ExPASy Molecular Biology Server

web site : www.expasy.org

→ go to SWISS-PROT data base

→ retreve the primary sequence of protein

OMPF_ECOLI

in FASTA format

→ view the 3D image of the protein using the

SWISS_3DIMAGE tool

→ run a similarity search for OMPF_ECOLI
using the BLAST program

→ compute pI/MW of the protein
(Primary structure analysis)

CHAPTER 16 Protein Folding and Design

"Protein Folding Problem"-

(i) How does the amino acid sequence of a protein specify its 3D structure?

(ii) How does an unfolded polypeptide chain acquire the form of the native protein?

Proteins fold by progressive stabilization of intermediates rather than by random search.

The essence of protein folding is the retention of partially correct intermediates.

Proteins are only marginally stable.

"Kinetic traps" : intermediates with favorable free energy, but not on the path to the final folded form.

Folding States of a Polypeptide Chain:

- native state (N)
- unfolded state (U)
- molden globule state (M)

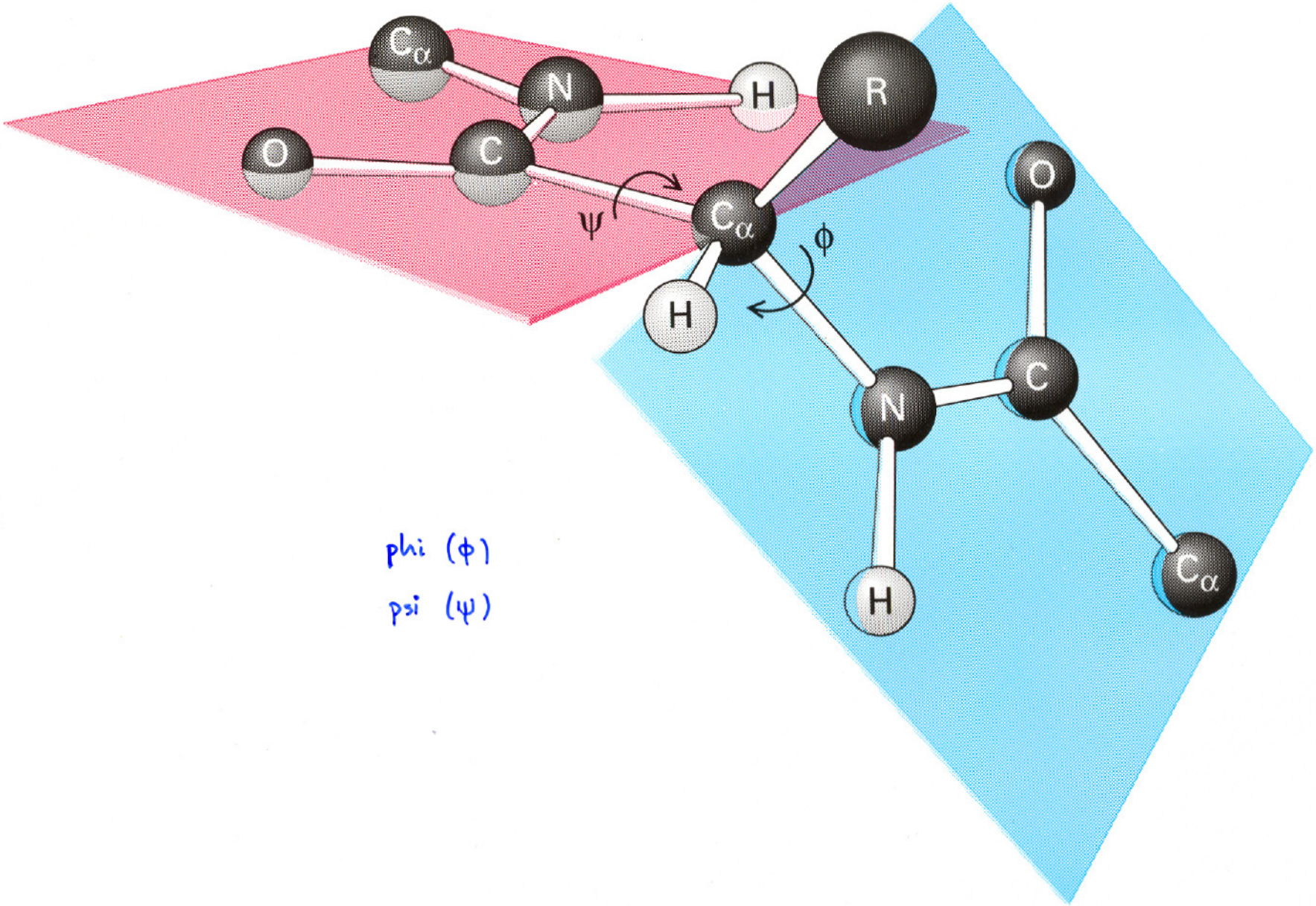
Molten globules contain native secondary structure but not tertiary structure.

