

## Part 2

- Protein folding: thermodynamics and kinetics  
thermodynamic stability – thermal stability
- Experimental techniques for characterizing folding  
intermediate states – transition states  
*in vitro* <-> *in vivo* folding: chaperone proteins

## Protein folding problem:

How does a protein pass from its unfolded state to its folded – biological active – state, which is (almost) unique ?

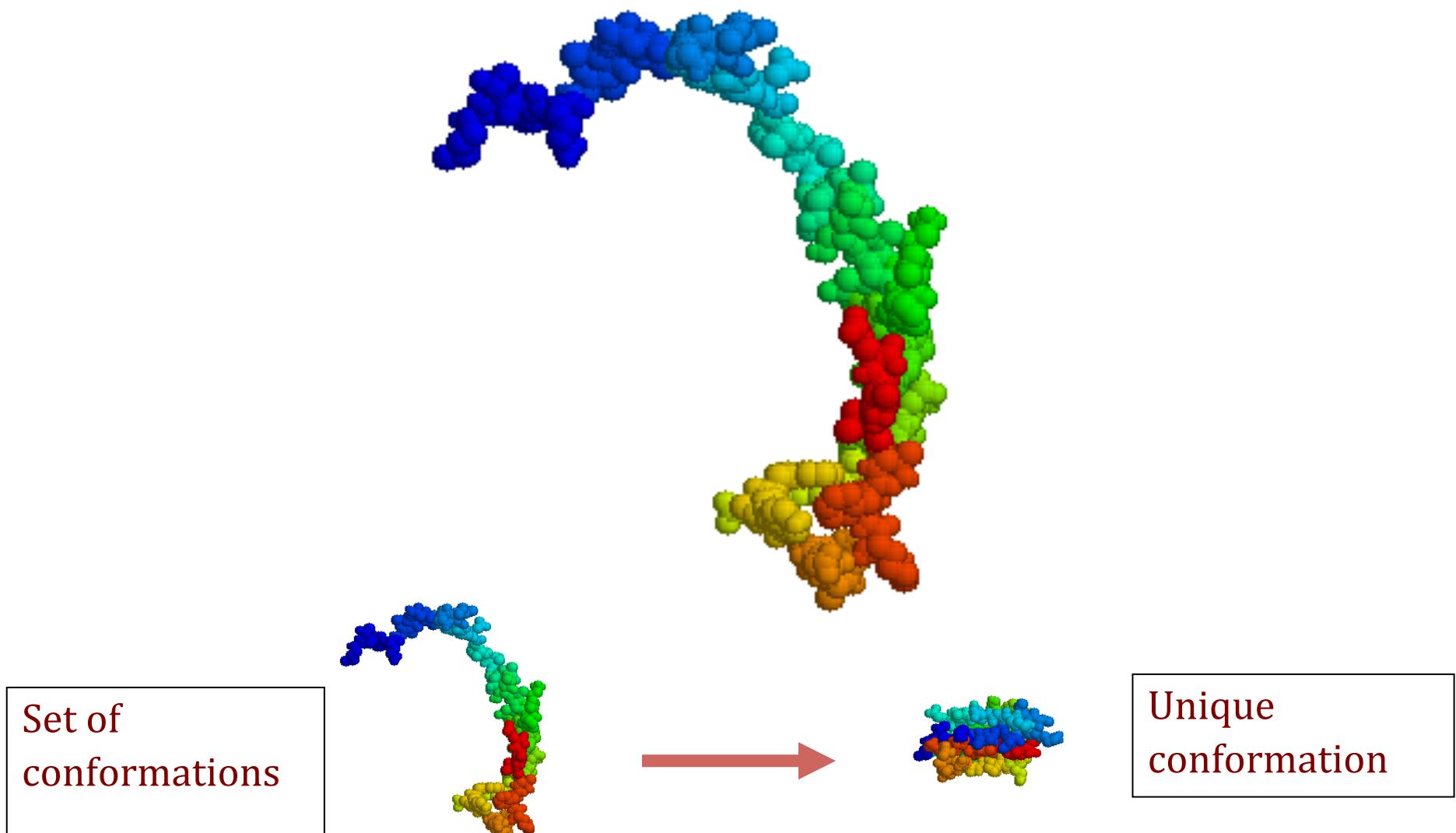
### What is the unfolded or denatured state?

- Corresponds to a disordered state - set of conformations of similar energies  
→ high conformational entropy, large enthalpy
- May contain residual structures: secondary structures or clusters of hydrophobic residues that form and deform – also disulphide bridges, which severely restricts the flexibility of the chain.
- There is not one denatured state - it depends on denaturing conditions (T, pH, denaturant, ...)

### What is the folded state?

- Structure of low(est ?) free energy (in what conditions?) - low entropy, low enthalpy - fixed structure that plays a clearly defined biological role - especially the active site, binding sites ....
- Exceptions: the function of some proteins is to be and remain unfolded e.g. repellent role to prevent the approach of certain molecules;
- Half exceptions: proteins that fold only when binding to a ligand, other protein, DNA or RNA

## In images:



# Thermodynamics of folding

More precisely:

D = denatured, unfolded state

N = folded, native state



Evolution towards the steady state

(at a slow or fast rate depending on the kinetics of the process)

Equilibrium position = point where the reaction rates are identical in the two directions (D  $\rightarrow$  N as N  $\rightarrow$  D)

Folding: D  $\rightarrow$  N: process favored under certain conditions of T, pH, ...., for example at room temperature

Unfolding: N  $\rightarrow$  D: process favored under different conditions of T, pH, .... for example at high temperature

But there is always a non-zero amount of N and D

# Thermodynamics of folding

In terms of:

Gibbs free energy  $G = H - TS$

$T$  = absolute temperature

$H$  = enthalpy:

$\Delta H = q$  at  $P$  constant ( $P$  = pressure,  $q$  = energy supplied to the system in the form of heat)

$S$  = entropy ; number of microstates associated with the state; measure of the disorder of matter and energy

$T \Delta S = q_{rev}$  = reversible part of the heat supplied to the system

Reversible process: in a reversible cyclic process, where the system is always kept at equilibrium (infinitesimal modifications), the system and the environment go back to their original state -> no entropy production.

# Thermodynamics of folding

Second principle of thermodynamics: the entropy of the universe tends to increase :  $\Delta S_{\text{tot}} > 0$

! Not necessarily of a subsystem, but of the subsystem + environment !!

G, H, S: state functions = physical properties that depend only on the initial and final step, not of the path followed.

$$\Delta G = \Delta H - T \Delta S = q - q_{\text{rev}} = q_{\text{irrev}} \quad (\text{P, T const})$$

- At equilibrium,  $\Delta G = 0$  (no  $q_{\text{irrev}}$ )
- For a spontaneous transformation at P, T const:  $\Delta G < 0$

Standard Gibbs free energy:  $\Delta G^0 = \Delta H^0 - T \Delta S^0$

=  $\Delta G$  when the reactants in their standard state are transformed into products in their standard state

Standard state for a substance in solution:  
concentration of 1M and a P=1 atm.

Useful because the  $\Delta G$  of a reaction depends not only on the reactants and products, but also their concentration.

# Thermodynamics of folding

Relation between  $\Delta G$  and  $\Delta G^0$  :

$\Delta G = \Delta G^0$  a term dependent on the concentration of reactants and products  
(here P const)

$$= \Delta G^0 + RT \ln Q$$

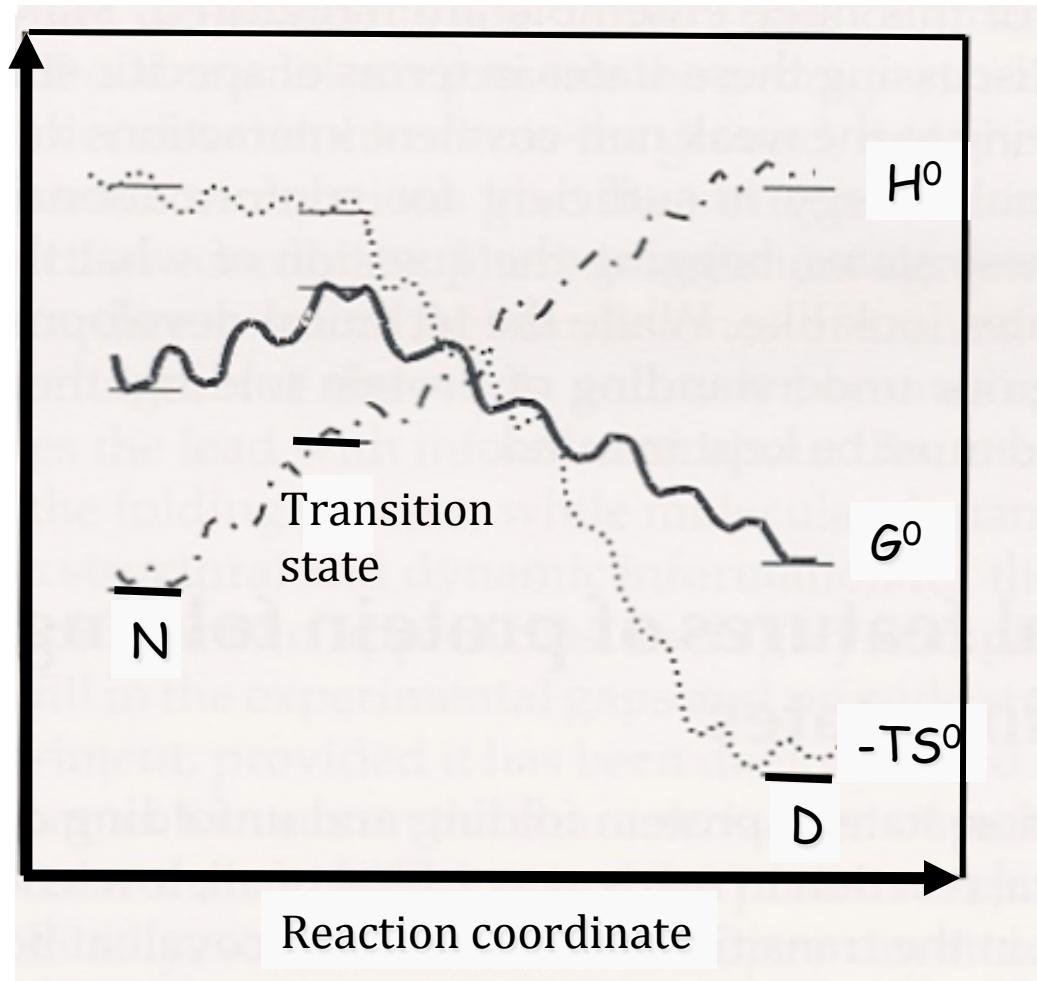
At equilibrium,  $\Delta G = 0$  and  $\Delta G^0 = -RT \ln K$   
with  $K$  = equilibrium constant =  $[N] / [D]$

- $\Delta G^0 = 0 \Rightarrow [N] = [D]$
- $\Delta G^0 < 0 \Rightarrow [N] > [D]$
- $\Delta G^0 > 0 \Rightarrow [N] < [D]$

$\Delta G^0$  depends on various parameters, T, pH, etc..

In a range of values of these parameters :  $\Delta G^0 < 0$ , N is more stable than D.  
In another interval,  $\Delta G^0 > 0$ , and D is more stable than N.

# Thermodynamics of folding



## State N:

H small (favorable): stabilizing interactions between residues

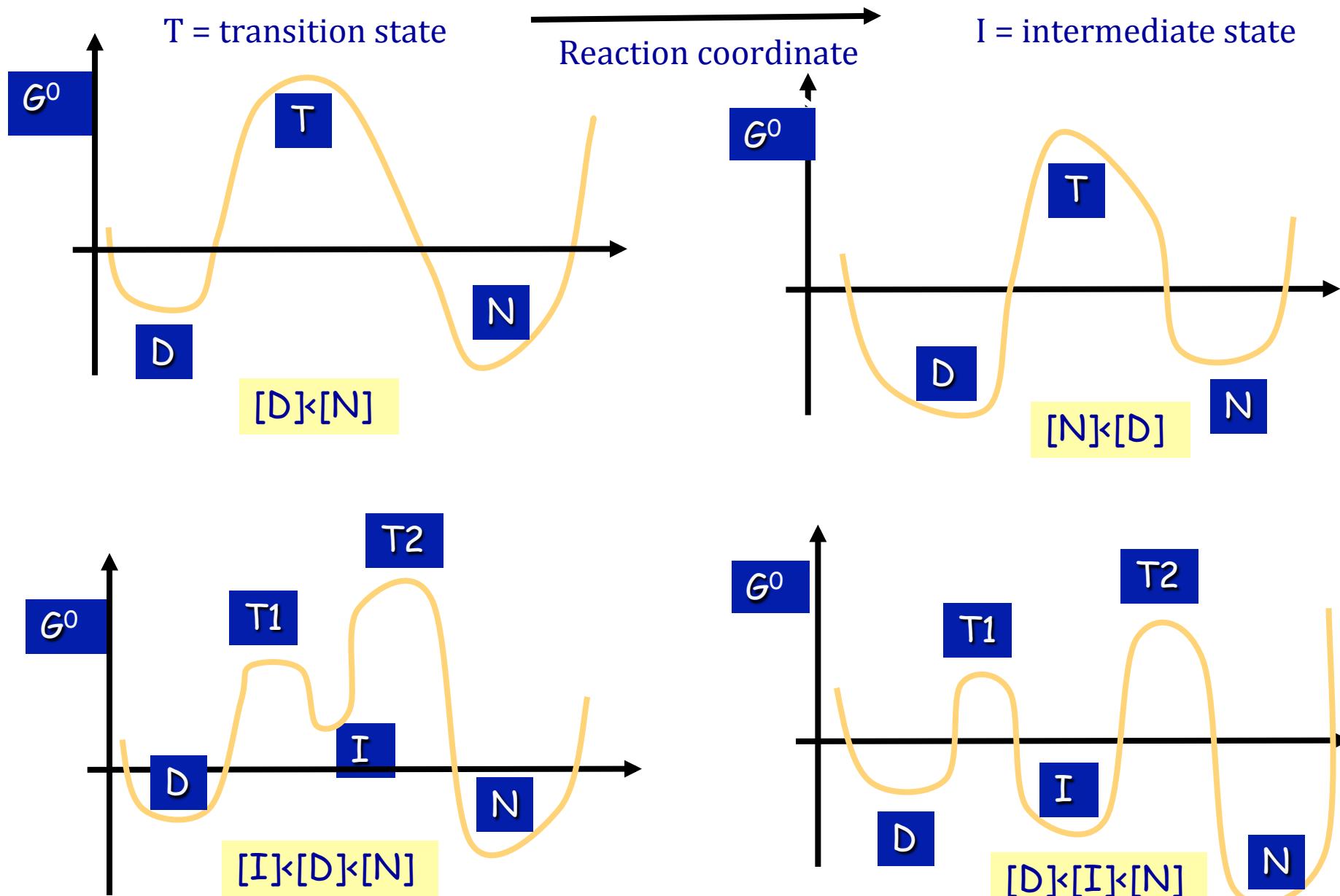
S small (unfavorable): low entropy

## State D:

H large (unfavorable): Little or no stabilizing interactions between residues

S large (favorable): large entropy

Transition state on the path between D and N ; located at the point where the decrease in entropy is not compensated by the decrease of energy.



