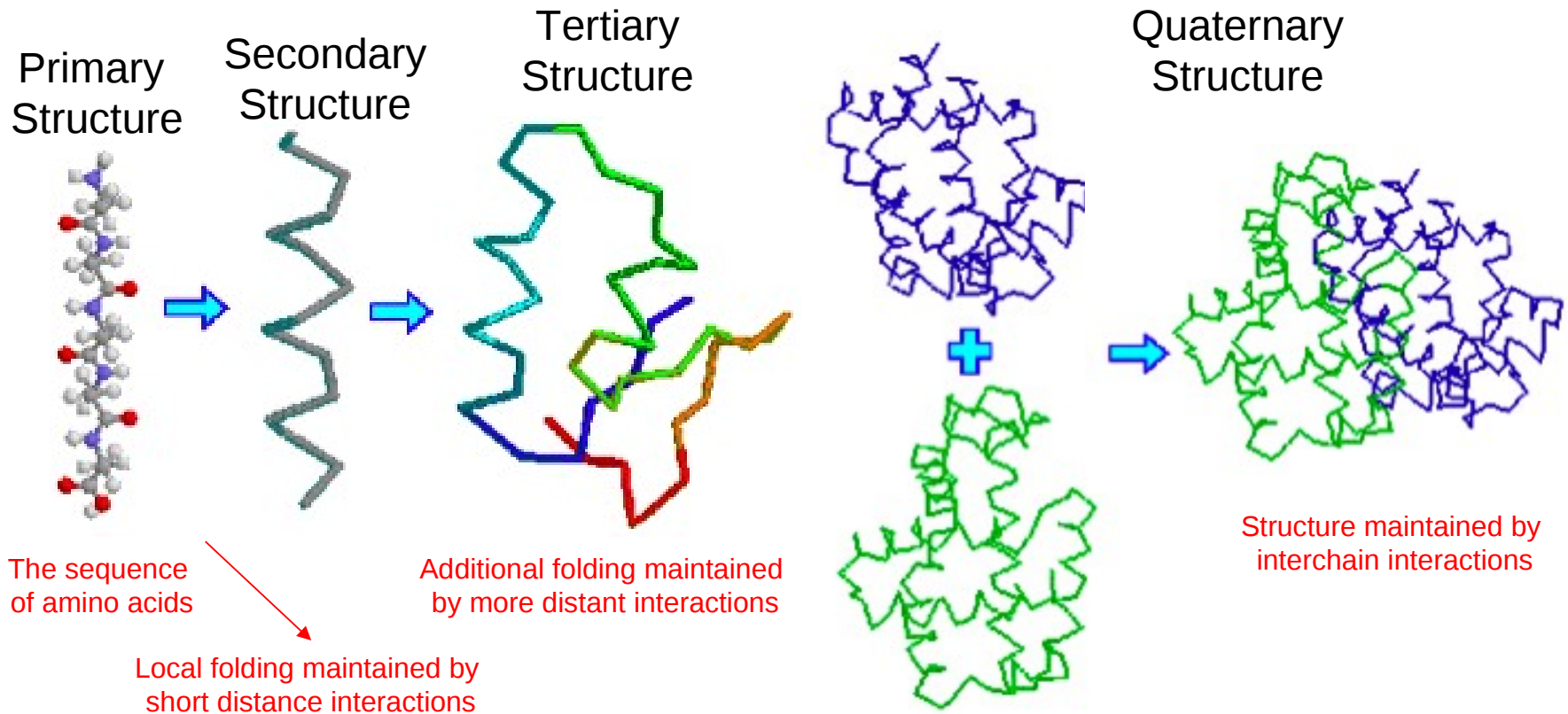


L-Tyrosine
(Tyr / Y)

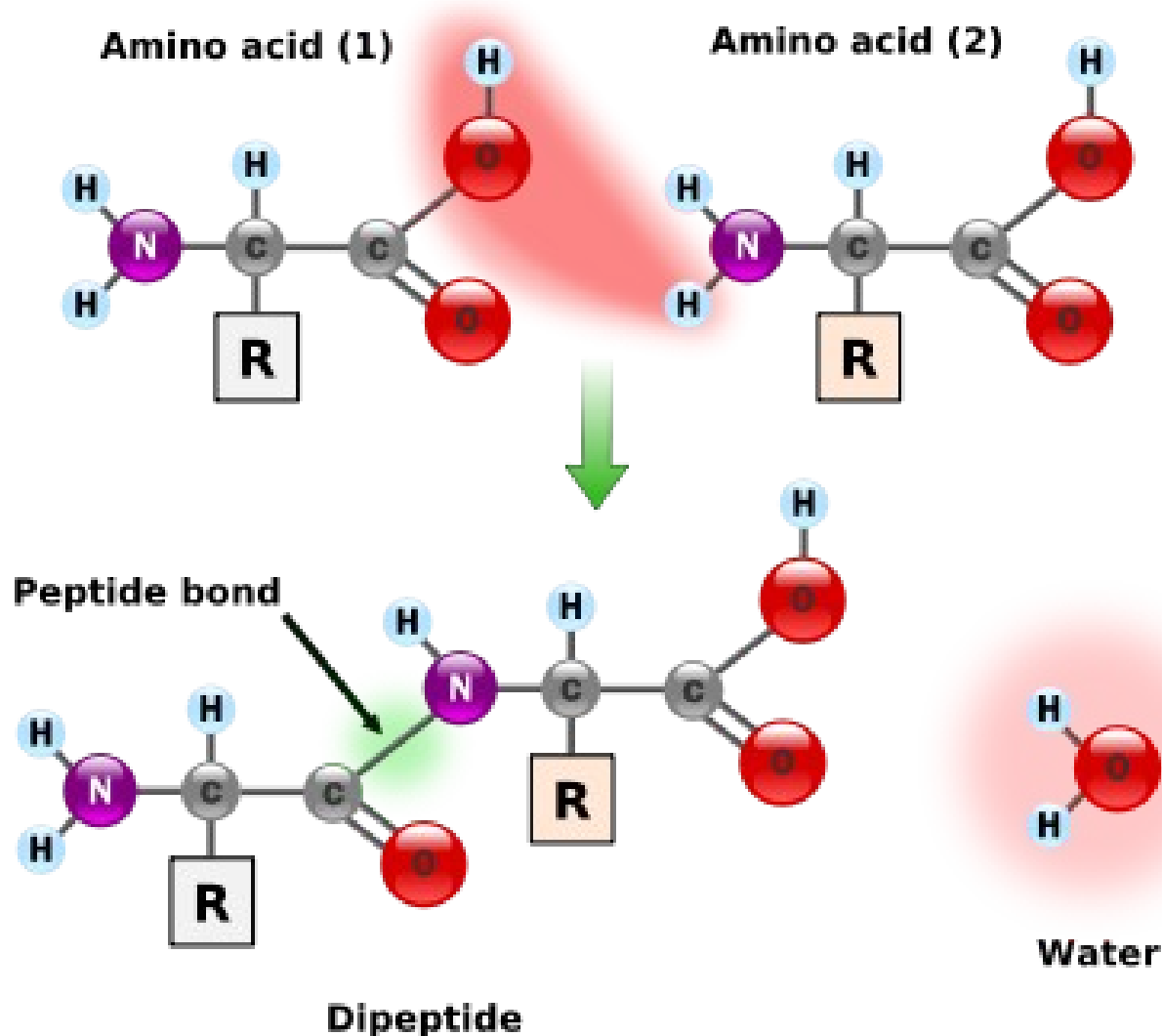


L-Valine
(Val / V)

