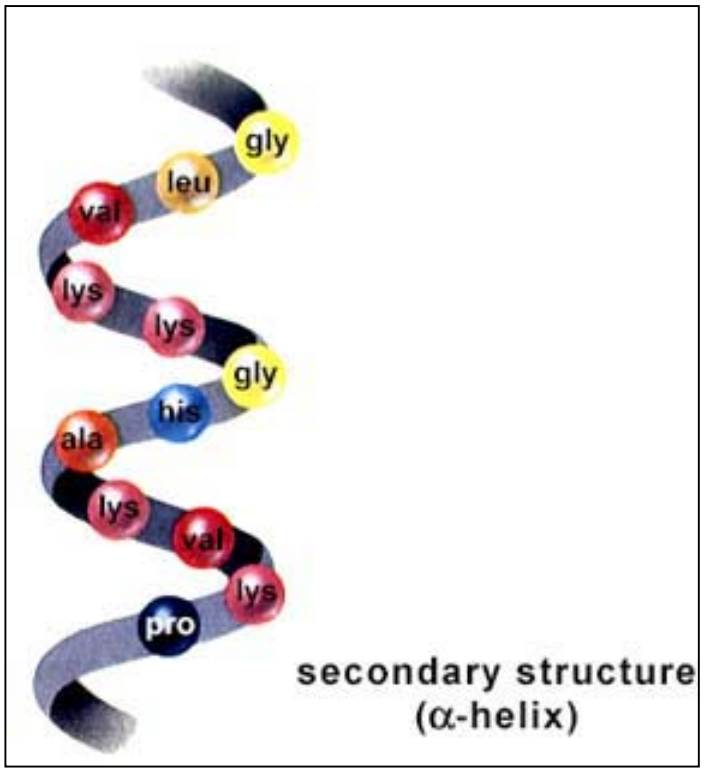
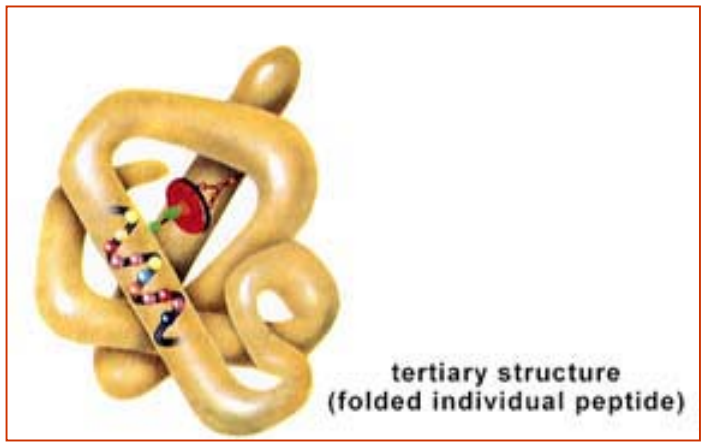
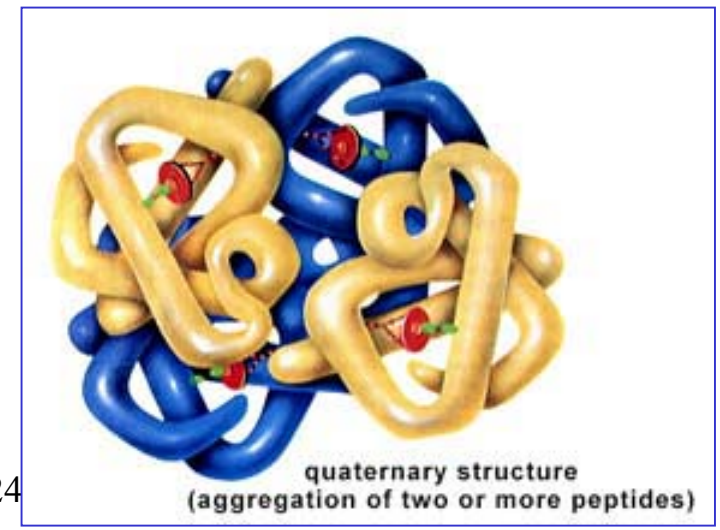
Protein Structures

