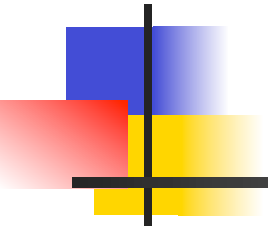
Protein Folding



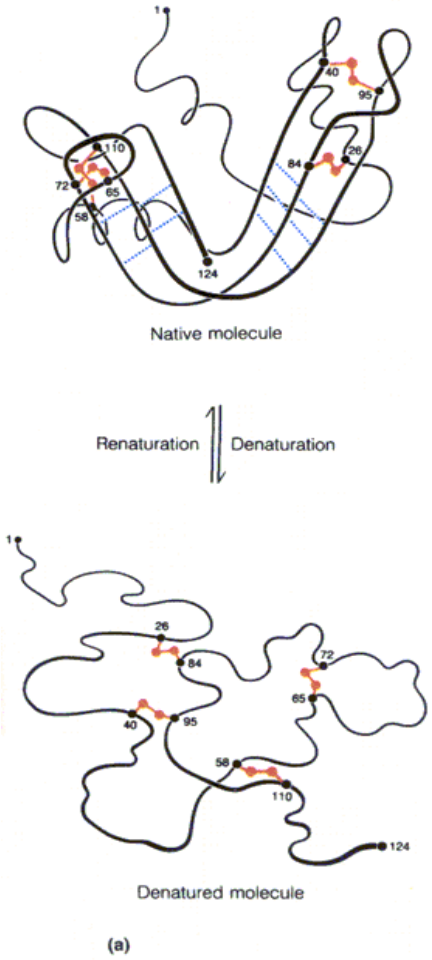
Protein folding

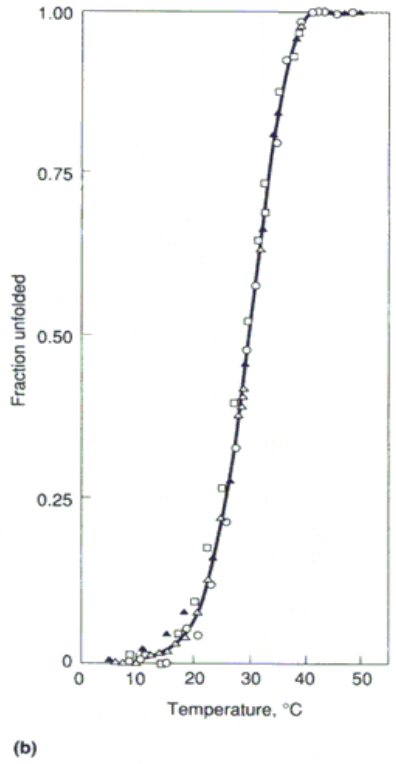
- The information on the 3D structure of a protein is carried in its aa sequence
- **Native** proteins can be **denatured** by changes in:
  - Temperature
  - Extreme pH
  - Solvents
  - Denaturating agents such as alcohols, urea, guanidinium chloride
- The denatured protein has lost the functions related to its 3D fold and the chain has a **random coil** conformation: freedom of rotation around all peptide and side chain bonds
- The transition unfolded - folded is usually reversible:

$$\mathbf{F \Leftrightarrow U}$$

Ribonuclease folding

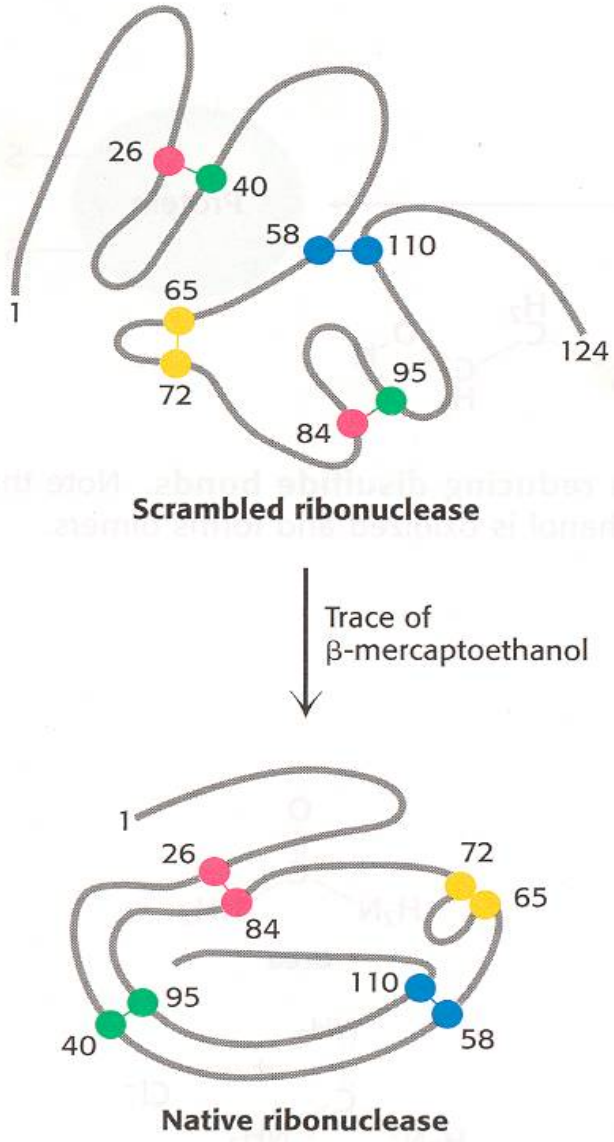
- Classic expt. with this enzyme that hydrolyses RNA
- Very important: the process is reversible - the enzyme “knows” how to re-fold





Restoration Of Activity

- Oxidation allowed
- RNAase activity regained
- Original, native confor-mation formed, with all correct S-S bonds reestablished.
- Probability of correct S-S bonds by chance, 1%.



Conclusions

- Native conformation of a protein is the state with the lowest Gibbs free energy.
- Proteins follow unique paths to attain a native state.
- Primary structure possesses sufficient information for proper folding.

Thermodynamics of folding

- As proteins spontaneously fold, must be $\Delta G < 0$
- This $\Delta G < 0$ is achieved by a number of factors that contribute to the classic equation:

$$\Delta G = \Delta H - T\Delta S$$

Where: ΔG = free energy, energy available for work

ΔH = enthalpy, a measure of bond formation (-)
or a measure of bond breaking (+)

T = temperature, oK

ΔS = entropy, a measure of increasing chaos (+)
or a measure of decreasing chaos (-)

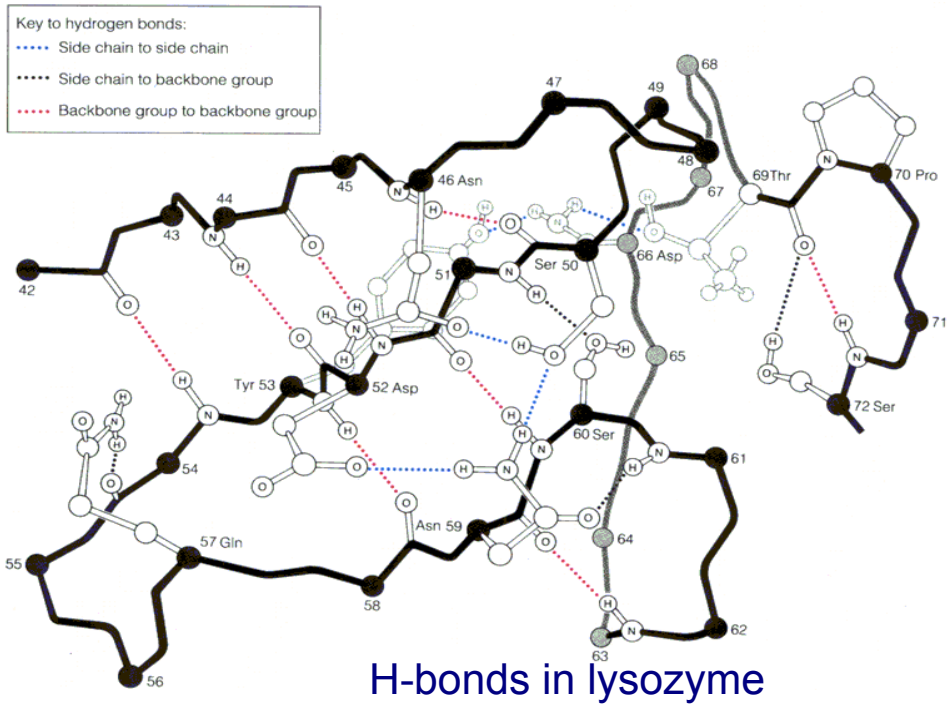
- The folding process leads to a <u>more ordered structure</u> of the protein, this is called <u>conformational entropy</u> and it works <u>against</u> folding. It follows that, in compensation:
  - 1. There must be a large negative ΔH through interactions between groups in the folded structure
  - 2. There must be other contributions to ΔS that leads to an increase in disorder

Interactions that favour folding:

- Charge-charge interactions
  - Salt bridges, for example Lys-Glu
  - Changes in pH can cause loss of charges, loss of salt bridges, hence denaturation
- Internal H-bonds
  - See aa Ser, Thr, Asn, Gln, His, CO and NH of peptide bonds
  - They can be:
    - Side chain - side chain
    - Side chain - backbone group
    - Backbone group - backbone group

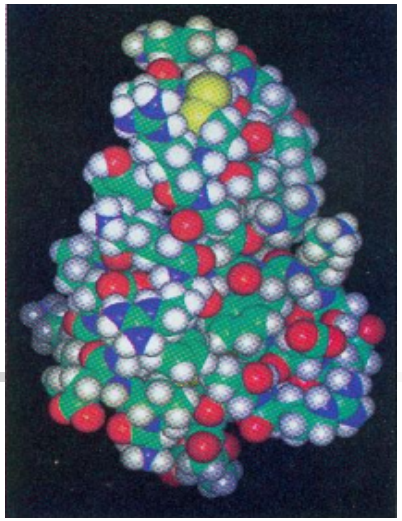
Group Type	Typical pK_a Range[a]
α-Carboxyl	3.5–4.0
Side chain carboxyls of aspartic and glutamic acids	4.0–4.8
Imidazole (histidine)	6.5–7.4
Cysteine (—SH)	8.5–9.0
Phenolic (tyrosine)	9.5–10.5
α-Amino	8.0–9.0
Side chain amino (lysine)	9.8–10.4
Guanidinyl (arginine)	~12

[a]Values outside these ranges are observed. For example, side chain carboxyls have been reported with pK_a values as high as 7.3.



H-bonds in lysozyme

cont.…



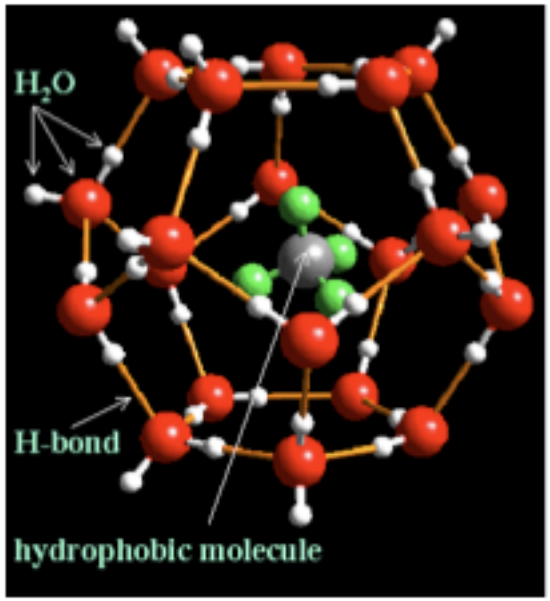
- van der Waals interactions:
  - Proteins are extremely well packed for max contact between side chains
  - These are very small contributions to stability BUT they are many and their sum is large
- Hydrophobic effect:
  - Hydrophobic side chains seek interactions and shielding from water in the protein interior
  - This causes an increase in ΔS due to the disordering of water clathrate from around the hydrophobic groups in the unfolded protein

Amino Acid	Scale of Engelman, Steitz, and Goldman[a]	Scale of Kyte and Doolittle[b]
Phe	3.7	2.8
Met	3.4	1.9
Ile	3.1	4.5
Leu	2.8	3.8
Val	2.6	4.2
Cys	2.0	2.5
Trp	1.9	−0.9
Ala	1.6	1.8
Thr	1.2	−0.7
Gly	1.0	−0.4
Ser	0.6	−0.8
Pro	−0.2	−1.6
Tyr	−0.7	−1.3
His	−3.0	−3.2
Gln	−4.1	−3.5
Asn	−4.8	−3.5
Glu	−8.2	−3.5
Lys	−8.8	−3.9
Asp	−9.2	−3.5
Arg	−12.3	−4.5

[a]Data from Engelman, D. M., T. A. Steitz, and A. Goldman, *Annu. Rev. Biophys. & Biophys. Chem.* (1986) 15:321–353.

[b]Data from Kyte, J., and R. F. Doolittle, *J. Mol. Biol.* (1982) 157:105–132.

Hydrophobic effect



Water molecules form a cage around apolar molecules called “clathrates”

- Water molecules surroundings apolar molecules form a lattice composed by H-bonded molecules with limited motions and orientations.