Protein Structure

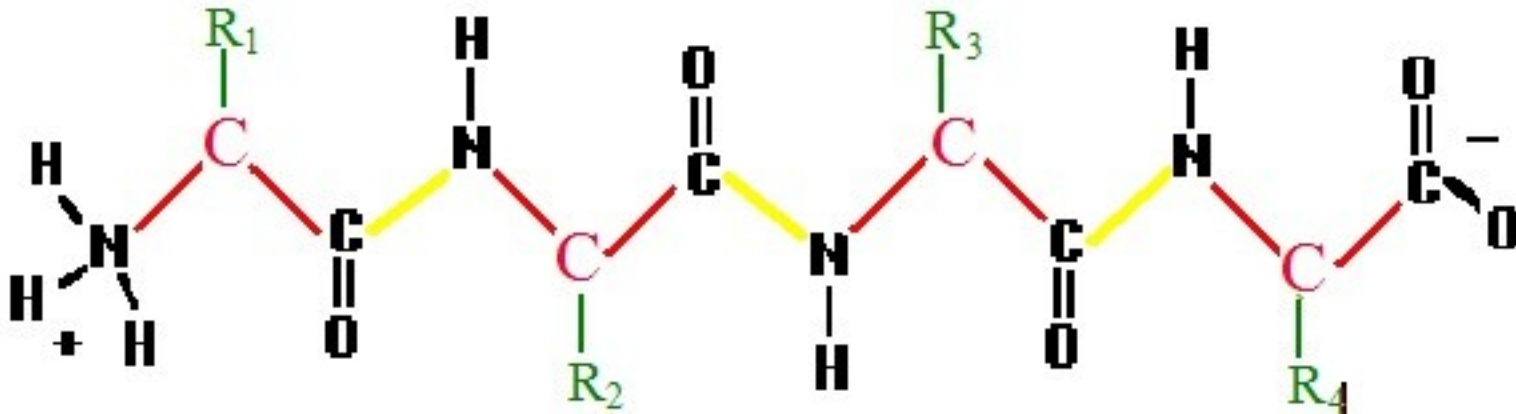


Primary Structure – Peptide Bond



Primary Structure – Polypeptide Chain

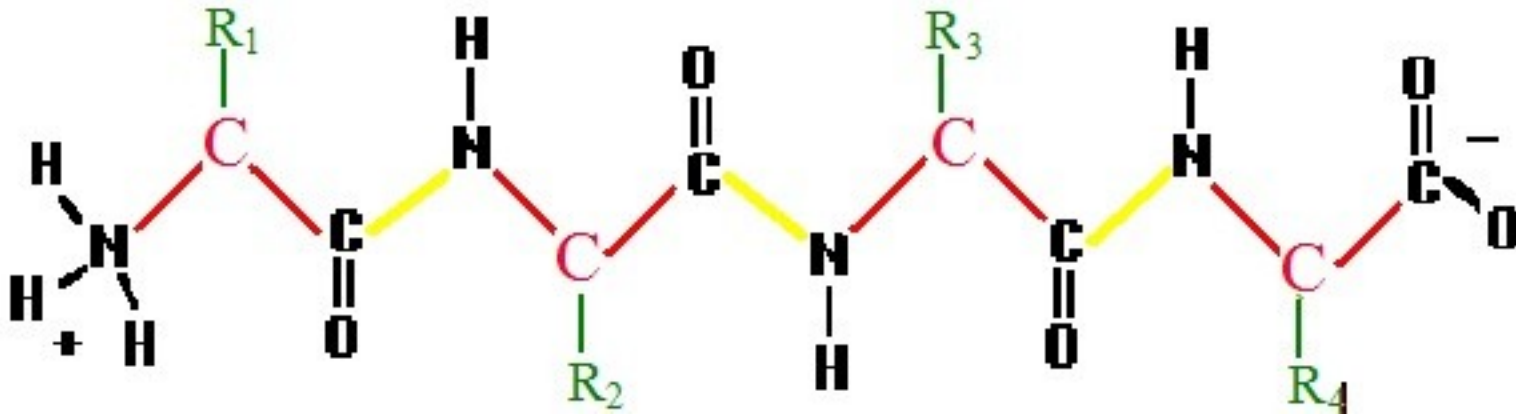
- A sequence of amino acids comes together to form a polypeptide chain
- The sequence of this polypeptide chain is called the primary structure of the protein



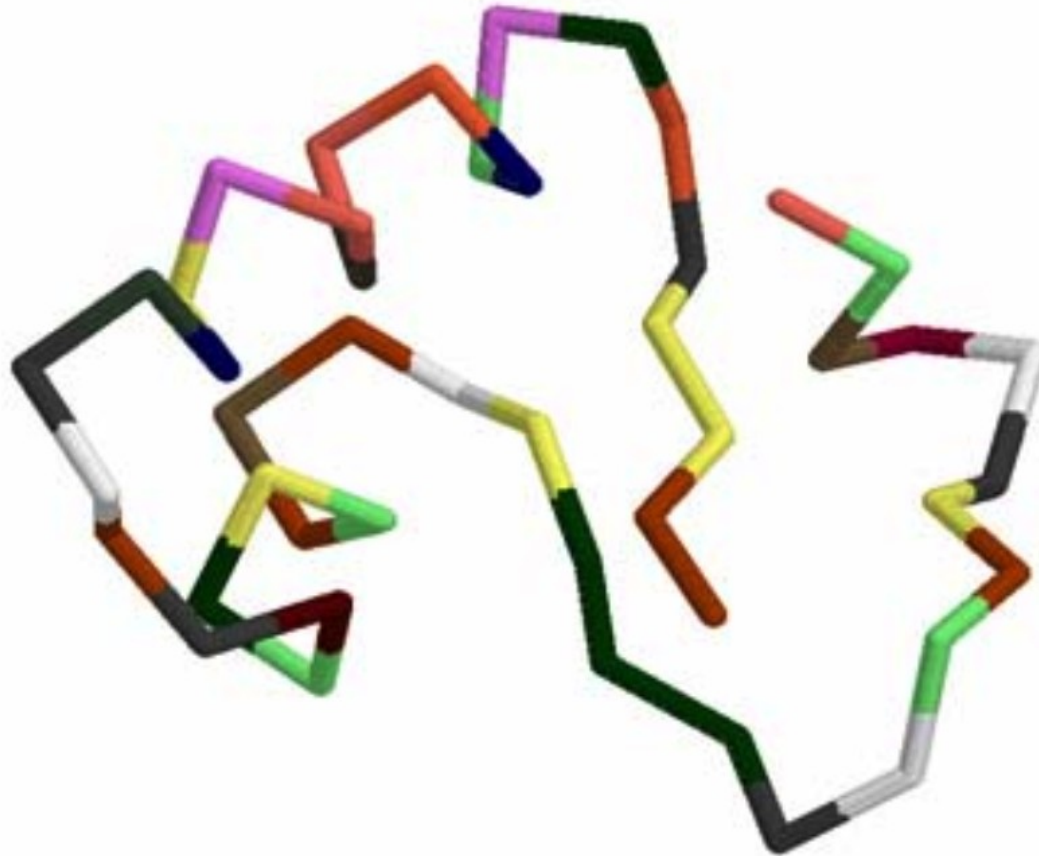
- Primary structure is only indirectly related to the function of the protein

Primary Structure – Polypeptide Chain

- To perform their function, the polypeptide sequence must fold into a specific shape 3-Dimensional shape – **native structure or native conformation**
- Carbon & Nitrogen atoms form the backbone (red & yellow) and the R groups (green) form side chains off the backbone.