The driving forces for the formation of molten globules are:

(i) the formation of hydrophobic interactions, and

(ii) the formation of stretches of secondary structure

Ramachandran plots display allowed conformations of the main chain

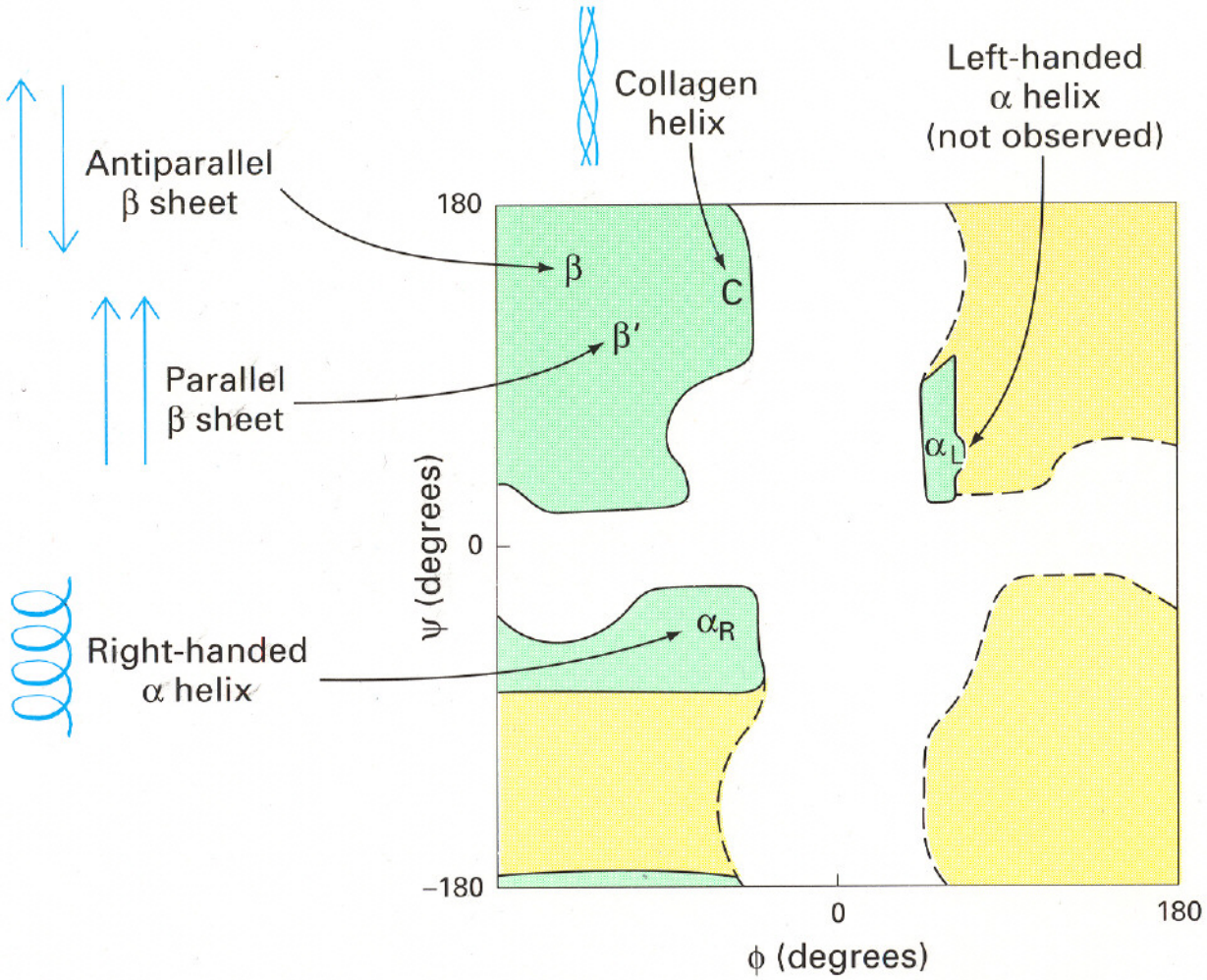


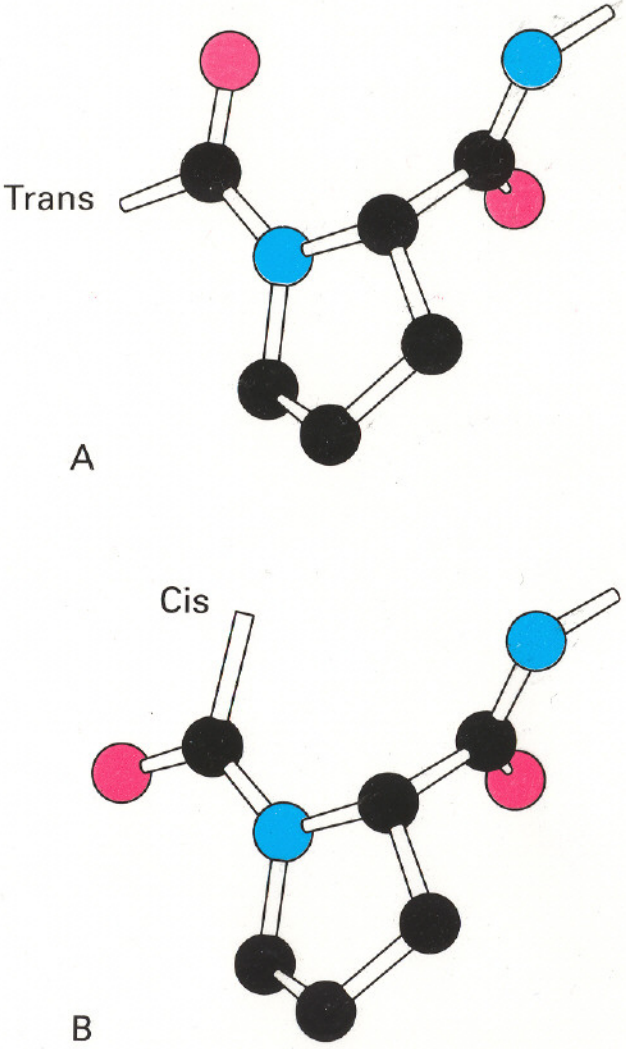
phi (ϕ)

psi (ψ)

Figure 16-5, page 420

Stryer: *Biochemistry*, Fourth Edition
© 1995 by W. H. Freeman and Company





Ramachandran plot showing allowed values of ϕ and ψ for L-alanine residues (green regions). Additional conformations are accessible to glycine (yellow regions).

Figures 16-6 and 16-7, page 421

Stryer: *Biochemistry*, Fourth Edition

© 1995 by W. H. Freeman and Company

The frequency of occurrence of amino acids in the secondary structure of proteins has provided insight into relations between sequence and conformation:

e.g.

(i) glutamate, alanine, leucine promote alpha-helices
(ii) valine, isoleucine promote beta-strands
(iii) glycine, asparagines, proline promote beta-turns

Predictions of the secondary structure based on local sequence have proved to be about 60 % accurate

Predictions of the secondary structure based on primary sequence have limited success, because:

(i) the conformational preferences of aa are not tipped all the way to one structure
(ii) tertiary interactions may be decisive in specifying the secondary structure of some protein sequences. Thus, a particular sequence can adopt an "X" conformation in one protein and a "Y" conformation in another protein.