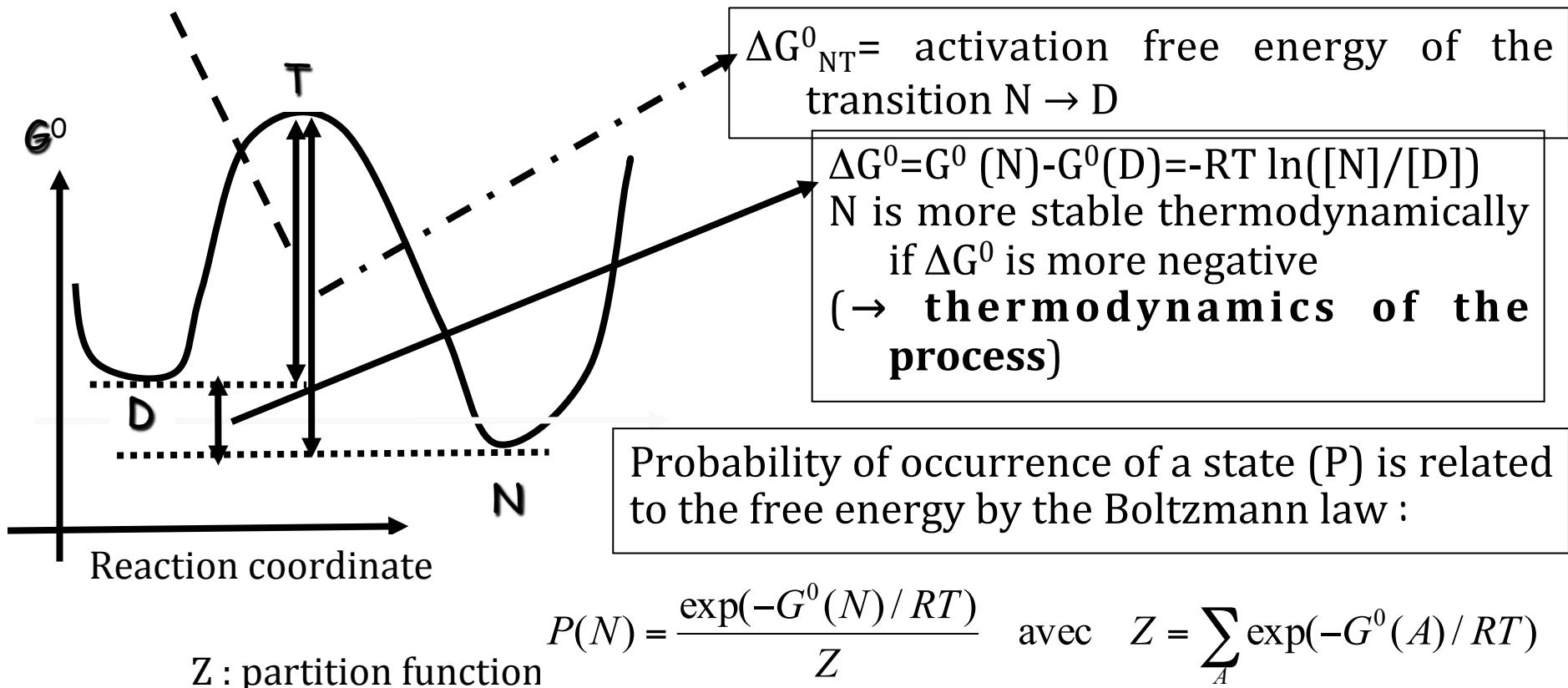
$\Delta G^0_{DT}$  = activation free energy =  $G^0(T) - G^0(D)$  of the transition  $D \rightarrow N$

$$= -RT \ln k_{D \rightarrow N} / k_0 = RT \ln t_{D \rightarrow N} / t_0$$

$k_{D \rightarrow N}$  = rate constant of the reaction  $D \rightarrow N$ ;  $k_0$  =  $k$  of « elementary » step

$t_{D \rightarrow N}$  = duration of the transition  $D \rightarrow N$ ;  $t_0$  =  $t$  of « elementary » step

The reaction is faster if the free energy barrier is small / i.e. if  $\Delta G^0_{DT}$  is small  
**(→ kinetics of the process)**



In general:

the folded state (N) can be unfolded (D) in a reversible manner by the addition of denaturants, by increasing the T, by varying the pH, increasing the pressure, or by cleaving the disulfide bridges.

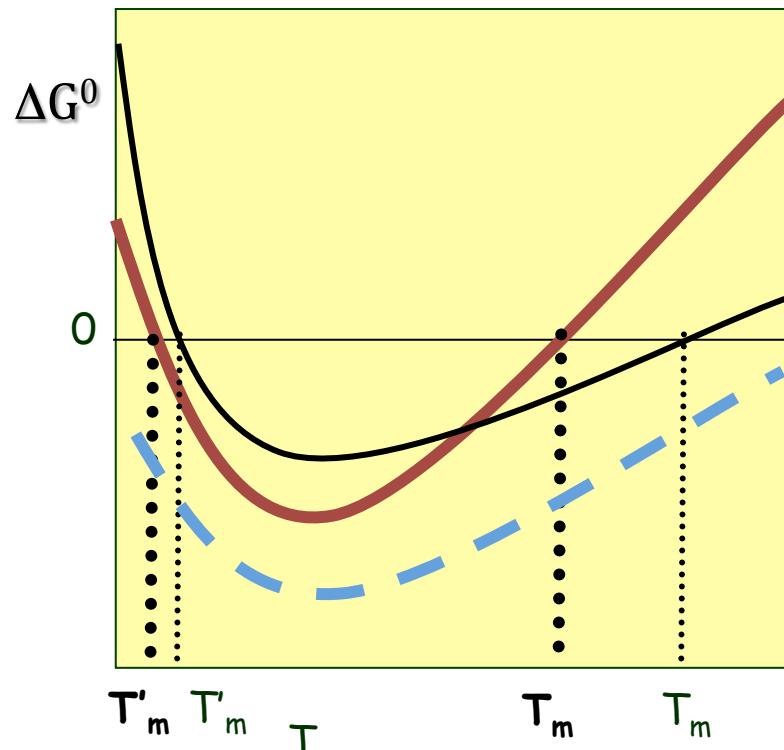
N states are only marginally more stable than the D states: free energy difference for typical small natural proteins between -5 and -10 kcal / mol.

The N state is stabilized by non-covalent interactions, including H-bonds, van der Waals, electrostatic, cation-pi, aromatic interactions, hydrophobic forces, ...

The state D is "stabilized" by the conformational entropy.

! Proteins of organisms living in extreme conditions (hot-cold-acid-salt) - called extremophiles – adopt their native structure in these conditions only - in any other, they unfold. These proteins are not very different from mesophilic proteins - their amino acid sequence is simply adapted to enhance the interactions that are strong in these conditions.

## Dependence of the free energy on the temperature T

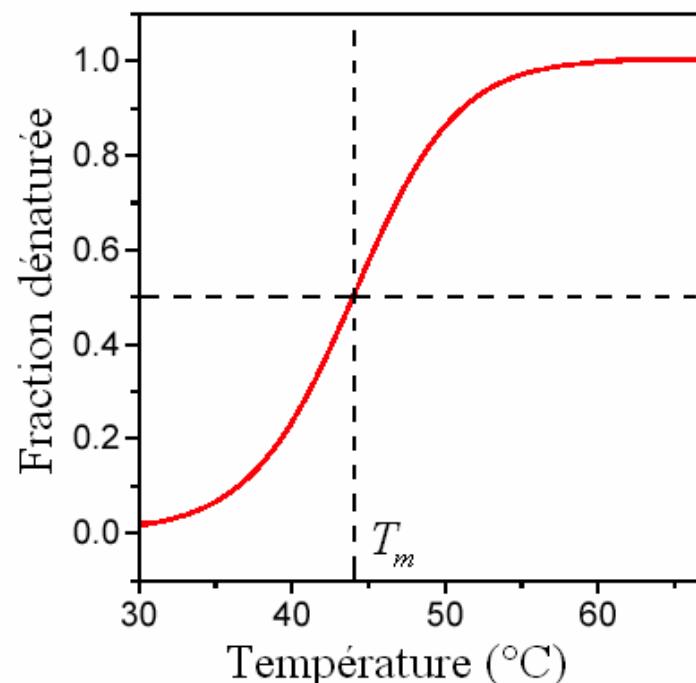


Reversed bell-like shape

T such that  $\Delta G^0(T)=0$  : Melting/denaturation temperatures  $T_m$ .

In general, a protein has 2  $T_m$ , one at high and one at low temperature.

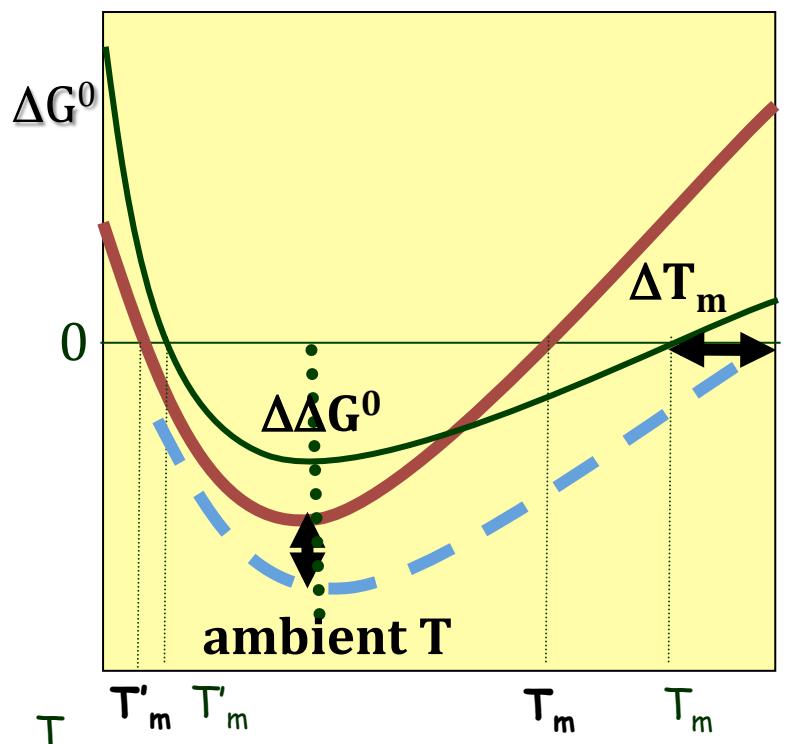
$$\Delta H^0(T_m) = T_m \Delta S^0(T_m)$$



at  $T = T_m$ , there are as many proteins in the states N and D :

$$\Delta G^0(T_m) = -RT_m \ln[N]/[D] = 0$$

Relation :  $T_m - \Delta G^0$  or thermal - thermodynamic stability



Thermodynamic stability is determined by the negativity of  $\Delta G^0$  at some T (often ambient T).

Thermal stability is determined by the magnitude of  $T_m$

In general: correlation between  $\Delta G^0$  (ambient T) and  $T_m$ :

The more negative  $\Delta G^0$ , the largest  $T_m$

Not always true, in one of the examples it is the opposite!!

When considering protein mutants:  
better correlation between  $\Delta\Delta G^0 = \Delta G^0$  (mutant)- $\Delta G^0$  (wildtype) and  
 $\Delta T_m = T_m$  (mutant) -  $T_m$  (wildtype)

(even better if few changes between wildtype and mutant).

But sometimes not: the strength of the interactions depends on the T, and this dependence varies with the type of interactions. For example, the salt bridges are more favorable at high T compared to hydrophobic forces.

## Relation : $\Delta T_m = \Delta \Delta G^0$ or thermal - thermodynamic stability

More precisely : Heat capacity at constant P:  $C_P^0 = \frac{dH^0}{dT} \Big|_{Pcte} = T \frac{dS^0}{dT} \Big|_{Pcte}$

$$\Rightarrow \begin{cases} H^0(T) = H^0(T_m) + \int_{T_m}^T C_P^0(T') dT' \\ S^0(T) = S^0(T_m) + \int_{T_m}^T C_P^0(T') \frac{dT'}{T'} \end{cases}$$

$$\begin{cases} car \quad S^0 = -\frac{dG^0}{dT} \Big|_{Pcte} \quad et \\ \frac{dH^0}{dT} = \frac{d(G^0 + TS^0)}{dT} = -S^0 + S^0 + T \frac{dS^0}{dT} \end{cases}$$

$$\Rightarrow G^0(T) = \underbrace{H^0(T_m) - TS^0(T_m)}_{||} + \int_{T_m}^T C_P^0(T') dT' - T \int_{T_m}^T C_P^0(T') \frac{dT'}{T'} \\ (T_m - T) S^0(T_m)$$

$$\Rightarrow \Delta G^0(T) = (T_m - T) \Delta S^0(T_m) + \int_{T_m}^T \Delta C_P^0(T') dT' - T \int_{T_m}^T \Delta C_P^0(T') \frac{dT'}{T'} \quad (N \leftrightarrow D)$$

Hypothesis:  $\Delta C_P^0$  varies few with T on the interval  $[T, T_m]$ )

$$\Rightarrow \Delta G^0(T) = (T_m - T) \Delta S^0(T_m) - \Delta C_P^0 \left[ (T_m + T \ln(\frac{T}{T_m}) - 1) \right]$$

## Relation : $\Delta T_m = \Delta\Delta G^0$ or thermal - thermodynamic stability

Consider mutant and wildtype proteins

Hypotheses:

$$1) \Delta C_P^0(\text{wildtype}) \approx \Delta C_P^0(\text{mutant})$$

$$2) \Delta S^0(\text{wildtype, at } T_m \text{ wildtype}) \approx \Delta S^0(\text{mutant, at } T_m \text{ mutant})$$

$$\Rightarrow \Delta\Delta G^0(T) \equiv \Delta G_{\text{mut}}^0(T) - \Delta G_{\text{wt}}^0(T)$$

$$\approx (T_m^{\text{mut}} - T_m^{\text{wt}}) \Delta S^0(T_m^{\text{wt}}) \equiv \Delta T_m \Delta S^0(T_m^{\text{wt}})$$