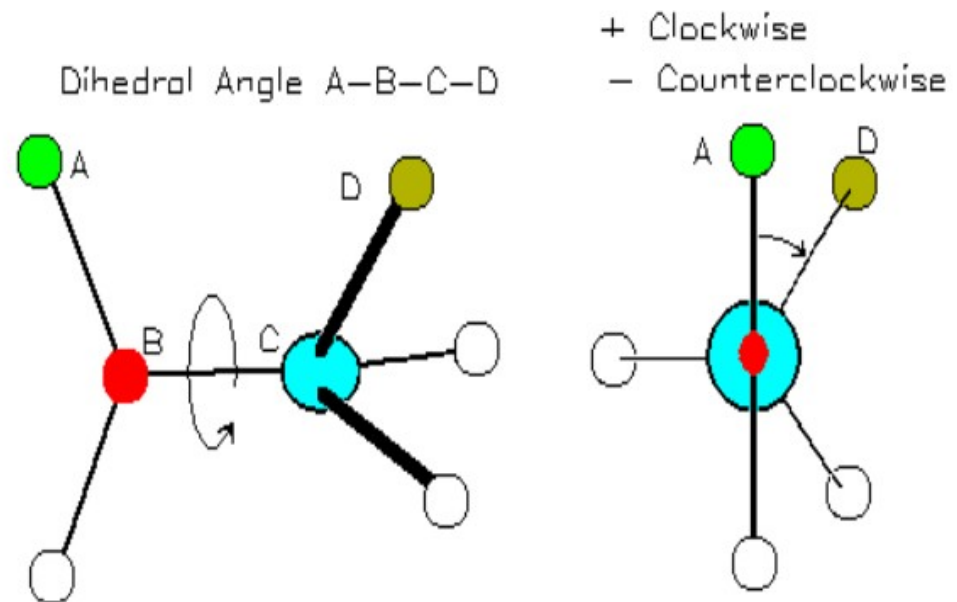
Protein Structure – Backbone



NWVLSTAADMQGVVTDGMASGLDKD

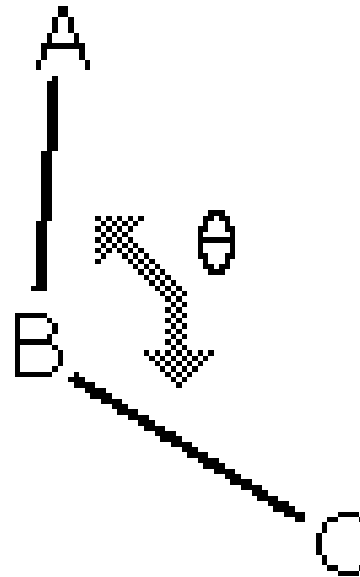
Secondary Structure

- Secondary structure refers to the local sub-structures within the peptide chain
- Determined by the backbone torsion/dihedral angles and specific main chain hydrogen bond pairing



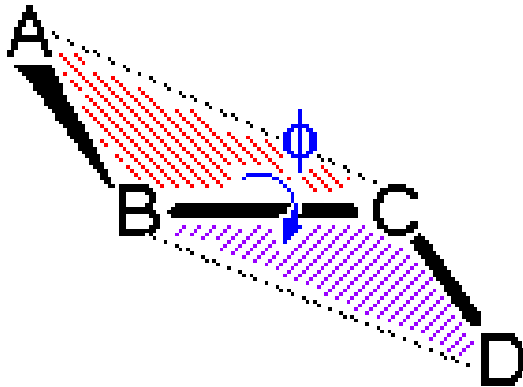
Secondary Structure – Covalent Bonds

- Electrostatic bonds between atoms
- Relative positions of atoms determined by:
 - bond length – geometric distance between atoms' nuclei
 - covalent angle – angle between two covalent bonds in a plane



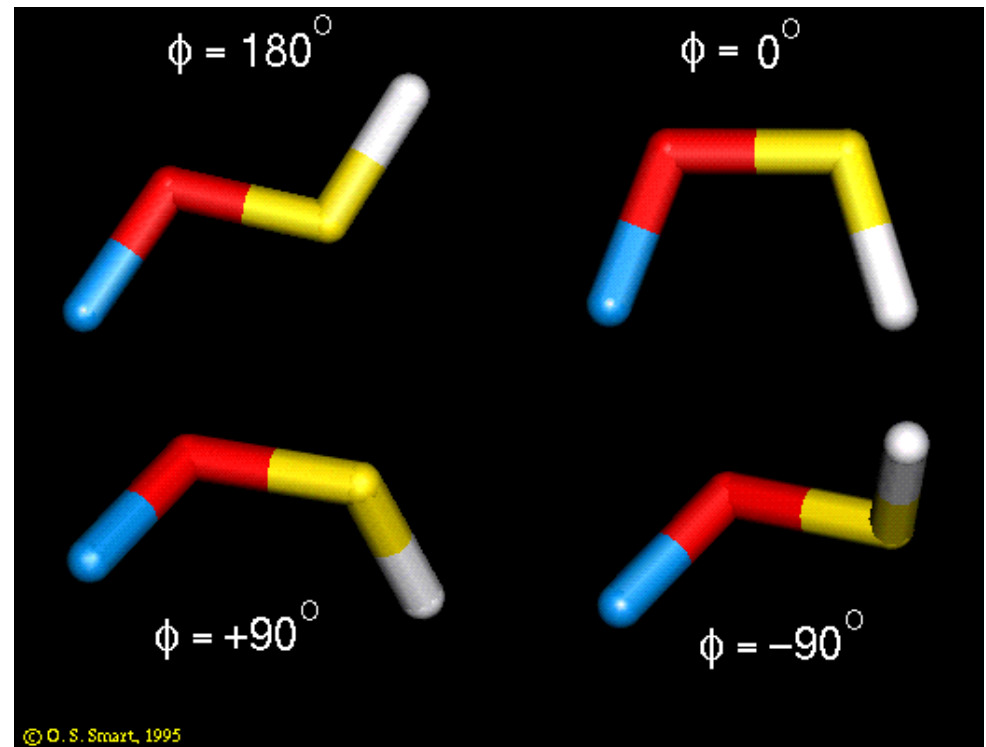
Secondary Structure – Dihedral Angles

- Covalent bond length and angles are generally fixed and well known. The key to the structure of the protein backbone are the dihedral (aka torsion) angles between planes



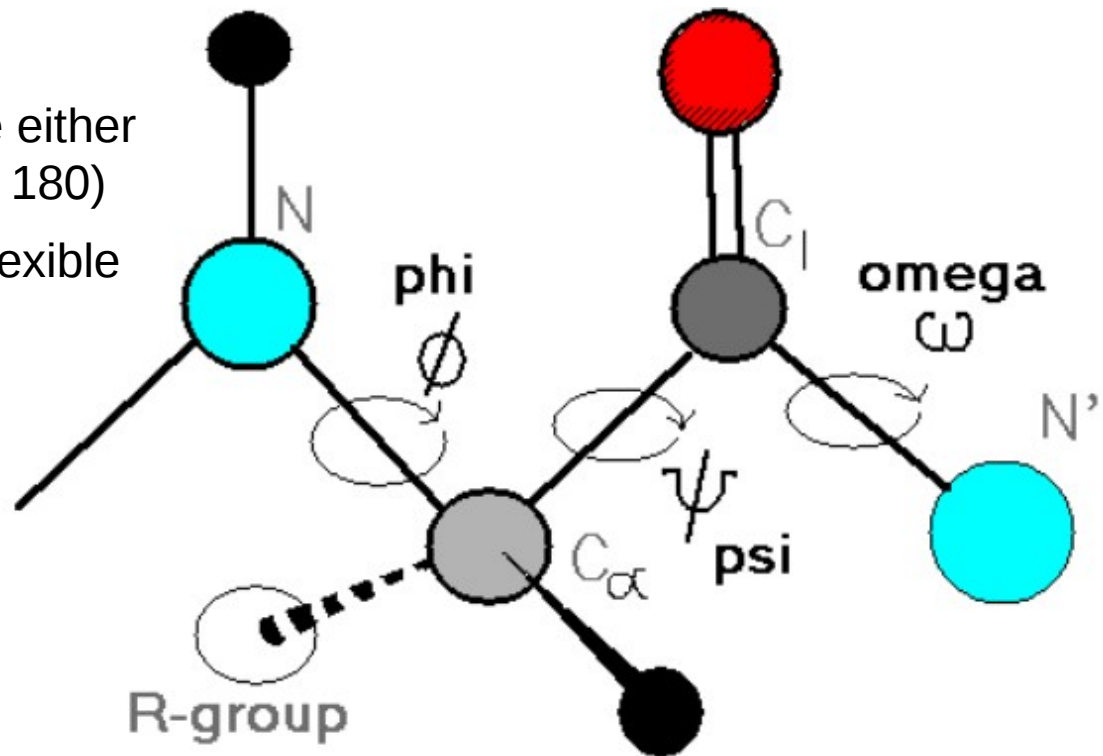
© O. S. Smart, 1995

Dihedral angle between two planes A-B-C and plane B-C-D



Secondary Structure – Dihedral Angles

- The backbone contains three repeating dihedral angles:
 - ϕ (phi) – between the amino group N and the alpha carbon, C α or C 2
 - ψ (psi) – between C α and the carboxylic carbon, C 1
 - ω (omega) – between C 1 and N
- In practice ω tends to be either 0 or 180 (almost always 180)
- This leaves ϕ and ψ for flexible folding of the chain



