BCHS 6229
Protein Structure and Function

Lecture 3 (October 18, 2011)

Protein Folding:

Forces, Mechanisms & Characterization

The folding problem

- One of the greatest unsolved problems of Science

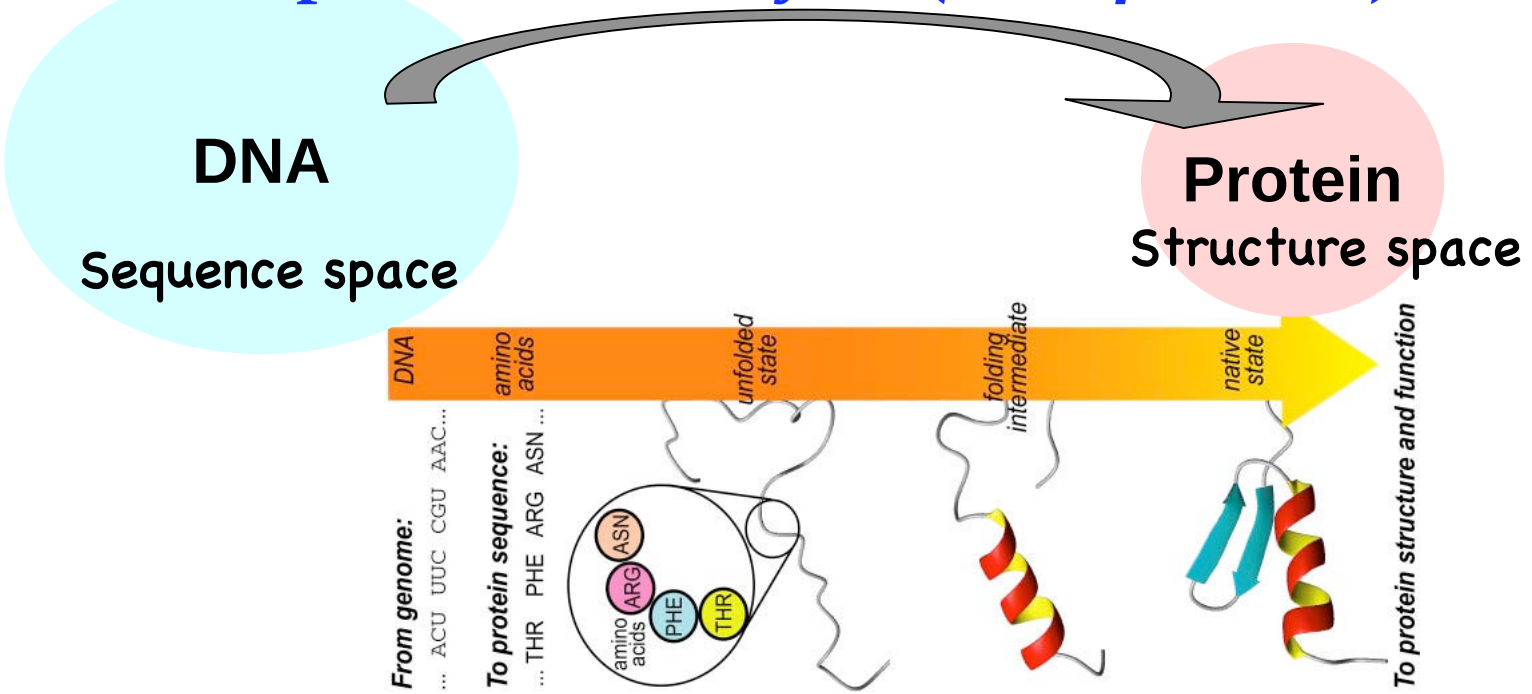
- The folding problem - Why so important?

- The mechanism(s) of protein folding
  What is meant by *mechanism*?

The folding problem

- Prediction of the 3D structure of a protein from its amino acid sequence
- Translate “Linear” DNA Sequence data to spatial information

a connection between the genome (sequence) and what the proteins actually do (their function).



Why solve the folding problem?

- Acquisition of sequence data relatively quick
- Acquisition of experimental structural information slow
- Limited to proteins that crystallize or stable in solution for NMR

Protein Stability

Chemical Stability of the covalent structure

Covalent changes, Irreversible

Conformational Stability of the folded state

Subject of our discussion

Measuring the conformational stability

Forces that stabilize protein structure: 1, 2, 3, etc..

1. The Hydrophobic Effect

Table 6-2 Hydropathy Scale for Amino Acid Side Chains

Side Chain	Hydropathy
Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8
Cys	2.5
Met	1.9
Ala	1.8
Gly	−0.4
Thr	−0.7
Ser	−0.8
Trp	−0.9
Tyr	−1.3
Pro	−1.6
His	−3.2
Glu	−3.5
Gln	−3.5
Asp	−3.5
Asn	−3.5
Lys	−3.9
Arg	−4.5

Source: Kyte, J. and Doolittle, R.F., *J. Mol. Biol.* **157,** 110 (1982).

Table 6-2 Fundamentals of Biochemistry, 2/e

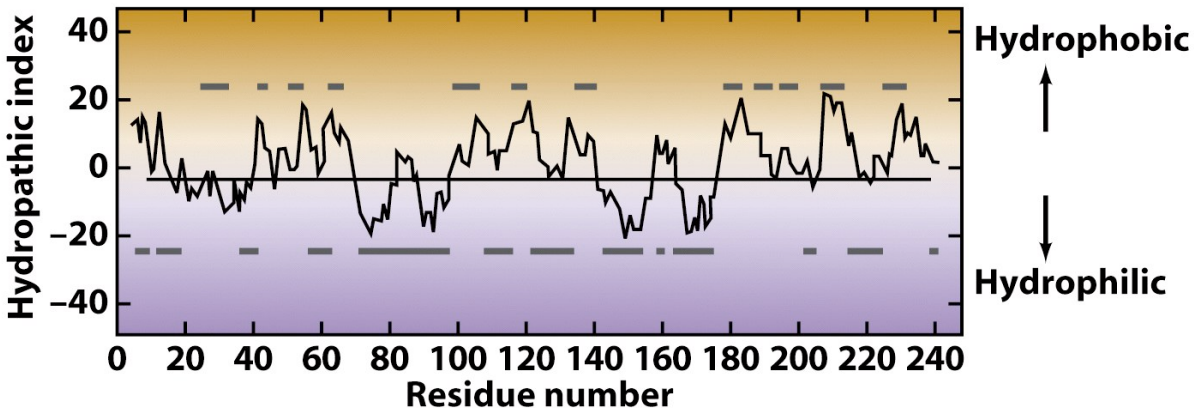


Figure 6-35 Fundamentals of Biochemistry, 2/e

2. Electrostatic Interactions

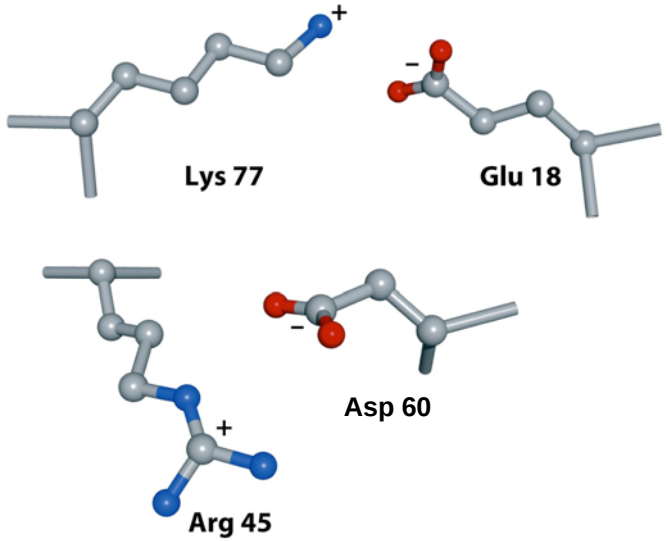
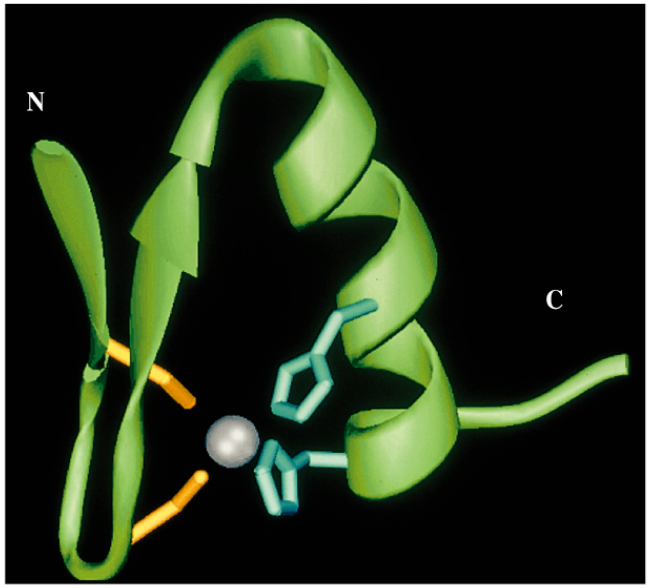


Figure 6-36 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Ion pair (salt bridge) of myoglobin

3. Chemical Cross-links



Zinc finger:
Nucleic acid-binding proteins

Protein folding dynamics

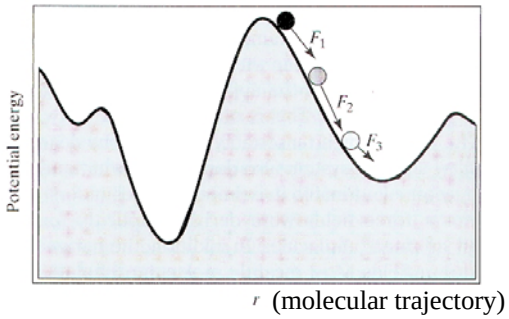
Key Concepts:

- Electrostatics, hydrogen bonds and van der Waals forces hold a protein together.
- Hydrophobic effects force global protein conformation.
- Peptide chains can be cross-linked by disulfides, Zinc, heme or other ligand compounds. Zinc has a complete *d* orbital, one stable oxidation state and forms ligands with sulfur, nitrogen and oxygen.
- Proteins refold very rapidly and generally in only one stable conformation.