Table 16-1
Relative frequencies of occurrence of amino acid residues in the secondary structures of proteins

Amino acid	*α helix*	*β sheet*	*β turn*
Ala	1.29	0.90	0.78
Cys	1.11	0.74	0.80
Leu	1.30	1.02	0.59
Met	1.47	0.97	0.39
Glu	1.44	0.75	1.00
Gln	1.27	0.80	0.97
His	1.22	1.08	0.69
Lys	1.23	0.77	0.96
Val	0.91	1.49	0.47
Ile	0.97	1.45	0.51
Phe	1.07	1.32	0.58
Tyr	0.72	1.25	1.05
Trp	0.99	1.14	0.75
Thr	0.82	1.21	1.03
Gly	0.56	0.92	1.64
Ser	0.82	0.95	1.33
Asp	1.04	0.72	1.41
Asn	0.90	0.76	1.28
Pro	0.52	0.64	1.91
Arg	0.96	0.99	0.88

After T.E. Creighton. *Proteins: Structures and Molecular Properties*, 2nd ed. (W.H. Freeman, 1992), p. 256.

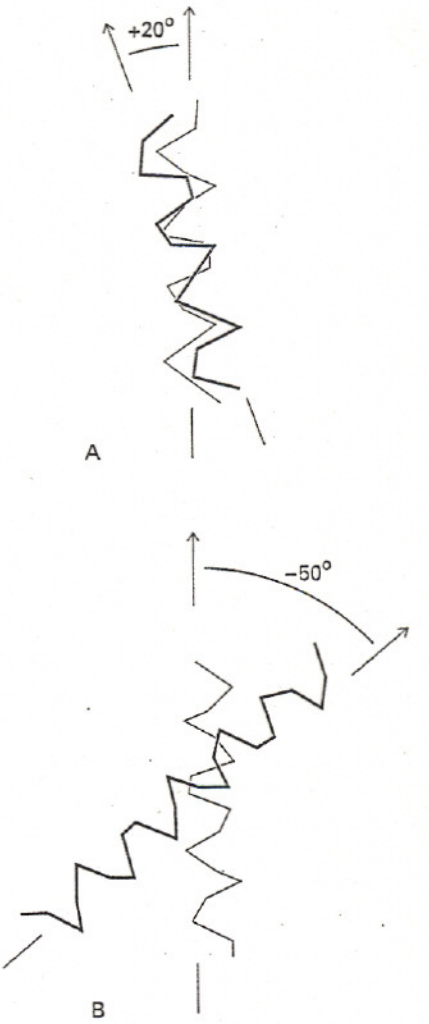


Figure 16-9
Two modes of packing of α helices. (A) +20°. (B) −50°. [After C. Chothia and A.V. Finkelstein. *Ann. Rev. Biochem.* 59(1990):1007.]

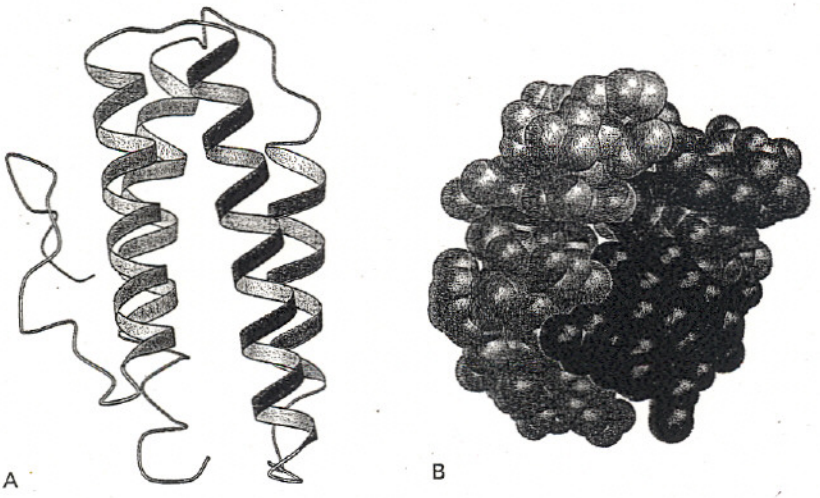


Figure 16-10
The four-helix bundle is the major structural motif of myohemerythrin, an oxygen-binding protein in worms. (A) Schematic diagram of the main-chain conformation. (B) Space-filling model of the four-helix bundle, shown in cross section. [Drawn from 2mhr.pdb. S. Sheriff, W.A. Hendrickson, and L.J. Smith. *J. Mol. Biol.* 197(1987):273.]