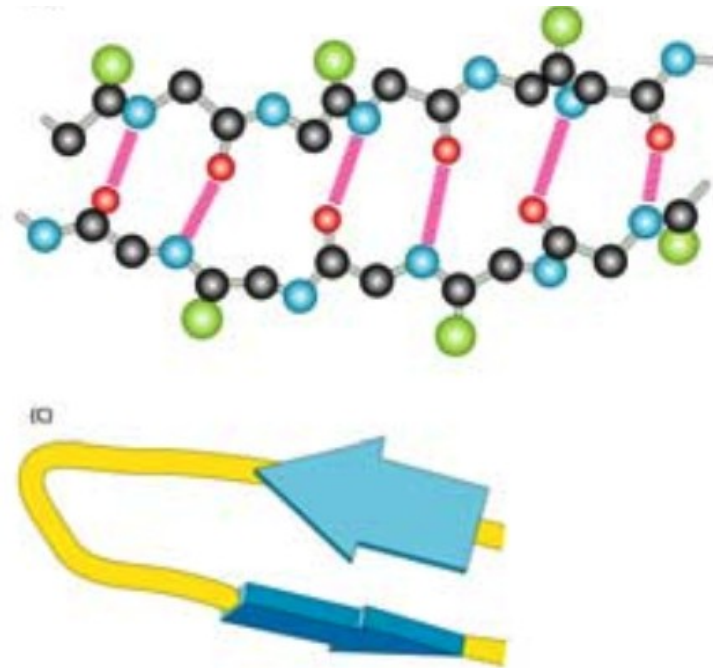
Secondary Structure -- α -helices & β -sheets