Uses of Potential Energy Function:

Energy minimization

Brings a total energy of a molecular conformation to a low energy well of the potential energy surface



Molecular Dynamics

Simulation: atoms and molecules are allowed to interact for a period of time by approximations of known physics, giving a view of the motion of the atoms

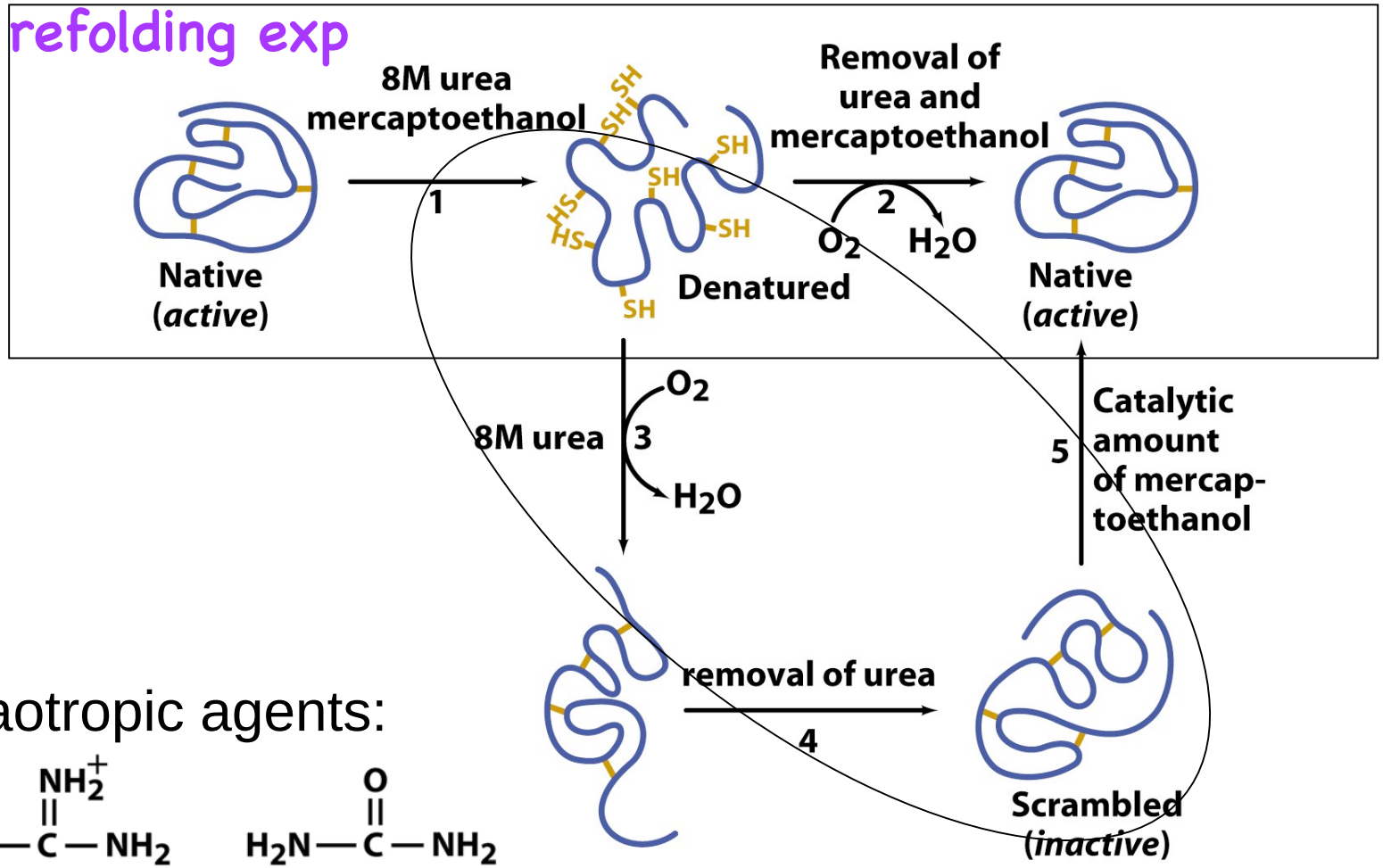
Monte Carlo Simulations

Simulation of a randomly distributed ensemble of trajectories (repeated random sampling)

Vibrational Analysis
X-ray refinement
Constraint/Restraint modeling
Free Energy Differences …

Protein denaturation and renaturation

An observation by Afinsen and colleagues



Chaotropic agents:

Guanidinium ion **Urea**

Unnumbered figure pg 159 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

by making the polar and non-polar residues have similar solubility

Anfinsen's Dogma

THE KINETICS OF FORMATION OF NATIVE RIBONUCLEASE DURING OXIDATION OF THE REDUCED POLYPEPTIDE CHAIN

By C. B. Anfinsen, E. Haber,* M. Sela,† and F. H. White, Jr.

LABORATORY OF CELLULAR PHYSIOLOGY AND METABOLISM, NATIONAL HEART INSTITUTE, NATIONAL INSTITUTES OF HEALTH

Communicated by John T. Edsall, July 31, 1961

Bovine pancreatic ribonuclease is completely reduced by treatment with mercaptoethanol in 8 *M* urea to yield a randomly coiled polypeptide chain containing eight cysteine residues.[1-3] Under optimal conditions of polypeptide concentration and pH, essentially complete reformation of the disulfide bonds of the native enzyme occurs in the presence of molecular oxygen.[2, 3] From chemical and physical studies of the reformed enzyme, it may be concluded that the information for the correct pairing of half-cystine residues in disulfide linkage, and for the assumption of the native secondary and tertiary structures, is contained in the amino acid sequence itself.

Preliminary to studies on the interactions involved in the refolding process, and to establish the order of chemical events during the formation of active protein,

Anfinsen C, Haber E, Sela M, White F jr. (1961) PNAS 47:1309-14.

Levinthal paradox

Consider a 100 residue protein. If each residue is considered to have just 3 possible conformations the total number of conformations of the protein is 3^{100}. Conformational changes occur on a time scale of 10^{-13} seconds i.e. the time required to sample all possible conformations would be $3^{100} \times 10^{-13}$ seconds which is about 10^{27} years. Longer than the age of the universe!

Therefore, proteins must fold in "pre-arranged pathways" and in a cooperative manner.

Levinthal, C (1968). "Are there pathways for protein folding?". Journal de Chimie Physique et de Physico-Chimie Biologique 65: 44–45
Zwanzig R, et al. (1992). "Levinthal's paradox". PNAS 89 (1): 20–22.

Mechanisms of the Protein Folding Reaction

1. Levinthal random sampling (paradox)
2. Sequential Folding Model
3. Nucleation/Growth Model
4. Diffusion-Collision-Adhesion
5. Framework Model
6. Hydrophobic Collapse Model
7. Jigsaw Puzzle Model

Cooperativity in protein-folding

How a globally optimal state can be found without a global search?

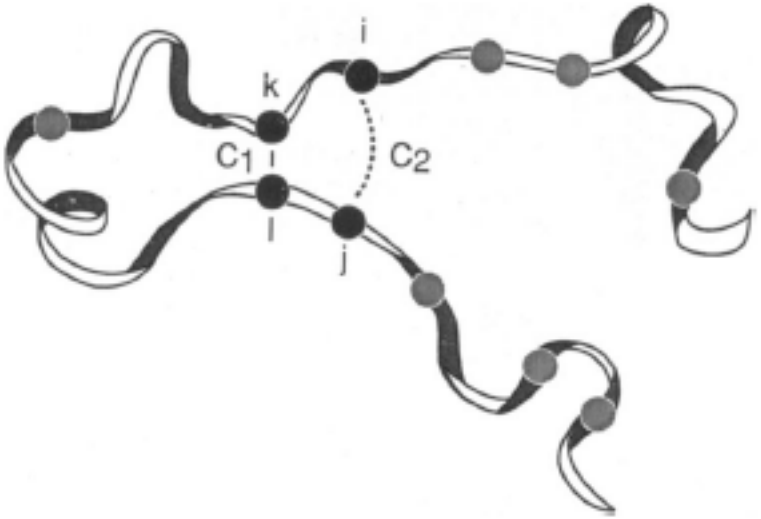
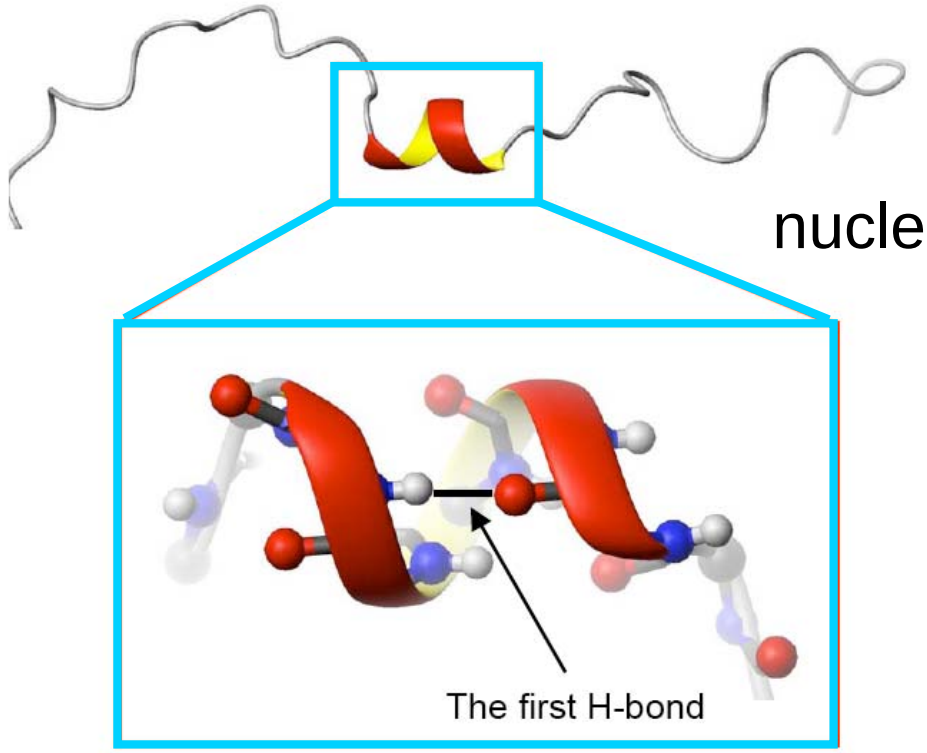


FIG. 3. Origin of cooperativity. Given constraint C_1 (k,l), monomer i explores mainly the possibility of pairing with j, not with all other H monomers; thus, this search is nonexhaustive. It is cooperative because the probability of forming contact $C_2 = (i,j)$ is much higher if C_1 is formed than in the absence of C_1 because the conformational entropy is higher for the latter.

Dill et al., PNAS (1993) 90:1942-1946

Coil-Helix transition

Paradigm for cooperativity in biopolymers

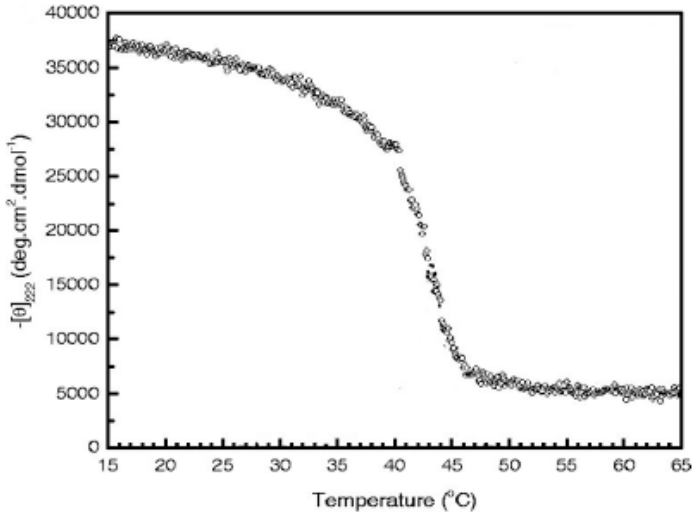


nucleation of α-helix

Initiation of a helical turn is much harder than appending another residue to a helical segment, due to higher entropic penalty.