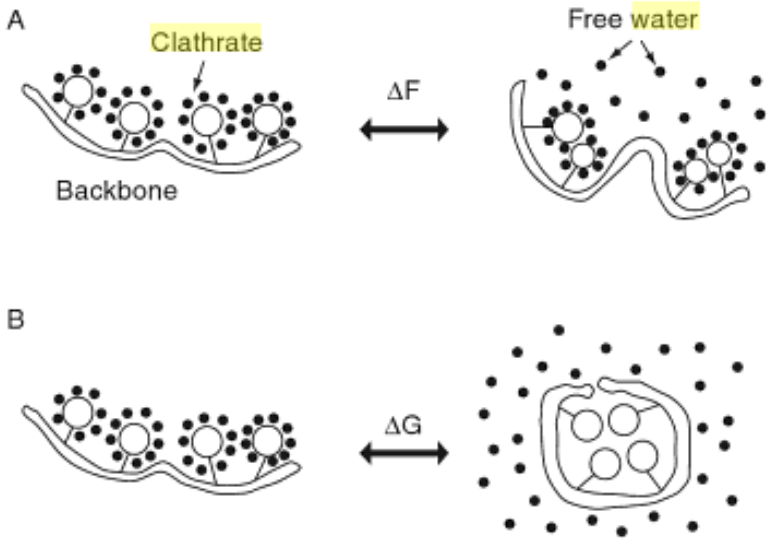
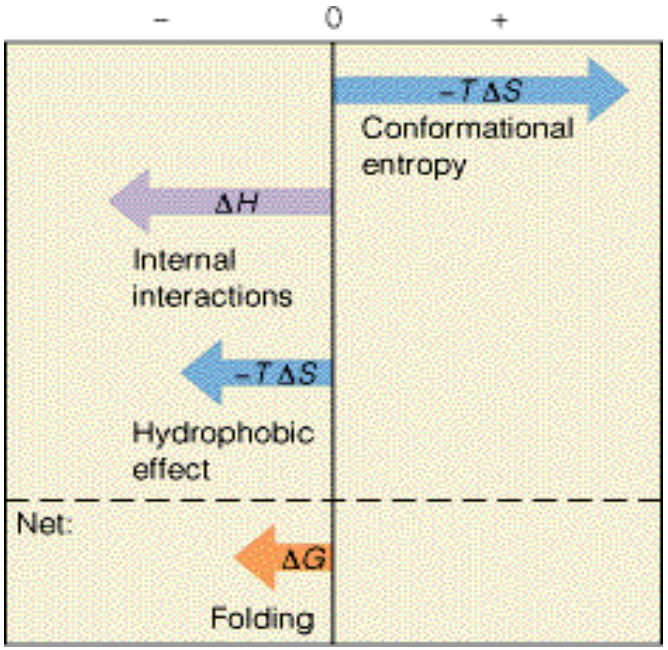


FIG. 10. Schematic diagrams illustrating the hydrophobic interaction. (A) Hydrophobic energy of formation, ΔF, when nonpolar sidechains interact in pairs and release some water molecules into the bulk solvent. (B) Hydrophobic energy of transfer, ΔG, when several sidechains form a core and most clathrate bound water molecules are liberated. Reused with permission from the *Journal of Chemical Physics*, 123, 054901 (2005). Copyright 2005, American Institute of Physics.

Contributions to ΔG

$$\Delta G = \Delta H - T\Delta S$$

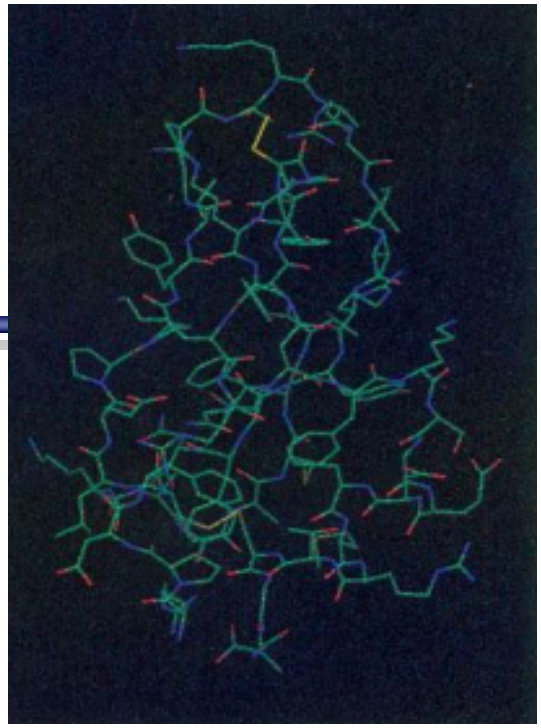
- From what we said, it follows that ΔH and ΔS can compensate to give an overall ΔG < 0



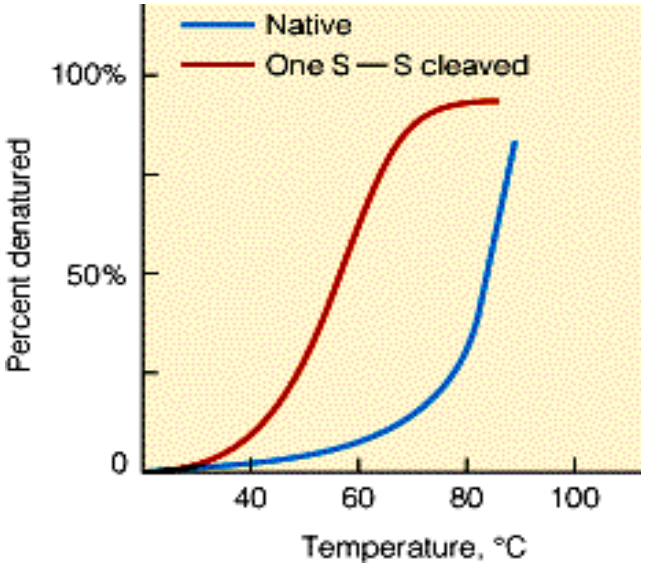
Protein	ΔG (kJ/mol)	ΔH (kJ/mol)	ΔS (J/K·mol)
Ribonuclease	−46	−280	−790
Chymotrypsin	−55	−270	−720
Lysozyme	−62	−220	−530
Cytochrome *c*	−44	−52	−27
Myoglobin	−50	0	+170

Note: Data adapted from P. L. Privalov and N. N. Khechinashvili, *J. Mol. Biol.* (1974) 86:665–684. Each data set has been taken at the pH value where the protein is maximally stable; all are near physiological pH. Data are for the folding reaction: Denatured ⇌ native.

Disulphide bonds



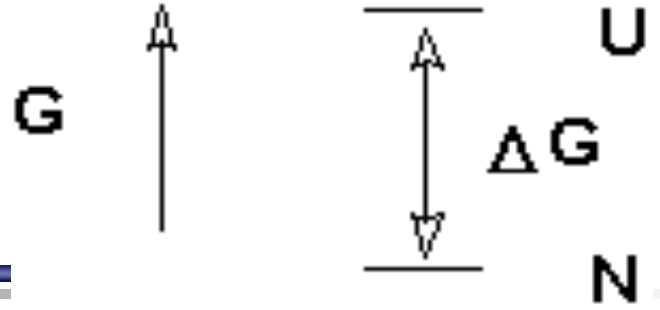
- Once folded, S-S bonds can futher stabilise
- There are 3 in BPTI (bovin pancreatic trypsin inhibitor)
- Elimination of 1 S-S (cys14-cys38) in BPTI cause a decrease of 21ºC in denaturation temperature (T_m from 80ºC to 59ºC)
- S-S bridges are usually found in proteins exported from the cells, for two reasons:
  - 1. The need in increase in stability
  - 2. The extracellular environment is oxidixing (SH + SH ➜ S-S + $2H^+$), while the intracellular environment is reducing (S-S + $2H^+$ ➜ SH + SH)



Protein Stability

- Protein stability is important for many reasons:
  - Providing an understanding of the basic thermodynamics of the process of folding,
  - increased protein stability may be a multi-billion dollar value in food and drug processing, and in biotechnology and protein drugs.
- Two relatively recent innovations, which have had major impact in the study of the thermodynamics of proteins were the development of very sensitive techniques, differential scanning calorimetry (especially by Privalov and Brandts) and site-directed mutagenesis.

Protein Stability



- The net stability of a protein is defined as the difference in free energy between the native and denatured state:
- Both G_N and G_U contribute to G
- The free energy may be readily calculated from the following relationships:

$$K = [N]/[U] = F_N/(1- F_N),$$

F_N = fraction folded, F_U=1-F_N

$$\Delta G = G_N - G_U = -RTlnK$$

- Decreasing the energy of the folded state or increasing the energy of the unfolded state have the same effect on ΔG.

Techniques for Measuring Stability

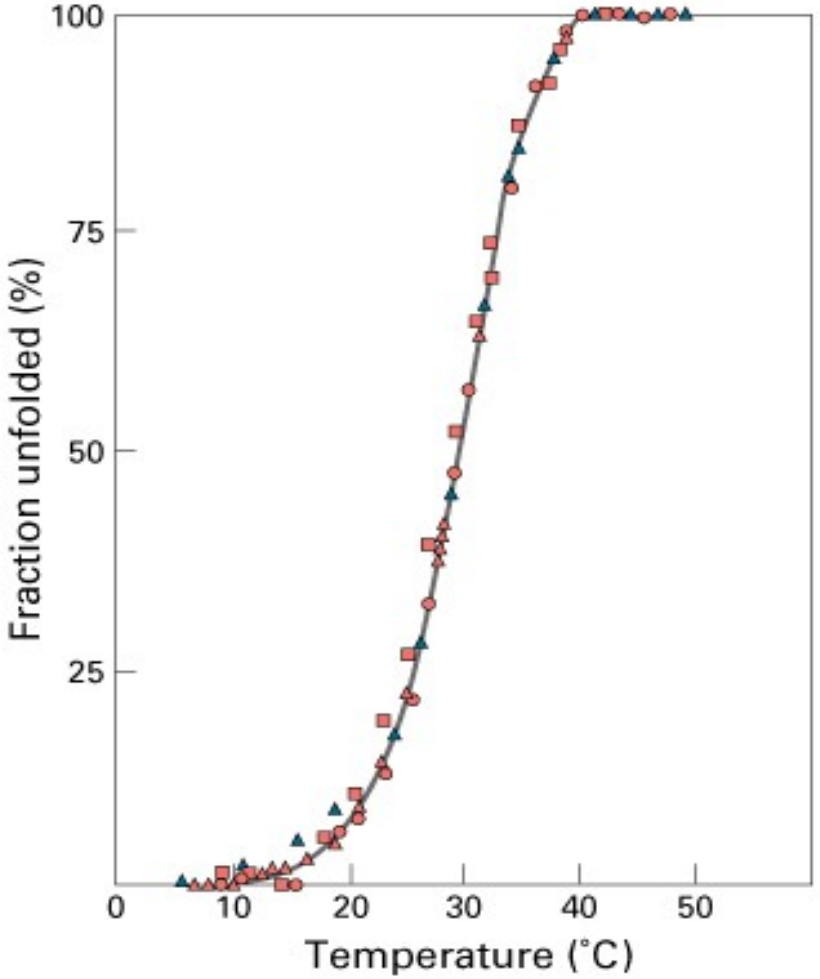
Any methods that can distinguish between U and F

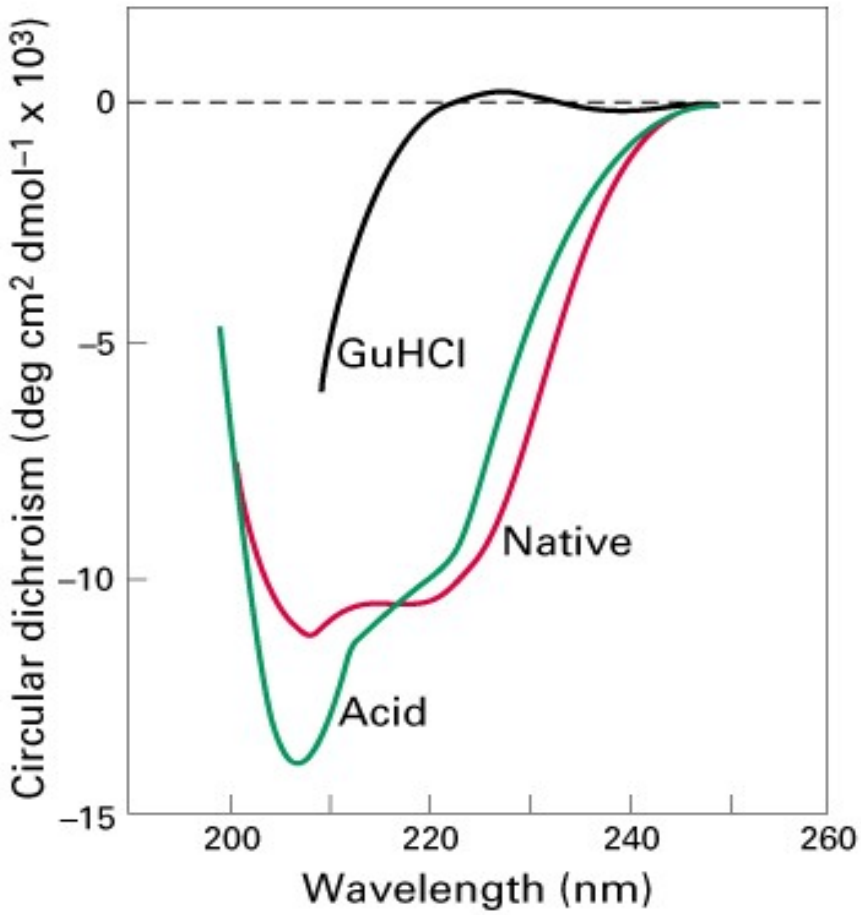
- Absorbance (e.g. Trp, Tyr)
- Fluorescence (Trp)-difference in emission max & intensity.
- CD (far or near UV) - (2^o or 3^o)
- NMR
- DSC (calorimetry)
- Urea gradient gels - difference in the migrating rates between F and U.
- Catalytic activity
- Chromophoric or fluorophoric probes