- Repeating values of ϕ and ψ along the chain result in regular structure

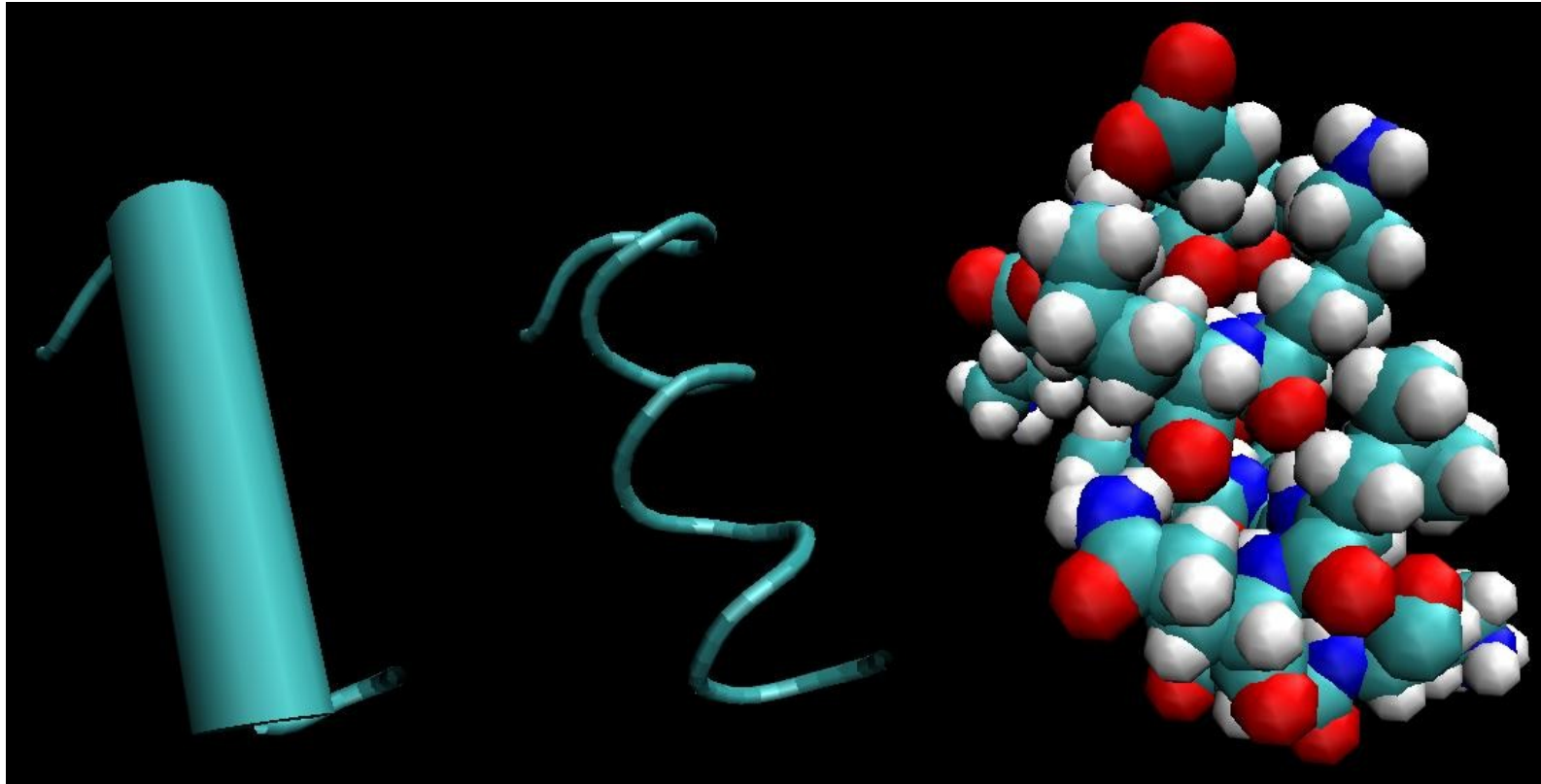
α -helix



β -sheet



Secondary Structure -- α -helix



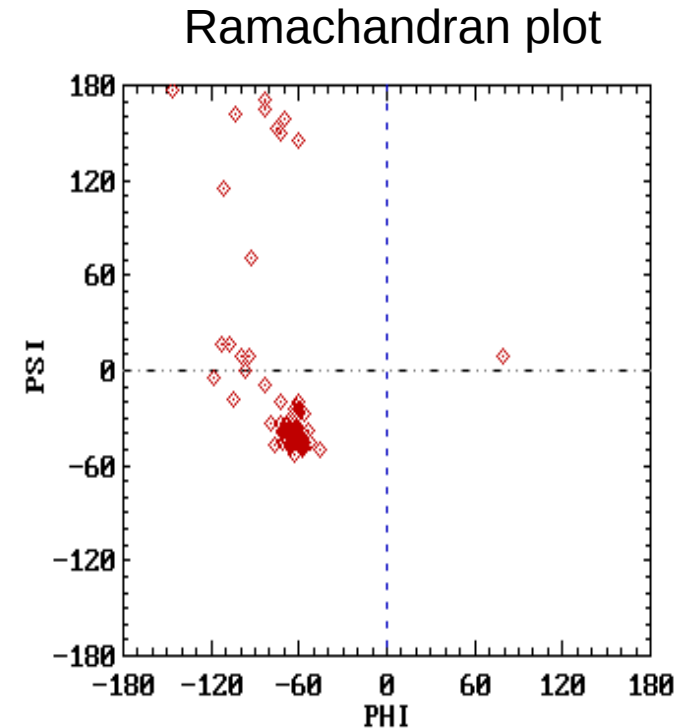
Cartoon
Representation

Ribbon
Representation

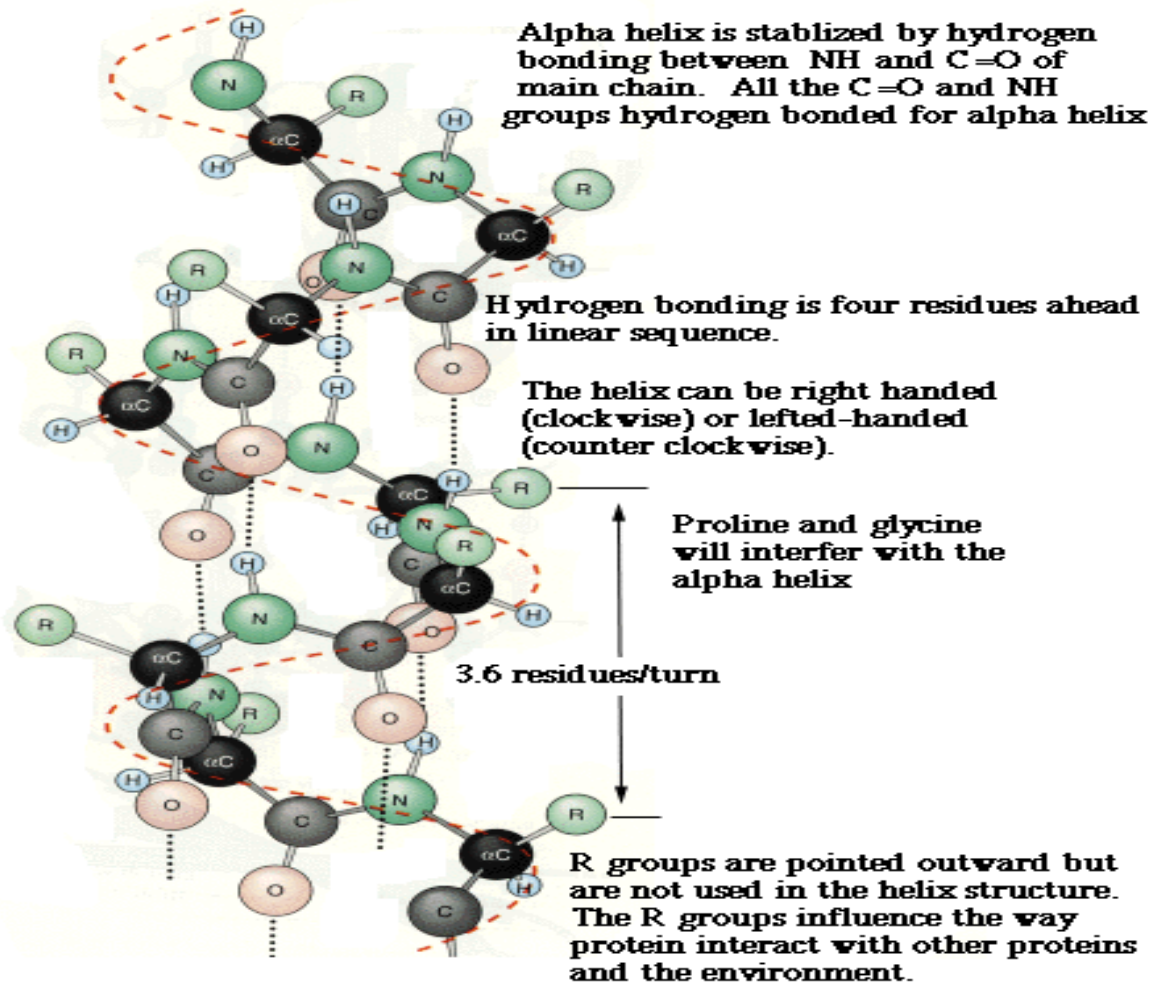
Van der Waals
Representation

Secondary Structure -- α -helix

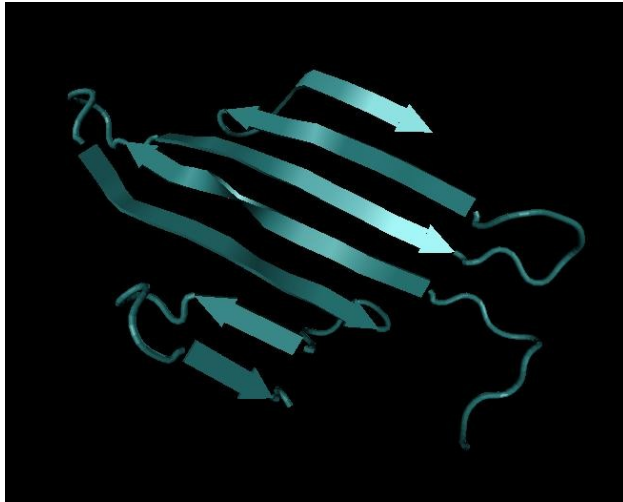
- Repeating values of $\phi \sim -57^\circ$ and $\psi \sim -47^\circ$ give a right-handed helical fold (the alpha-helix)
- The structure of cytochrome C-256 shows many segments of helix and the Ramachandran plot shows a tight grouping of ϕ, ψ angles near to $-50, -50$