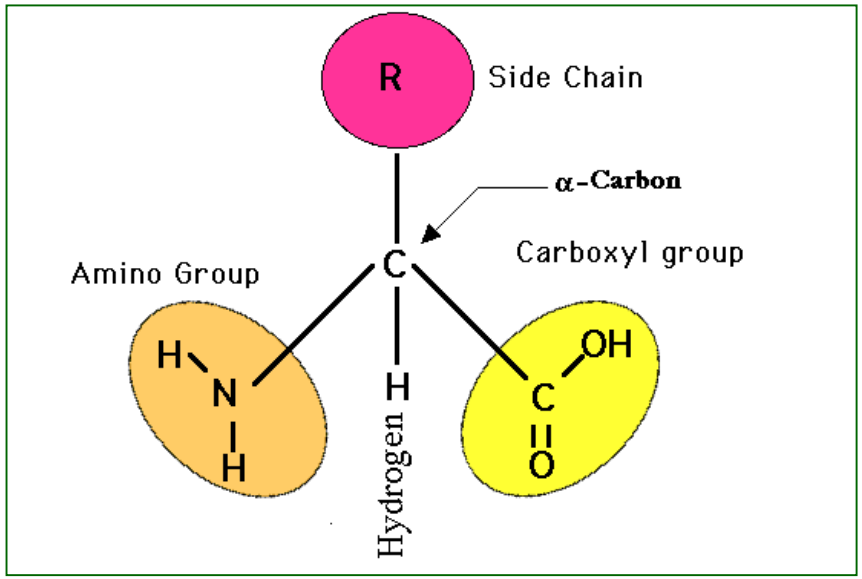
primary structure
(amino acid sequence)







All 3 figures are cartoons of an amino acid residue.



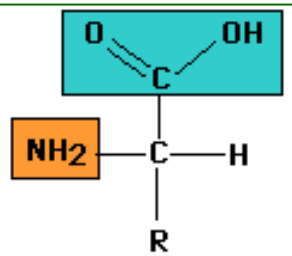
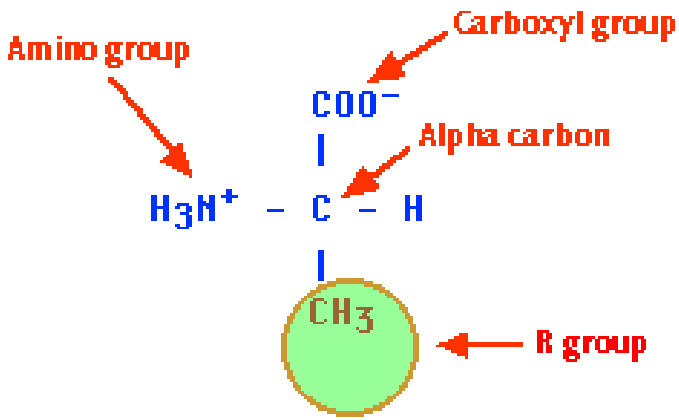
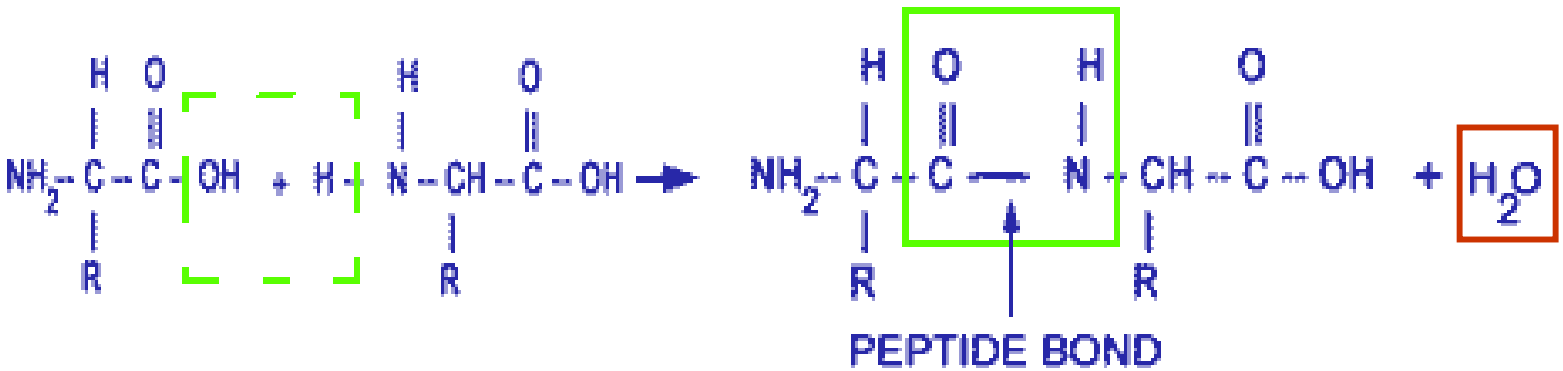