Two-state Unfolding of Protein

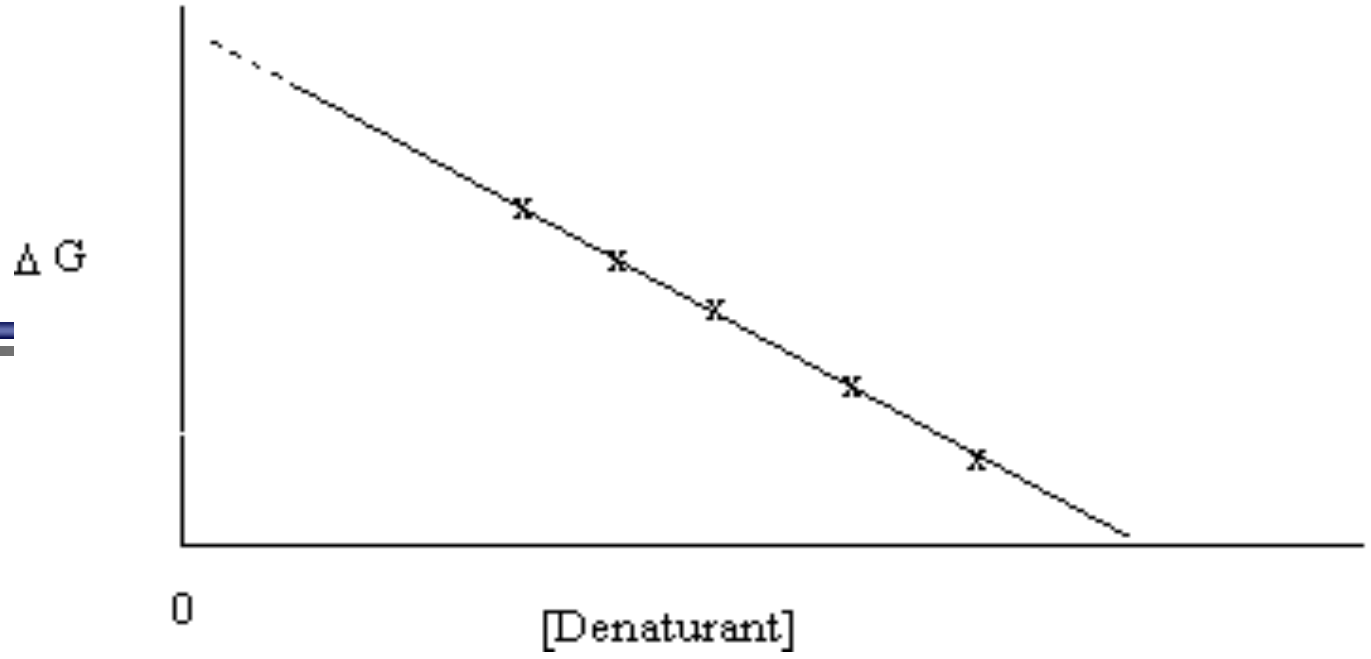
- $K_{eq}=[N]/[U]= ([\theta]_{obs}- [\theta]_D)/([\theta]_N- [\theta]_D) = F_N/(1- F_N)$

  F_N = fraction folded





Denaturants



- It is common to extrapolate the data for the unfolding transition as a function of denaturant to 0 M to give the value in water (e.g. G(H_2O)).

  $\Delta G_{D\text{-}N} = \Delta G^{H20}_{D\text{-}N} - m_{D\text{-}N}$ [denaturant]

  $\Delta G^{H20}_{D\text{-}N}$ is about –5 to –10 kcal/mol

- The extrapolation can have large errors.

Equilibrium Unfolding Measurements

Protein denaturation was obtained incubating the protein with different amounts of guanidinium hydrochloride (GdHCl) for 12 h at 4°C. Refolding of fully unfolded samples was achieved by diluting the denaturant concentration with buffer. The analysis of the fluorescence and circular dichroism unfolding transitions were performed according to a single pathway following the scheme:

$$N \underset{}{\overset{K}{\longleftrightarrow}} U$$

where N and U represent the native and unfolded protein fractions respectively. The fractions of total protein in the native (f_N) and ufolded state (f_U) were calculated ($f_N + f_U = 1$) for each denaturant concentration and the equilibrium constant K_U and the free energy change ΔG_{obs}^{U} were derived from the equations:

$$K_U = f_U / (1- f_U) = f_U / f_N \qquad (1)$$

$$\Delta G_{obs}^{U} = - RT \ln K_U \qquad (2)$$

According to linear extrapolation method the unfolding free energy is correlated to the denaturant concentration by the equation:

$$\Delta G_{obs}^{U} = \Delta G_{obs}^{H2O} – m\,[GuHCl] \qquad (3)$$

where m is the molar cosolvent term, ΔG_{obs}^{H2O} is the the free energy extrapolated to 0 M [GuHCl] and [GuHCl] is the denaturant concentration.

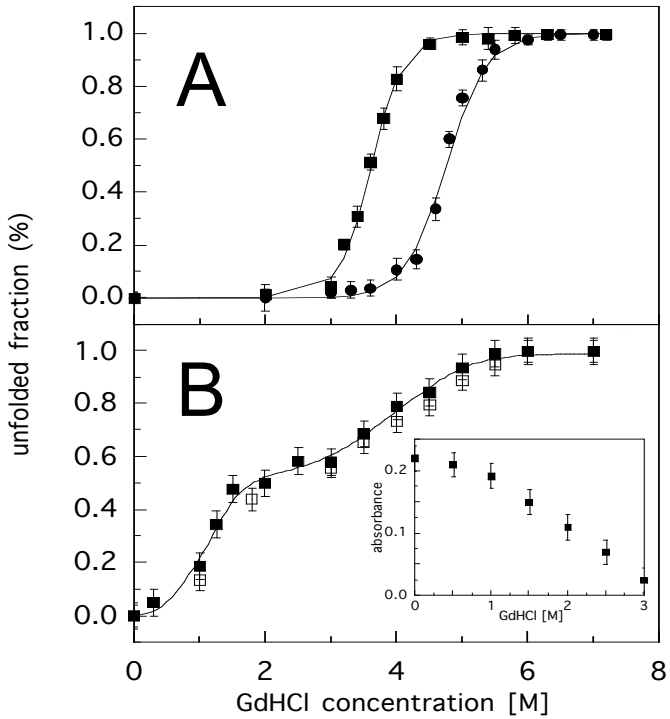
The analysis of the rop mutants fluorescence unfolding transitions required a double step denaturation pathway following the scheme:

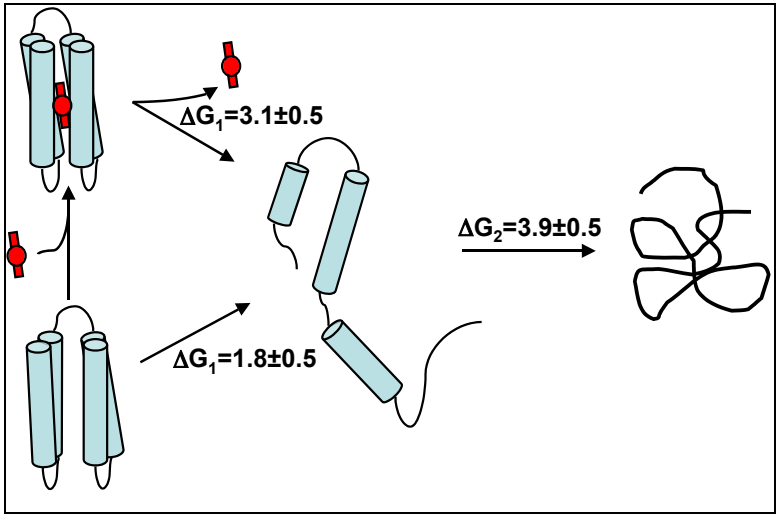
$$N \overset{K_1}{\longleftrightarrow} I \overset{K_2}{\longleftrightarrow} U$$

In this scheme, N, I and U represent the native, intermediate and unfolded protein species, respectively, while K_1 and K_2 are the two equilibrium constants.

A: two state unfolding

B: three state unfolding: an intermediate is detected. Inset: heme signal as function of denaturant concentration





Model for the unfolding pathway of rop mutants with and without heme consistent with circular dichroism and fluorescence data.

Principles of calorimetry

- In every chemical process, including enzymatic reactions and ligand-protein interactions, there is development or absorption of heat.
- The variation of enthalpy ΔH, that is the change in heat linked to intra- and inter-molecular interactions, is measured
- The thermodynamic information that can be achieved concern phenomena related to the structure/function of proteins, nucleic acids, lipids, the binding of drugs and other ligands.
- There are basically two types of experiments::
  - Isothermal Titration Calorimetry (ITC)
  - Differential Scanning Calorimetry (DSC)

Differential Scanning Calorimetry (DSC)

- This technique can measure the energy associated to the conformational changes of proteins and enzymes.
- The biomolecules do not need any treatment, the technique simply measures the excess of thermal capacity
- In practice the changes in heat that are reversibly generated upon controlled increase or decrease of temperature.
  - Two parameters are measured
    - 1. Tm or melting temperature, the temperature of the transition due to the phenomenon that is being observed DSC.
    - 2. Heat associated tothe transition

Experimental procedure for DSC

- A DSC instrument and consists of two cells, 1 mL. which are thermally isolated from the environment.
  - reference cell with buffer
  - Sample cell with buffer + protein / enzyme / substrate to be measured. Typical concentrations are 1 mg / mL
- Initially the two cells are balanced to a start temperature, eg. 5 ° C.

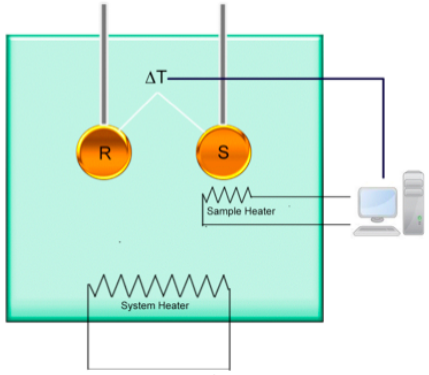


Figure 2. Diagram of a Differential Scanning Calorimeter (DSC) showing reference and sample cells with the ability to measure a temperature difference between the two cells. The ΔT is used to control a current into the sample cell heater to raise the temperature of the sample to match that of the reference. The system heater raises (i.e. scans) the temperature of the calorimeter cells at a constant rate, e.g. 1 deg/min.

- Then the temperature is slowly raised up with the control of the computer, generally of 1 degree per min. and any temperature differences between the reference and sample cells is measured.
- When an endothermic transition occurs, part of the heat given to the cell is used by the molecule to make the endothermic transition happen , rather than to raise the temperature of the sample cell.
- Therefore, a difference in temperature between the reference and sample cells is recorded. The instrument controller provides to an additional heating to the sample cell to balance the temperature of the two cells.

Experimental procedure for DSC

- The excess heat that must be given to the sample cell represents the excess of heat required by the endothermic transition.
- The result is a peak, and the extent of the peak, according to the Kirchhoff equation, is the enthalpy of the transition:

$$C_p = (\delta H / \delta T)_p$$

- In the case study of protein folding (refolding) of the protein, you can measure:

$$\Delta C_P^{den} = C_P^D(T) - C_P^N(T)$$

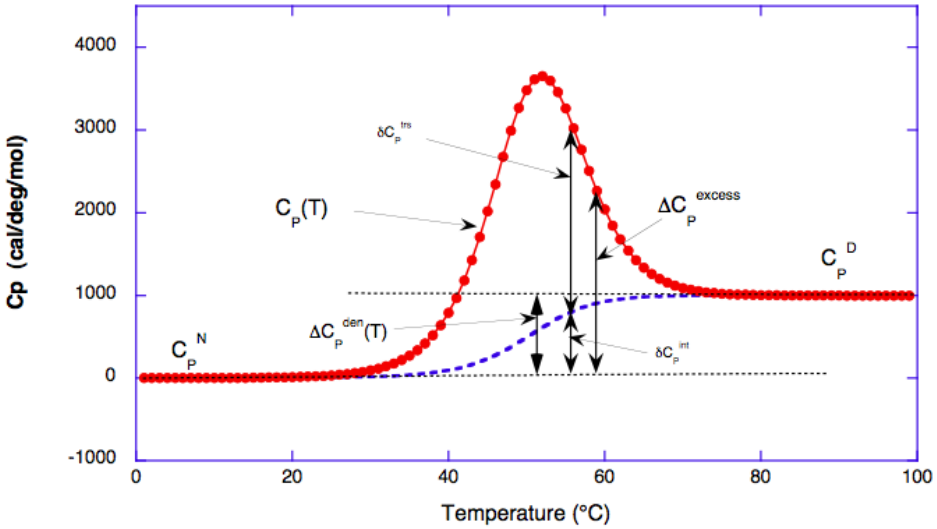


Figure 4. Simulation of the heat capacity observed in a DSC scan of a protein with a T_m of 50 °C, a ΔH of 50 kcal/mol, and a ΔC_P of 1000 cal/deg/mol. The C_P curve which is experimentally observable is shown in red. The grey lines show the extrapolation of the baselines into the transition region, and the blue dashed curve shows the sigmoidal transition in the intrinsic heat capacity (which cannot be observed experimentally). In this example, the heat capacity of the N species is set to 0. In general the heat capacity of the native species will differ significantly from zero and it will be temperature dependent.