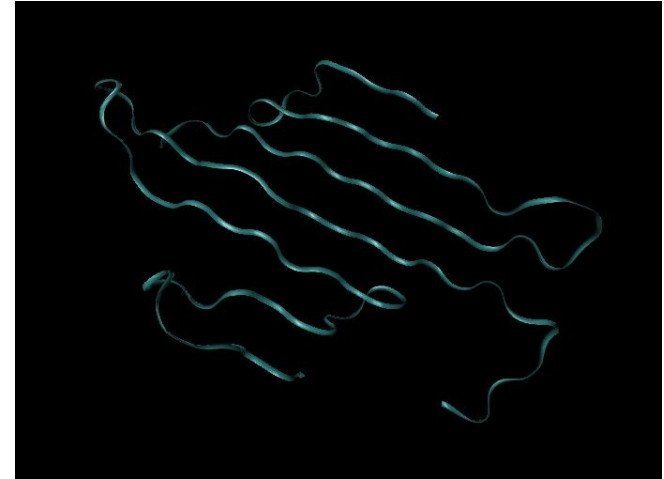
Secondary Structure -- α -helix



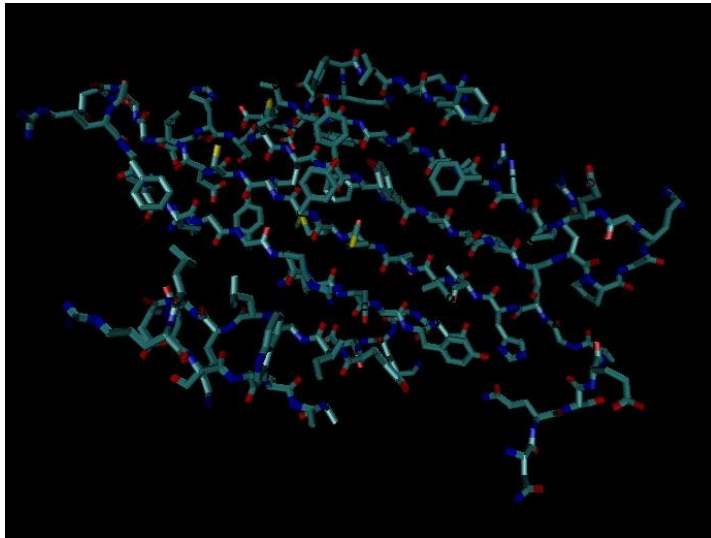
Secondary Structure -- β -sheets



Cartoon
Representation



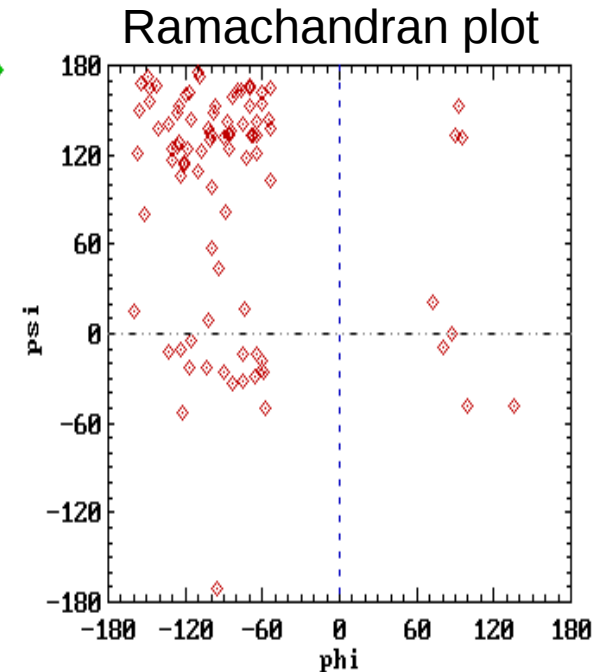
Ribbon
Representation



Bond
Representation

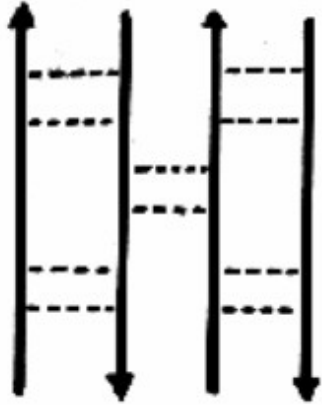
Secondary Structure -- β -sheets

- repetitive values in the region of $\phi = -110$ to -140 and $\psi = +110$ to $+135$ give extended chains with conformations that allow interactions between closely folded parallel segments (beta sheet structures)
- The structure of plastocyanin is composed mostly of beta sheets and the Ramachandran plot shows a broad range of values in the $-110, +130$ region

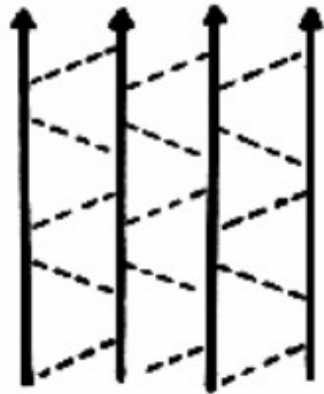


Secondary Structure -- β -sheets

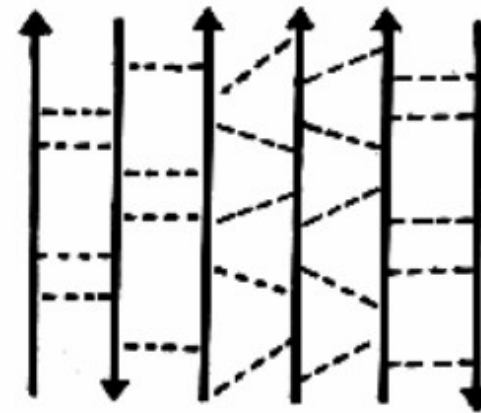
Antiparallel beta-sheet



The different types of beta-sheet. Dashed lines indicate main chain hydrogen bonds.