Cooperative transition has a sigmoidal profile

Melting curve of a coiled-coil protein in aqueous medium



Cooperativity results *roughly* in a two-state system

The free energy of protein folding

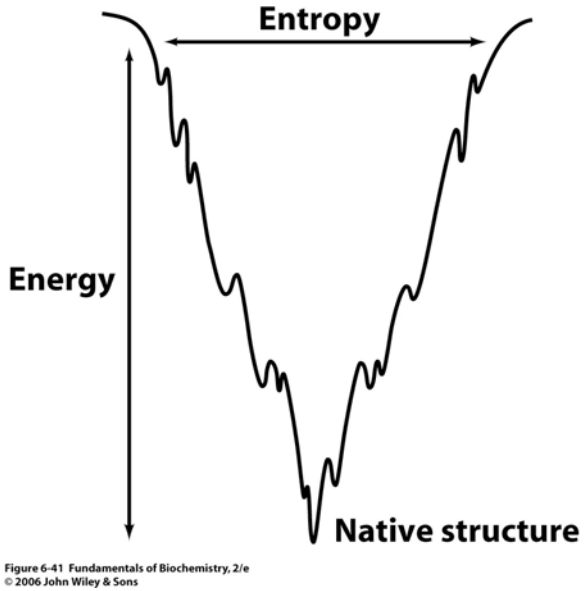
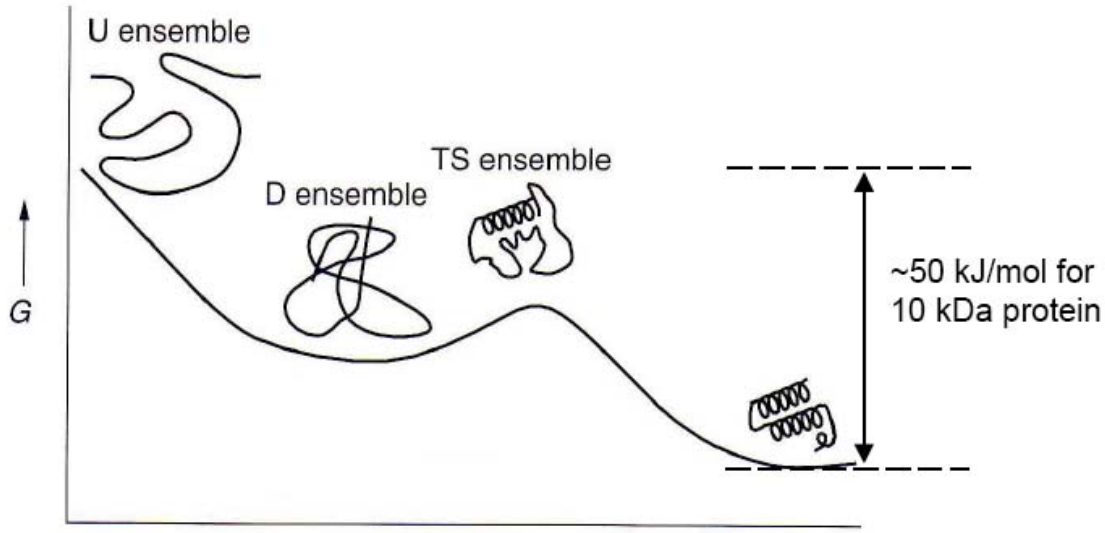


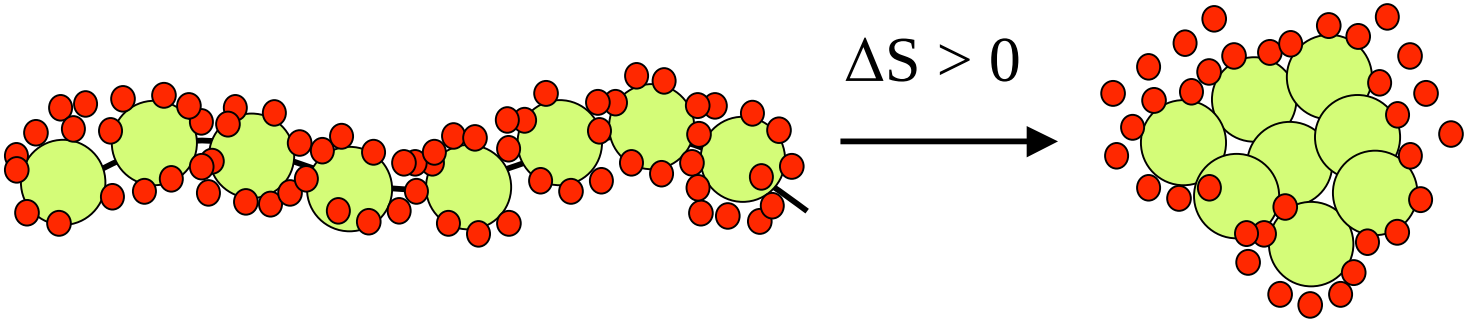
Figure 6-41 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Folding funnel
Energy-entropy relationship for protein folding



From Alan Fersht. “Structure and Mechanism in Protein Science”

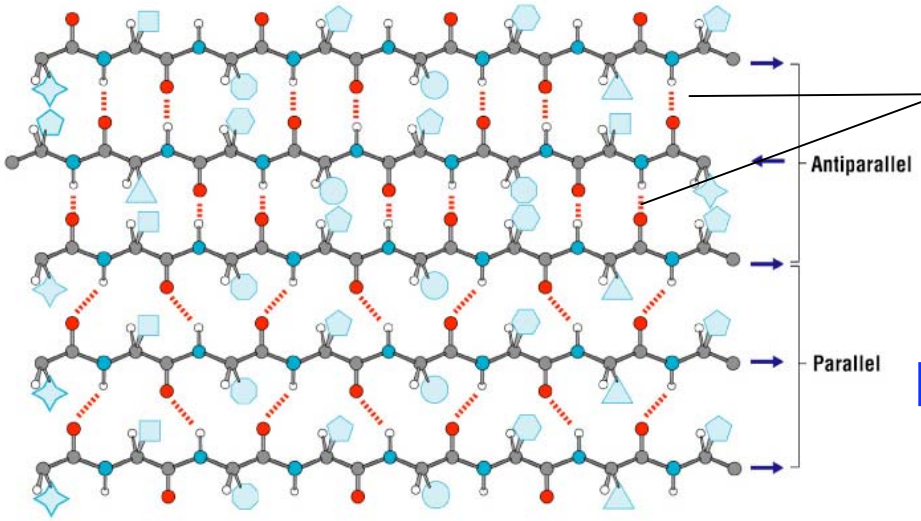
Hydrophobic interaction: entropic



More Hydrocarbon-Water Interfacial Area,
More Water ordered

Less Hydrocarbon-Water Interfacial Area,
Less Water ordered

Hydrogen Bonds: enthalpic



H-bonds

H-bonds within protein

&

H-bonds with water (hydration)

$\Delta H_{H\text{-bonds}}$ = -(4~20) kJ•mol^{-1}

Other Enthalpic interactions?

-

-

-

-

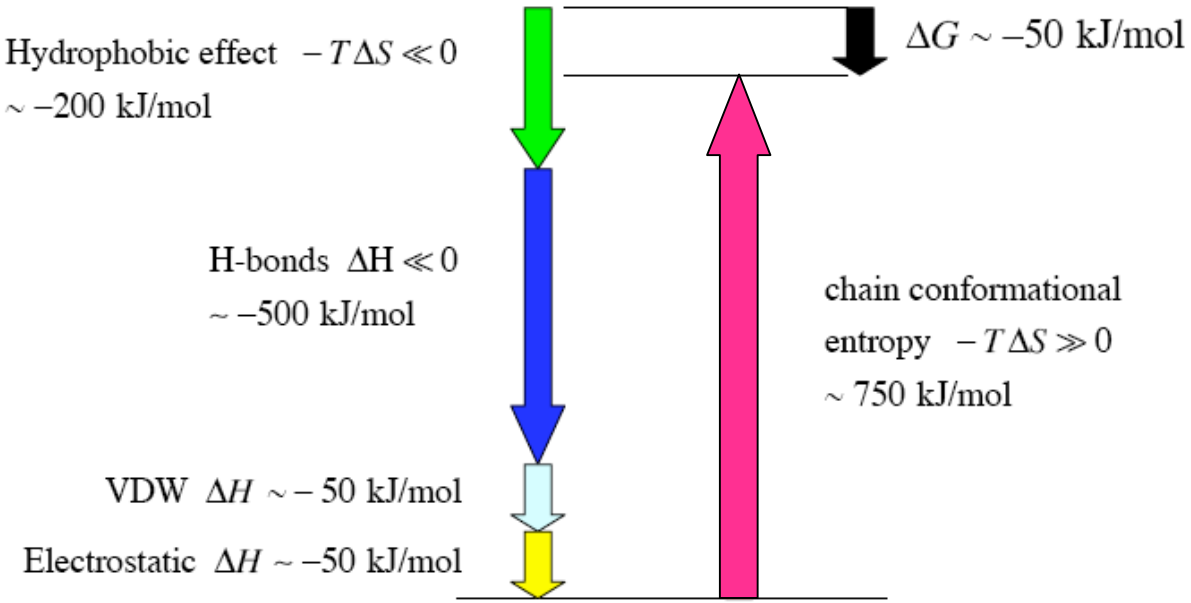
Dissecting the free energy of protein folding

Unfolded ↔ Folded

$\Delta G = \Delta H - T\Delta S$ $\Delta G =\sim -50$ kJ/mol

Near freezing T, entropy of H2O arc non-polar residues i less different from those around polar residues

Heating makes ΔH less negative



Protein folding is a balance of forces

- Proteins are only marginally stable
- Free energies of unfolding ~5-15 kcal/mol
- The protein fold depends on the summation of all interaction energies between any two individual atoms in the native state

Denatured (D) ↔ Native (N)

$$\Delta G_{D\text{-}N} = \Delta H_{D\text{-}N} - T\Delta S_{D\text{-}N}$$

- Also depends on interactions that individual atoms make with water in the denatured state

Thermodynamics of unfolding

- Denatured state has a high configurational entropy

$$S = k_B \ln W$$

Where W is the number of accessible states

k_B is the Boltzmann constant

- Native state conformationally restricted
- Loss of entropy balanced by a gain in enthalpy

Define Entropy from a Probability Distributions

- Number of way of arranging N particles in n_i groups:

$$W = \frac{N!}{n_1! n_2! n_3! n_4! \cdots n_N!}$$

- Natural log is chosen because

$$k_B \ln(W_1 W_2) = k_B \ln(W_1) + k_B \ln(W_2)$$

Measuring the conformational stability

For a two-state unfolding

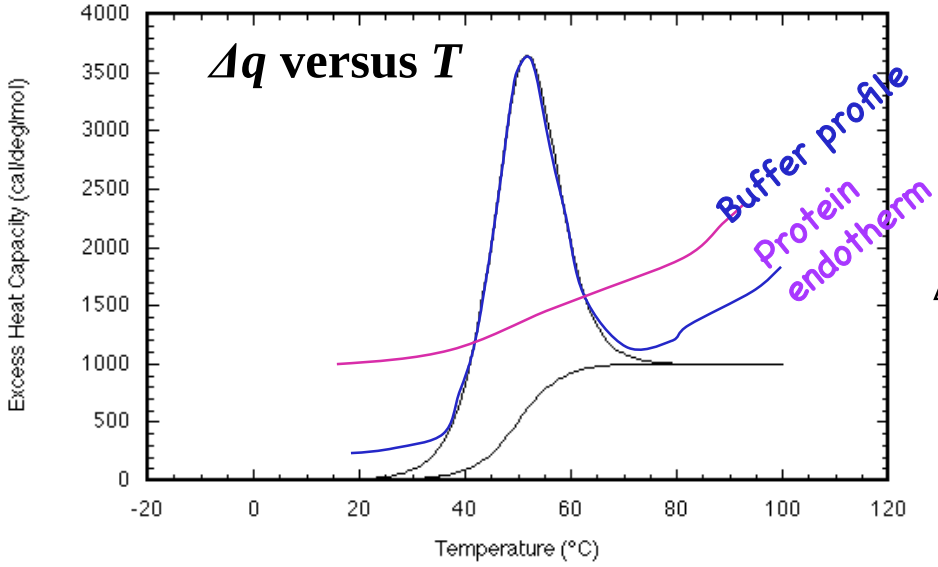
Folded (F) ↔ Unfolded (U)