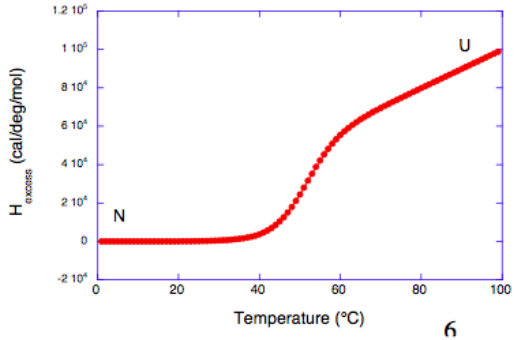
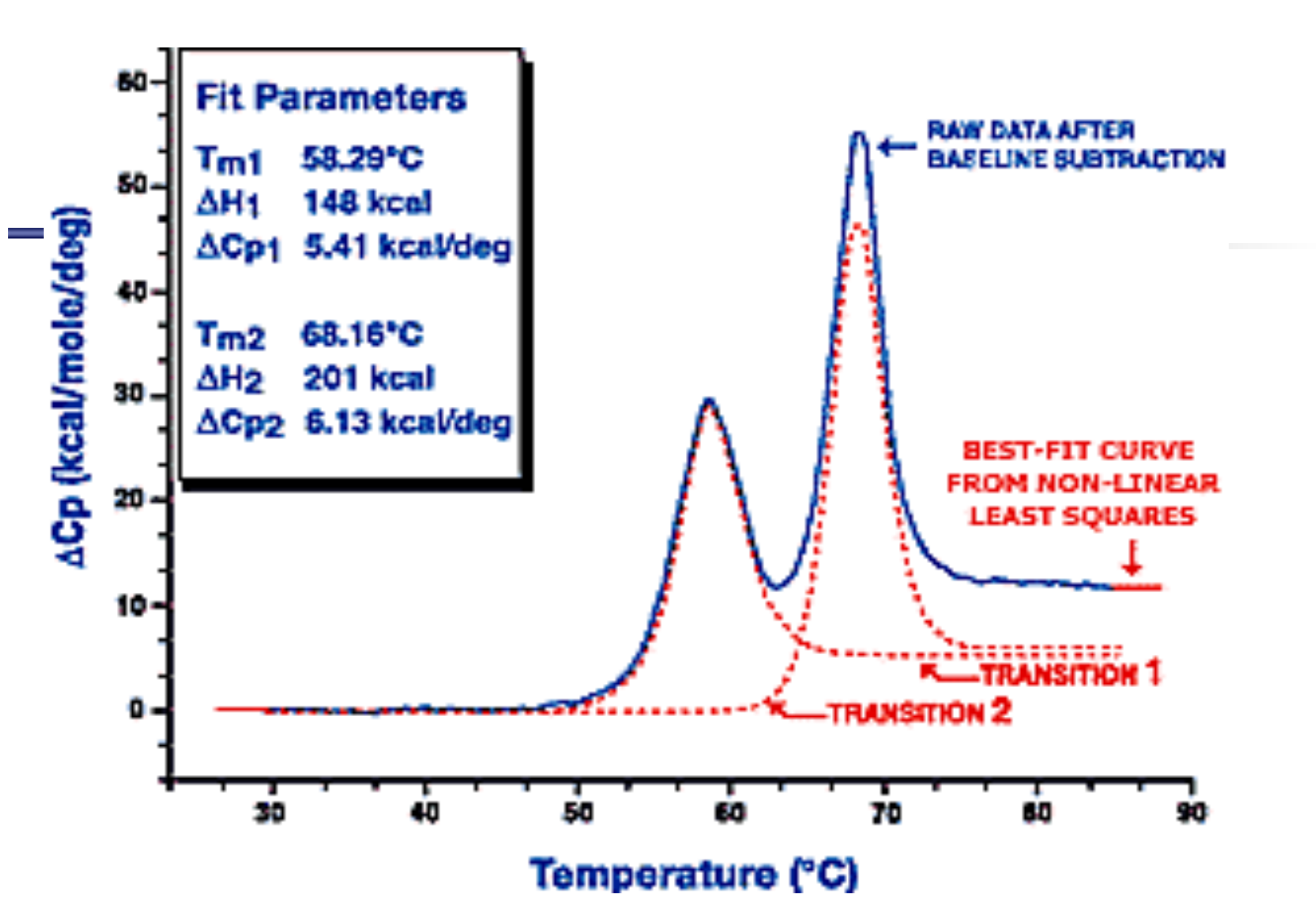
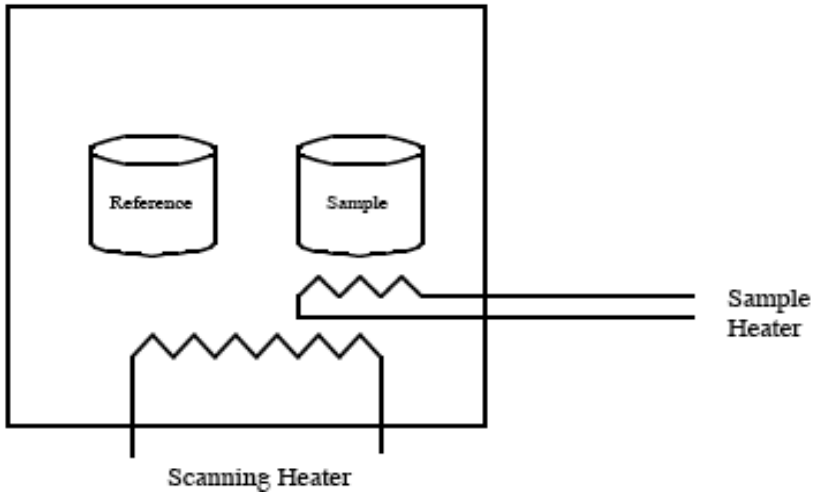


Figure 3. The excess enthalpy of a sample due to a transition with a Tm at 50 °C and a ΔH of 50 kcal/mol, and a ΔC_P of 1000 cal/deg/mol.


Fit Parameters
Tm1 58.29°C
ΔH1 148 kcal
ΔCp1 5.41 kcal/deg
Tm2 68.16°C
ΔH2 201 kcal
ΔCp2 6.13 kcal/deg
RAW DATA AFTER BASELINE SUBTRACTION
BEST-FIT CURVE FROM NON-LINEAR LEAST SQUARES
TRANSITION 1
TRANSITION 2
ΔCp (kcal/mole/deg)
Temperature (°C)
60
50
40
30
20
10
0
30
40
50
60
70
80
90

DSC instrument



- http://www.microcal.com

Thermal Unfolding of Barnase

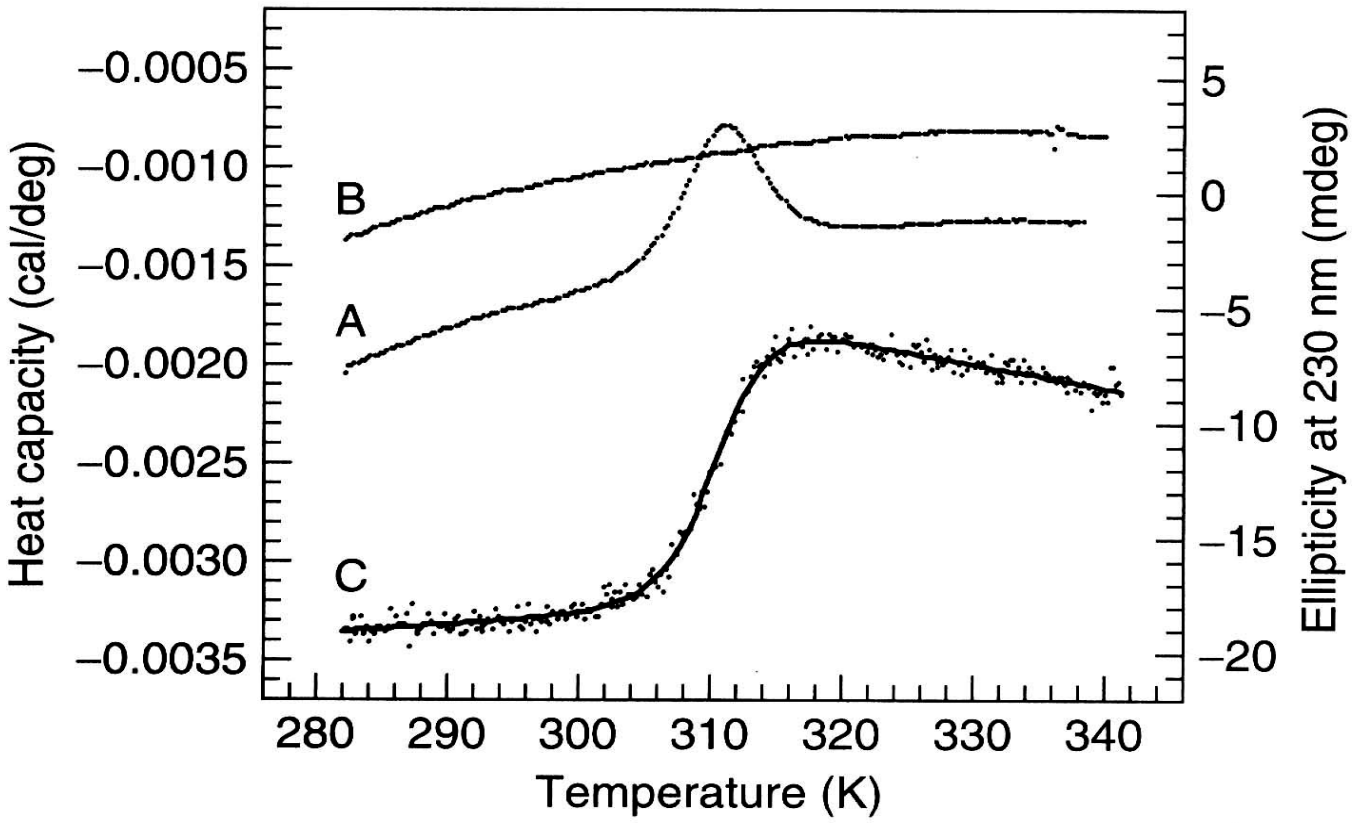
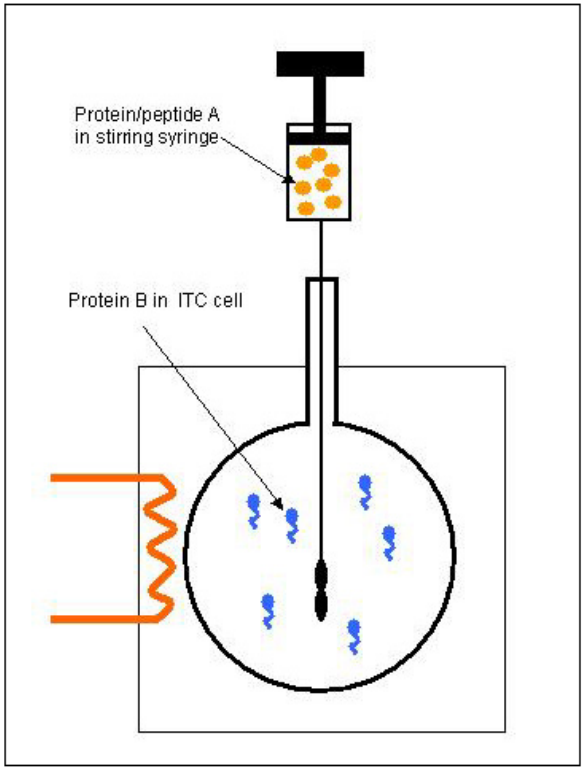


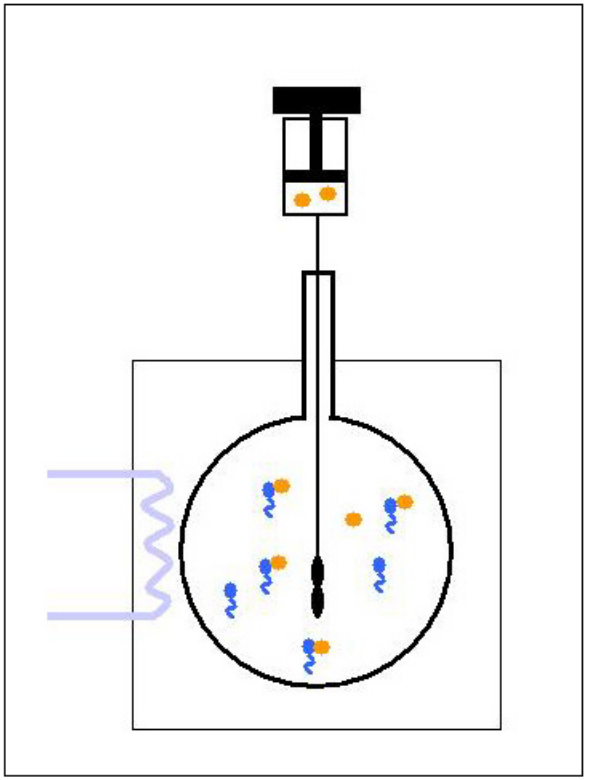
Figure 17.1 Thermal unfolding of barnase measured by calorimetry and spectroscopy. The heat capacity of barnase (trace A) was measured using differential scanning calorimetry with a baseline (trace B) of buffer versus buffer: T_m is 310.9 ± 0.01 K, $\Delta H_{D-N(cal)}$ = 98.4 ± 0.2 kcal/mol, and $\Delta H_{D-N(vh)}$ = 98.1 ± 0.3 kcal/mol. The ellipticity at 230 nm in the circular dichroism (trace C) under identical conditions fits T_m = 310.5 ± 0.1 K, and $\Delta H_{D-N(vh)}$ = 93 ± 3 kcal/mol (equation 17.5).