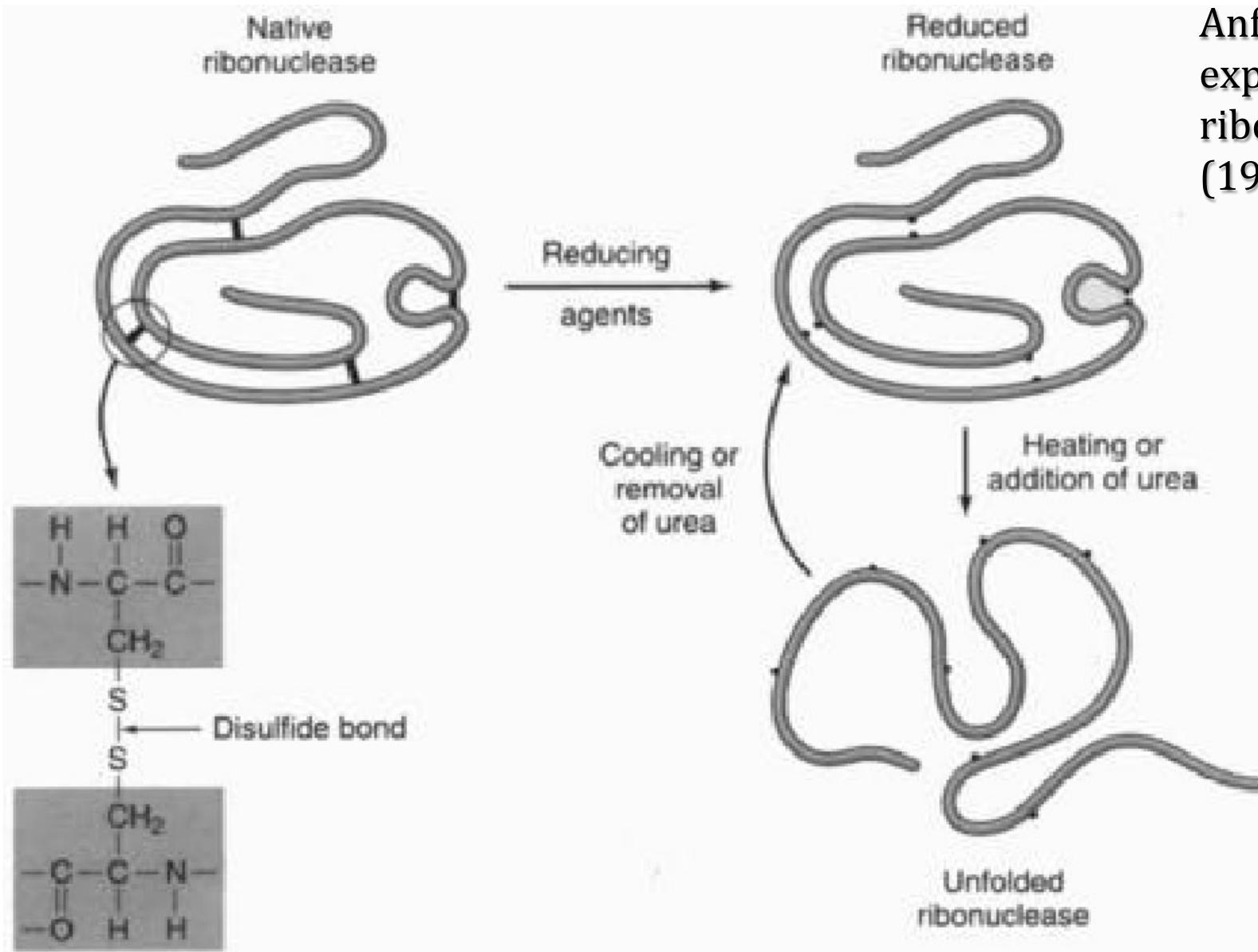
Thus under these hypotheses :  $\Delta\Delta G^0(T)$  and  $\Delta T_m$  are proportional

But these hypotheses are only partially verified in general.

They are better verified if:

- $T$  and  $T_m$  are close
- Not too much difference in sequence and structure between wildtype and mutant.

## Protein folding is a reversible process



Anfinsen  
experience on  
ribonuclease A  
(1973)

## Protein folding is a reversible process

Christian Anfinsen and his group have shown that ribonuclease A, which has its secondary and tertiary structure destroyed by urea (denaturing solvent) and mercaptoethanol (destroys disulfide bridges), folds spontaneously and regains its activity when these denaturing agents are removed.

-> Proteins can renature if they are not too large (~ contain a single domain) and have not undergone substantial chemical changes after folding in vivo.

! Renaturation can be prevented by precipitation, aggregation (difficulties increase with the size - in general).

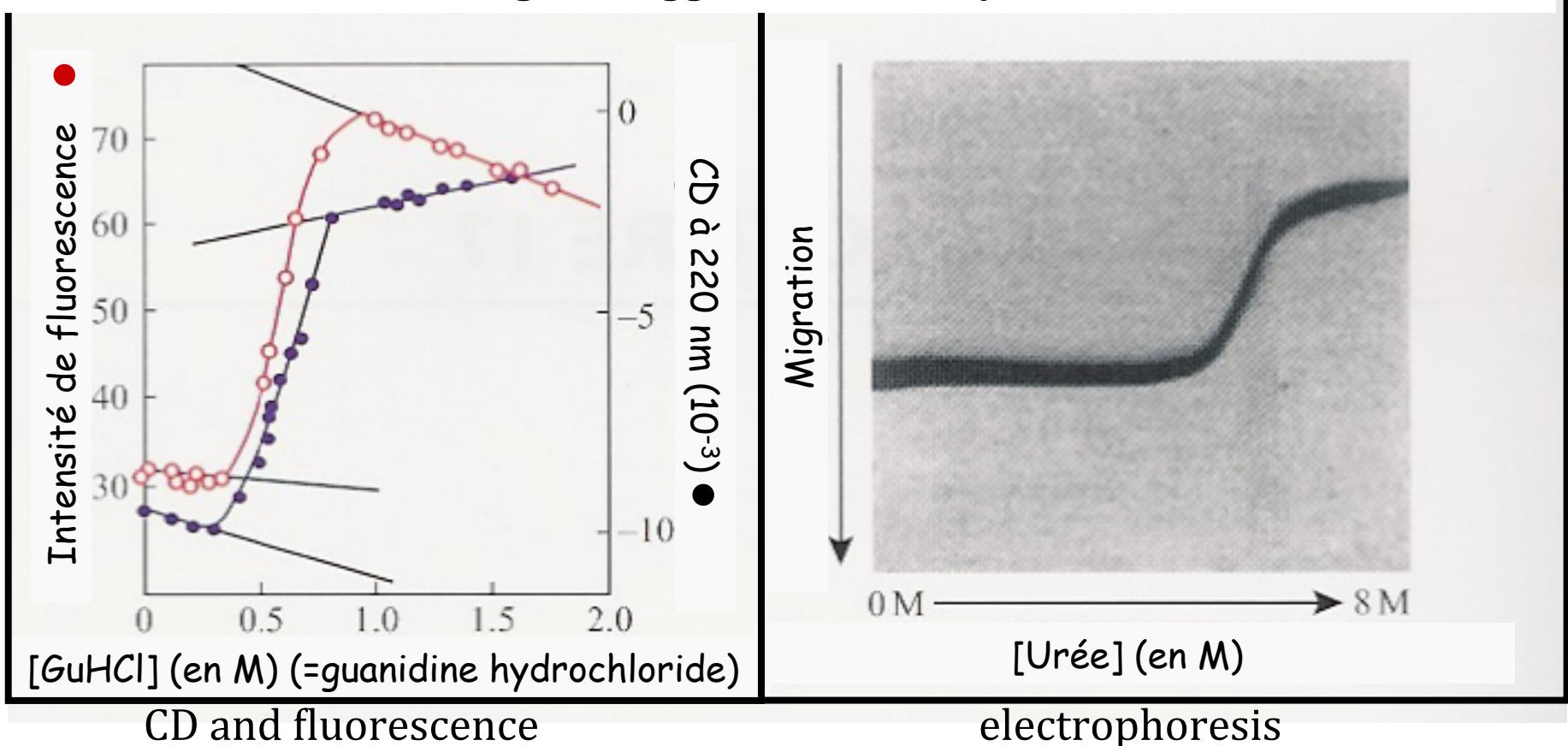
Reversibility is important: it shows that all the information necessary to build the 3D structure of a protein is contained in its amino acid sequence. And that the structure itself is thermodynamically stable (! Again for a protein not too modified and not too big)

## Protein folding is a cooperative process

Denaturation of small proteins: usually a cooperative transition

=> when certain key interactions are created, others are easily formed.

Experimental denaturation in vitro: an abrupt change in several characteristics of the molecule is observed (sigmoidal shape - S shape) - the narrow transition region suggests that many residues are involved.



**Electrophoresis** is a method for separating electrically charged particles by differential migration under the action of an electric field. The fundamental characteristic for electrophoresis is obviously the particle charge, since it determines the direction, speed and distance of migration.

=> pH of the medium has a significant influence. Also, the particle size has some influence, especially when the support is a porous matrix : the smaller the particles are compared to the pore size, the faster they migrate more.

**Fluorescence measurements:** UV laser light pulses are sent to a protein - this light is partially absorbed by aromatic residues - partially converted into heat, and partially re-emitted as photons of larger wavelength. This fluorescence is detected and provides information on the location of aromatic residues, because its brightness is influenced by its environment.

## Protein folding is a cooperative process

$$\Delta G^0(1+2+\dots+N) < \Delta G^0(1+2) + \Delta G^0(1+3) + \dots + \Delta G^0((N-1)+N)$$

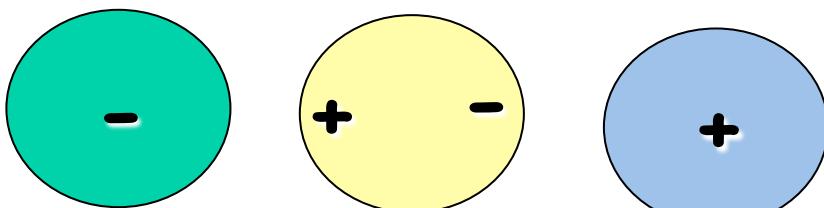
Overstabilization when adding a new partner to the interaction.

!!! *<-> Non-additivity of the interactions: interactions reinforce one another (cooperativity). Sometimes they can weaken one another (negative cooperativity or anti-cooperativity)*

### 1) Enthalpic cooperativity:

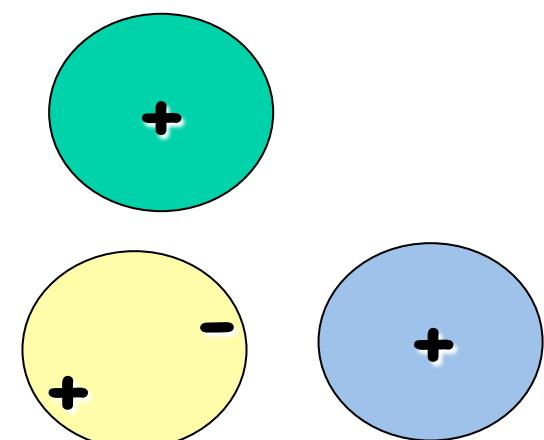
$$\Delta H^0(1+2+\dots+N) < \Delta H^0(1+2) + \Delta H^0(1+3) + \dots + \Delta H^0((N-1)+N)$$

For example:



Polarizable molecule  
between a positively and a  
negatively charged molecule  
=> Cooperativity

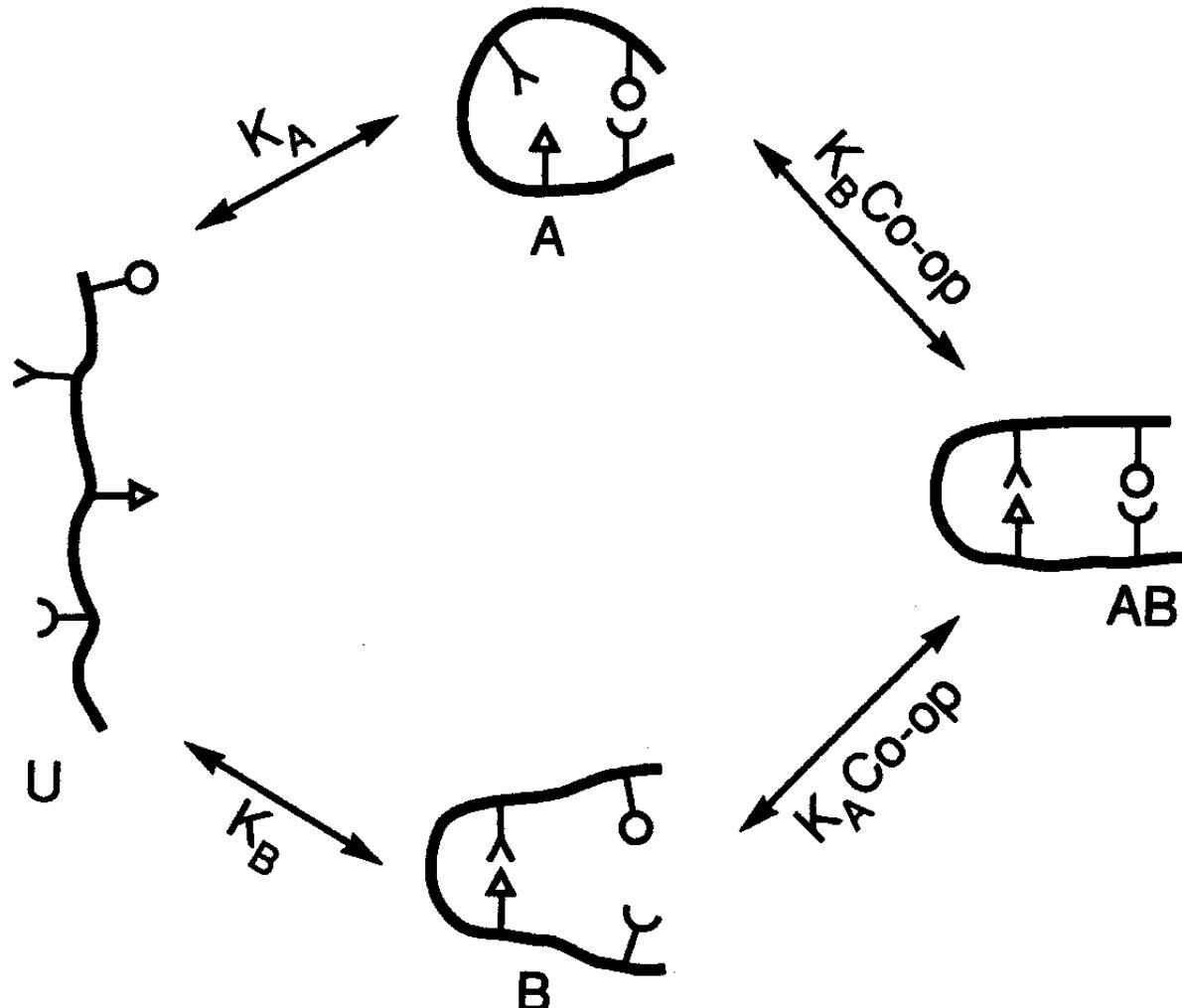
Polarizable molecule between two positively charged molecules  
=> Anti-cooperativity