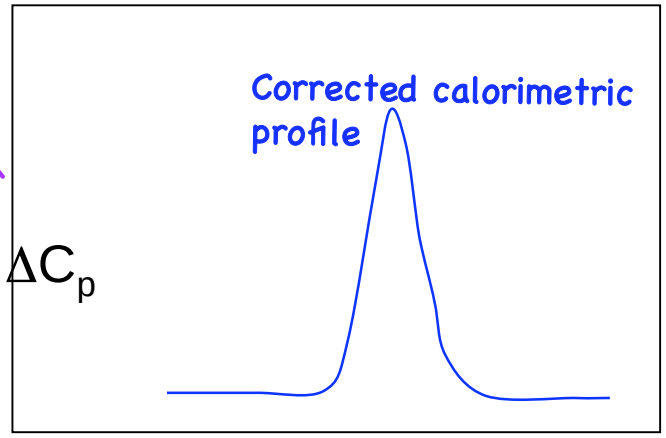
Native (N) ↔ Denatured (D)

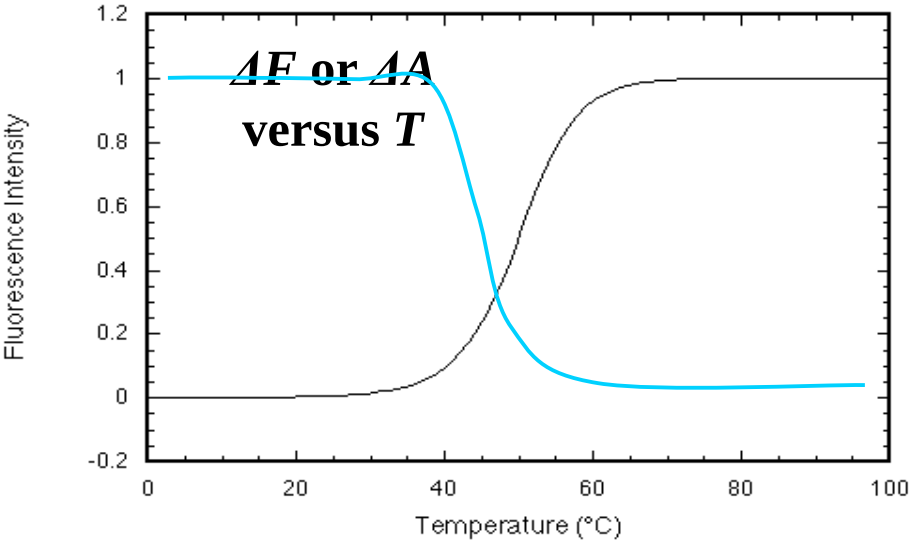
Selecting a technique to follow unfolding

UV difference spectroscopy
Fluorescence and Circular dichroism (CD)
Biological activity
Optical rotatory dispersion (ORD)
Nuclear Magnetic Resonance (NMR)
Viscosity and other hydrodynamic methods, etc…

Measuring thermal denaturation

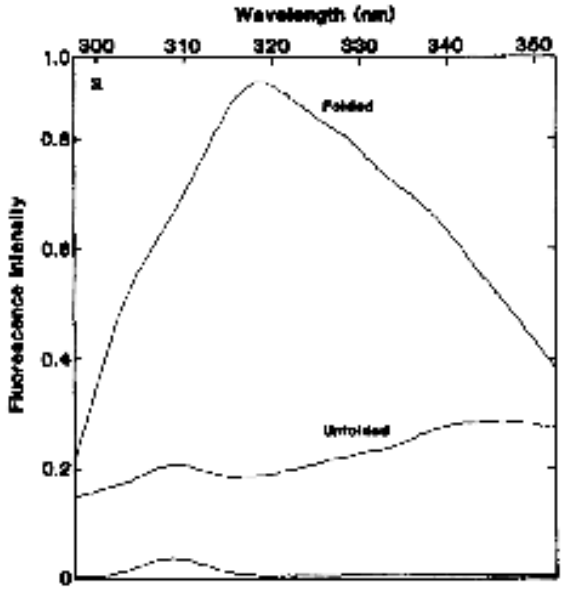


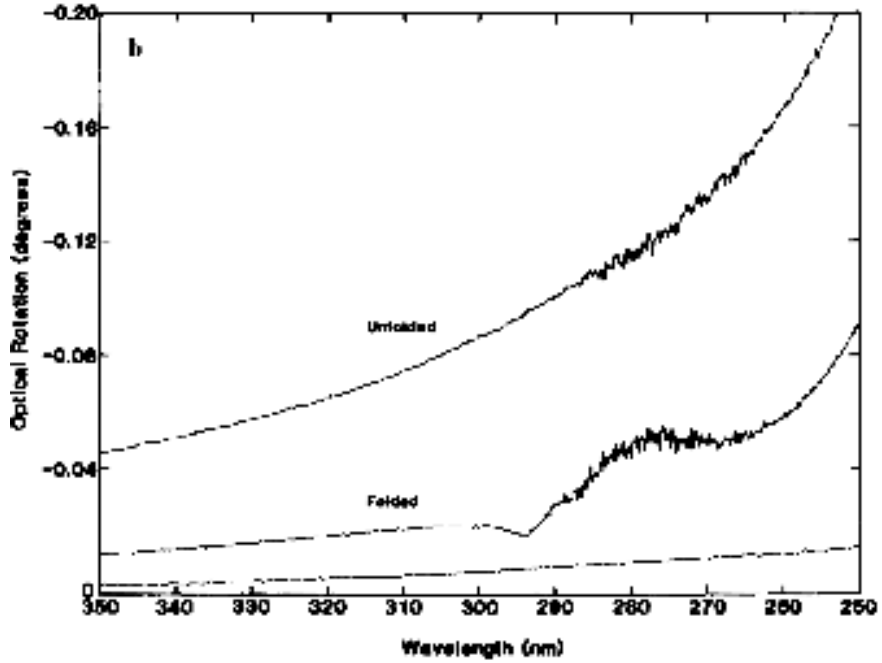




fraction folded
vs.
fraction unfolded

Example: Spectra of folded and unfolded RNase T1



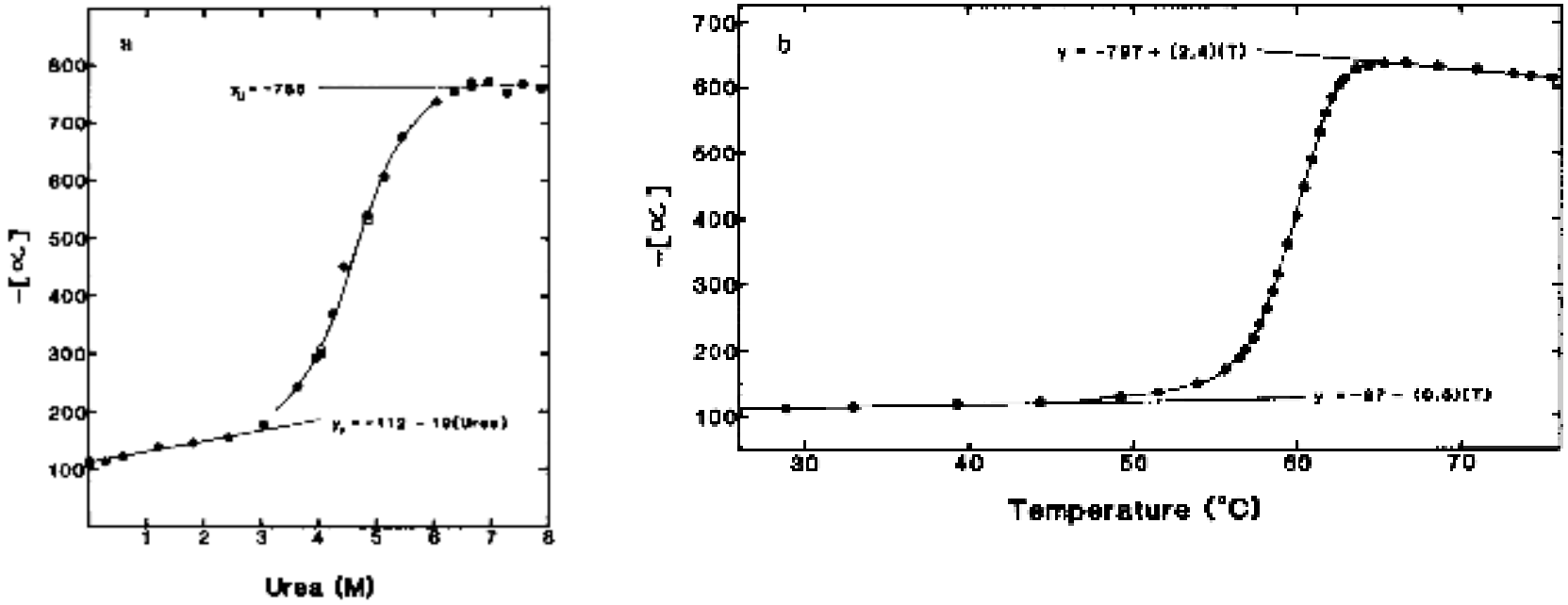


Intrinsic fluorescence emission spectra upon excitation at 278 nm

ORD spectra

Determining unfolding curves

Equilibrium and reversibility
Urea and GdmCl / Thermal unfolding



Urea- and heat-induced unfolding transitions of RNase T1

y: physical parameters
pre-transition region
transition region
post-transition region

Analyzing unfolding curves

Two-state folding mechanism

Many small single domain proteins are closely approximated by the two state model and have few intermediates.