Fig. General formula for an amino acid molecule. "R" represents the variable groups that are attached to this basic molecule to make up the 20 common amino acids



Peptide bonds in chains of residues



Angles ϕ and ψ in the polypeptide chain

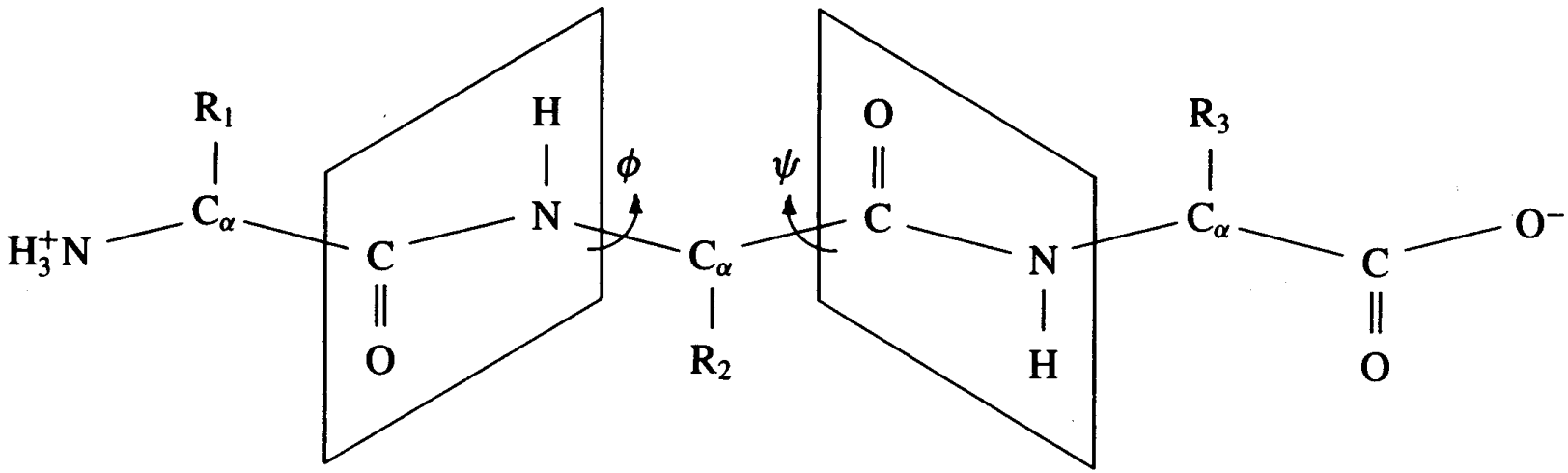


FIGURE 1.2

A polypeptide chain. The R_i side chains identify the component amino acids. Atoms inside each quadrilateral are on the same plane, which can rotate according to angles ϕ and ψ.

Active Sites

Active sites in proteins are usually hydrophobic pockets/crevices/troughs that involve sidechain atoms.