## Protein folding is a cooperative process

2) Entropic cooperativity:  $\Delta S(A+B) < \Delta S(A) + \Delta S(B)$

2 types of entropy: conformational (rotation/ translation) and solvation



Loss of entropy for one of the two interactions (rotation, translation, and solvation for hydrophobic residues), automatically implies entropy loss for the second interaction

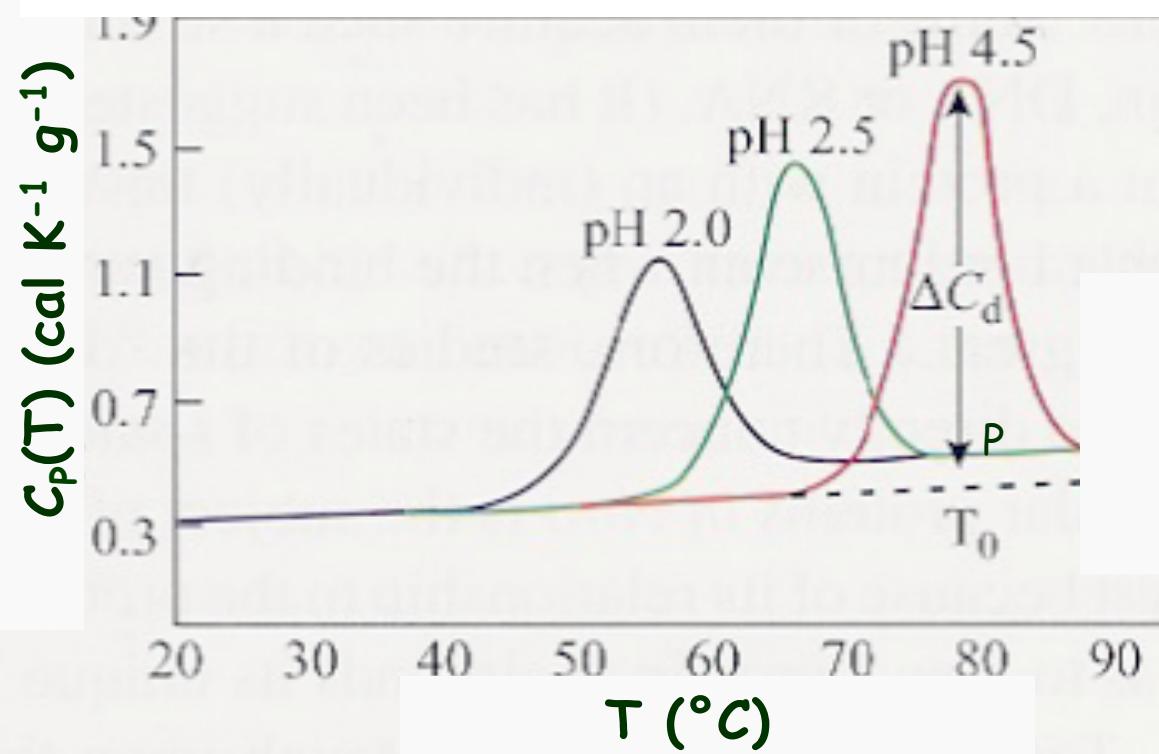
Little additional entropy loss upon formation of the second interaction.

## Protein folding is an all-or-none transition (for a small protein)

i.e. only the initial and final states (N and D) are in detectable amount ; semi-denatured states (I - intermediate) are virtually absent.

In fact I states exist in small quantities, because protein cannot denature without going through intermediate states - but these are transient and never or rarely accumulate (for small proteins and in usual conditions)

May depend on the conditions - some can induce intermediary states



Calorimetric study of denaturation by heat at  $\neq$  pH. Change of  $C_p$  (here per gr) as a function of T:

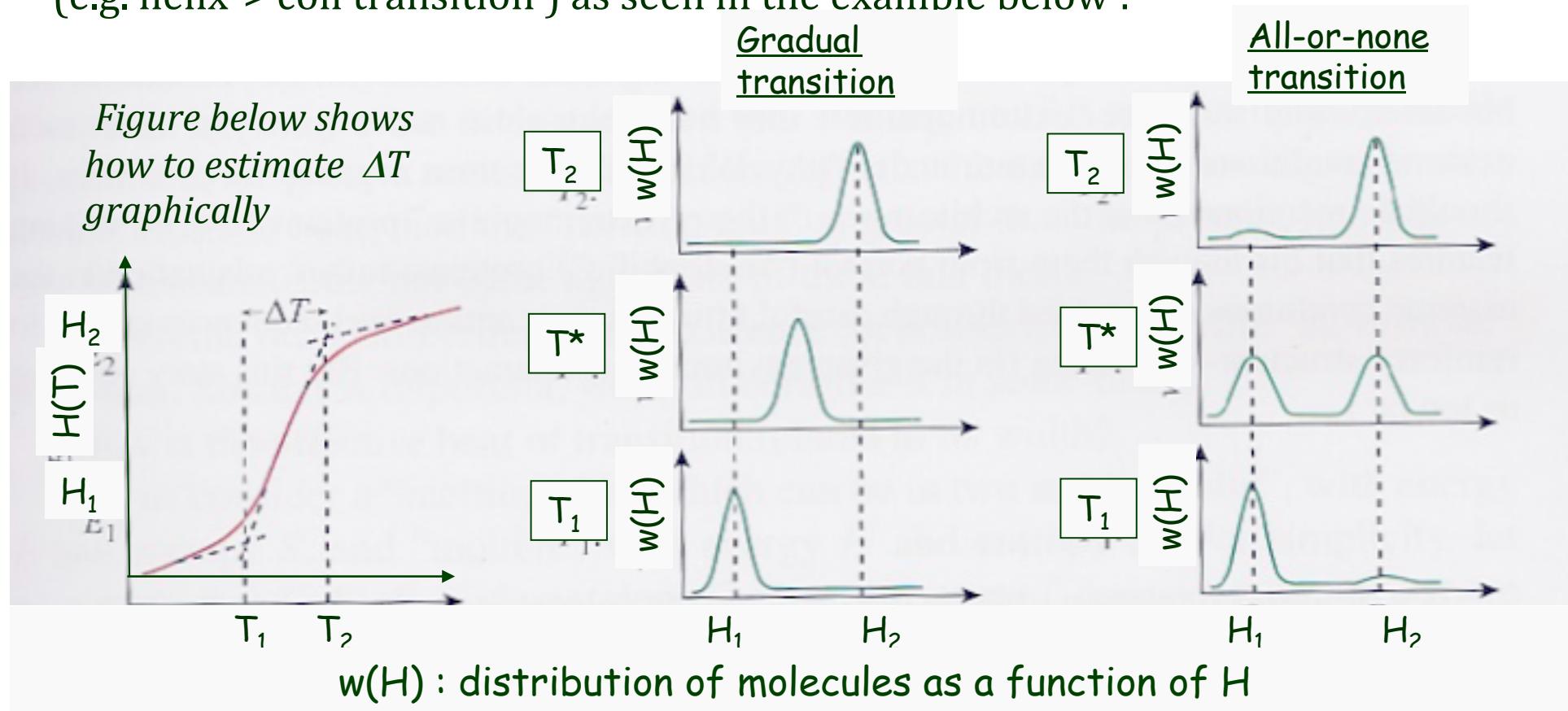
$$\Delta C_P^0 = \Delta \frac{dH^0}{dT} \Big|_{Pcte} \approx \frac{\Delta q}{\Delta T}$$

$\Delta T$ : peak width; area under the peak: heat ( $\Delta q$ ) absorbed by 1 gr of protein

## Protein folding is an all-or-none transition (for a small protein)

**Remark:** A sigmoidal dependence of  $H^0$  in T (or in any other observable parameter), or equivalently a peak dependence of  $C_P$  in T, shows that the transition is cooperative, but only suggests that it is all-or-nothing.

Indeed, it can be either all-or-nothing (e.g. protein denaturation) or gradual (e.g. helix-> coil transition ) as seen in the example below :



## Protein folding is an all-or-none transition (for a small protein)

- ⇒ A peak in  $C_p$  (or a sigmoidal dependence in T of H in the form of heat transfer) is not sufficient to show that denaturation is all-or-none
- ⇒ Indeed, it may involve only a part of structure, a "(un)folding unit".
- ⇒ What is done to detect whether or not a transition is all-or-nothing

Boltzmann law: Probability of occurrence of state D ( $P_D$ ) is given by:

$$P_D = \frac{\exp[-G_D^0 / RT]}{\sum_{i=\text{all states}} \exp[-G_i^0 / RT]} \approx \frac{\exp[-G_D^0 / RT]}{\exp[-G_D^0 / RT] + \exp[-G_N^0 / RT]} = \frac{1}{1 + \exp[\Delta G^0 / RT]}$$

where  $\Delta G^0 = G_D^0 - G_N^0$

$$\Rightarrow \frac{dP_D}{dT} = P_D(1 - P_D) \frac{\Delta H^0}{RT^2}$$

$\left\{ \begin{array}{l} \text{because } S = -\frac{dG}{dT} \Big|_{P \text{ cte}} \\ \frac{d(G / RT)}{dT} = \frac{dG / dT}{RT} - \frac{G}{RT^2} = \frac{-S}{RT} - \frac{H - TS}{RT^2} = \frac{-H}{RT^2} \end{array} \right.$

Middle of the transition:  $\Delta G^0=0$ ,  $P_D=1/2 \Rightarrow T_m = \Delta H^0 / \Delta S^0$

$$\Rightarrow \frac{dP_D}{dT} \Big|_{T=T_0} = \frac{\Delta H^0}{4RT_m^2}$$

$$\approx \frac{\Delta P_D \Big|_{\text{extrapolated}}}{\Delta T} = \frac{1}{\Delta T} \quad (\text{graphical resolution})$$

$$\Rightarrow \Delta H^0 = \frac{4RT_m^2}{\Delta T}$$

$T_0, \Delta T$  calculated from the form of the transition curve

## Protein folding is an all-or-none transition (for a small protein)

On the other hand: measure the heat absorbed in the calorimeter:

$$\Delta q_{TOT} / N$$

where  $\Delta q_{TOT}$  is the heat absorbed by the set of N molecules

So:

- when  $\Delta H^0 = \Delta q_{TOT} / N$ : denaturation of the whole protein is all-or-none
- when  $\Delta H^0 < \Delta q_{TOT} / N$ : denaturation unit is smaller than the whole protein
- when  $\Delta H^0 > \Delta q_{TOT} / N$ : denaturation unit is larger than the whole protein – i.e. it is an aggregate of protein molecules rather than a single protein

This is van 't Hoff criterion

This is the hot denaturation (melting). The existence of cold denaturation has been shown (for some proteins - for all? ; sometimes water freezes first, so difficult to show).

- Cold denaturation is also a all-or-none transition.
- Due to the hydrophobic interactions that become less/not favorable at low T

$$!!! \quad \Delta G(T) = \Delta H(T) - T \Delta S(T) \quad !!!!!$$

## How does protein folding proceed ?

### Levinthal paradox (1967)

Estimation of  $N$ , the number of conformations accessible to a protein

Say :  $L = 150$  residues, and counting  $n \approx 3$  conformations per residue

$\Rightarrow N$  is of the order of:  $N = n^L \approx 3^{150} \approx 10^{72}$

(we can take  $n = 10$ , the conclusion remains unchanged)

Say: duration of elementary transitions is of the order of  $\tau \approx 10^{-12}$  sec

$\Rightarrow$  the time required to sample all conformations is:  $t = N \tau \approx 3^{150} 10^{-12}$  sec  
 $\approx 10^{48}$  years

But real folding takes :  $t_{\text{real}} \approx 1$  msec to 1 sec!

$\Rightarrow$  Apparent paradox

Number of conformations sampled by a natural protein during refolding is of the order of:  $N_{\text{real}} = t_{\text{real}} / \tau \approx 10^9-10^{12} \ll 10^{72}!!$

## How does protein folding proceed ?

Levinthal calculation is an overestimate :

- it is unrealistic that a protein must adopt all possible conformations before finding its native state – more realistically, it can perform a random walk in the conformational space towards its final structure
- → takes less time but it does not solve the paradox: it is too long compared to the time observed.
  - Resolution of the paradox (by Levinthal by itself, in the paper where he presented his paradox) : there are (one or several) folding pathway(s)

If there are multiple pathways: possibility to have key points where all pathways cross  
For example, formation of certain secondary structures, or some contacts.