Folded (F) ↔ Unfolded (U) (1)

$f_U = (y_F - y) / (y_F - y_U)$ (2)

where $(f_F + f_U = 1, \quad y = y_F f_F + y_U f_U)$

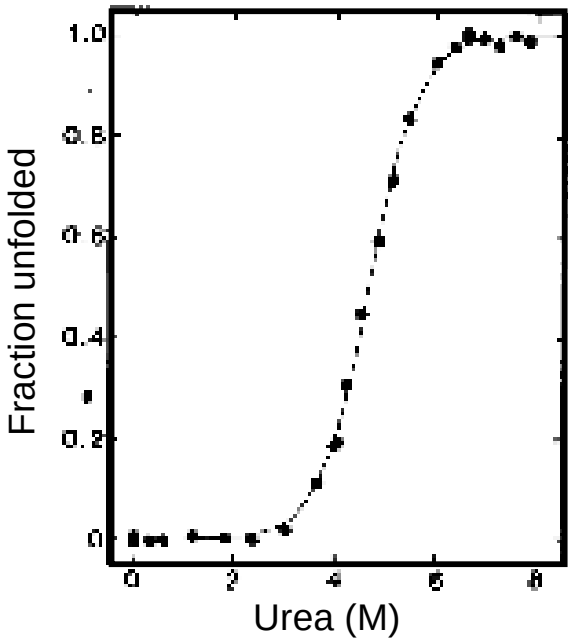
$K = f_U/(1 - f_U) = f_U/f_F = (y_F - y)/(y - y_u)$ (3)

$\Delta G = -RT \ln K = -RT \ln [(y_F - y)/(y - y_u)]$ (4)

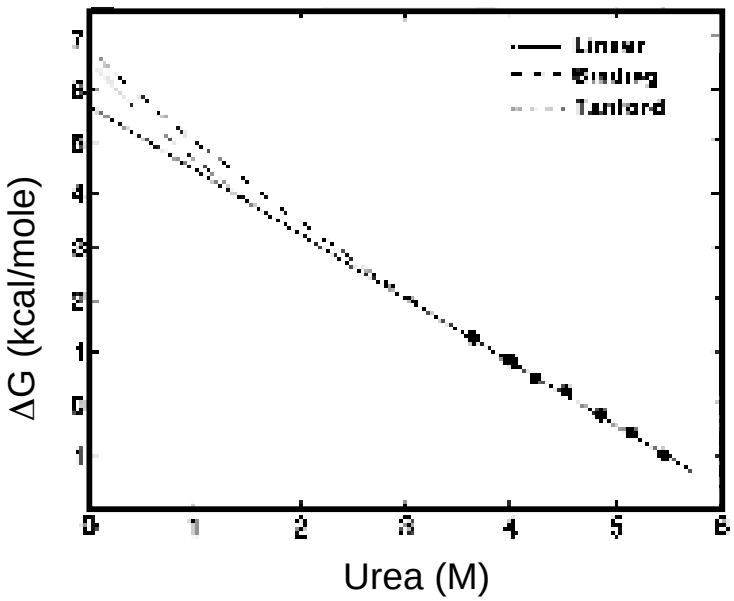
$\Delta G = \Delta G(H_2O) - m[\text{denaturant}]$ (5)

Analyzing unfolding curves

Urea and GdmCl unfolding



Fraction of RNase T1 unfolded, f_U, as a function of [Urea]



ΔG for RNase T1 unfolding, as a function of [Urea]

Analyzing unfolding curves

Thermal unfolding

$$d(\ln K)/d(1/T) = -\Delta H/R \qquad (6)$$

$$\ln K = -\frac{\Delta H^{\ominus}}{RT} + \frac{\Delta S^{\ominus}}{R}$$ van't Hoff equ.

van't Hoff plot (lnK vs. 1/T) of protein unfolding transitions are found to be non-linear --- indicates that ΔH varies with T

$$d(\Delta H)/d(T) = C_p(U) - C_p(F) = \Delta C_p \qquad (7)$$

Recall $\Delta G^{\ominus} = \Delta H^{\ominus} - T\Delta S^{\ominus}$ $\Delta G^{\ominus} = -RT \ln K$

$$\Delta G(T) = \Delta H_m(1-T/T_m) - \Delta C_p[(T_m - T)] + T \ln (T/T_m)] \quad (8)$$

Specific Heats

- Property defined by the amount of heat required per unit mass to raise the temperature by one degree.

- 1) Constant Volume C_v

$$C_v = \frac{1}{m}\left(\frac{\partial Q}{\partial T}\right)_v = \frac{1}{m}\left(\frac{\partial U}{\partial T}\right)_v = \left(\frac{\partial u}{\partial T}\right)_v$$

- **2) Constant Pressure C_p**

$$C_p = \frac{1}{m}\left(\frac{\partial Q}{\partial T}\right)_p = \frac{1}{m}\left(\frac{\partial H}{\partial T}\right)_p = \left(\frac{\partial h}{\partial T}\right)_p$$

Differential Scanning Calorimetry (DSC)

direct measurement of T_m, ΔH, ΔC_p from increased in heat transfer occurring with unfolding as a F(T)

Constant pressure

(versus bomb-calorimeter)

Measures ΔH

$C_p = dq/dT$

$C_p = dH/dT$

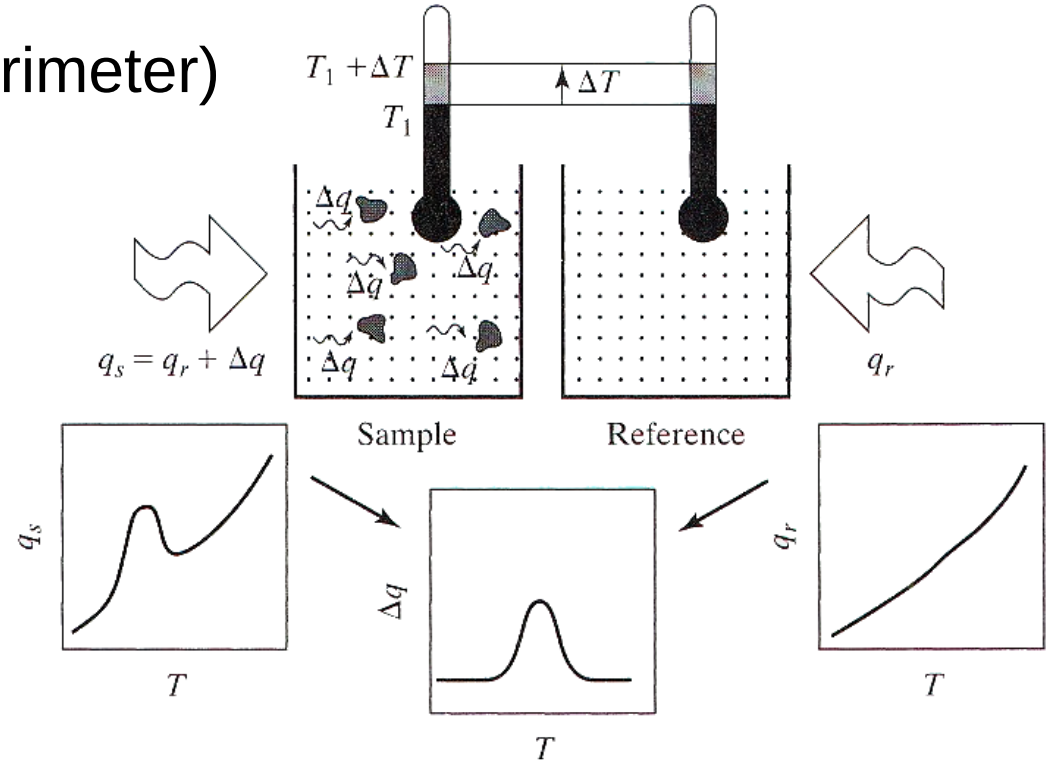


Figure 2.7 Experimental setup for a differential scanning calorimetry experiment. The amount of heat required to increase the temperature by the same increment (ΔT) of a sample cell (q_s) is higher than that required for the reference cell (q_s) by the excess heat absorbed by the molecules in the sample (Δq). The resulting DSC scans with the reference subtracted from the sample shows how this excess heat changes as a function of temperature.

Melting Points

Knowing $\Delta G=0$ at equilibrium:

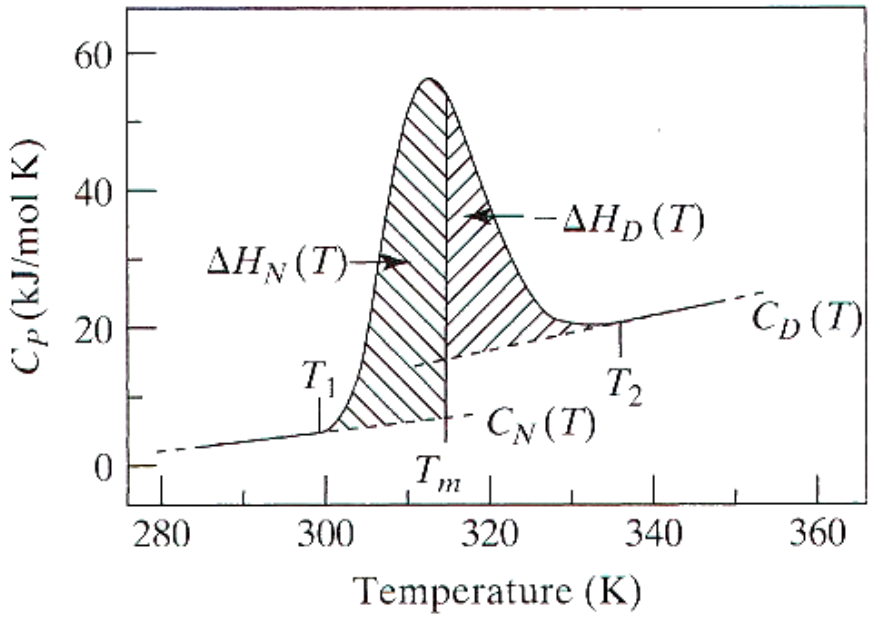
$$\Delta S_m^0 = \Delta H_m^0 / T_m$$

So now at any Temperature:

Change in *Cp*

$$\Delta H^0(T) = \Delta H_m^0 + \int_{T_m}^{T} \Delta C_p \partial T$$

$$\Delta S^0(T) = \Delta H_m^0 / T_m + \int_{T_m}^{T} \frac{\Delta C_p}{T} \partial T$$



For reversible two state model
Note that $C_D > C_N$

Gibbs Free Energy at the melting point is:

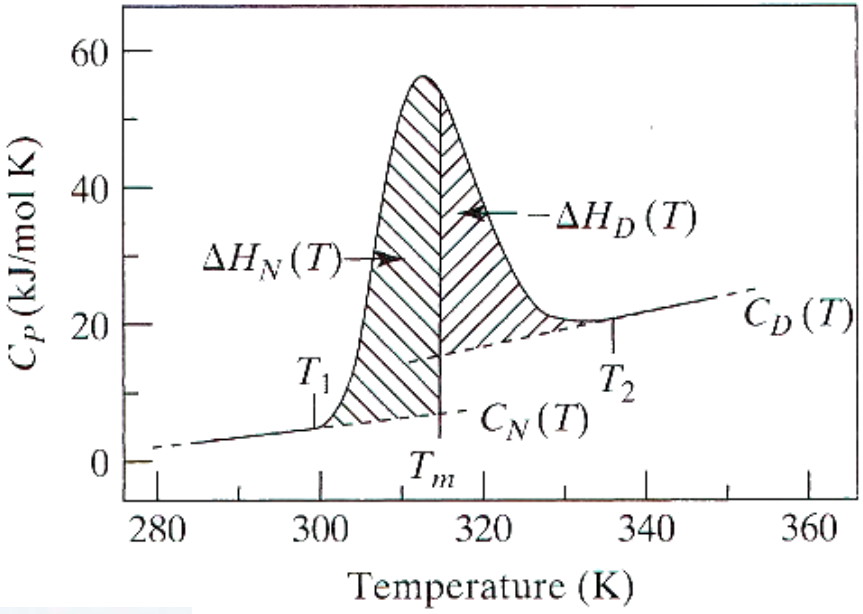
$$\Delta G^0(T) = \Delta H^0(T) - T\Delta S^0(T)$$

Enthalpy and Fractional Unfolding

Parallel to the van't Hoff equation:

$$\Delta H^0_{vH}(T) = -RT^2 \frac{d \ln K(T)}{dT}$$

So change in enthalpy is related to the fractions of native and denatured forms of the protein.:



$$\Delta H^0_{vH}(T) = -RT^2\left(\frac{\Delta C_n(T)}{\Delta H^0_n(T)} - \frac{\Delta C_d(T)}{\Delta H^0_d(T)}\right)$$

ΔH_{vH} is determined from the temperature dependence of K_{eq}

The van't Hoff Relationship

- Methodology of finding dH and dS from experimental data.

$$\ln K_{eq} = -\frac{\Delta G^\circ}{RT} = -\left(\frac{\Delta H^\circ - T\Delta S^\circ}{RT}\right)$$

$$\ln K_{eq} = \frac{1}{R}\left(\Delta S^\circ - \frac{\Delta H^\circ}{T}\right)$$