Isothermal Titration Calorimetry (ITC)

- The Isothermal Titration Calorimetry (ITC) is a technique particularly suitable for measuring the amount of heat developed or absorbed when a ligand or a substrate binds to a protein or an enzyme
- The measurement of this heat lead to the determination of the entire thermodynamic profile of the reaction with a single experiment:
  - Binding constant: K_B o K_{eq}
  - Stoichiometry of the reaction (how many molecules bind to how many sites): n
  - Entalpy of the reaction: ΔH
- The instrument consists of a syringe containing the ligand (or substrate or inhibitor) that is added in aliquots to a cell containing the protein
- Upon protein-ligand interaction, heat is released or absorbed by the cell, and it is measured by the instrument
- Upon different ligand additions, the protein becomes saturated and less heat is released or absorbed, and the signal decreases

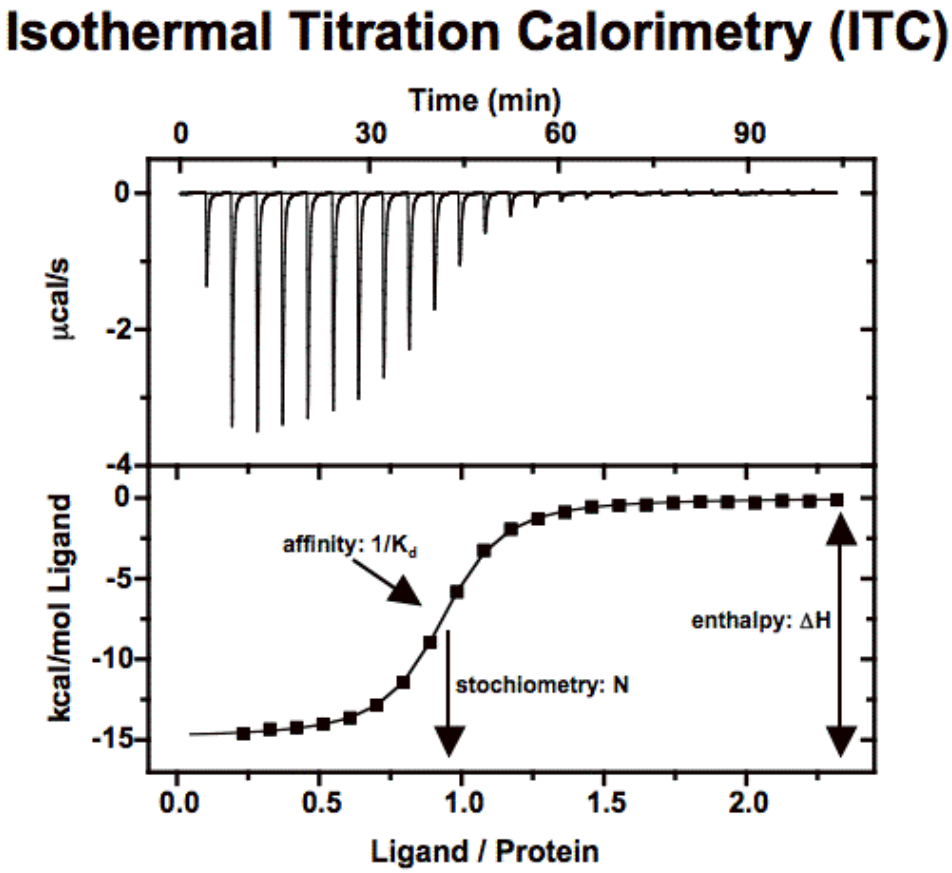
Experimental procedure for ITC



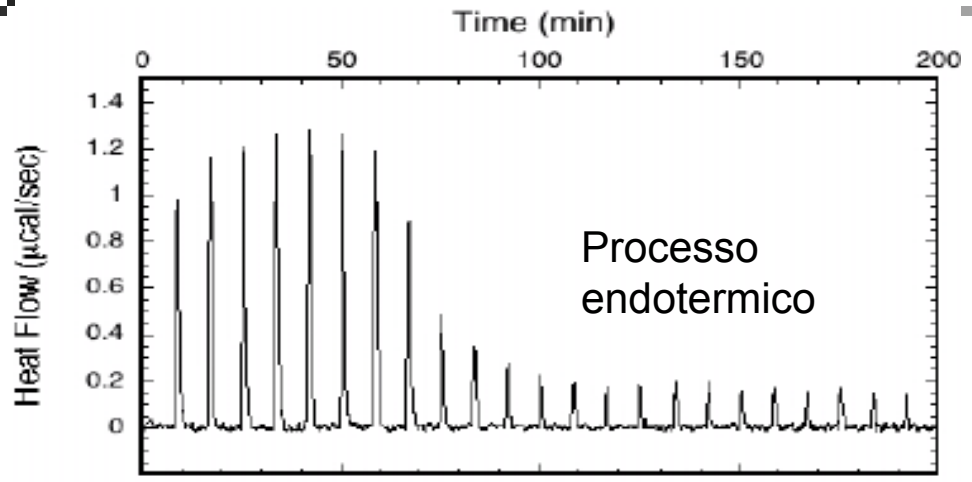


Experimental procedure for ITC

- Typical experiment with 18 injections of the ligand in the cell with the protein.
- The area of each peak corresponds to the total heat released and each injection (Fig. Above).
- The graph of the heat released for each ligand/protein molar ratio results in a binding isotherm (Fig lower)
- The instrument then provides a fitting that produces the parameters indicated in the graph.



ITC instrument



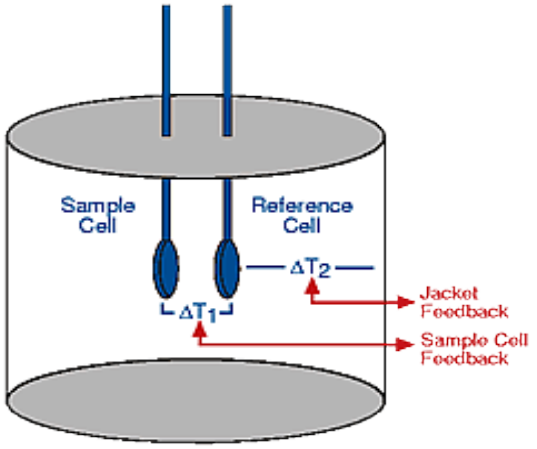


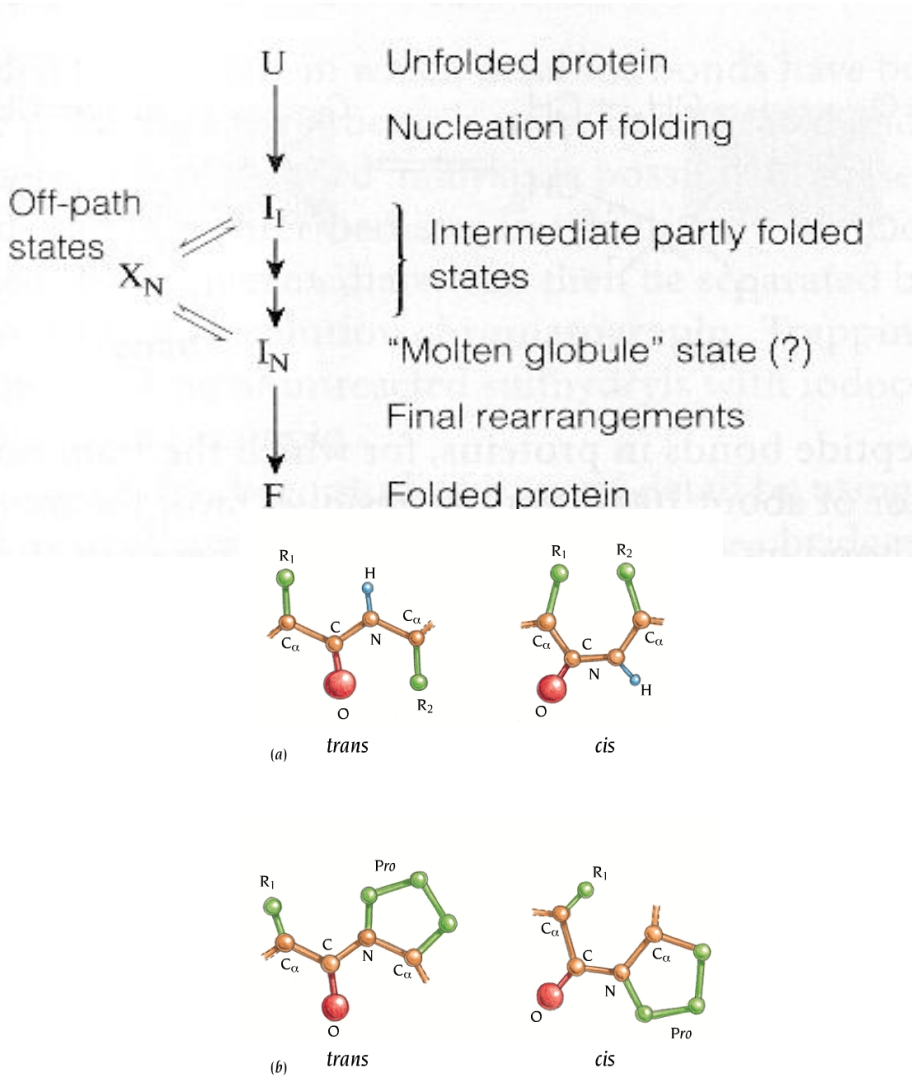
Figure 2.

Kinetics of folding

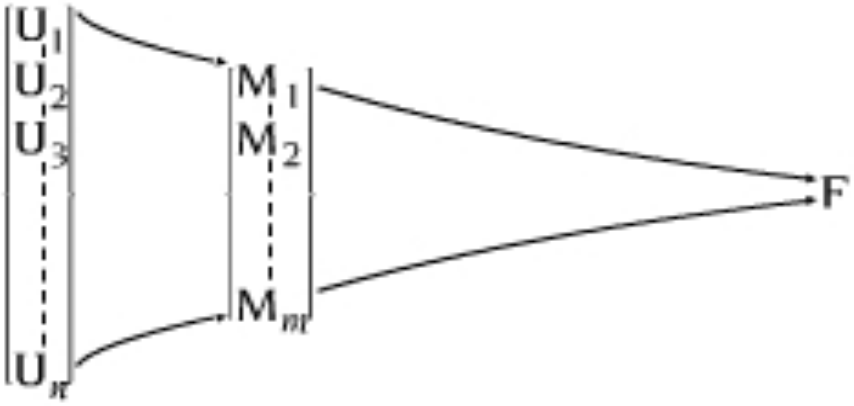
- Protein folding occurs in < 1 sec!
- 1968: Levinthal's paradox
  - Ribonuclease, 124 aa:
    - ➜ ~ 10^{50} conformations possible
    - ➜ trial of 1 conformation = 0.1 ps (picosec, 10^{-13} s)
    - ➜ it will take 10^{30} years to try them all (NB: 10^{6} = 1 million yeas, here is 10^{30} years)
    - ➜ impossible!
- It has been established that protein folding is rapid and it involves well-defined intermediates
  - Intermediate of folding = pathway model of folding

Pathway model for protein folding

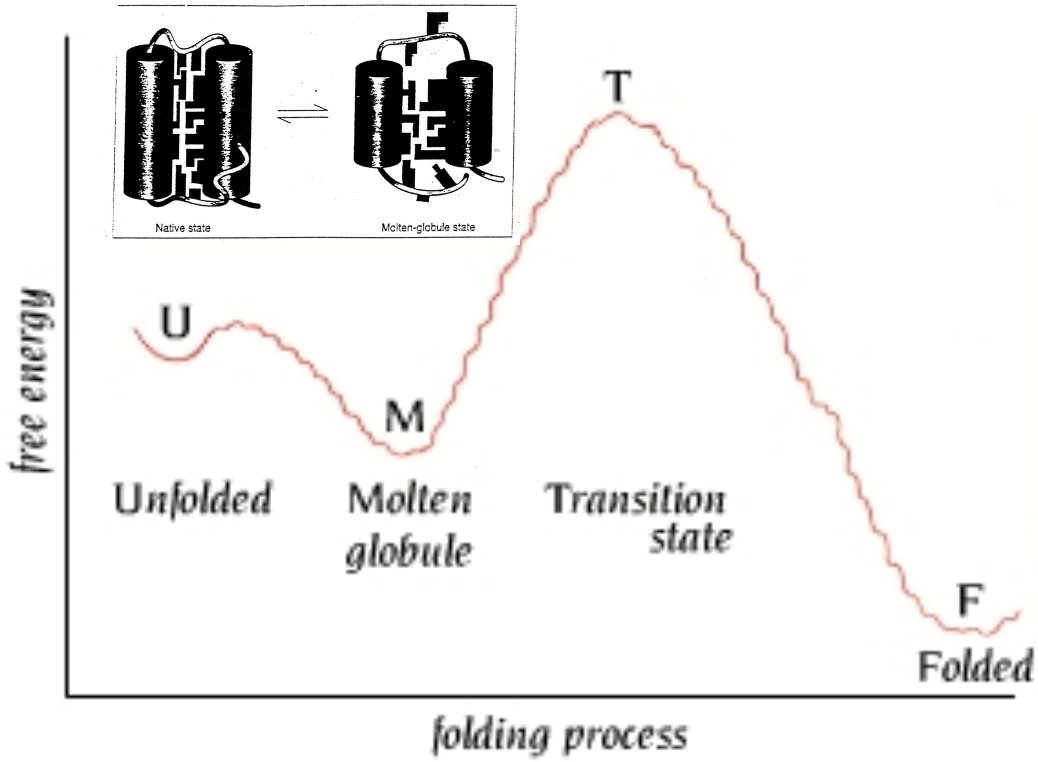
- Nucleation is the critical step:
  - It is more difficult to begin an α-helix than to extend it
  - Remember: you need 4 residues in a certain geometry to give the first stabilising H-bonds in an α-helix
  - Nucleation starts at more than one point, then all the partially folded structures get funnelled towards a final folded state with minimal energy
- Molten globule:
  - Important intermediate state, that is a compact structure where 2ary and 3ary structure is present, but internal hydrophobic residues have not yet been optimised
- 'Off path' states:
  - Traps that delay protein folding
  - An example is the common error than can occur in the *cis-trans* isomeration of the amide bond next to a Pro.
  - The trans form is preferred in most peptide bonds (roughly 1000:1 ratio in trans:cis populations). However, X-Pro peptide groups tend to have a roughly 3:1 ratio