## How does protein folding proceed ?

Reversible process => essentially independent of initial conditions - otherwise denatured protein could not fold in vitro -

This excludes the hypothesis that the correct folding is achieved during the synthesis of the polypeptide chain on the ribosome

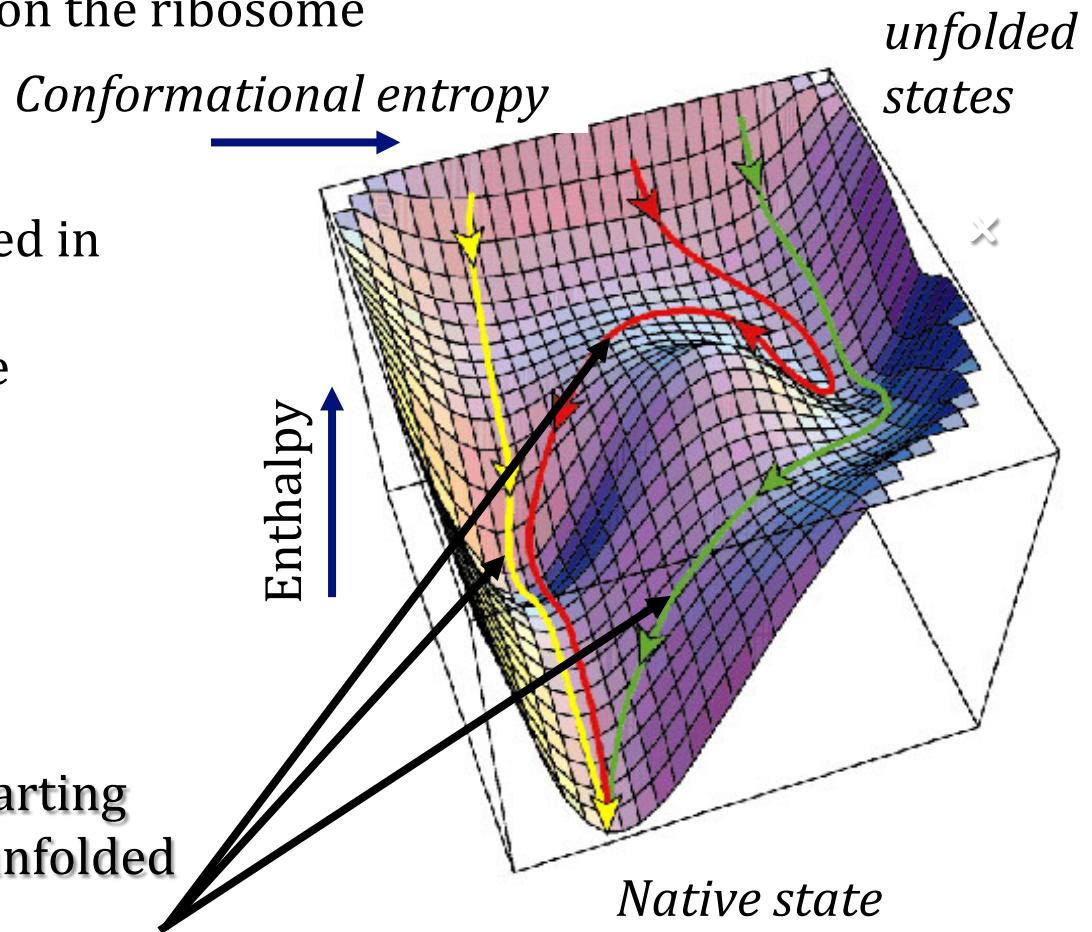
Folding funnel

conformational space is represented in the form of a funnel.

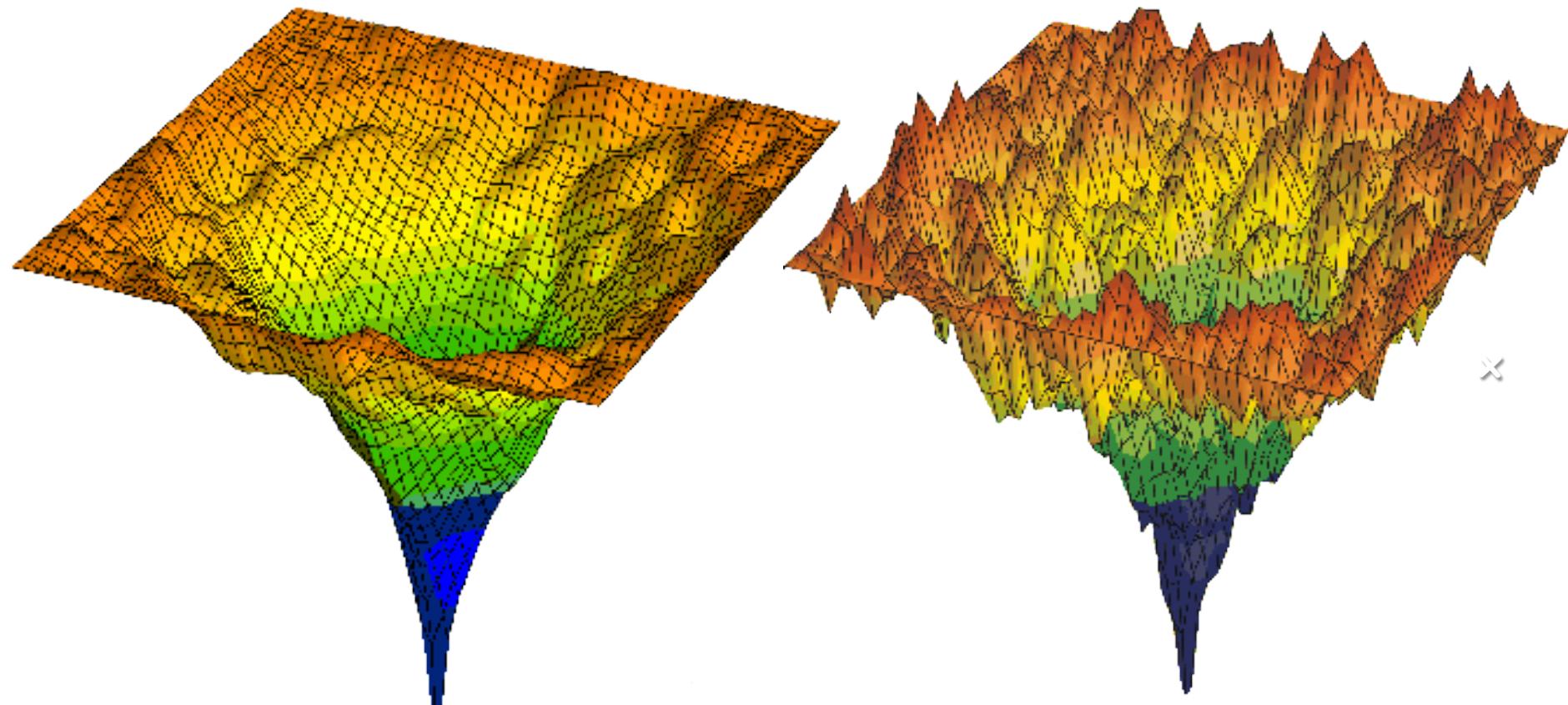
=> Ensures independence from the initial conditions

=> The number of conformations sampled is a small fraction of the possible conformations.

folding paths starting from different unfolded conformations



## How does protein folding proceed ?



The folding funnel of a given protein can vary according to the environmental conditions - here solvated and unsolvated (simulated).

So the folding pathway and possible folding intermediates may depend on the environment.

## Models of the protein folding mechanism

These different models are not necessarily exclusive or completely different.  
They may depend on the environmental conditions, type of protein, protein size,  
....

### ● **Nucleation-condensation model** (Zimm & Bragg, 1959)

Folding proceeds in 2 stages: a nucleation stage followed by a rapid propagation of the structure. The time limiting step is the nucleation process. Takes into account the cooperativity of folding.

### ● **Hierarchical model** (Baldwin, 1975; Schulz, 1977)

Unique folding pathway with segments of the structure formed and assembled at different levels.

Hierarchy in folding: formation of the primary → secondary → tertiary → quaternary structures

### ● **Diffusion – collision model** (Karplus & Weaver, 1976).

Nucleation occurs simultaneously at several points along the polypeptide chain. The structures generated diffuse and coalesce to form microstructures having a native conformation. The folding is accomplished via several successive steps of diffusion / collision.

## Models of the protein folding mechanism

### ● **Folding in modules** (Wetlaufer, 1981; Chothia, 1984)

(Sub) structural domains are independent folding units that could fold independently and then come together to give the native structure

### ● **Framework model** (Ptitsyn, 1991)

Assumes that the secondary structure is formed at an early stage of folding before the 3D structure. The elements of secondary structure diffuse, collide and merge to form the tertiary structure.

### ● **Hydrophobic collapse model** (Levitt, Warshel, 1975; Dill 1990).

The first event of protein folding is a hydrophobic collapse taking place before the formation of secondary structures. This collapse leads to the folding of the protein and the stabilization of the 3D structure.

### ● **Hydrophobic zipper model** (Dill et al., 1993)

The formation of secondary structure segments occurs simultaneously with the hydrophobic collapse.

## Models of the protein folding mechanism

### ● Jigsaw Puzzle model (Harrison, Durbin, 1985)

Folding is seen as a puzzle, with multiple paths to the unique native solution.

Thus, no consensus between the different models...

### What are the folding intermediates?

Folding is usually "all-or-nothing" => no stable intermediates, which accumulate.



But there are unstable structural intermediates, which are transient (less stable if the free energy minimum is less deep). Not easy to detect because :

- often several possible folding pathways,
- the transient species are sparsely populated and in rapid equilibrium with other structures.

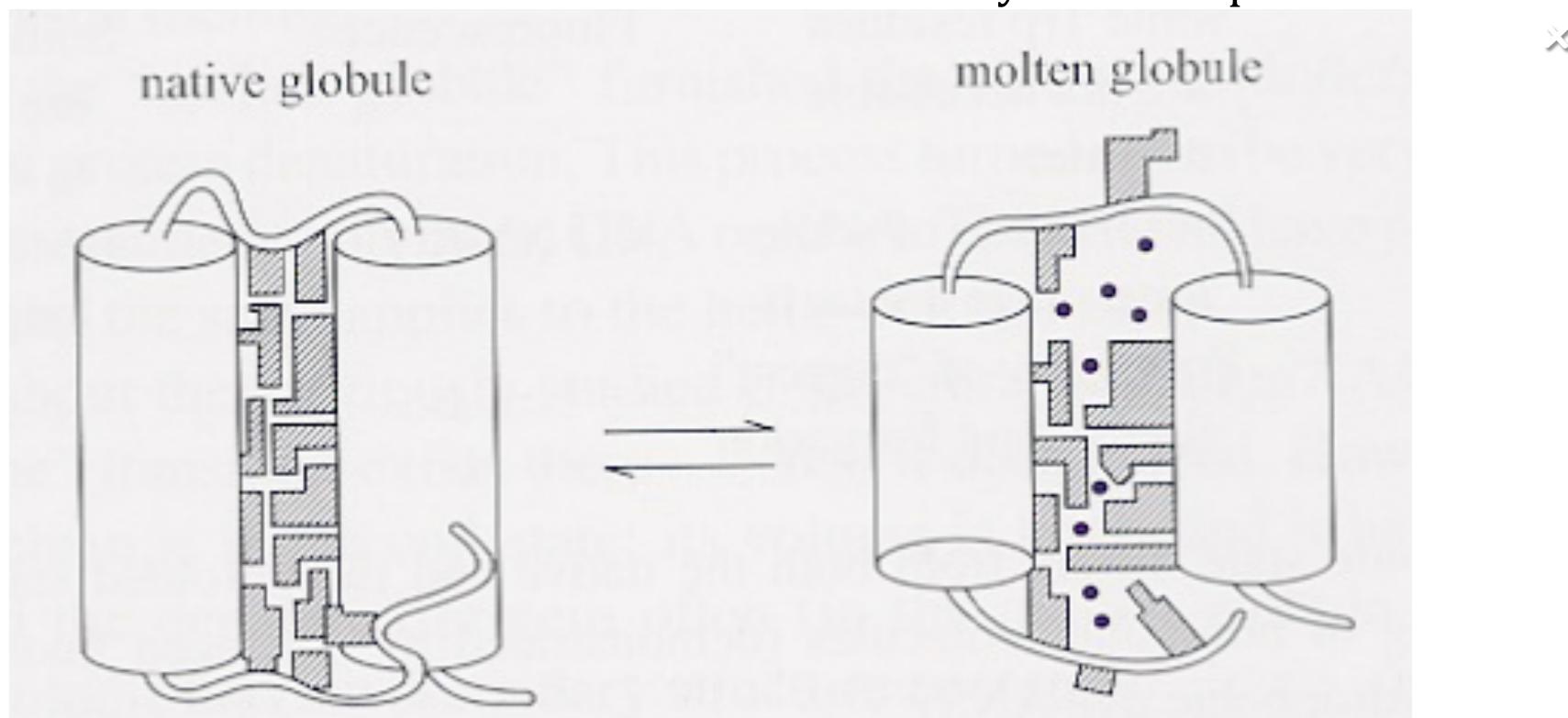
Sometimes possible to accumulate transient folding intermediates under specific denaturing conditions, e.g. at low pH.

=> existence of several refolding steps: an early stage of fast kinetics and a late stage of slow kinetics, responsible for structural rearrangements.