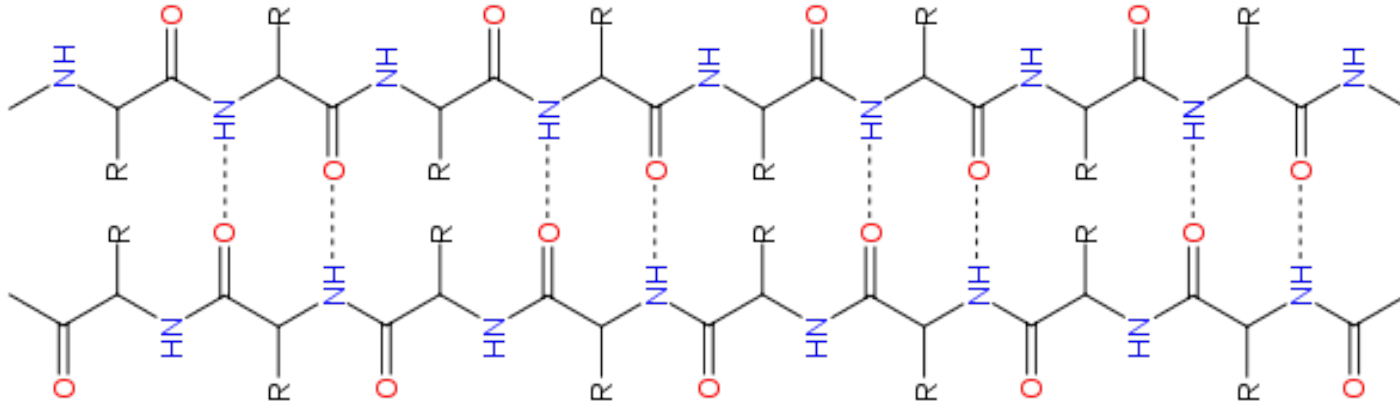
Parallel beta-sheet



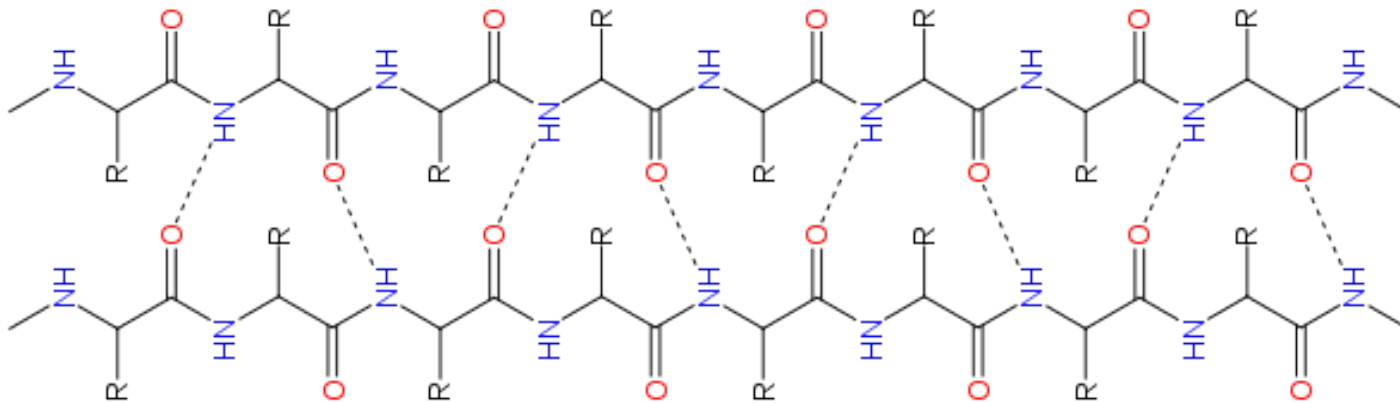
Mixed beta-sheet

Secondary Structure -- β -sheets

Antiparallel beta-sheet

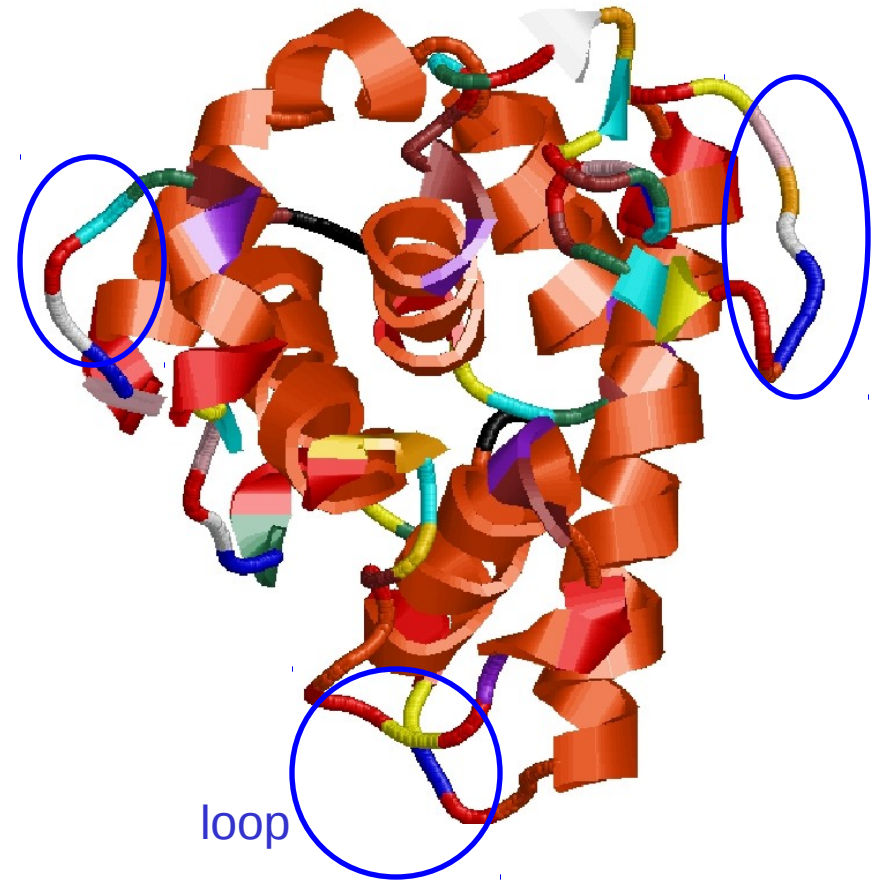


parallel beta-sheet



Secondary Structure – Loops and Coils

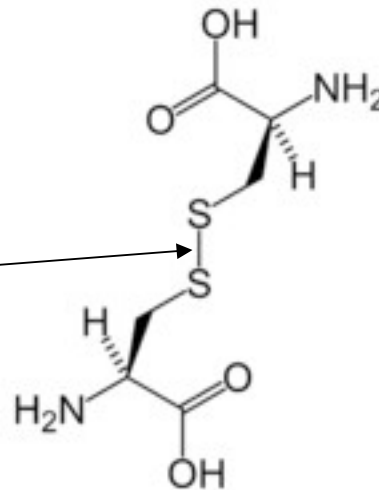
- The sections of the peptide chain that link the alpha-helices and beta-sheets are referred to as turns and loops
- Other secondary substructure classifications exist, but are rarely seen in practice
- Sub-units that do not fit into any other classification are known as random coils



Tertiary Structure

- Tertiary structure is the structure formed by bringing together the various sub-chains in the Secondary structure together

- Tertiary structure is determined by non-bonding interactions and the disulfide bonds



- When in an aqueous solution, hydrophobic residues are pushed to the center and hydrophilic residues come to the surface