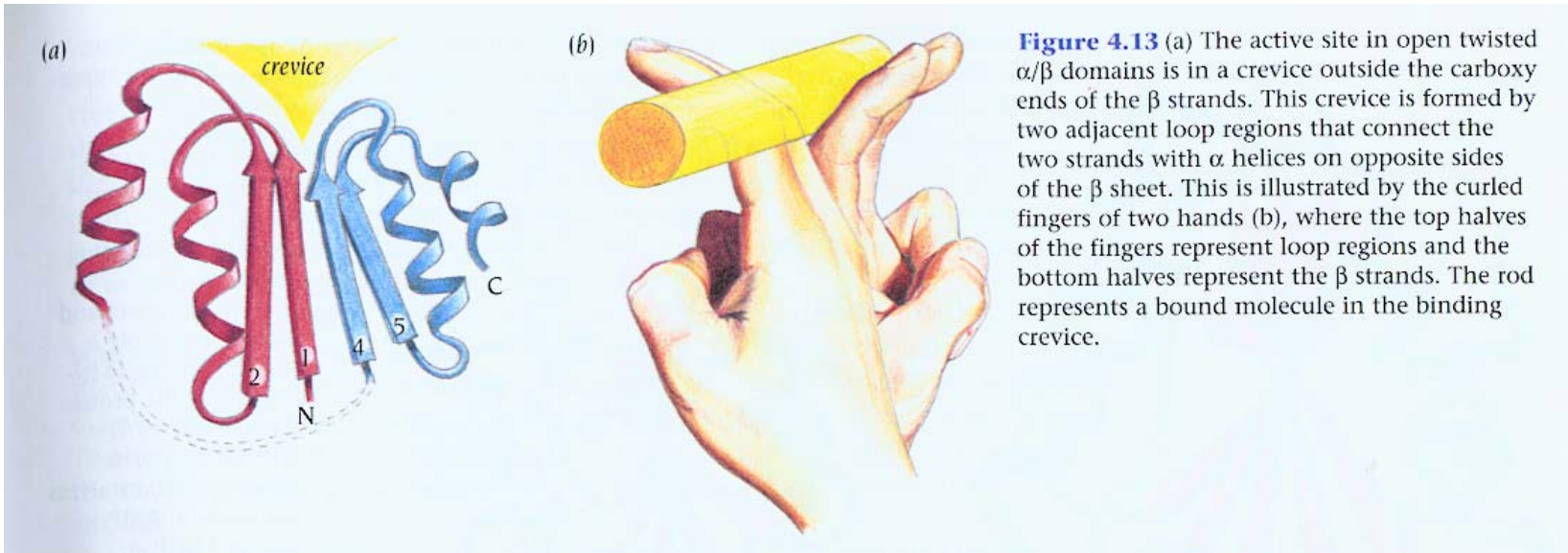
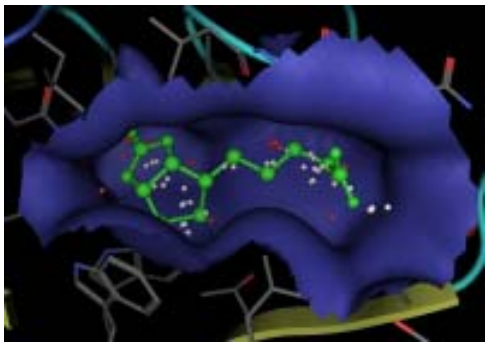
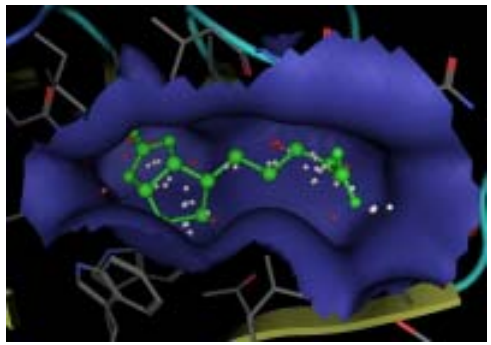
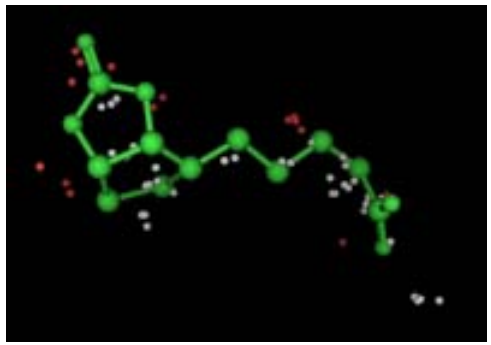


Figure 4.13 (a) The active site in open twisted α/β domains is in a crevice outside the carboxy ends of the β strands. This crevice is formed by two adjacent loop regions that connect the two strands with α helices on opposite sides of the β sheet. This is illustrated by the curled fingers of two hands (b), where the top halves of the fingers represent loop regions and the bottom halves represent the β strands. The rod represents a bound molecule in the binding crevice.