## Folding intermediates

Example of folding intermediate - observed in most proteins if denatured softly. Called molten globule

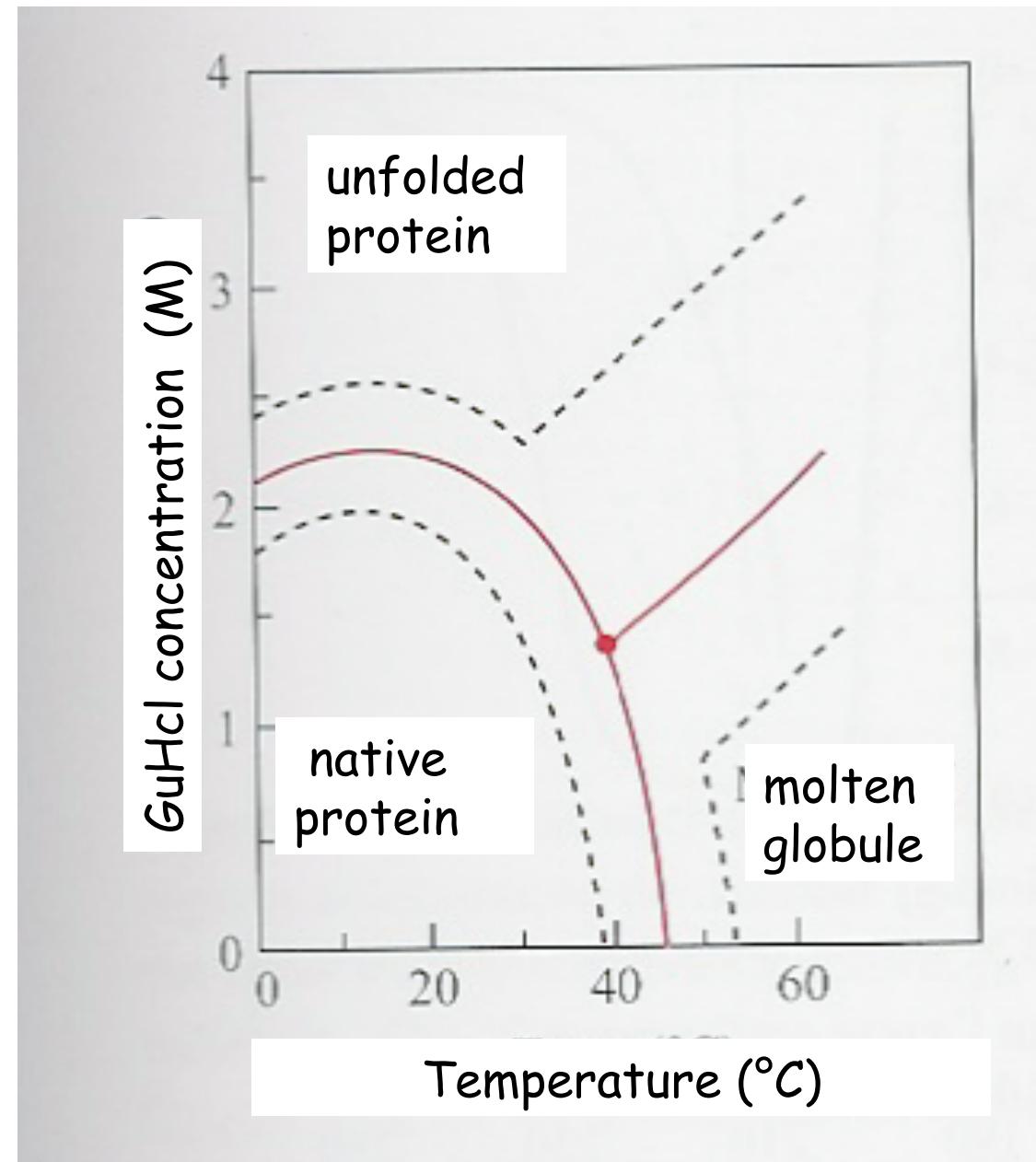
= Collapsed state on the pathway between native and denatured states - in some denaturation conditions - has largely the secondary structure of the native state - difference with the native state lies in less dense packing, presence of water molecules and increased mobility of the loops.



## Phase diagram

Phase diagram of conformational states of lysozyme at pH 1.7 in a solution of denaturant (guanidine dihydrochloride).

Transition unfolded - molten globule is wider (= less cooperative) than the other transitions

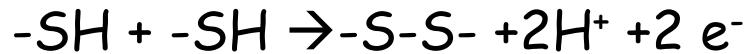


## Experimental techniques to detect folding intermediates

The oldest method : uses disulfide bridges as a probe.

Principle: the interactions between 2 Cys towards disulfide bridge formation/breakage is a redox process:

oxidation:



reduction:



⇒ the formation or breakage of these bridges requires the presence of electron acceptors or donors, respectively. By varying the number of these, it is possible to experimentally control the process by quenching the formation or breaking of the disulphide bridges and thus trapping possible intermediates.



But: restriction of the method to small proteins with disulphures bridges - and the possibility to build non-native disulfide bridges, possibly off-pathway.

Generally, the most productive pathways involve first the formation of disulphide bridges that are buried in the folded structure, and then those at the surface.

# Experimental techniques to detect folding intermediates

Folding of  
BPTI  
(bovine  
trypsin  
inhibitor) at  
pH 8.7 -  
Approximate  
structure  
determined  
by NMR -  
thickness of  
the arrows  
indicates the  
rate.



Off-pathway folding intermediate are observed

## Experimental techniques to detect folding intermediates

### Fast mixing techniques

- 1) Protein in denaturing (renaturing) conditions
- 2) Gradual dilution of the protein solution that contains high (low) concentrations of denaturant - or gradual change of pH - or gradual increase in T

So gradual change from denaturing to renaturing conditions (or the converse)

Stop folding => possible folding intermediates are trapped.

Possibility of several consecutive stops.

=> intermediates along the unfolding or folding pathway.

The structural characterization of these trapped intermediates is done by various techniques: NMR, fluorescence, infrared spectroscopy, CD, and mass spectroscopy.

=> gives information on the structures formed after a time of the order of 1 nsec to 1 msec

# Experimental techniques to detect folding intermediates

## Fast mixing techniques

Example coupled to CD: applied in 1984 to the folding of ribonuclease S - then to other proteins.

In most experiments, a burst phase is observed, which takes place after less than  $\sim 1$  msec folding – contains already a significant proportion of secondary structure.

Example coupled to infrared spectroscopy:

- Applied to the unfolding of ribonuclease A :  $\beta$ -sheet seems to break after the intrusion of water molecules in hydrophobic core.
- Applied to the folding of a small helical peptide. Clear results: the helix forms in about 10 nsec. The helix-unfolded state interconversion is very fast on the time scale of the formation of tertiary contacts.

## Experimental techniques to detect folding intermediates

### Fast mixing techniques

Example coupled with fluorescence: apomyoglobin.

Protein solution was cooled to -10 ° C without freezing, and then heated with a laser pulse. The water absorbs the pulse almost instantaneously, and the solution is heated to about 20 ° C in a few nsec. This initiates protein folding.

It is observed that the apomyoglobin collapses to form a compact globule after a few msec, with some helices already formed.

→ in this case significant effect of local interactions along the chain.

These results are consistent with the idea that (some of) the secondary structure forms before tertiary contacts (at least for some proteins under certain conditions).

In addition, it does not seem necessary that a peptide sequence has a high propensity for helical conformation in aqueous solution to serve as a nucleation site, because the time scale for helix formation is much shorter than the time required to form tertiary conformations. These helices that form and unfold must, however, be stabilized at a given time by tertiary interactions.

## Experimental techniques to detect folding intermediates

### Fast mixing techniques coupled to H<sup>+</sup> exchange

*Labeling method which uses the exchangeable H<sup>+</sup> of the protein as a probe of the structural change.* It exploits the exchange reaction of the amide protons (NH groups) of the backbone with solvent protons, and the fact that this reaction is much slower in the folded than in the unfolded state, because in the folded protein these protons are e.g. involved in H-bonds in secondary structures or inaccessible to solvent.

⇒ a proton becomes resistant to exchange with the solvent in a given stage of the folding process, due to its participation in a secondary structure or because it becomes buried in the core.

To measure the time when a given proton becomes resistant to exchange, the protein is unfolded in a denaturing solution containing D<sub>2</sub>O (deuterium = 1 proton + 1 neutron)  
→ all the amide protons that can exchange become deuterated. Then, folding is re-initiated (by e.g. dilution of the denaturant), and proceeds during a certain time, e.g. ~ 10 msec.

After this time, the partially folded protein is exposed to a labeling pulse of H<sub>2</sub>O, which leads to the protonation of the amide groups in the protein regions that are still unstructured. After that, the protein is allowed to fold completely.  
Finally the presence of H<sup>+</sup>/D<sup>+</sup> is analyzed.

## Experimental techniques to detect folding intermediates

### Fast mixing techniques coupled to H<sup>+</sup> exchange

The first H<sup>+</sup> exchange experiments coupled with NMR were performed in 1988 on ribonuclease A and cytochrome C.

In ribonuclease A, stable secondary structure elements are formed within 1.5 sec refolding.

In cytochrome C, the observed intermediates appear much earlier in the folding process: they are trapped after 20 msec. In this intermediate, N-and C-terminal helices are already formed. As these helices are in contact in the native structure, they are probably already present in the intermediate state.

These techniques have been refined by varying the pH of the labeling pulse, which alters the rate of exchange of H<sup>+</sup>.

An important limitation of these techniques is that only the amide protons that are protected from exchange in the native structure can be analyzed.

Moreover, it is impossible to determine by this method that a region of the chain adopts a non-native structure in the early stages of folding.

## Experimental techniques to detect folding intermediates

### Fast mixing techniques coupled to H<sup>+</sup> exchange

H<sup>+</sup> ↔ D<sup>+</sup> experiments can be coupled to NMR spectroscopy but also to mass spectroscopy - these two techniques provide complementary information.

Mass spectroscopy: in the case of H<sup>+</sup> ↔ D<sup>+</sup> exchange, the property that the deuterated protein molecules have a larger mass is exploited.

This technique does not give specific information of the structure at each site (unlike NMR); rather allows us to distinguish situations in which e.g. all the amide protons exchanged:

- completely in 50% of the proteins and not in the other 50%
- with a probability of 50% in all proteins.

Mass spectroscopy proceeds by sending ionized proteins through a magnetic field and measuring the deviation (dependent on mass) they undergo -> separation of proteins according to their mass.

## Experimental techniques to detect folding intermediates

### Analysis of protein fragments/peptide models

- Cut protein in pieces
- Analyze of the intrinsic stability (that is marginal!!) of the fragments in solution – at low T or in an organic solvent such as TFE (trifluoroethanol) (~ hydrophobic environment mimicking the interior of a protein)

Examples:

- Myohemerythrin is a completely helical structure. Several peptides tested preferentially adopt a helical structure in solution.
- For plastocyanin, a  $\beta$ -sandwich structure, the peptides tested showed relatively few secondary structures in solution → folding is not initiated by the formation of secondary structures in this case.

Experiments in which the sequence is circularly permuted: often, the modified protein reaches its final structure

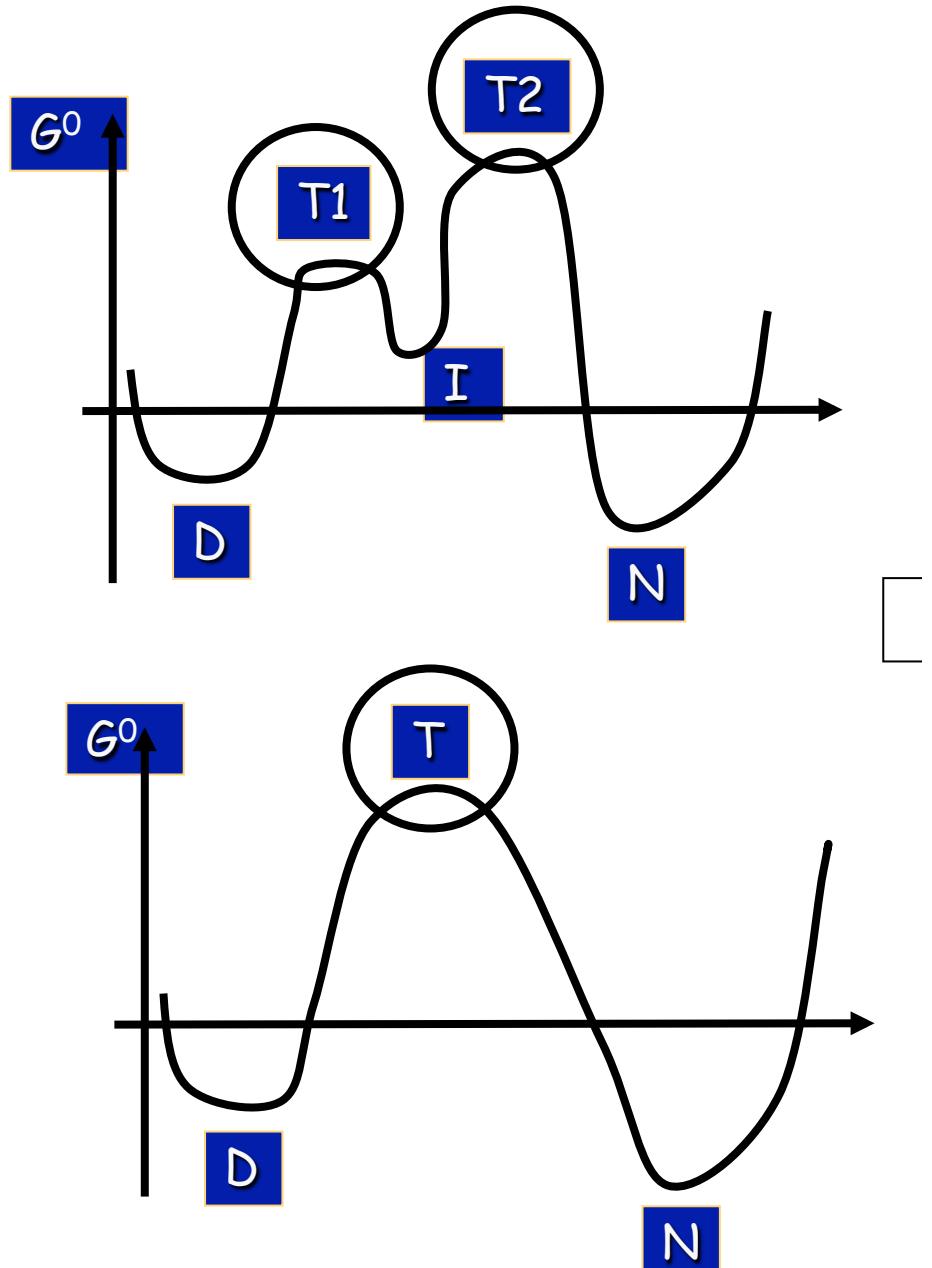
→ eliminates folding mechanisms where structure formation propagates along the entire polypeptide chain. But this result is perhaps not general!

## Experimental techniques to detect transition states

Impossible to use the same type of experiments to characterize transition states and intermediate states because:

- Intermediates correspond to free energy minima, so they are detectable directly - even if they are only transient (shallow minimum)
- Transition states correspond to a maximum => impossible to trap. They are characterized through kinetic measurements.

We consider here the simple case without intermediates - but can be generalized



## Experimental techniques to detect transition states

### Protein engineering technique

Provides information on the interactions formed in the transition state - but only for one interaction at a time!

Principle:

- Specify an interaction for which you wish to know whether or not it is formed in the transition state.
- Mutate one of the residues that form this interaction in order to break it - but the mutation must not be too disruptive to prevent altering the native structure.
- This is followed by denaturation or renaturation experiments (by heat or denaturant) to measure the stability and reaction rates of wildtype and mutant proteins.