Folding motifs or **supersecondary structures**, which arise by the association of alpha-helices and beta-strands, play a fundamental role in protein folding:

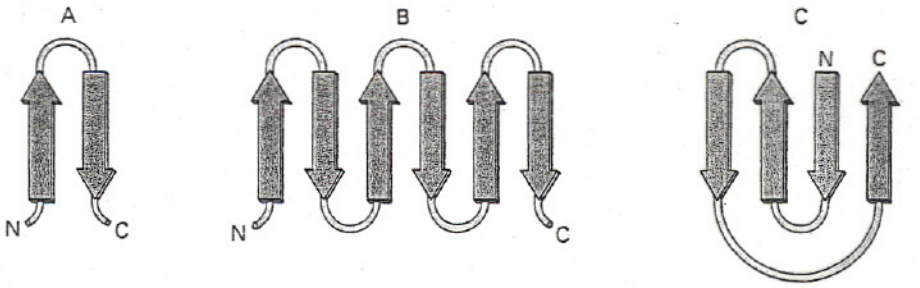
(i) four-helix bundle

(ii) β hairpin

(iii) β sheet

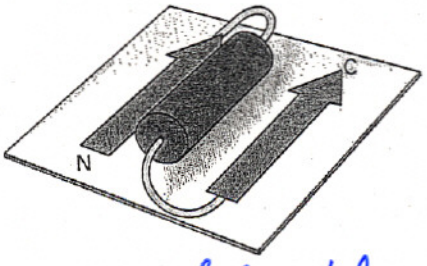
(iv) Greek key motif

(v) βαβ motif



β hairpin β sheet Greek key

Figure 16-11
Schematic diagram of β structures. (A) β Hairpin. (B) β Sheet. (C) Greek key. [After C. Branden and J. Tooze. *Introduction to Protein Structure* (Garland, 1991), p. 21.]



bαb motif

Figure 16-12
Schematic diagram of a βαβ motif. The β strands are shown in green, and the α helix joining them is shown in red. [After C. Branden and J. Tooze. *Introduction to Protein Structure* (Garland, 1991), p. 24.]

Partially folded intermediates can be detected, trapped, and studied at atomic resolution:

(i) Rapid-kinetic studies: folding intermediates are generated by abruptly changing the external conditions to favor the refolding of an unfolded protein

(ii) Trapping the disulfide-bonded intermediates: Folding of unfolded proteins containing disulfide bonds can be initiated by removing the reducing agent

(iii) Pulse-label NMR studies: An unfolded protein in a D2O-denaturing solution is refolded by diluting the sample in D2O to lower the concentration of denaturant

(iv) Synthesis of folding kernels: By removing parts of a protein, we can discern the roles of different regions in the folding process

(v) Site-directed mutagenesis

(vi) Protein design

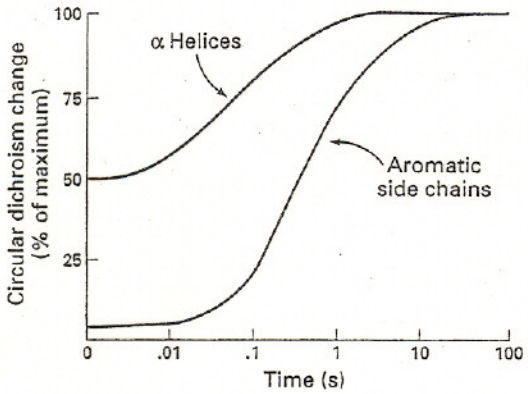


Figure 16-13
α Helix formation is more rapid than tertiary-structure rearrangements of aromatic side chains in the folding of cytochrome *c*. The kinetics of these structural changes were determined from circular dichroism measurements at 222 nm and 289 nm, respectively. [After G. Elöve, A.F. Chaffotte, H. Roder, and M.E. Goldberg. *Biochemistry* 31(1992):6879.]

Figure 16-14
The sequence of formation of disulfide bonds in proteins can be determined by trapping free cysteine residues with iodoacetate. The *S*-carboxymethyl derivative of cysteine is stable.