Active Sites

Left PDB 3RTD (streptavidin) and the first site located by the MOE Site Finder. **Middle** 3RTD with complexed ligand (biotin). **Right** Biotin ligand overlaid with calculated alpha spheres of the first site.

Amino Acid Classes

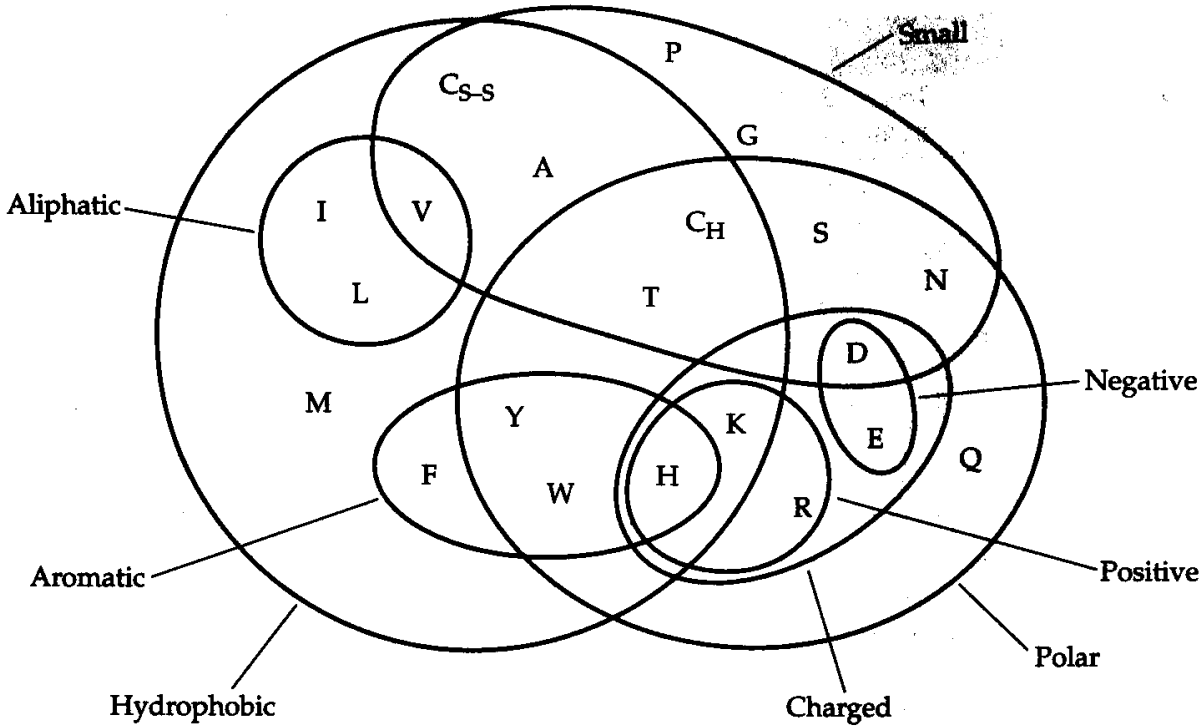


FIGURE 1.9 A Venn diagram showing the division of the 20 primary amino acids into overlapping categories according to size, structure of the side chain, polarity, charge, and hydrophobicity. Note the placement of cysteine in two places, as reduced cysteine (C_H) and as cystine (C_{S-S}). See Table 1.2 for the one-letter abbreviations of the amino acids. Modified from Taylor (1986).