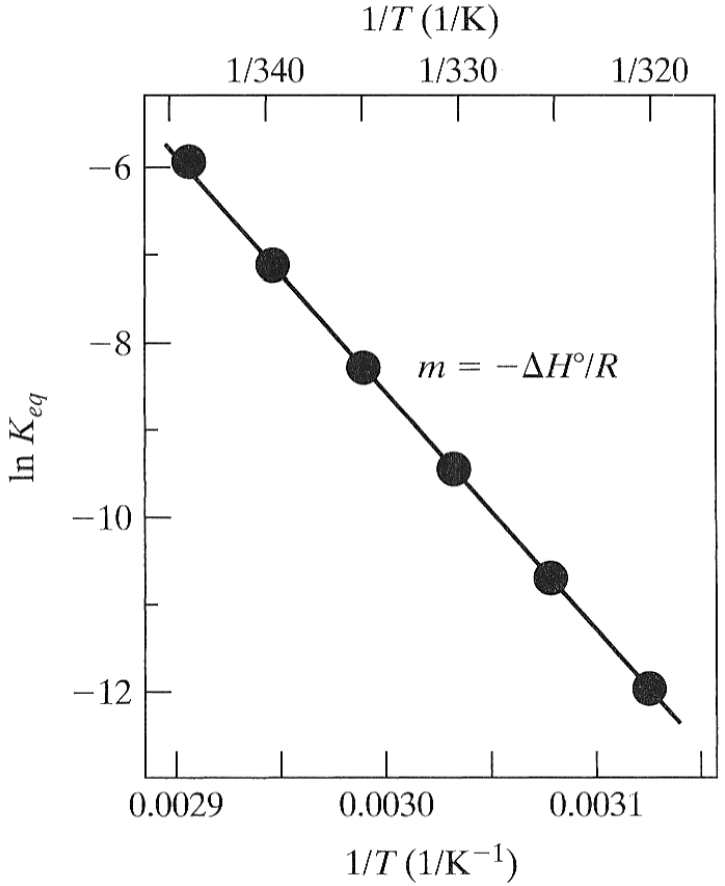


Figure 2.6 Van't Hoff plot for the thermal unfolding of the coiled coil domain of the GCN4 transcription factor. [Adapted from Holtzer et al. (2001), *Biophys. J.* **80**, 939–951.]

High ΔC_p changes enthalpy significantly with T

- For a two state reversible transition N↔D

$$\Delta H^0(T) = \Delta H^0_m + \int_{T_m}^{T} \Delta C_p \partial T$$

$$\Delta H_{D\text{-}N(T2)} = \Delta H_{D\text{-}N(T1)} + \Delta C_p(T_2 - T_1)$$

- When $T_2 > T_1$, & $\Delta C_p\ (C_D - C_N)$ is positive and the enthalpy becomes more positive

- i.e. $\Delta H_{D\text{-}N}$ favors the native state

High ΔC_p changes entropy with temperature

- For a two state reversible transition N↔D

$$\Delta S^0(T) = \Delta H^0_m/T_m + \int_{T_m}^{T} \frac{\Delta C_p}{T} \partial T$$

$$\Delta S_{D\text{-}N(T2)} = \Delta S_{D\text{-}N(T1)} + \Delta C_p ln[T_2 / T_1]$$

- When $T_2 > T_1$, & ΔC_p $(C_D - C_N)$ is positive the entropy becomes more positive

- i.e. $\Delta S_{D\text{-}N}$ favors the denatured state

Free energy of unfolding

- For

$$\Delta G_{D\text{-}N} = \Delta H_{D\text{-}N} - T\Delta S_{D\text{-}N}$$

- Gives

$$\Delta G_{D\text{-}N(T2)} = \Delta H_{D\text{-}N(T1)} + \Delta C_p(T_2 - T_1) - T_2(\Delta S_{D\text{-}N(T1)} + \Delta C_p ln[T_2 / T_1])$$

- As temperature increases $T\Delta S_{D\text{-}N}$ increases and causes the protein to unfold and midpoint is where $\Delta G_{D\text{-}N} = 0$

Calculating the Equilibrium Mid Point

$$\Delta G^0(T) = \Delta H^0(T) - T\Delta S^0(T)$$

$$\Delta H^0(T) = \Delta H^0_m + \int_{T_m}^{T} \Delta C_p \partial T$$

$$\Delta S^0(T) = \Delta H^0_m / T_m + \int_{T_m}^{T} \frac{\Delta C_p}{T} \partial T$$

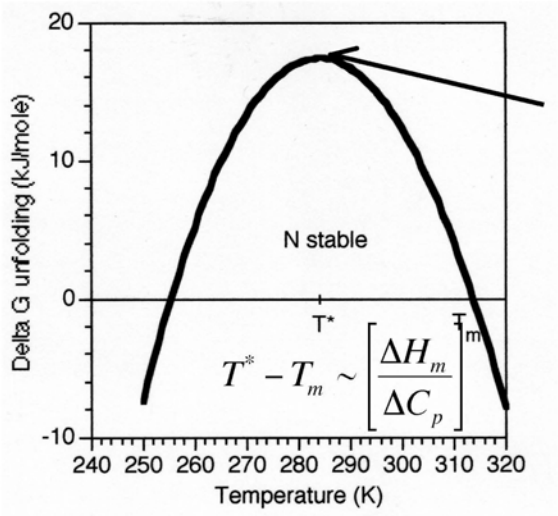
case (b)ΔC_p is constant ($\neq 0$);

$\therefore \Delta H^\circ$ and ΔS° are temperature dependent

$$\frac{d\Delta H^\circ}{dT} = \Delta C_p \Rightarrow \Delta H^\circ(T) = \Delta H^\circ(T_o) + \Delta C_p(T - T_o) \quad T_o = \text{reference temp.}$$

$$\frac{d\Delta S^\circ}{dT} = \frac{\Delta C_p}{T} \Rightarrow \Delta S^\circ(T) = \Delta S^\circ(T_o) + \Delta C_p \ln\left[\frac{T}{T_o}\right]$$

$$\therefore \Delta G^\circ(T) = \Delta H_m\left[\frac{T_m - T}{T_m}\right] - \Delta C_p\left[T_m - T\left(1 - \ln\left[\frac{T}{T_m}\right]\right)\right] \text{ with } T_o \equiv T_m$$



$$\Delta G(T^*) \cong \left[\frac{\Delta H^2_m}{2T_m \Delta C_p}\right]$$

stability curve given by ΔH_m, T_m and $\Delta C_p(>0)$

quadratic approximation

Cold unfolding

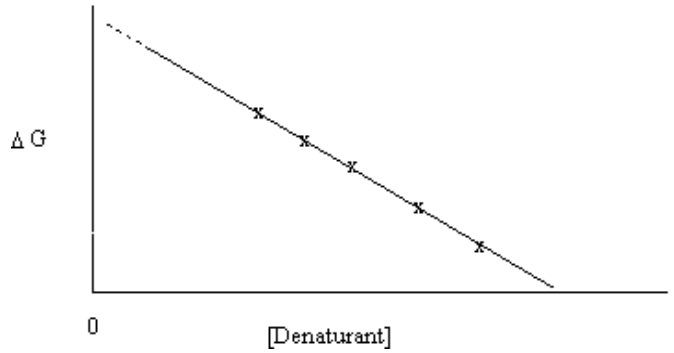
- Due to the high value of ΔC_p
- Lowering the temperature lowers the enthalpy decreases

$$T_c = T^2_m / (T_m + 2(\Delta H_{D\text{-}N} / \Delta C_p))$$

i.e. $T_m \sim 2\,(\Delta H_{D\text{-}N}) / \Delta C_p$

Solvent denaturation

- Guanidinium chloride (GdmCl) $H_2N^+=C(NH_2)_2 \bullet Cl^-$
- Urea H_2NCONH_2
- Solubilize all constitutive parts of a protein
- Free energy transfer from water to denaturant solutions is linearly dependent on the concentration of the denaturant



- Thus free energy is given by

$$\Delta G_{D\text{-}N} = \Delta G^{H2O}{}_{D\text{-}N} - m_{D\text{-}N}\,[denaturant]$$

Two state transitions in multi-state reactions
&
Rate determining steps

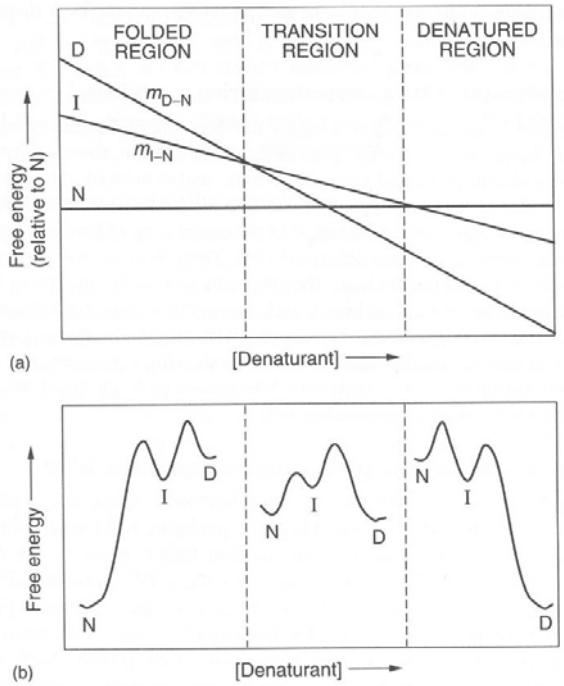


Figure 17.5 Apparent two-state transitions in multistate reactions. (a) Free energies of D and I are plotted against [denaturant], relative to that of N, assuming that simple linear equations such as 17.6 are followed. Typically, in the absence of denaturant, D is the most unstable, and I the next most unstable. The free energy of D decreases most rapidly with increasing denaturant because D has the greatest exposed surface area. Consequently, the relative energies of the three states change with increasing [denaturant]. In the transition region, the equilibrium is effectively between D and N, because I is present at low concentration. (b) The relative energies of D, I, and N.

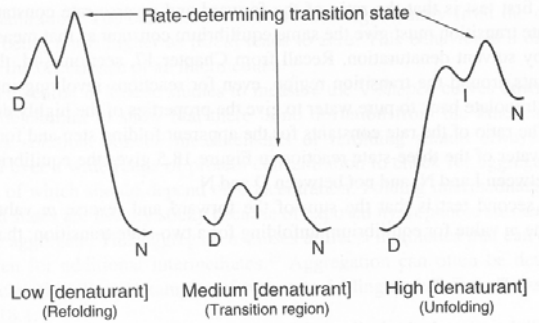


Figure 18.4 Free energy profiles for a folding reaction in which there is an intermediate that is always at higher energy than D or N. The kinetics are indistinguishable from those of a two-state system.