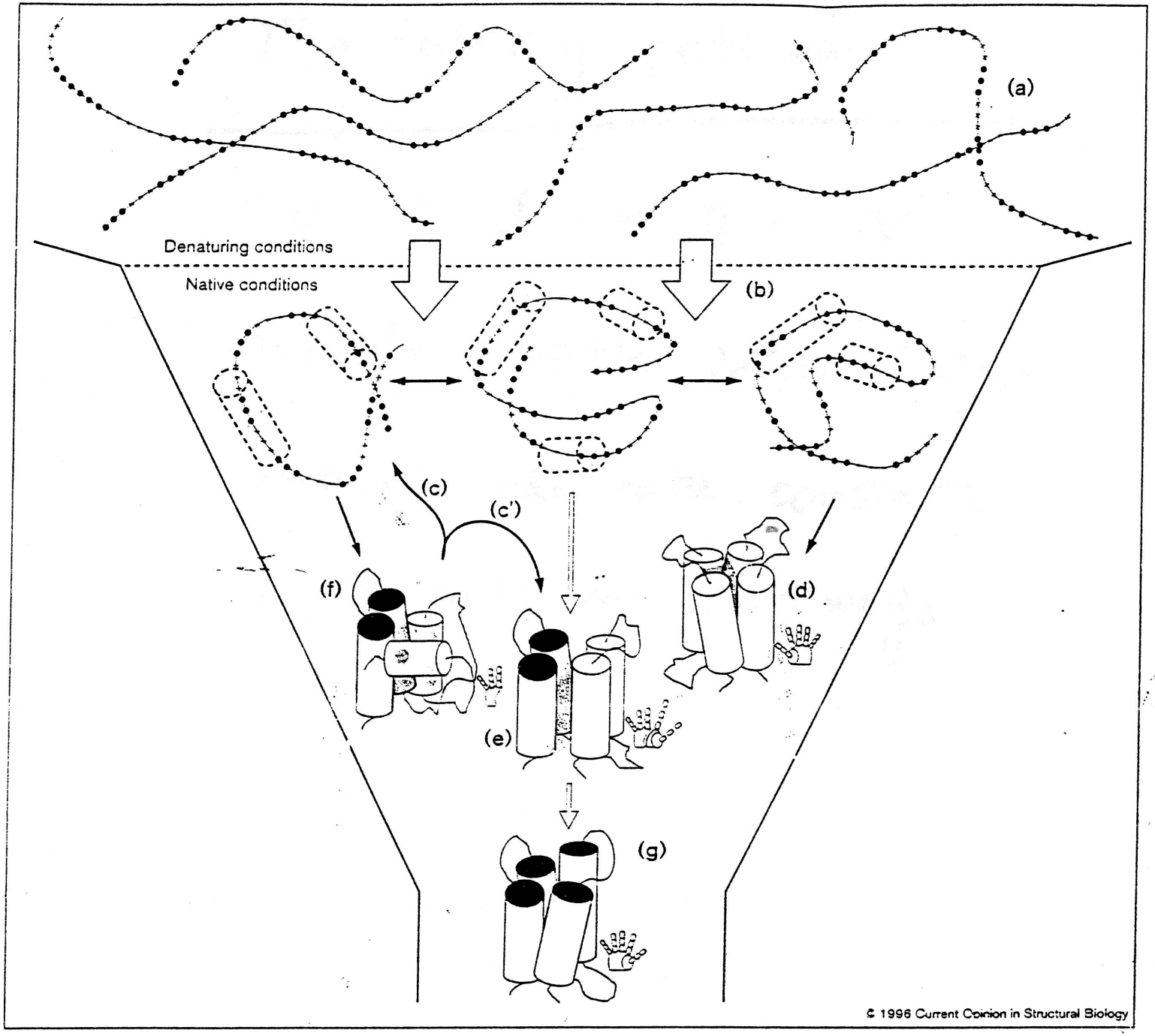


Unfolded State



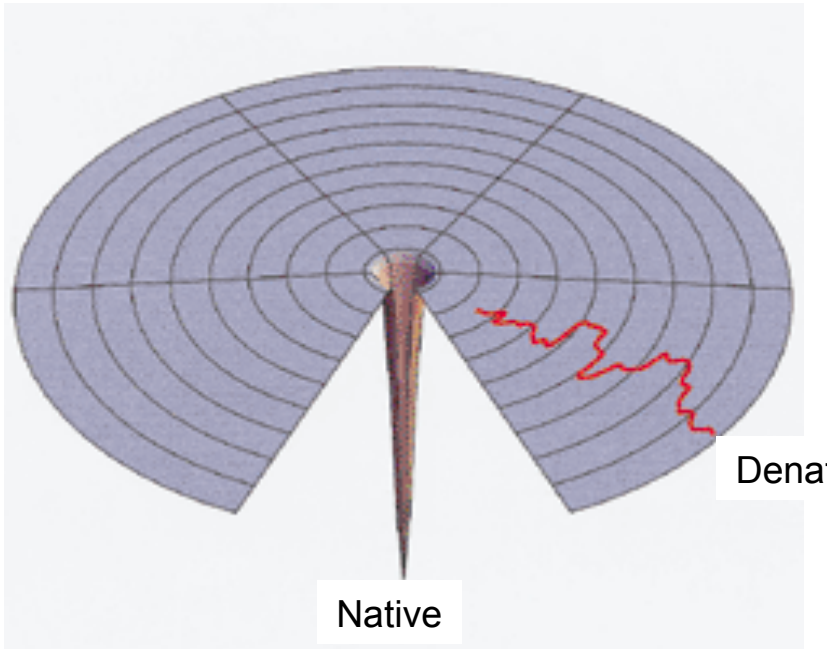
- The unfolded state is an ensemble of a large number of molecules with different conformations.



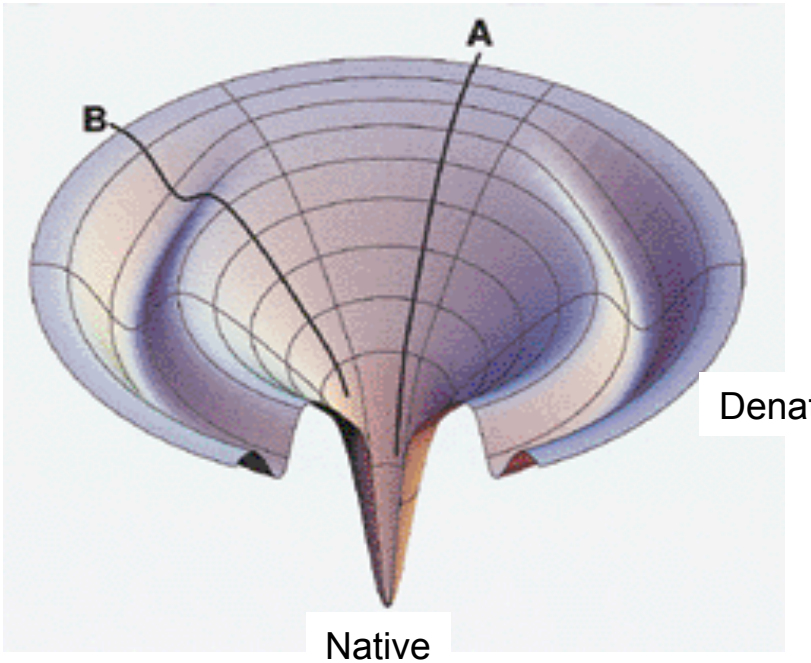

(a)
Denaturing conditions
Native conditions
(b)
(c)
(c')
(f)
(d)
(e)
(g)
© 1996 Current Opinion in Structural Biology

Energy surfaces for protein conformations

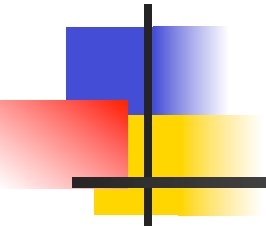
Golf course model
Levinthal landscape



Funnel model
Energy landscape

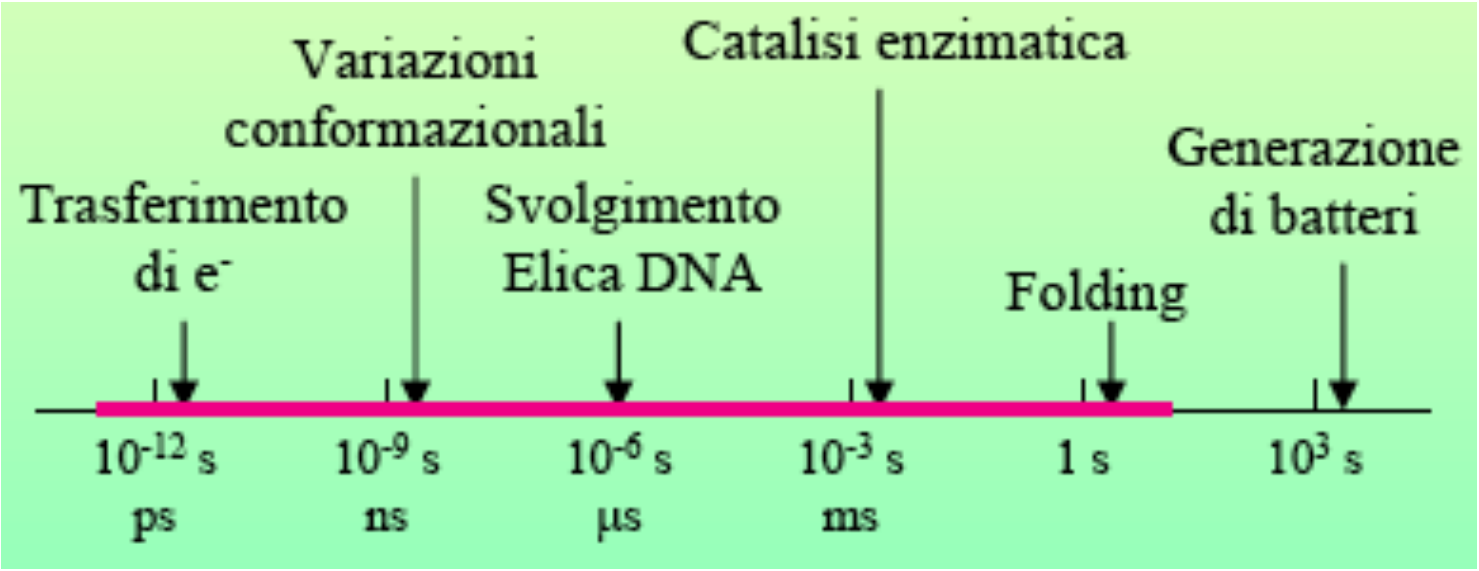


Methods for the study of kinetics of protein folding



Continuous and Stopped Flow

Timescale


Variazioni conformazionali
Catalisi enzimatica
Trasferimento di e-
Svolgimento Elica DNA
Generazione di batteri
Folding
10-12 s
ps
10-9 s
ns
10-6 s
µs
10-3 s
ms
1 s
103 s

The “continuous flow” method

- Two syringes are connected by a mixing chamber which continues into a tube. The syringes of the solutions are mixed and the reaction takes place during the path in the tube.
- The reaction will get older and older as it goes on.
- For example. For a constant flow of 10 m / s, the length of 1 cm the solution will have 10 ms.
- There will be a dead time that will match to the time taken by the solutions to go from the mixing chamber to the nearest measuring point.
- Typically this time is about 45-10 microseconds.

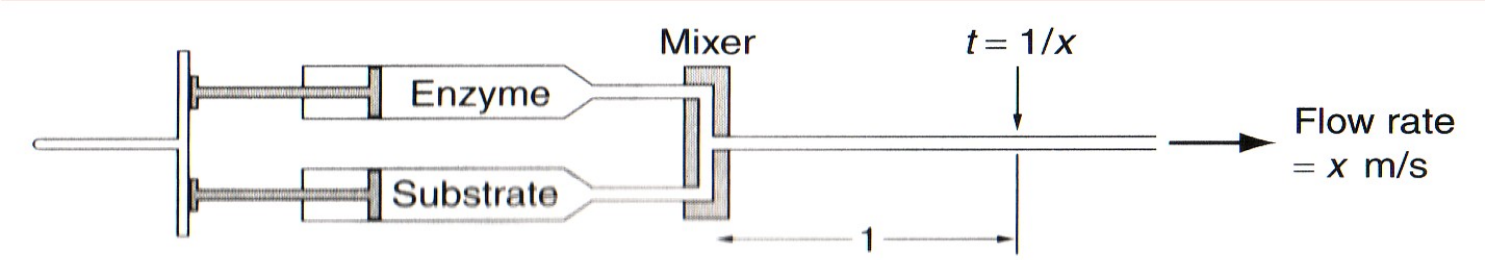
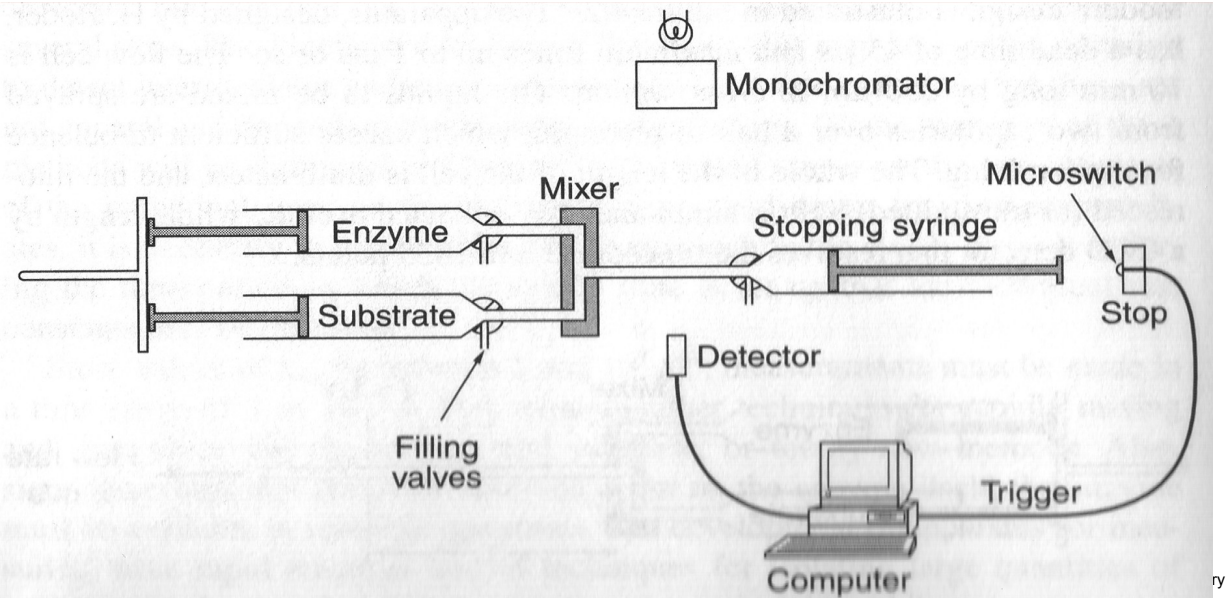
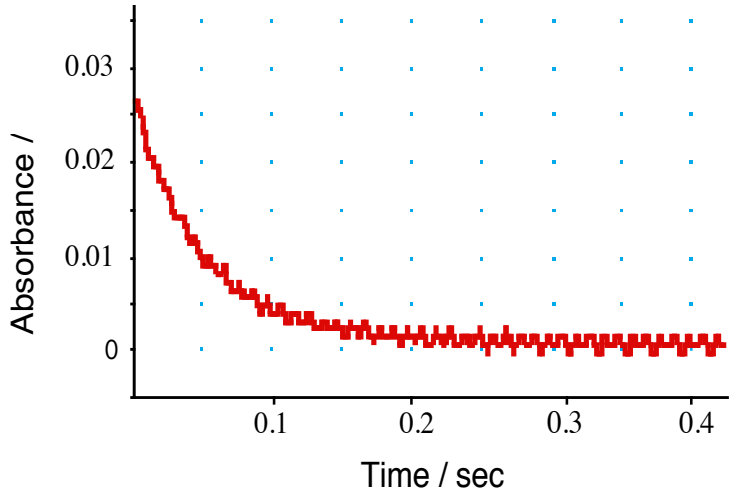


Figure 4.1 Continuous-flow apparatus.