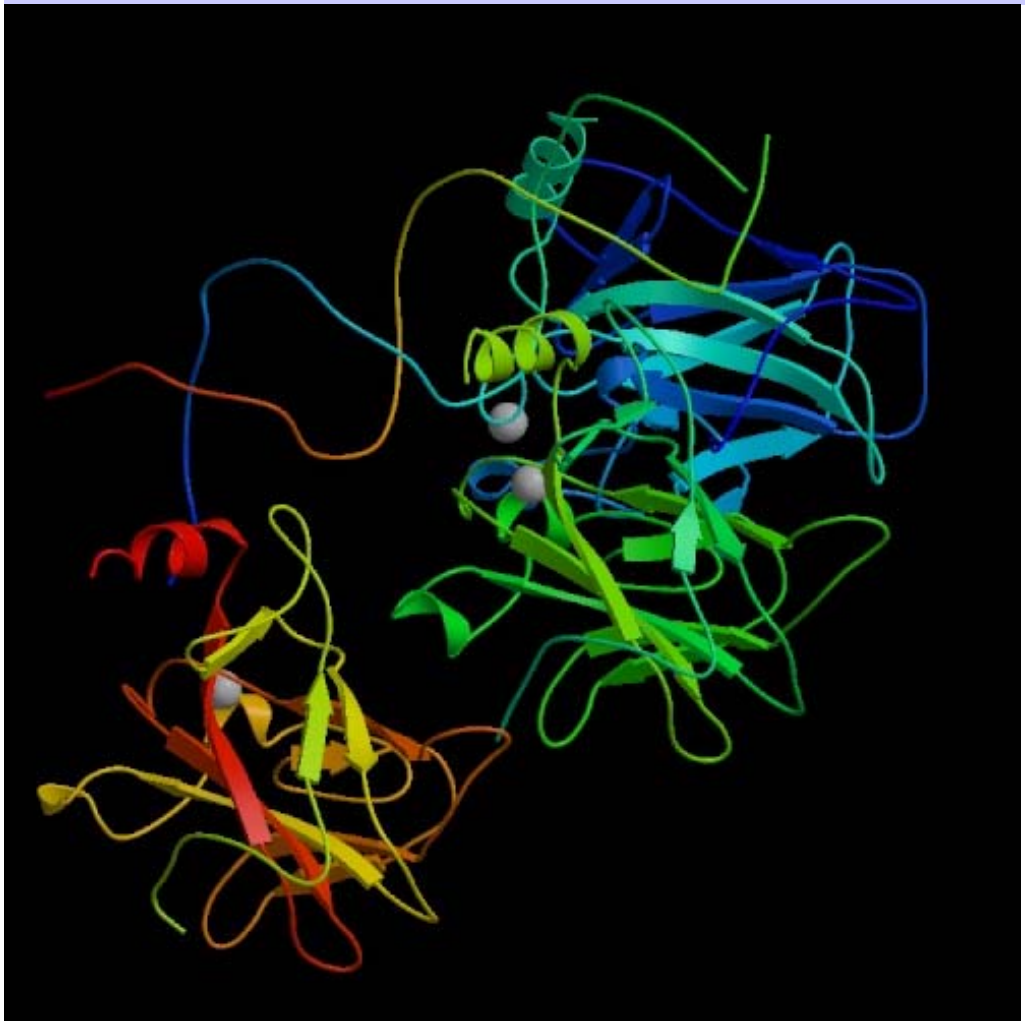
Protein Structure: experimental determination

- X-ray crystallography
- Nuclear Magnetic Resonance Spectroscopy (NMR)

- X-ray crystallography methods are more accurate.
- Different stable crystal structures are possible
- NMR methods model the protein in its natural state, i.e., in solvent.