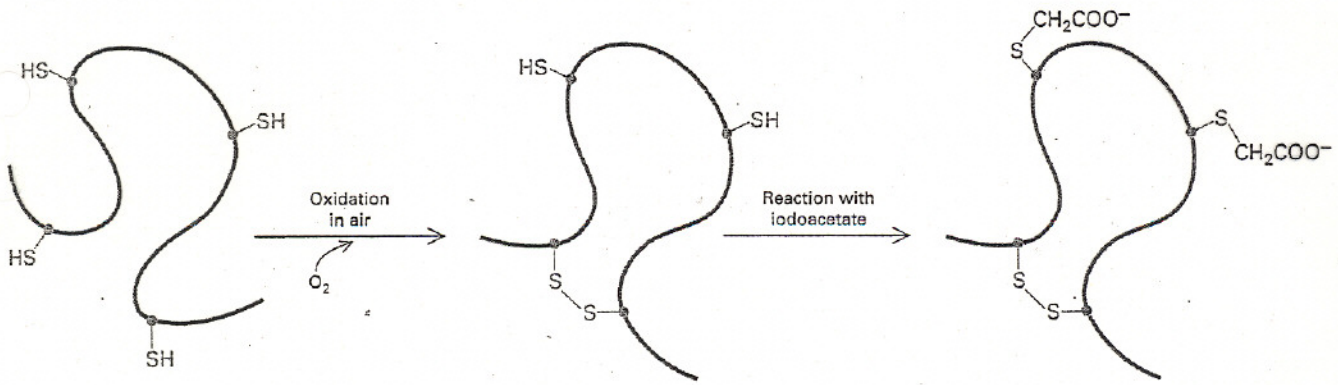
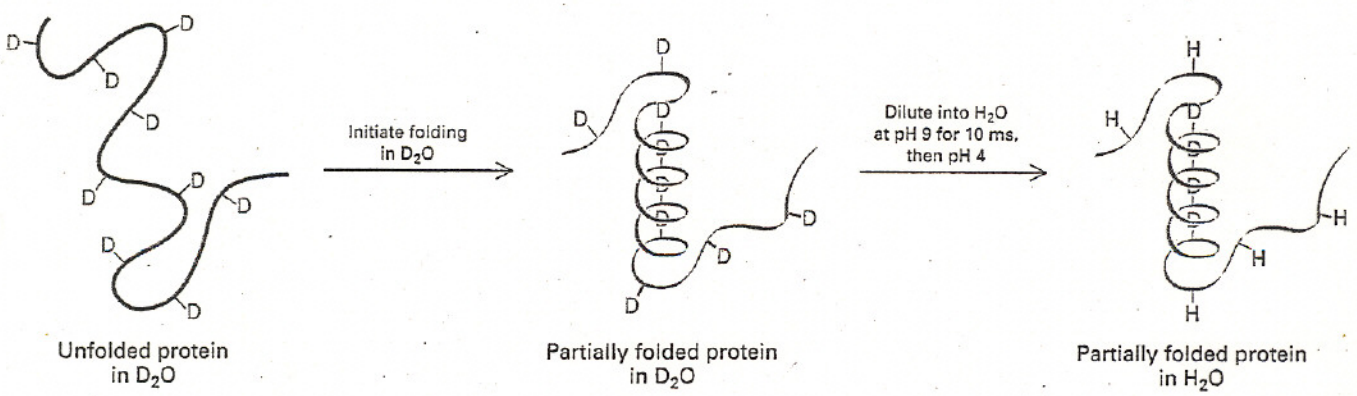


Figure 16-15
Pulsed hydrogen-deuterium exchange can be used to monitor the acquisition of secondary structure in protein folding.



3D Structure Predictions of Proteins

Predictions are difficult because:

(i) a polypeptide chain has a vast number of potential conformations

(ii) proteins are only marginally stable

Significant progress is now being made in this area for the following reasons:

(i) the structures of an increasingly large number of proteins are now available

(ii) DNA sequencing is providing a wealth of aa sequence information

(iii) Sophisticated computer programs are taking advantage of the rapidly enlarging sequence and structure databases

(iv) Predictions can be tested experimentally

If two proteins are more than 40 % identical in sequence, their backbone conformations are very likely to be nearly the same.

Two proteins with significantly different sequences can have the same backbone structure, if their hydrophobicity patterns are alike.