The “stopped flow” method

- In this case, the syringes are much smaller and contain 50-200 μL each.
- In this case after mixing, the solution flows and mechanically stops (stopped-flow) and the absorption or fluorescence of the species that are formed or consumed are observed as a function of time.





An example...

Controllo

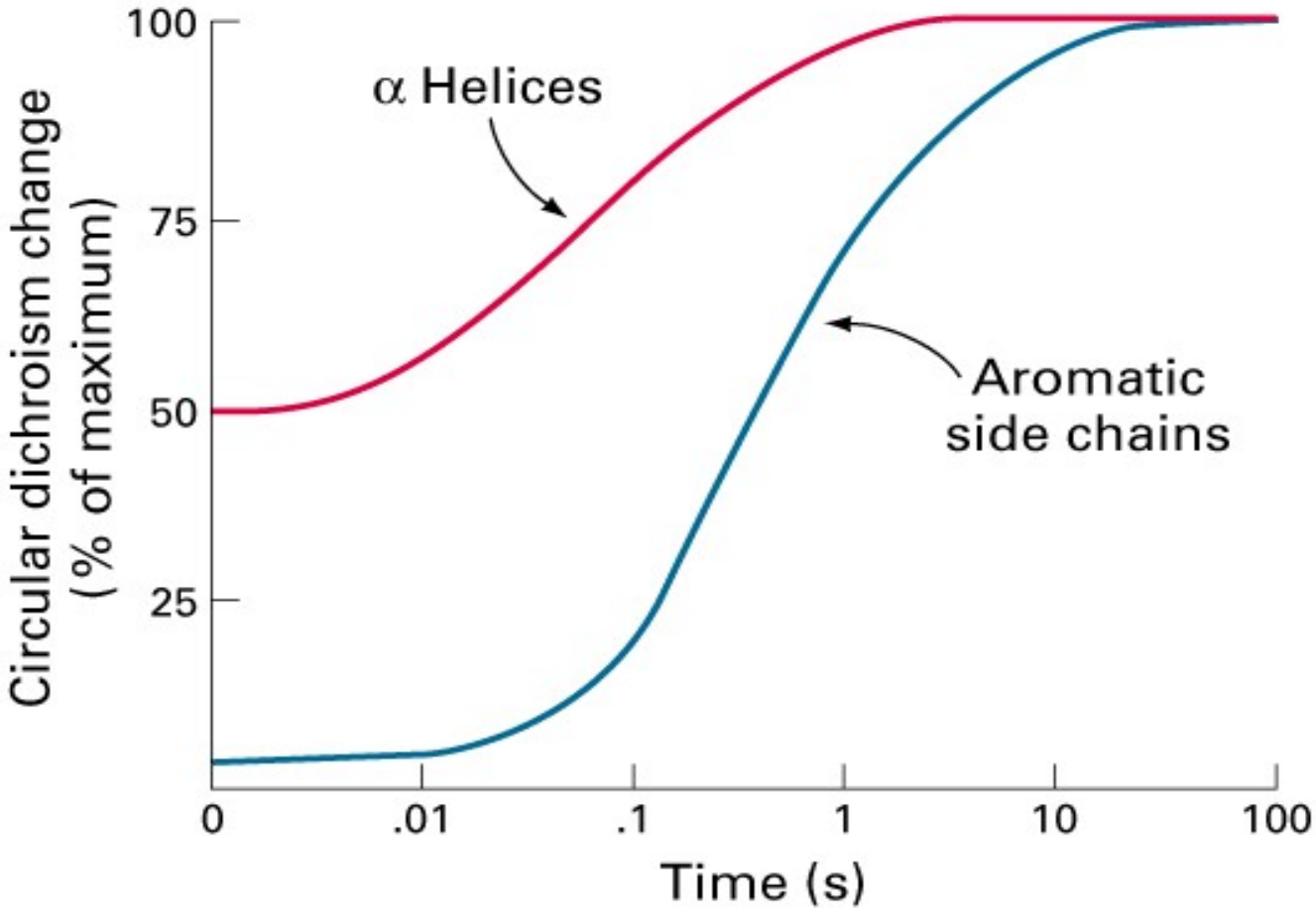
Siringhe
Mixer
Lettura

Sorgente
Detector

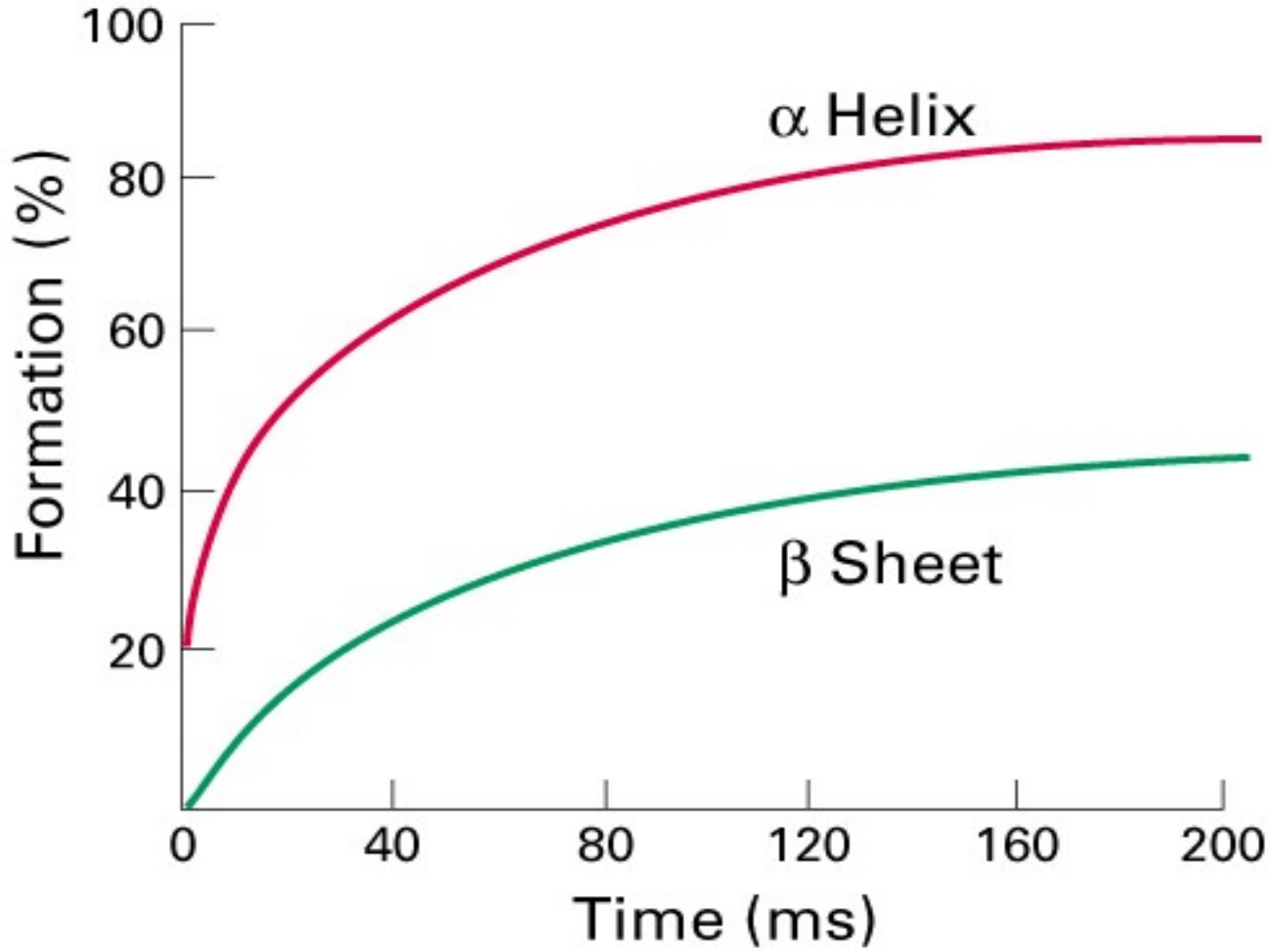
- http://www.hi-techsci.com

Folding of Cytochrome c

- α−helix formation is more rapid than tertiary structure rearrangements of aromatic sidechains in the folding of cytochrome c.
- The kinetics of these changes were determined by CD at 222 and 289 nm

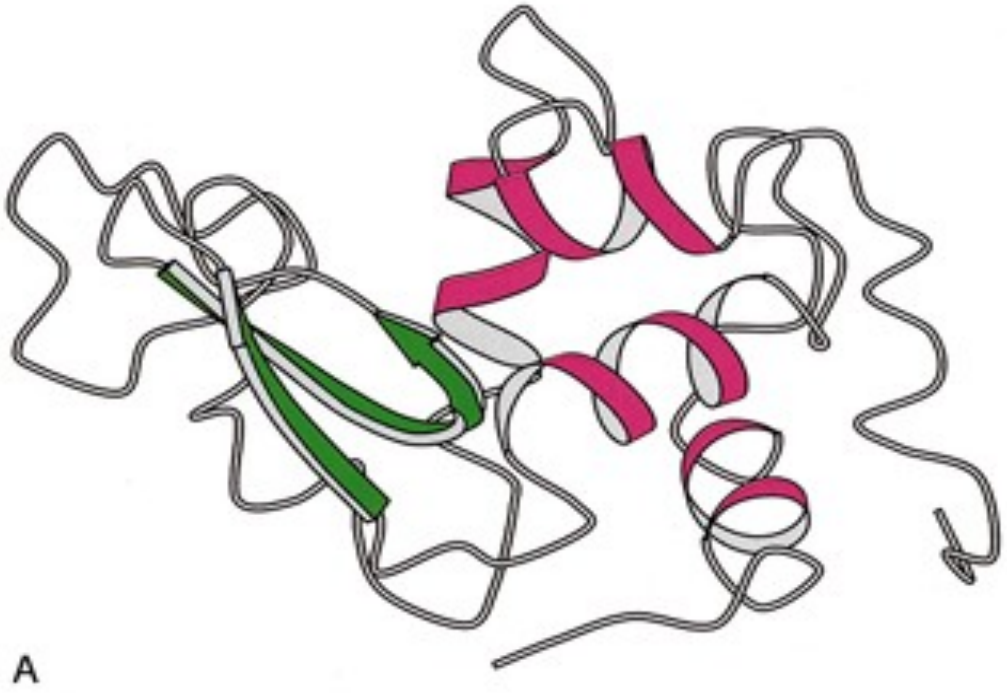


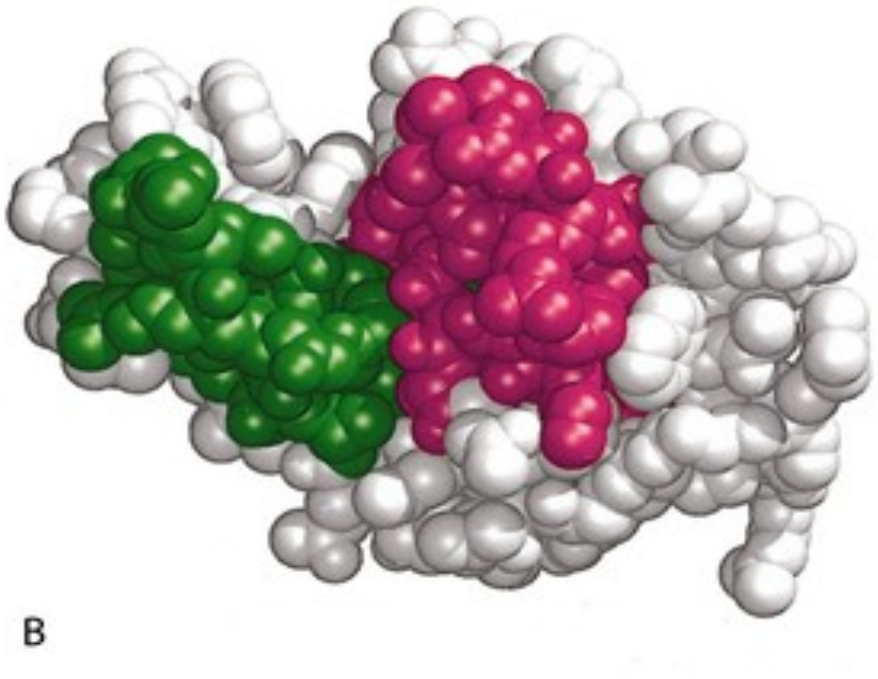
Folding of Lysozyme



- In the refolding of lysozyme, the helix domain is formed before the β-sheet.
- Proton exchangeability was measured at different times after the initiation of folding.