## Experimental techniques to detect transition states

Protein engineering technique

Renaturation experiments

**Thermodynamic measurements** (performed at equilibrium):  
the difference between the folding free energy of the wildtype and mutant proteins  $\Delta\Delta G_{D-R}$  is measured from the equilibrium constants K:

$$\Delta\Delta G_{D \rightarrow N}^0 \equiv \Delta G_{D \rightarrow N}^0(\text{wt}) - \Delta G_{D \rightarrow N}^0(\text{mut}) = -RT \ln(K_{D \rightarrow N}^{\text{wt}} / K_{D \rightarrow N}^{\text{mut}})$$

### **Kinetic measurements:**

the protein in a denaturing solution is rapidly diluted in native solutions. The rate constants k of folding give the change in activation free energy  $\Delta\Delta G_{D-T}$  between mutant and wildtype proteins :

$$\Delta\Delta G_{D \rightarrow T}^0 \equiv \Delta G_{D \rightarrow T}^0(\text{wt}) - \Delta G_{D \rightarrow T}^0(\text{mut}) = -RT \ln(k_{D \rightarrow T}^{\text{wt}} / k_{D \rightarrow T}^{\text{mut}})$$

- Same thing can be done in denaturing rather than renaturing conditions:  
 $\Delta\Delta G_{N-T}$  instead of  $\Delta\Delta G_{D-T}$
- and the procedure can be generalized in the presence of intermediates

## Experimental techniques to detect transition states

### Protein engineering technique

We define:

$$\Phi \equiv \Delta\Delta G_{D \rightarrow T}^0 / \Delta\Delta G_{D \rightarrow N}^0$$

Hypothesis: the mutation does not affect the unfolded state (not always true)

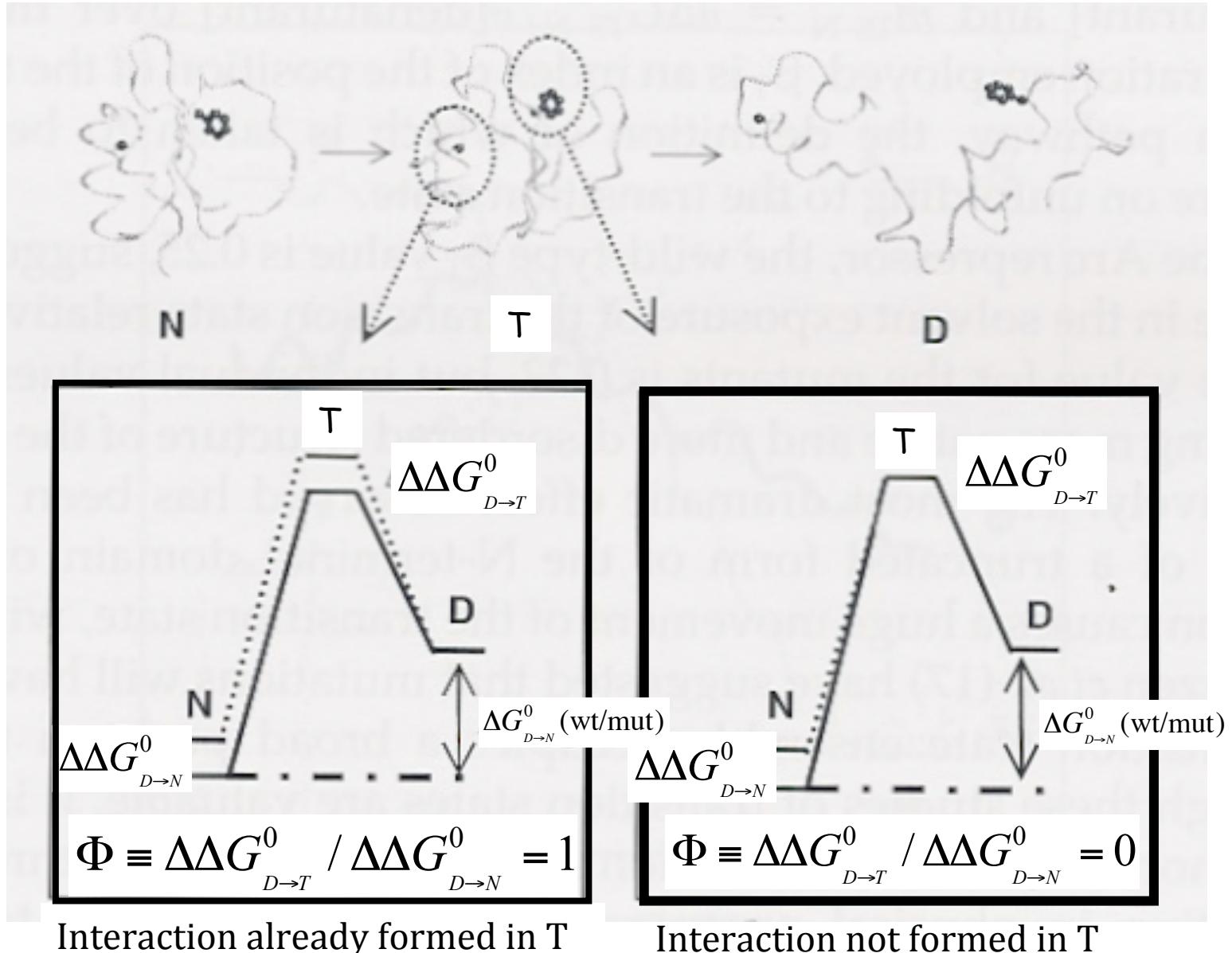
But by construction: it always affects the native state (breaks an interaction)  
and sometimes the transition state

Two extreme cases:

- 1) The mutation affects both the native and transition states =>  $\Phi = 1$   
-> the interaction studied is already formed in the transition state.
  - 2) The mutation only affects the native state =>  $\Phi = 0$   
-> the interaction studied is not formed in the transition state.
- + all intermediate cases where  $0 < \Phi < 1$

# Experimental techniques to detect transition states

Protein  
engineering  
technique



## Experimental techniques to detect transition states

### Protein engineering technique

For example:

- If the mutation of a residue in a helix destabilizes by the same amount both T and N states
  - =>  $\Phi = 1$  and the helix is already formed in state T.
- If the mutation destabilizes N but not T:
  - =>  $\Phi = 0$  and the helix is unfolded in the state T.

Often fractional  $\Phi$  values are found. They are difficult to interpret because they can have different origins:

- weakened interactions;
- parallel folding pathways leading to mixtures of T states.

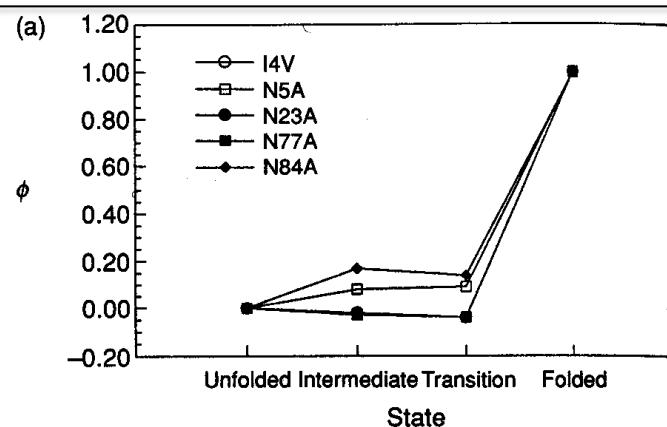
But fortunately, in practice, the simple cases  $\Phi \approx 1$  and  $\Phi \approx 0$  are the most common outside the hydrophobic core

⇒ Most structural elements are folded either completely or not at all.

Two sequences adopting the same structure do not always fold along the same pathway and do not always have the same I and T states.

# Experimental techniques to detect transition states

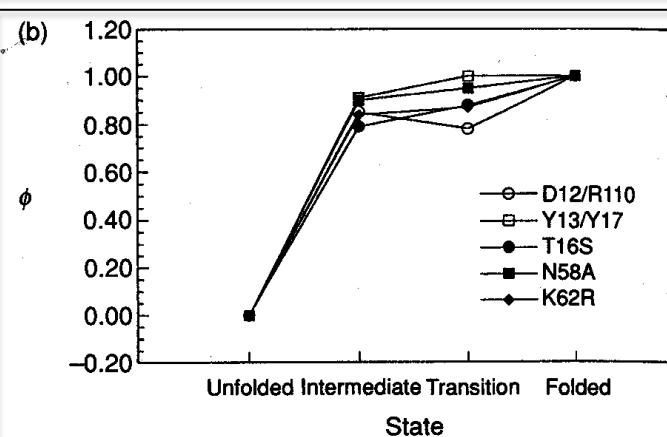
## Protein engineering technique



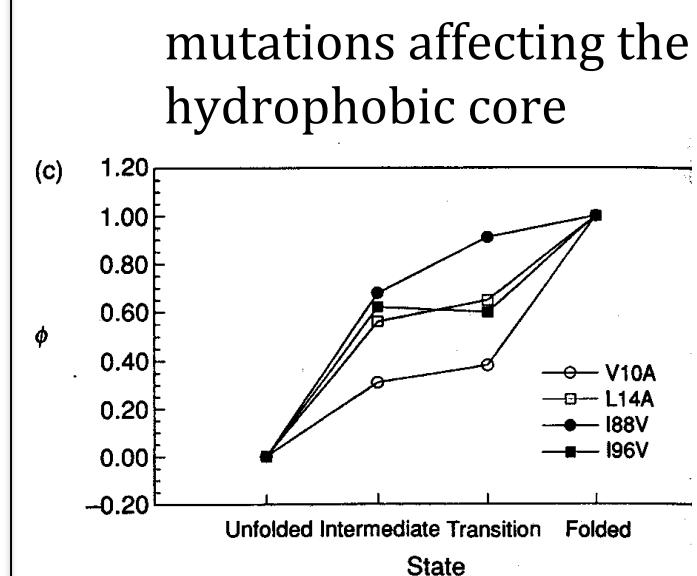
mutations  
affecting  
surface  
loops

Example :

$\Phi$  In barnase,  
for mutations  
in 3 different  
sites



mutations  
affecting  $\alpha$ -  
helices



mutations affecting the  
hydrophobic core

The results do not favor a single model, but rather show that different proteins fold differently (with however common points).

For example:

- *Chymotrypsin inhibitor 2* : 2-state behavior, all-or-nothing. By the methods of protein engineering and peptide models, it has been shown that the nucleus is only weakly formed in the D state and forms in the T state, and that the secondary and tertiary structures develop in parallel.

⇒ Folds according to the hydrophobic collapse model, rather than according to models with preformed secondary structures. Assumed to be a general feature of small proteins with no I states.

- *Apomyoglobin* : folding has a kinetic folding intermediate that is compact, contains some helices, is formed on the time scale of 1 msec, and is preceded by the formation of helices

⇒ Folds according to a model with preformed secondary structure elements.

In addition, the folding mechanism may depend on the experimental conditions.

## The protein structure is robust with respect to changes in amino acid sequence

Usually, two proteins adopt roughly the same structure if they have 20-25% sequence identity.

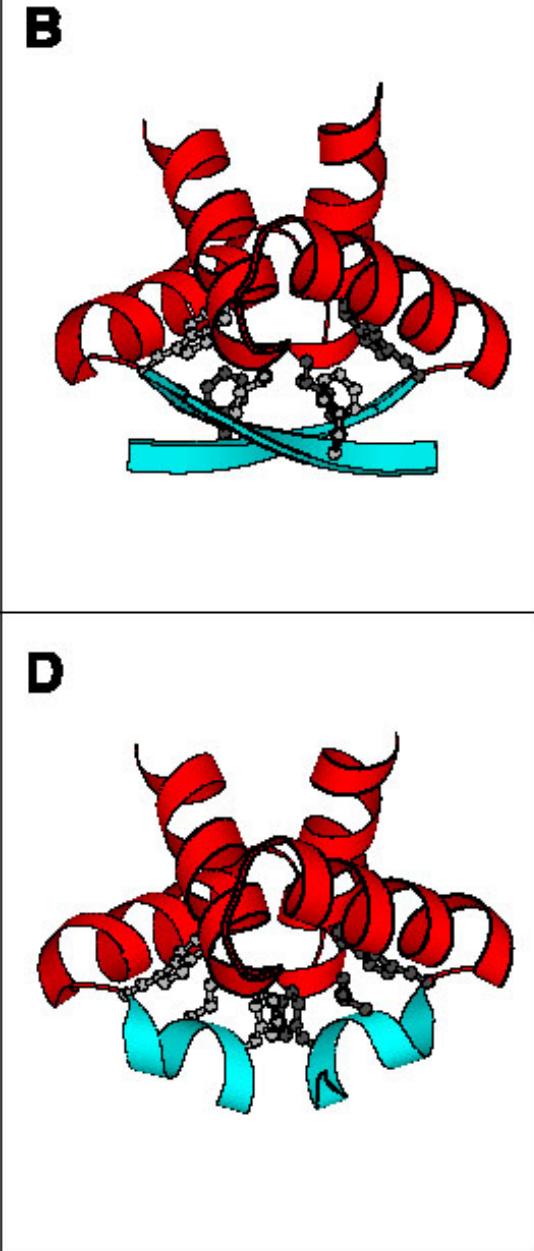
Exceptions... sometimes a single mutation modifies the structure (cf conformational illnesses).

For example : ARC repressor

Wildtype protein:  $\beta$ -sheet

Double-site mutant Asn 11  $\rightarrow$  Leu, Leu 12  $\rightarrow$  Asn :  
 $\alpha$ -helix

Single-site mutant Asn 11  $\rightarrow$  Leu : hesitates between the 2 structures :  
Evolutionary bridge between the 2 structures



## ***In vivo <-> in vitro folding***

### ***In vivo: chaperone proteins***

In vivo, proteins are experiencing during their whole existence problems with denaturation and aggregation:

- they are synthesized on the ribosome and must fold properly.
- cellular environment is so dense that the proteins can easily form interactions with other molecules and not refold
- they can unfold due to fluctuations of T and chemical composition.