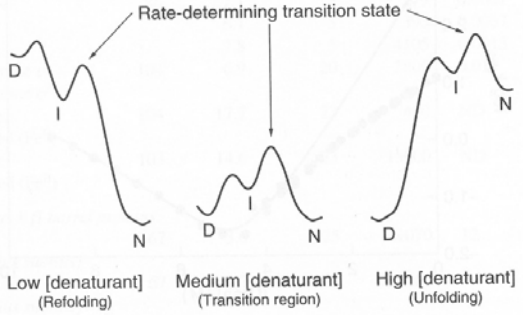
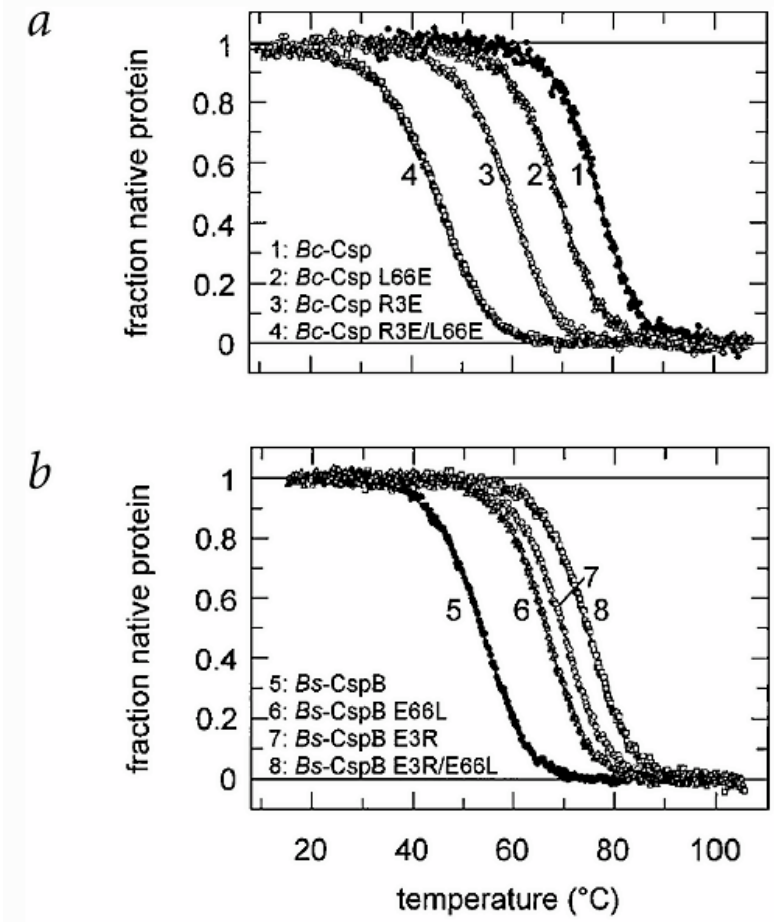


Figure 18.5 Free energy profiles for a folding reaction in which there is an intermediate that is at higher energy than D or N in the transition region but is more stable than D at low [denaturant]. The step D → I is often fast, and the kinetics can be deceptively simple.

$$k = k_B T \frac{\kappa}{h} \exp \frac{\Delta S^{\ddagger}}{R} \exp \frac{-\Delta H^{\ddagger}}{RT}$$

Transition midpoint temperatures, T_m, define thermal unfolding curves but are not useful estimates of conformational stability (why?). However, the T_m remains useful for comparing closely related proteins (*e.g.* homologs, single site mutants)



Thermal stability curves.
a, Unfolding transitions of wild type Bc-Csp and three destabilized variants.
b, Unfolding transitions of wild type Bs-CspB and three stabilized variants in 100 mM Na-cacodylate/HCl, pH 7.0, at protein concentrations of 4 muM.

The fractions of native protein obtained after a two-state analysis of the data are shown as a function of temperature. The continuous lines show the results of the analysis.

Perl D et al.(2000) Nat. Struct. Biol. 7, 380-3

Protein folding research

- ***Static***: 3-D prediction (theoretical approach)

- ***Dynamic***: protein folding (misfolding and refolding) mechanism
  - Experimental approach: *in vivo* and *in vitro* (most studies)
  - Theoretical approach (in vitro): molecular dynamics, *Monte Carlo* simulation

Protein misfolding and refolding

Most proteins except membrane proteins are soluble in *in-vivo* and *in-vitro* aqueous systems

In vivo and in vitro experiment

Inclusion body: protein aggregate

Medical and industrial implications: loss of biological functions

Including misfolding, aggregation, unexpected multimerization

Misfolded proteins can be refolded to regain their biological function

Protein disulfide Isomerase

Proteins *in vivo* fold faster than *in vitro*…

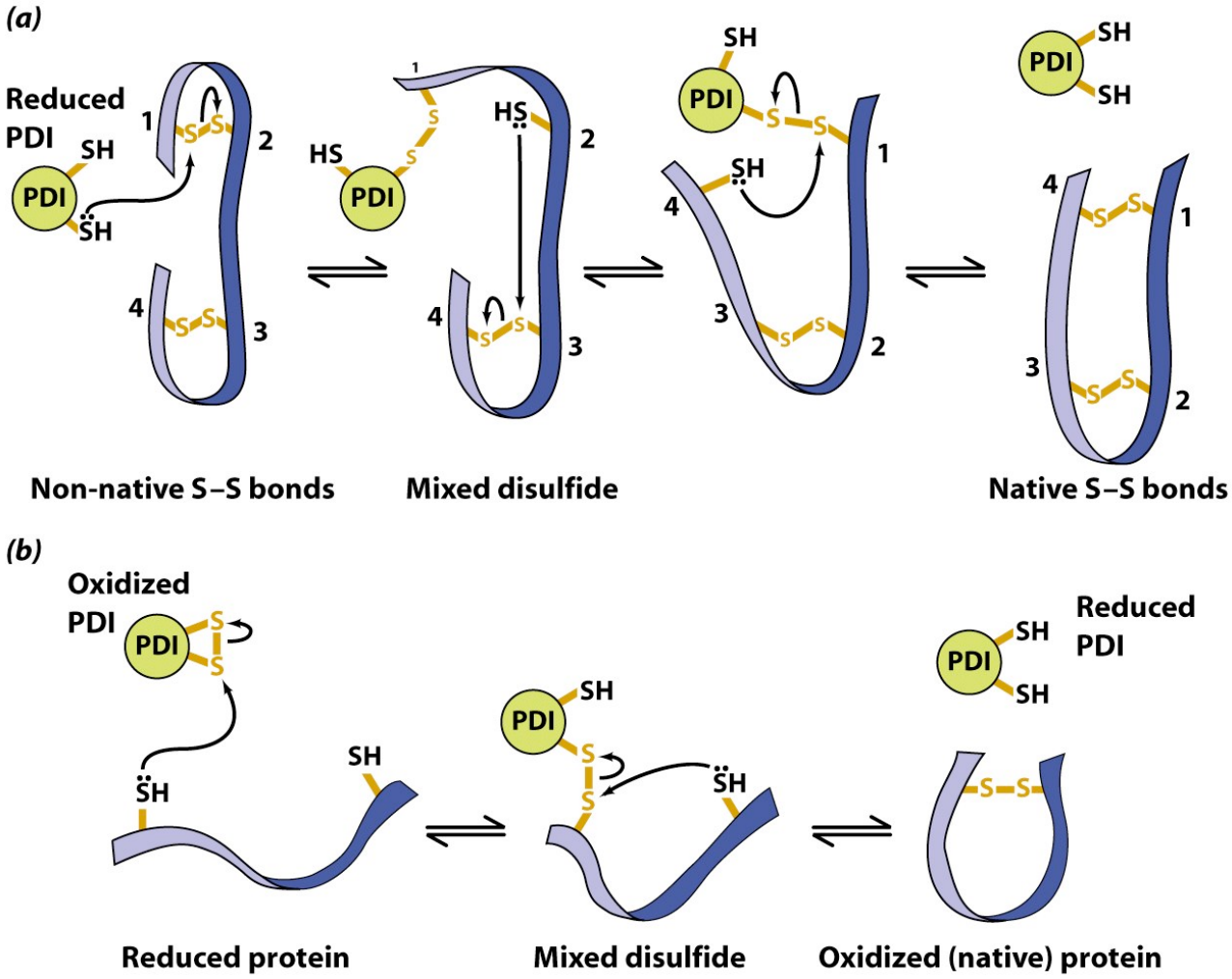
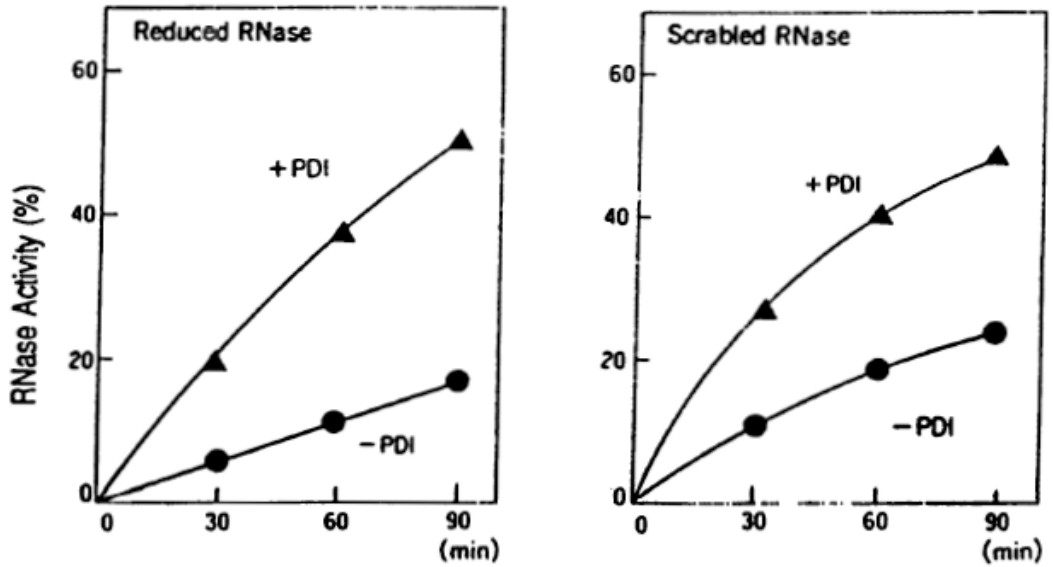


Figure 6-42 Fundamentals of Biochemistry, 2/e

Refolding System

	final conc.
Substrate protein	200 μg/ml
Sodium phosphate buffer, pH 7.5	100 mM
Glutathione	0.2 mM
Reduced glutathione	2 mM
PDI	20 μg/ml

Results



Molecular chaperons

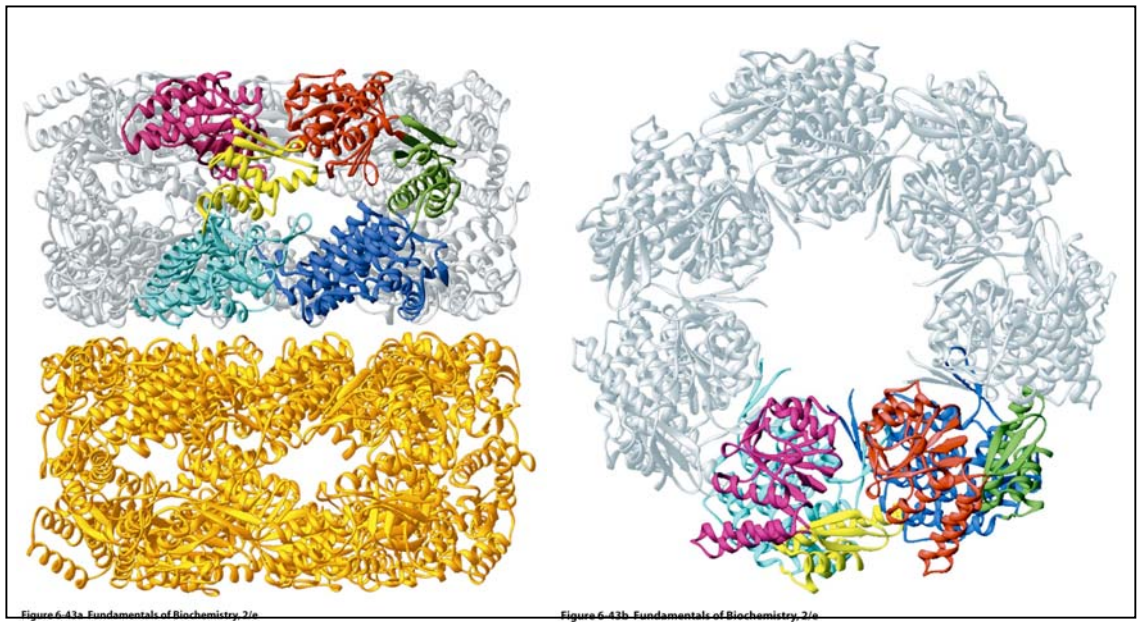
Molecular chaperones:

(1) Hsp70 proteins function as monomer

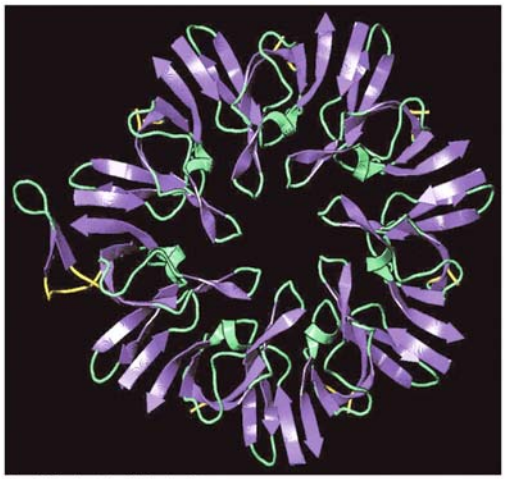
(2) Chaperonins, large multisubunit proteins (Hsp60 & Hsp10)

(3) Hsp90 proteins for the folding of proteins involved with signal transduction

(4) Trigger factor



GroEL



GroES

ATP binding and Hydrolysis coordinate the conformational changes in GroEL/ES

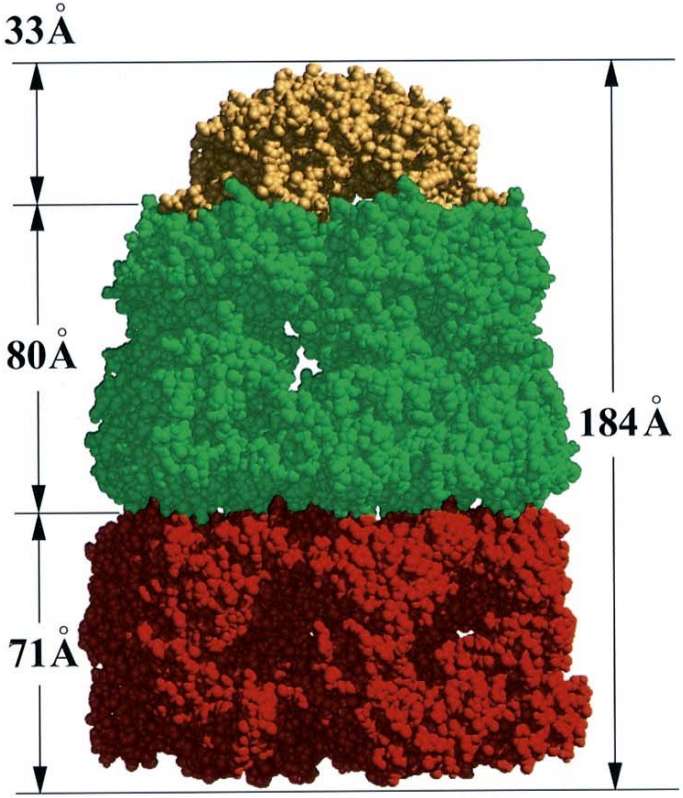
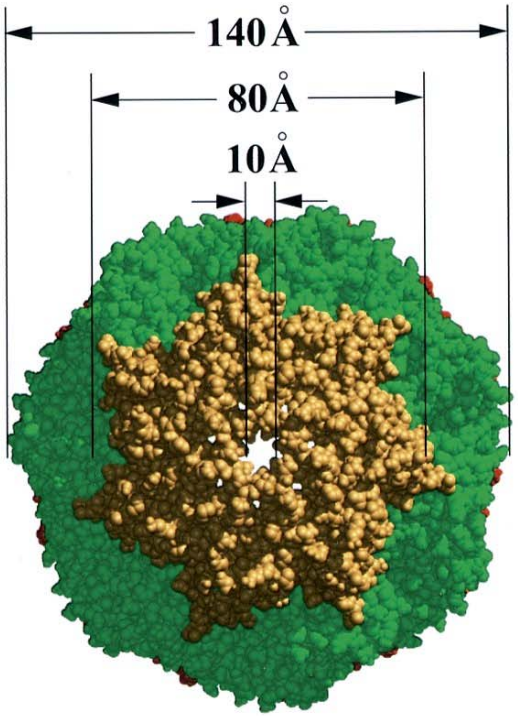


Figure 6-45ab Fundamentals of Biochemistry, 2/e



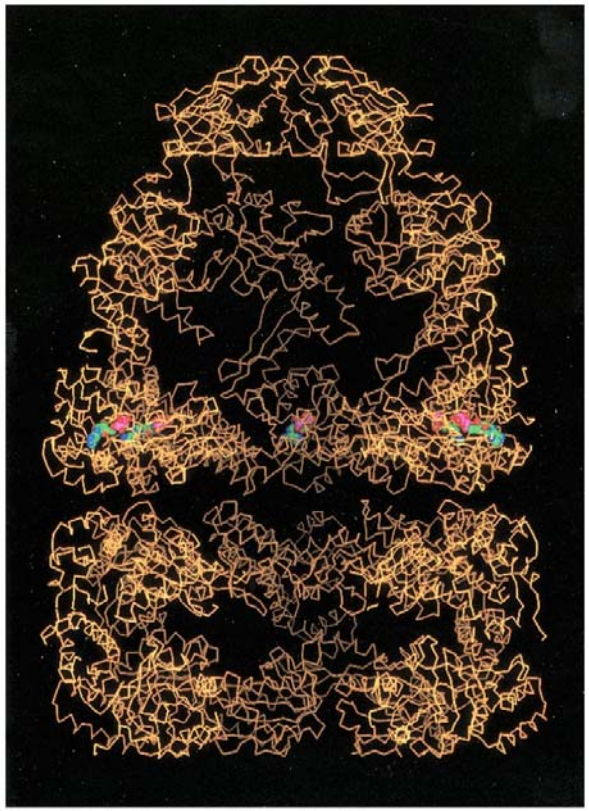
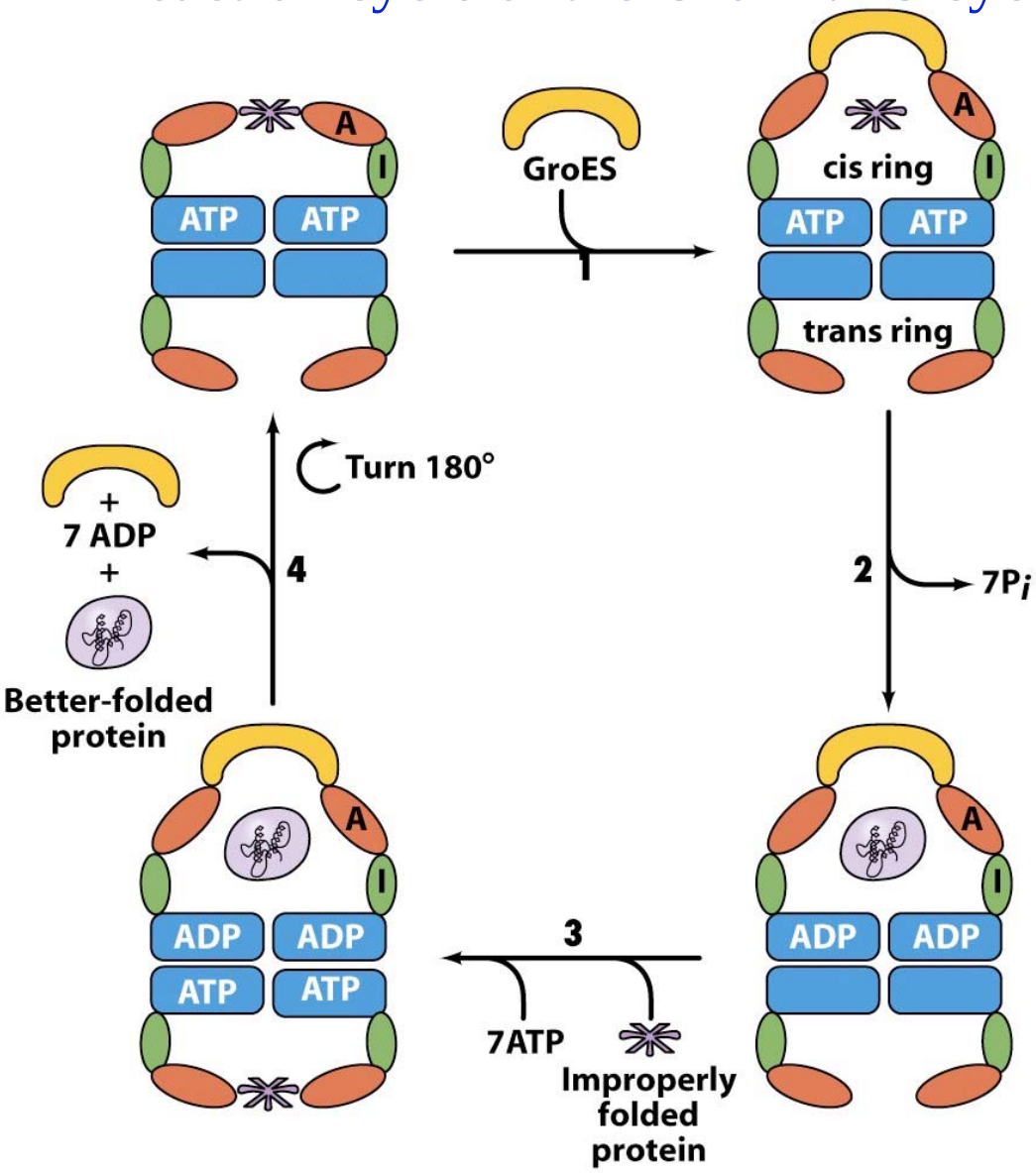


Figure 6-45c Fundamentals of Biochemistry, 2/e

Reaction cycle of the GroEL/ES cycle



1. GroEL ring binding 7 ATP and a substrate (improperly folded protein). Then it binds a GroES cap to become the ***cis*** ring.

2. The ***cis*** ring catalyzes the hydrolysis of its 7 ATP.

3. A 2nd substrate binds to the ***trans*** ring followed by 7 ATP.

4. The binding of substrate and ATP to the ***trans*** ring conformationally induces the cis ring to release its bound GroES, 7 ADP, and the better folded substrate.The trans ring becomes the cis ring.

Figure 6-46 Fundamentals of Biochemistry, 2/e

in-vitro protein folding

Application

- Biopharmaceutical is a typical example, (e.g. insulin)
- Prevent protein misfolding (e.g. amino acid effect on protein folding)
- Refold misfolded protein to regain its biological function

How long does protein folding take?

- μs, ms, s, min…Depending on protein size, temp, etc.

Monitor protein folding

- Spectrofluorometer, Circular dichroism, Stop flow system, PAGE (SDS gel and native gel), etc.