Folding of Lysozyme

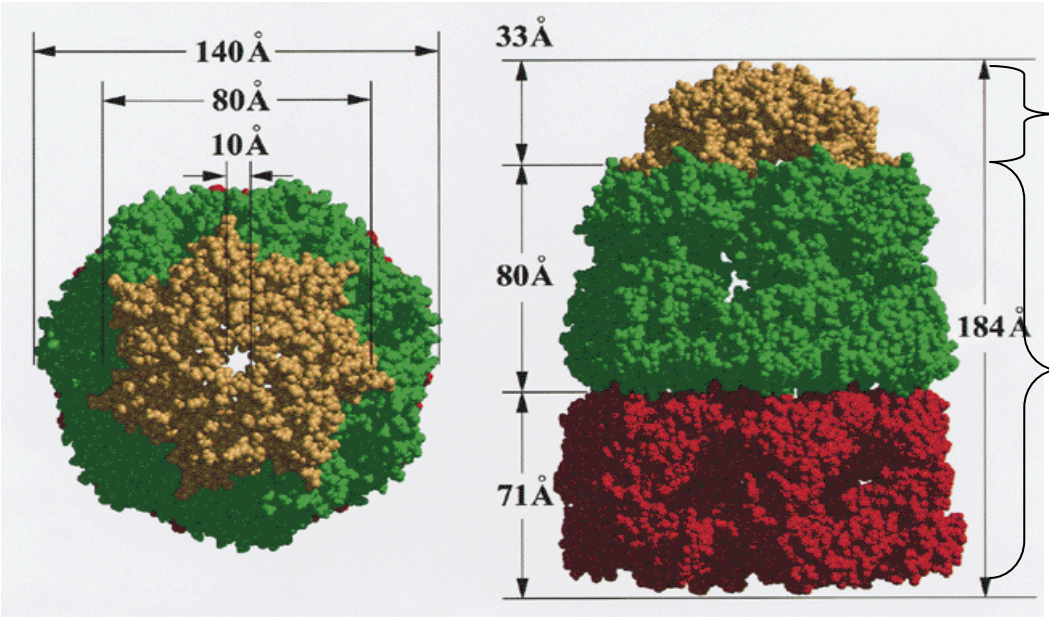




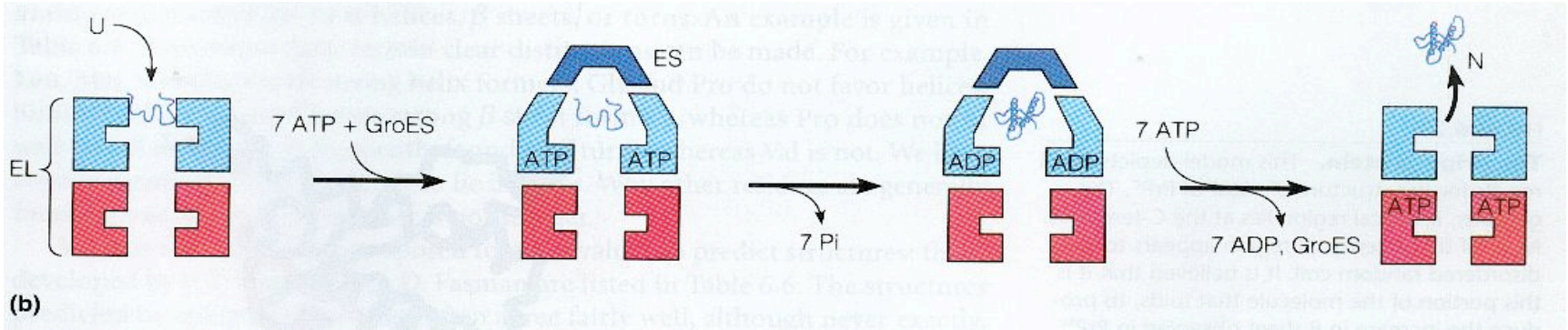
- The alpha helix domain is folded faster than the beta domain.

Assisted folding: Chaperonins

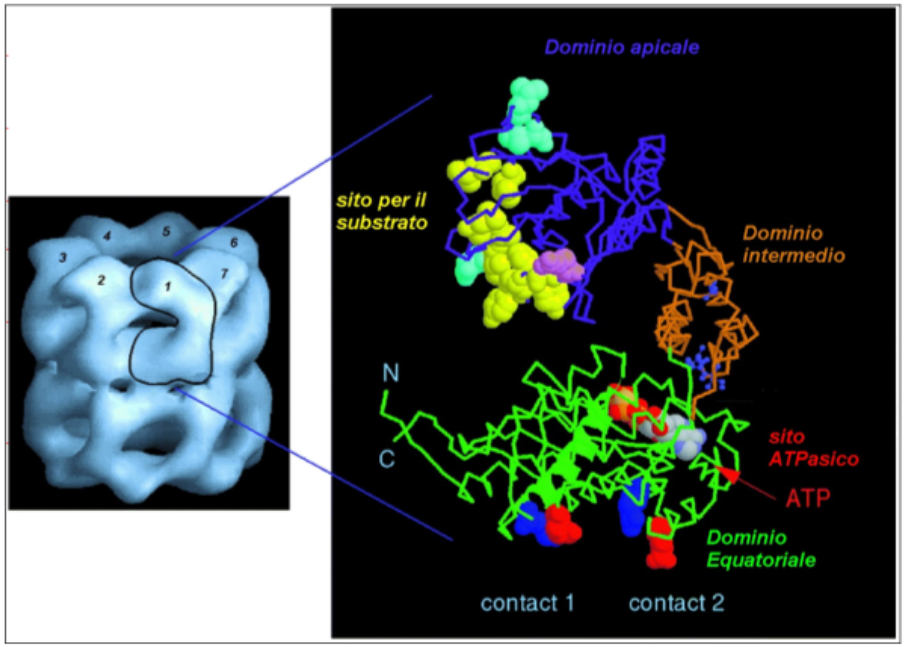
- Problems can occur during folding:
  - Improper folding
  - Aggregation
- Chaperone proteins assist protein folding *in vivo*
- Example: GroEL-ES complex from *E.coli*
- Mechanism:



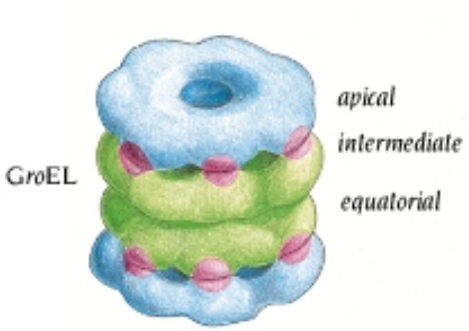
Top view Side view

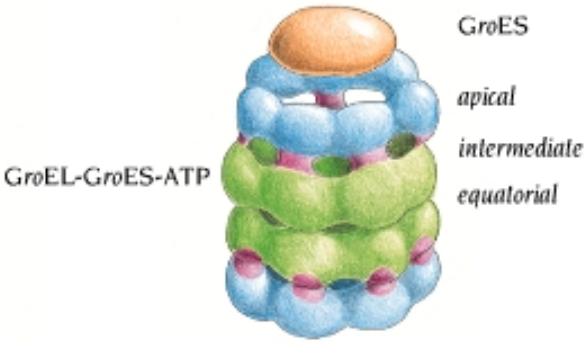


Assisted folding: Chaperonins

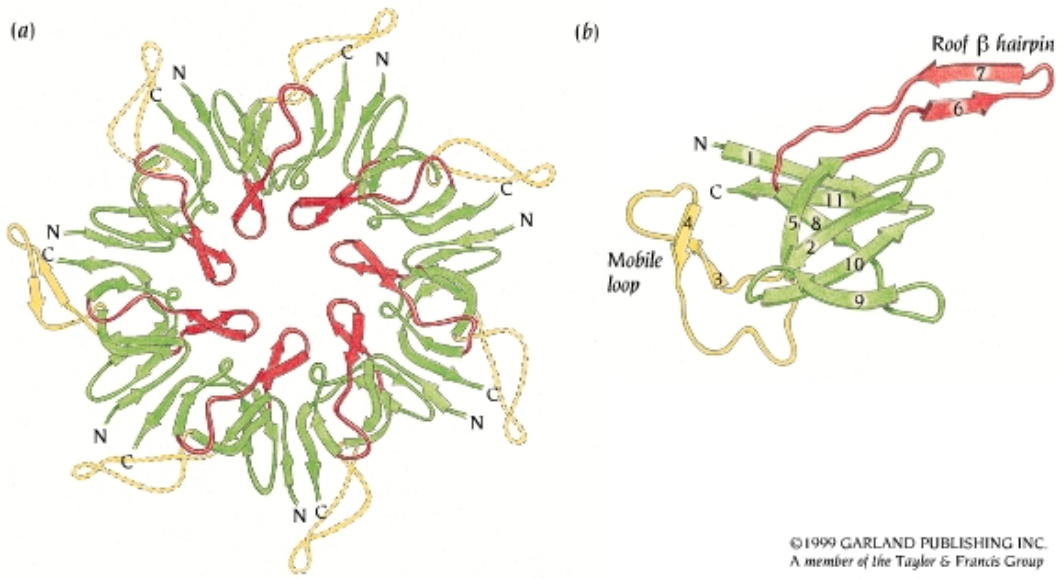


- GroEL consists of 14 subunits forming 2 rings (each one consisting of 7 subunits)
- Each subunit has 3 domains





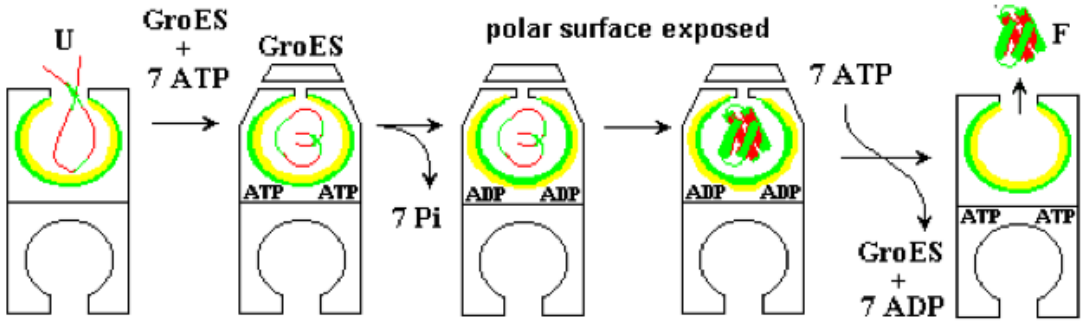
Assisted folding: Chaperonins



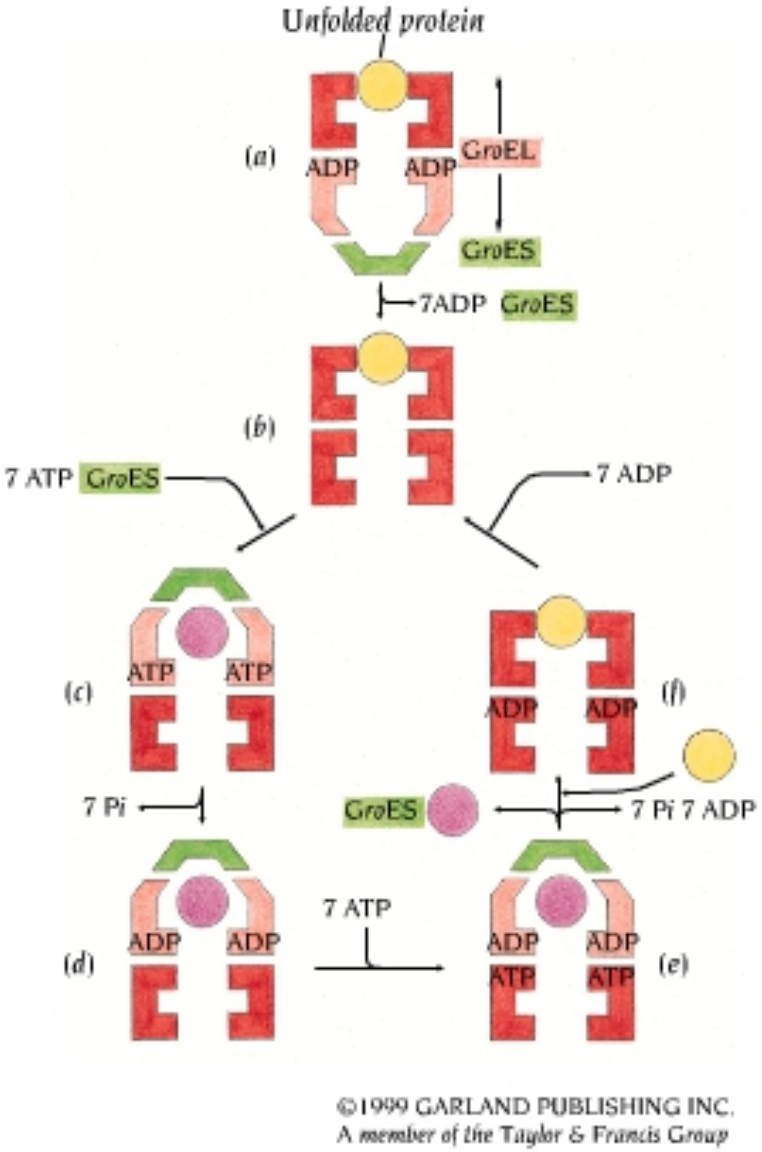
- GroES consists of 7 subunits forming 1 rings
- Each subunit has:

1.a core beta-barrel (green)
2.a long hydrophobic loop interacting with GroEL (yellow)
3. a loop forming a beta-hairpin (red) closing the cavity

Assisted folding: Chaperonins



Yellow: hydrophobic surface
Green: polar surface



Proteins are dynamic systems

- Proteins are not static, they are dynamic systems
  - Class I motions occur even in the crystal (they can limit resultion)
  - Class II and III motions occur in solution and are critical to function (binding, kinetics, movement etc.)

Class	Type of Motion	Approximate Range	
		Amplitude (nm)	Time (s)
1	Vibrations and oscillations of individual atoms and groups	0.2	10^{-15}–10^{-12}
2	Concerted motions of structural elements, like α helices and groups of residues	0.2–1	10^{-12}–10^{-8}
3	Motions of whole domains; opening and closing of clefts	1–10	$\geq 10^{-8}$