Functional motifs in proteins can also be identical by scanning aa sequences

<u>Design and synthesis of novel proteins</u>

(i) de nove synthesis tests our understanding of fundamental principles

(ii) can we use our knowledge of protein folding to optimize the properties of naturally occurring proteins?

(iii) Can we construct proteins with novel functions?

Molecular chaperones

Roles:
- bind nascent polypeptide chains
- bind denatured proteins
- bind proteins before their transport across membranes

Chaperones are slow ATPases / bind to hydrophobic regions of unfolded target proteins

General mechanism of action:

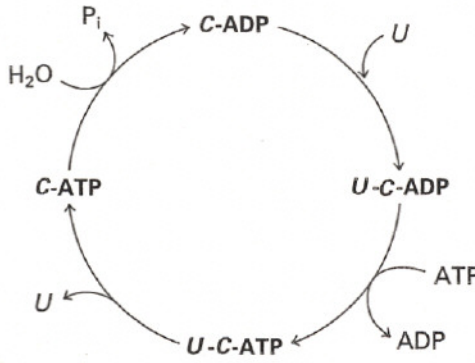
chaperone/ADP complex binds unfolded protein → → release of ADP and binding of ATP → → release of protein → hydrolysis of ATP → ...

Main chaperone families:

(i) Hsp70 protein (eg. Hsp70, Bip, Dnak)

(ii) Hsp60 family (eg. GroEL, Hsp60) → chaperonin family

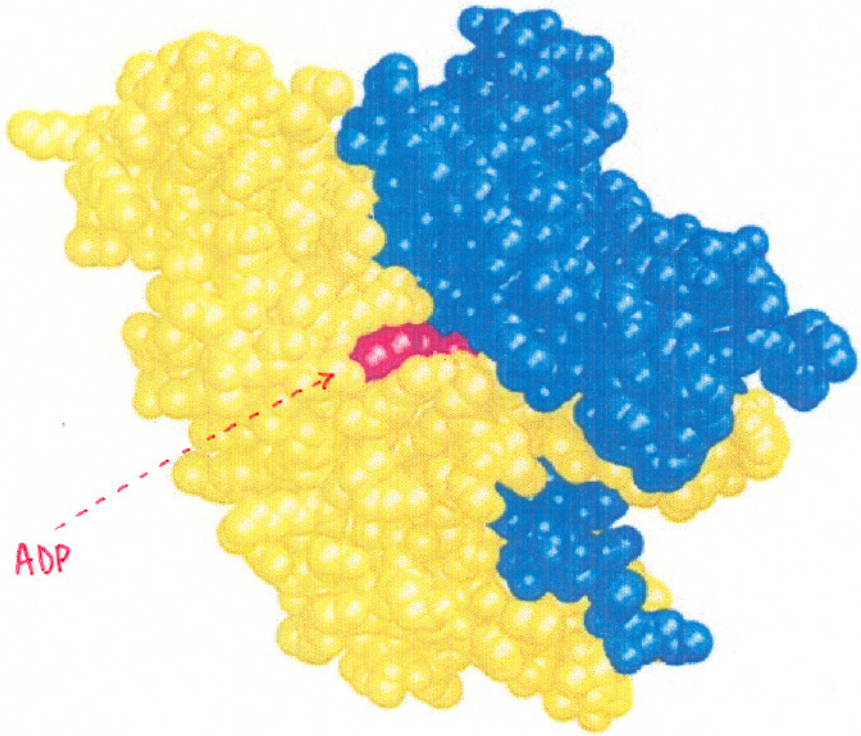
(iii) Hsp90 proteins (e.g. Hsp83)



C = chaperone;
U = unfolded peptide segment

Table 35-1
Heat-shock proteins

Hsp60 (chaperonin-60) family
GroEL (in bacterial cytosol)
Hsp60 (in mitochondrial matrix)
Rubisco binding protein (in chloroplasts)
Hsp70 (stress-70) proteins
Hsp70 (in mammalian cytosol)
BiP (in ER of eukaryotes)
Grp75 (in mitochondria)
DnaK (in bacterial cytosol)
Hsp90 (stress-90) proteins
Hsp83 (in eukaryotic cytosol)
Grp94 (in mammalian ER)
HtgP (in bacterial cytosol)



Structure of the ATPase fragment derived from Hsc70

An hsp70 class heat-shock protein

Fig 35-10