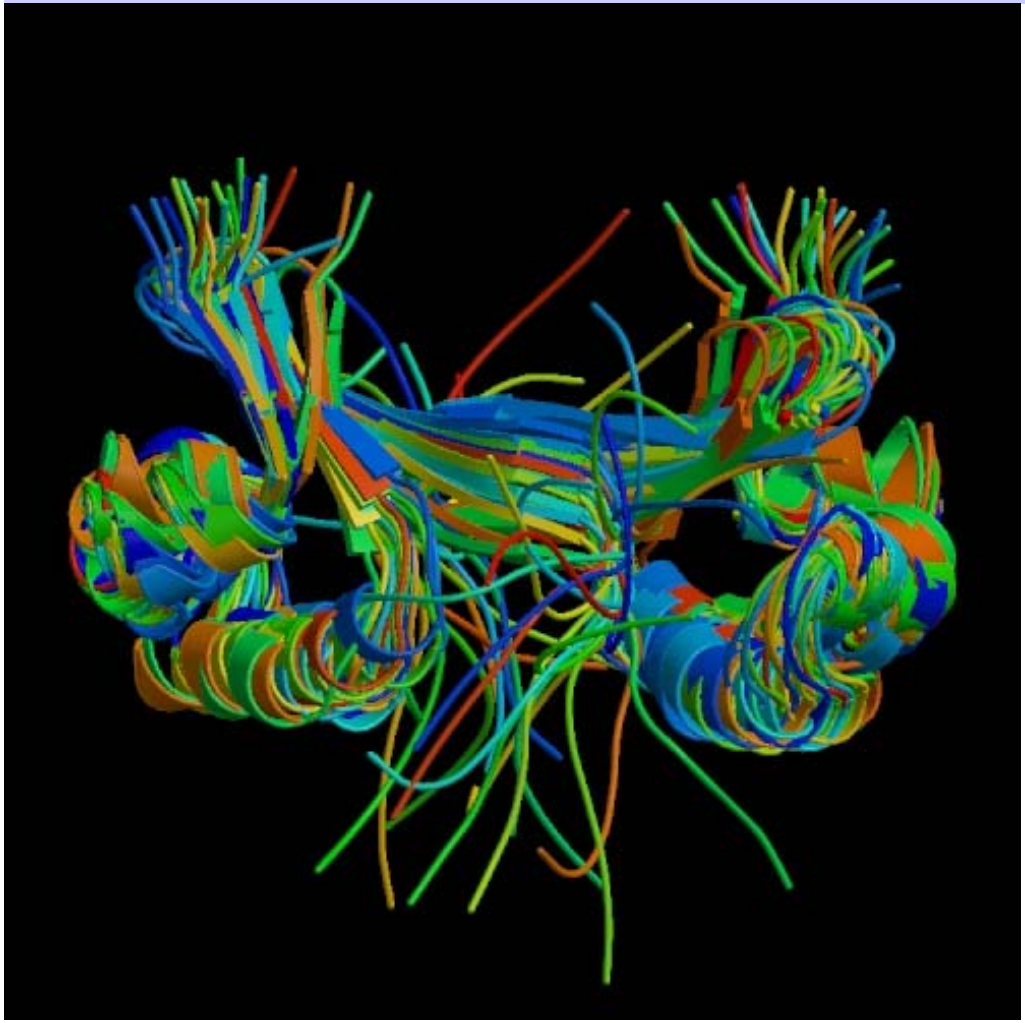
- Sequence ⇒ Structure
- Very similar sequences can have dissimilar structures
- Very dissimilar sequences can have similar structures

p53 Structure (1tsr)



X-ray crystallography
2.2 A° resolution

Lambda CRO Repressor (1cop): Structure



NMR Methods:
20 models

Protein Folding

Unfolded

Rapid (< 1s)

Molten Globule State

Slow (1 – 1000 s)

Folded Native State

- Infinitely many configurations:
  How to find minimum energy configuration?
  Find approximate minimum energy configuration.

Protein Structure Determination: Methods

- Homology-based Methods
  - Use structure of homologous proteins
- Protein Threading
  - Assuming the protein assumes a known structure, compute its energy.
- *Ab initio* Methods
  - List all forces between all atoms and the resulting energy of the system. Find min energy configuration
- Combinations Methods
  - Use substructures from database and then assemble

Software for protein structures

- SWISS-MODEL
  - 1st Pass: use homology method
  - 2nd Pass: refines using biochemical info
- VAST: compares structures
- DALI: aligns structures
- TOPITS: searches in PDB
- RASMOL: visualizing PDB structures
- Cn3D: visualize MMDB structures (from NCBI)
- Amber & CHARMM: Energy functions & minimization software

Protein Threading

- Input: Sequence S, Structure model M, Score functions f, g
- Output: Threading of sequence S to model M
- Model:
  - core segments: $C_1, C_2, \ldots, C_m$,
  - their lengths: $l_1, l_2, \ldots, l_m$
  - Min and Max lengths of connecting loop regions:
    - $n_1, n_2, \ldots, n_m$
    - $x_1, x_2, \ldots, x_m$
- Total Score: depends on the placement of residues of S according to model M and the energy of resulting structure
- Optimization of total score: branch-&-bound method