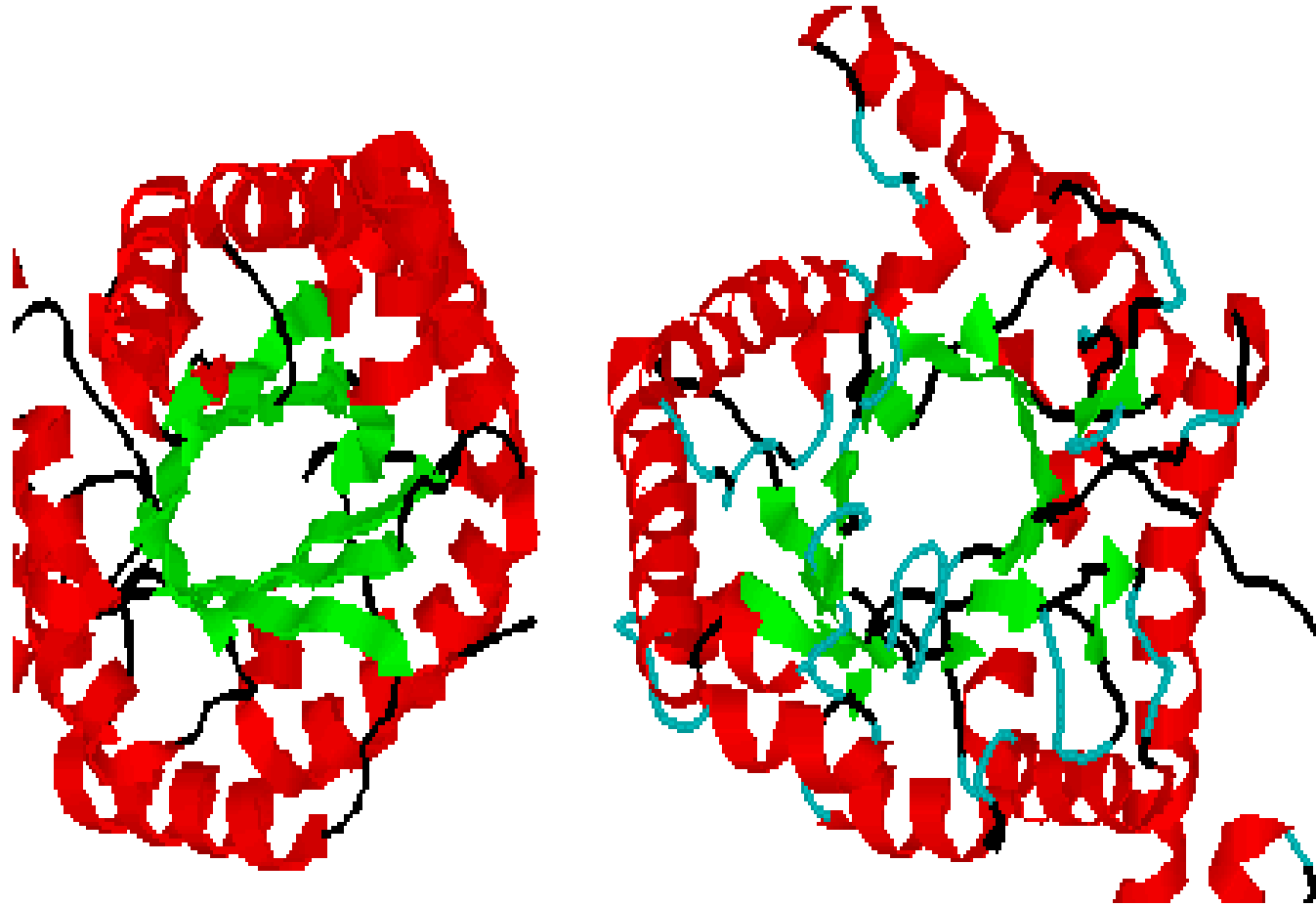
Tertiary Structure -- Mostly α -helices



Tertiary Structure -- Mostly β -sheets

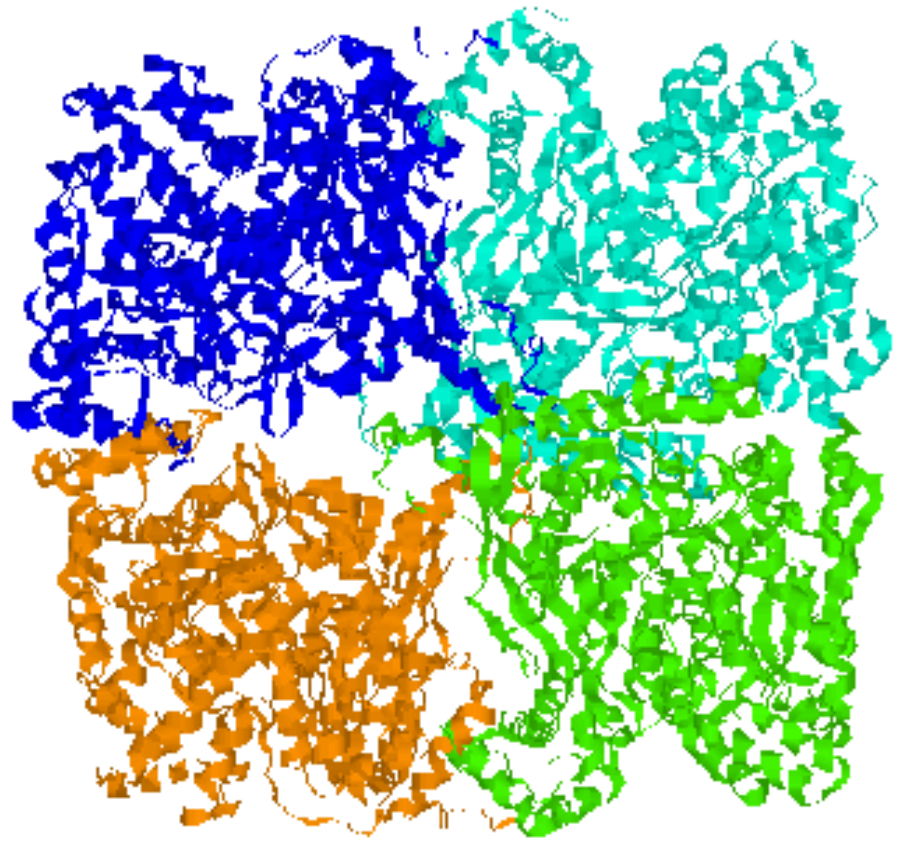


Tertiary Structure -- Mixed



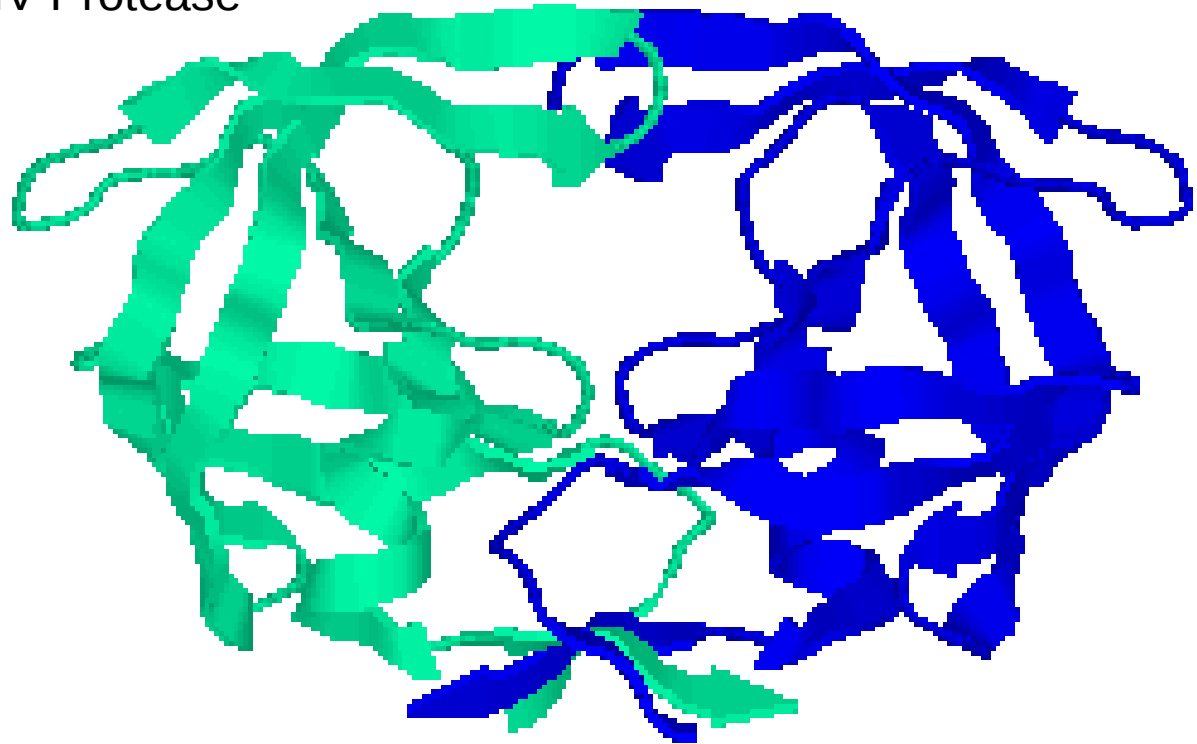
Quaternary Structure

- Combination of two or more polypeptide chains to form a complete unit
- The interactions between the chains are no different from those in tertiary structure, but are distinguished only by being inter chain rather than intra chain



Quaternary Structure

- Some proteins are composed of identical subunits (chains)
- Example –
the dimer of HIV Protease

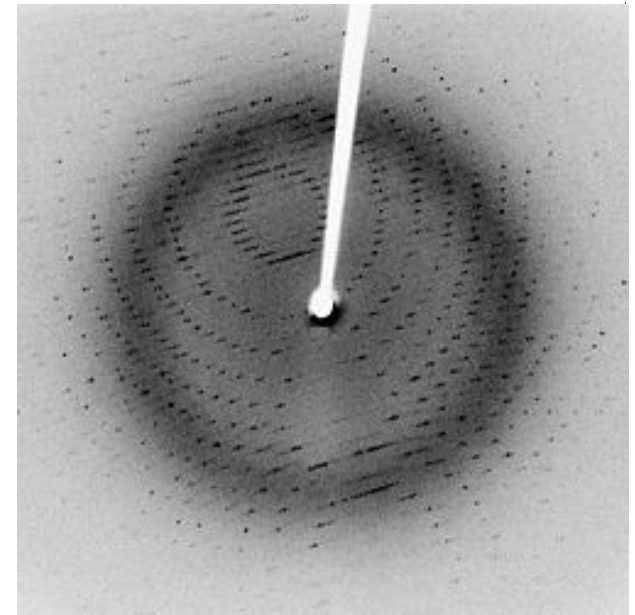


Structure Determination

- Directly determined by the primary structure and the chemical environment the protein is in (aqueous or oily)
 - Cannot reliably determine the native structure based solely on the primary structure
 - Computational (dry-lab) methods are an ongoing area of research
 - Experimental (wet-lab) methods are effective, but costly and time consuming
 - X-ray crystallography
 - NMR
 - Electron Diffraction
-

X-Ray Crystallography

- Most common and highest resolution method
- Pass X-rays through a crystallized version of the protein
- The resulting diffraction of the X-rays can be used to determine the location of each atom
- Goal is to approximate the x, y, and z coordinates of every atom in the structure
- Only provides static structural information under the specific experimental conditions used



http://en.wikipedia.org/wiki/X-ray_diffraction

Nuclear Magnetic Resonance (NMR) Spectroscopy.

- More recent technique for determining native structure
- Uses a magnetic field to force the charged atomic nuclei into alignment
- Data collected when atoms return to unaligned state can be used to determine structure
- Unlike X-ray crystallography, it can be performed on proteins in solution
- Provides statistics on flexibility
- Reports an average structure that reproduces *most* of the statistics
- Only effective on relatively small proteins



<http://publications.nigms.nih.gov/structlife/chapter3.html>

Protein Structure Repositories

- Protein Data Bank (PDB) -- <http://www.rcsb.org>
- Currently ~55,000 structures
 - Majority determined using X-ray crystallography
 - Growing number using MNR and other methods
- Use a 4-digit protein ID

to find a structure in the PDB

For ex: 1A7N

- PDB file format documentation

<http://www.wwpdb.org/docs.html>

The screenshot shows the RCSB Protein Data Bank website. The browser address bar displays the URL <http://www.rcsb.org/pdb/home/home.do>. The page header includes the RCSB PDB logo and navigation links. A blue banner at the top right states "An Information Portal to Biological Macromolecular Structures" and "As of Tuesday Feb 03, 2009 there are 55660 Structures". The main content area features a sidebar with navigation links, a central article titled "A Resource for Studying Biological Macromolecules", and a right-hand news section with updates from February 2009 and December 2008.

Visualizing Protein Structures



<http://www.ks.uiuc.edu/Research/vmd/>