Protein Threading

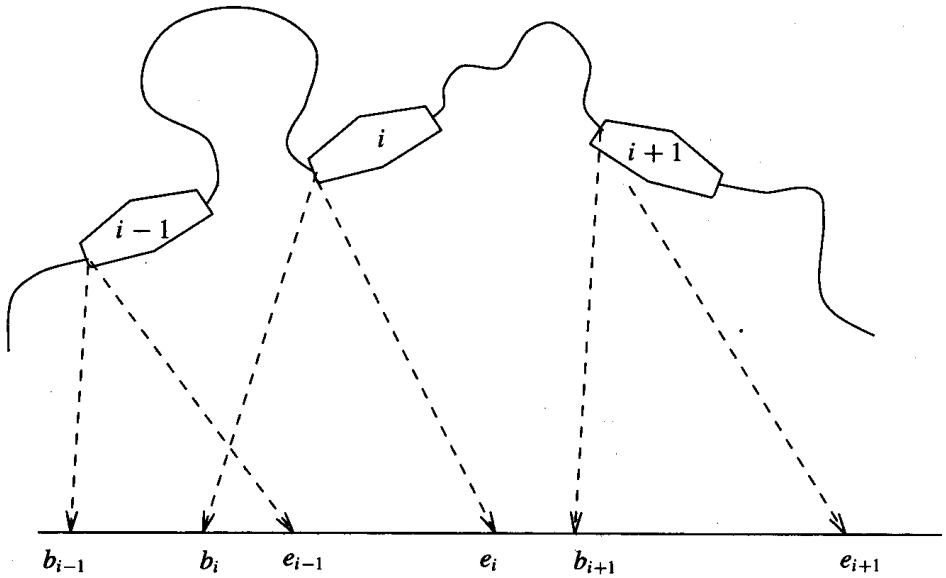


FIGURE 8.3

Sets of threadings. Core segments are the polygons labeled by $i-1$, i, and $i+1$. Between them are loop regions. The line at the bottom is the sequence being threaded, and the dashed arrows show the intervals where the first position of each core segment can assume values in the sequence, hence depicting sets of possible threadings.

Lattice Models: HP Model
Simplified assumptions

- Assume that all atoms may only occupy lattice points. Lattices: 2D-grid, 3D-grid, Hex-grid
- Only backbone (central C atoms) is relevant, and side chain placements not considered.
- Every amino acid is either hydrophilic (P) or hydrophobic (H)
- If 2 H residues are topological neighbors, then they contribute -1 unit to the free energy of configuration.
- If residues are not both H residues or they are connected neighbors, they do not contribute to free energy.
- Native structure: self-avoiding, minimum free energy structure (Thermodynamic hypothesis)

Extensions to HP Model

- Side chain consists of 1 “bead”.
- Assume different free energy values for
  H-H, H-P, and P-P.
- Consider effect of non-local interactions (distance 2 units)

Ab initio Methods

- Energies involved
  - Bond energies (deviation from opt bond length)
  - Angles (deviation from opt due to bending)
  - Dihedral angle (deviation from opt)
  - Lennard-Jones potential
  - Coulomb potential
  - Van der Waal's energy
  - Urey-Bradley term

Project Presentations

- YOU NEED TO BE PRESENT FOR ALL SEMINARS!
- YOU NEED TO BE ON TIME!

Presentation Schedule

- **Nov 21**:
  - 9:30 AM: Teiresias Project (C. Zhan & N. Zhao)
  - 9:50 AM: GYM Project (Y. Sun & Z. Deng)
  - 10:10 AM: Phage Design (J. Yan)
  - 10:25 AM: Probe Design (D. Cazalis)
- **Nov 26**:
  - 9:30 AM: Structure Pattern Discovery (T. Milledge)
  - 9:50 AM: Geometric Hashing (C. D'Cunha & M. Hu)
  - 10:10 AM: Primer Design (X. Wei)
  - 10:25 AM: Eco-Informatics (C. Yang & Y. Wang)
- **Dec 3**:
  - 9:30 AM: Multiple Sequence Alignment (L. Li)
  - 9:45 AM: HIV Data Analysis (P. Buendia)
  - 10:00 AM: Microarray Data Analysis (S. Rodriguez)
  - 10:15 AM: Pattern Discovery in Microarray Data (E. Wu & Z. Dai)