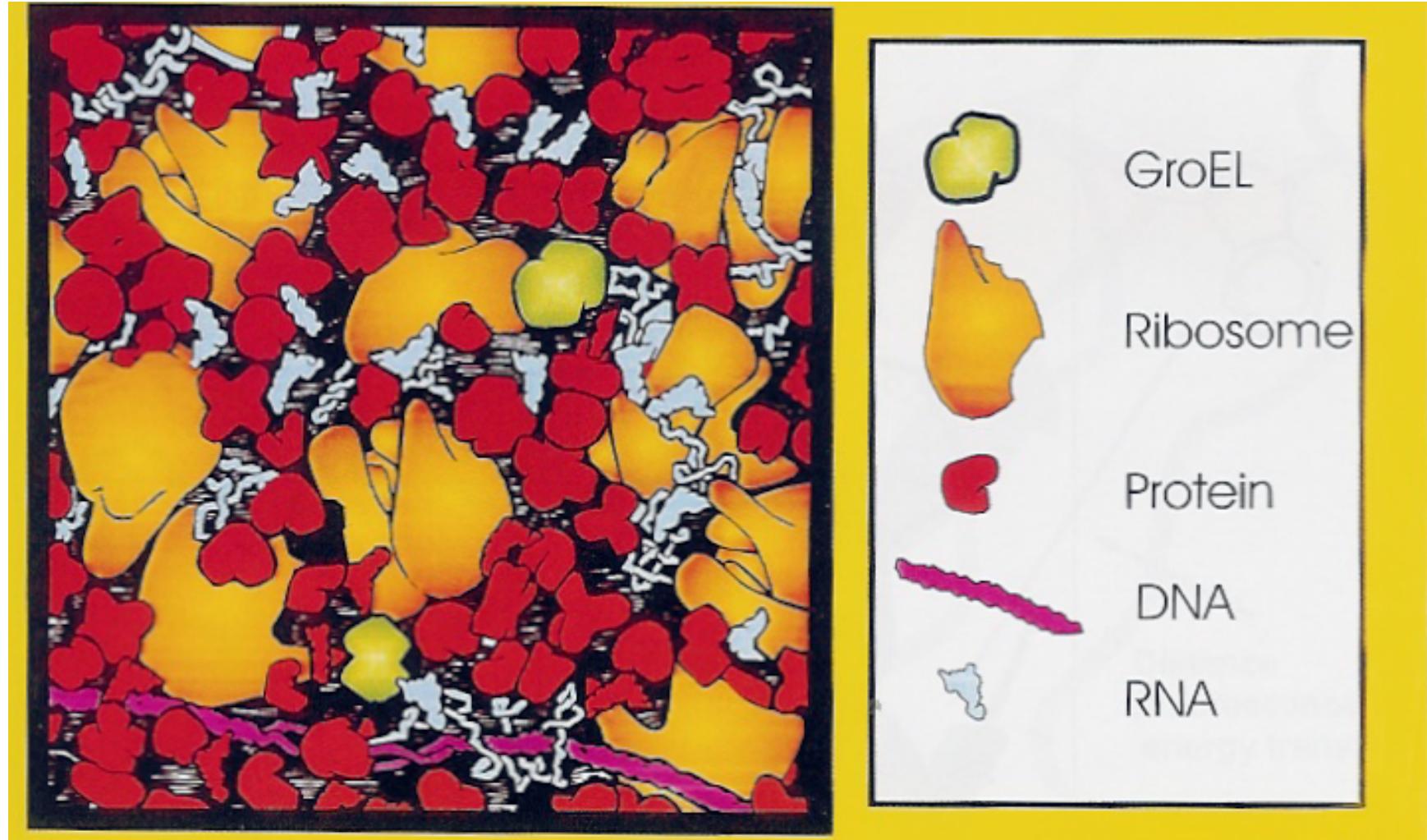
⇒ Development in the cell of "folding machines" called chaperone proteins.

General definition: any protein that interacts transiently with, and stabilizes, an unstable conformation of another protein and facilitates its folding, its oligomeric assembly, its interaction with other cellular components, or its intracellular transport.

Example of a protein "folding machine" : GroEL. It is a chamber formed of flexible subunits that bind to proteins that are in non-native states.

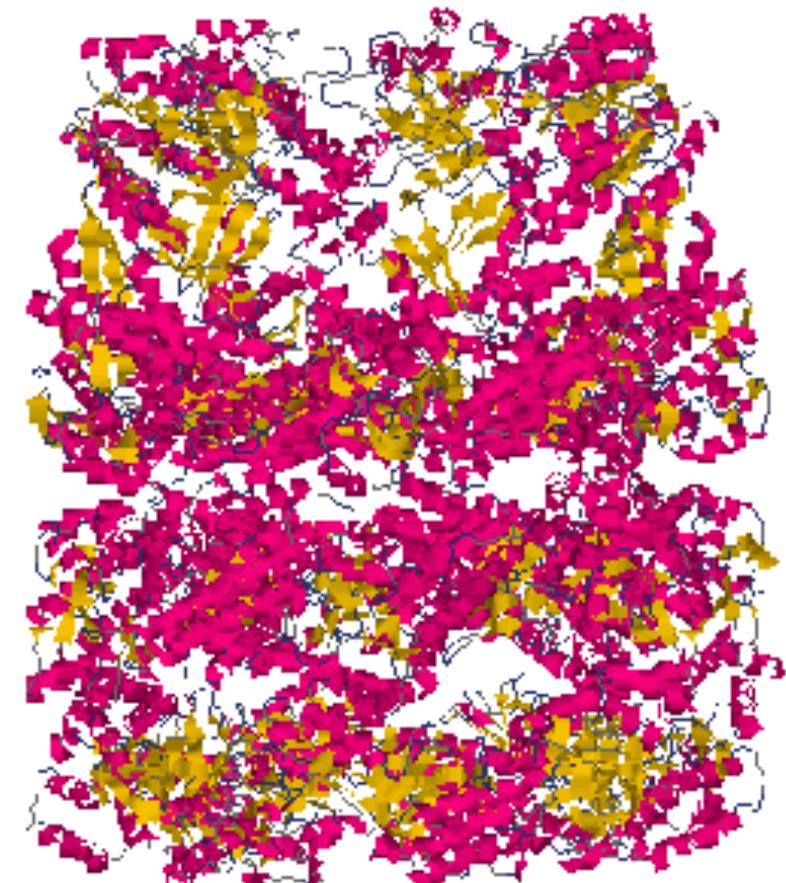
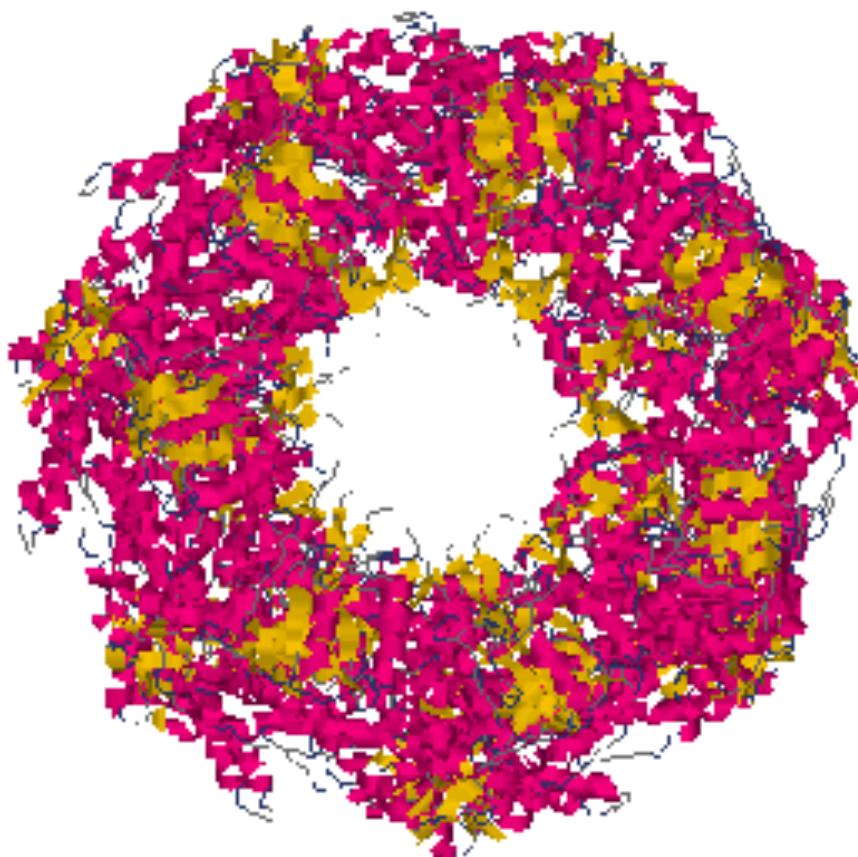
## *In vivo: chaperone proteins*

In vivo: difficulties to fold properly because of the large concentration in the cellular medium. Indeed :



## *In vivo:* chaperone proteins GroEL

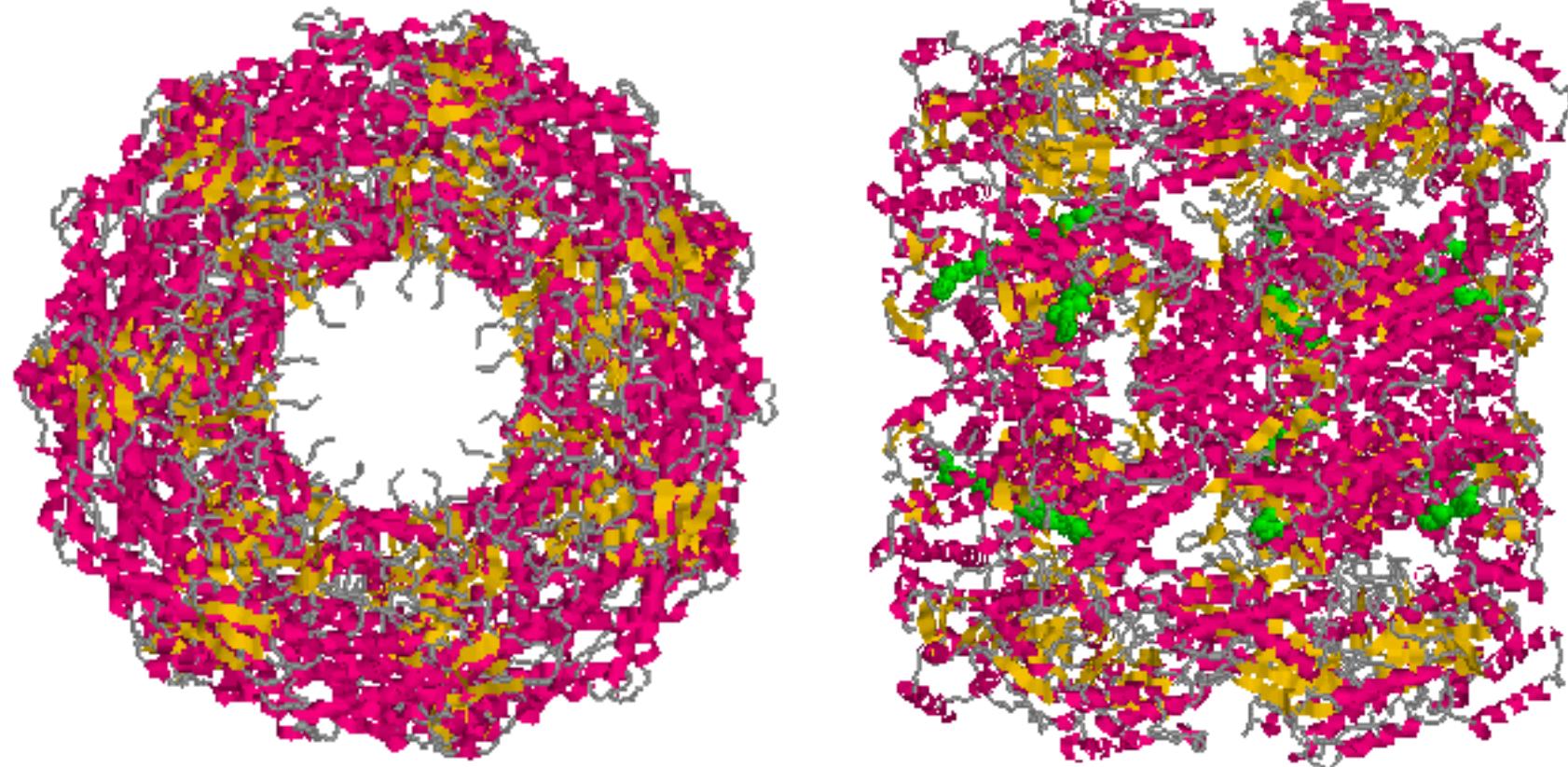
GroEL in its inactivated state: tetradecamer of identical subunits



## *In vivo:* chaperone proteins GroEL

GroEL binds strongly to denatured proteins due to the high affinity of its hydrophobic pocket for hydrophobic side chains => facilitates the total unfolding of the protein upon binding.

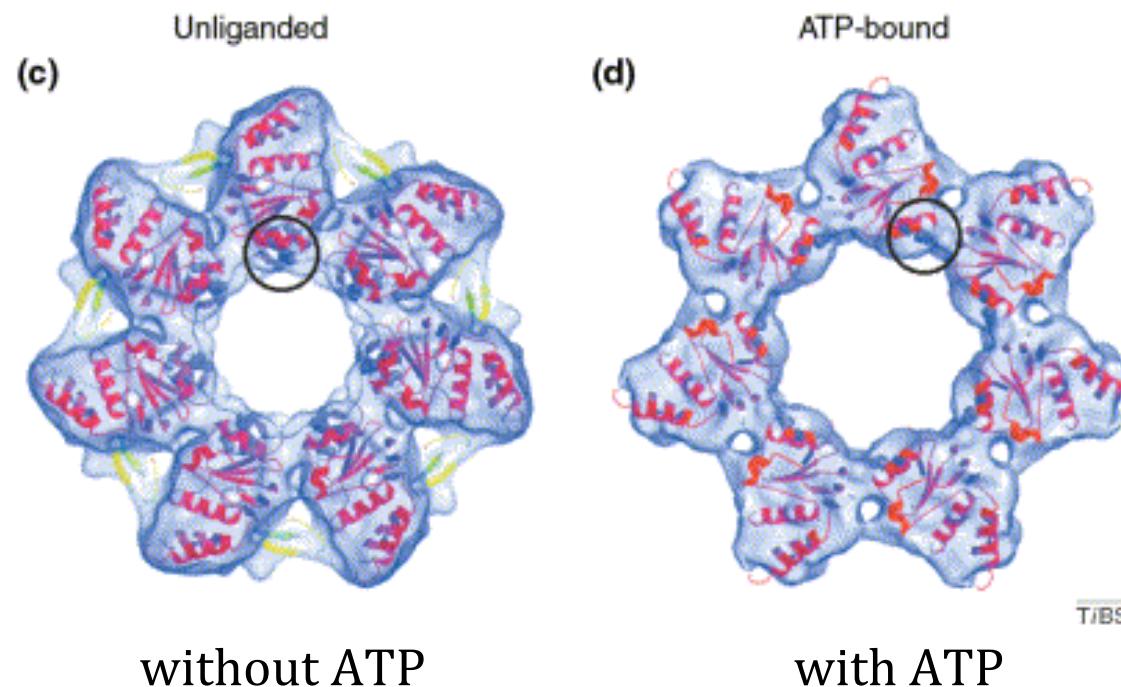
A number of ATP molecules (green) bind to GroEL, initiating a conformational change that extends the cylinder:



## *In vivo:* chaperone proteins GroEL

Another view of the conformational change, looking inside the GroEL cylinder

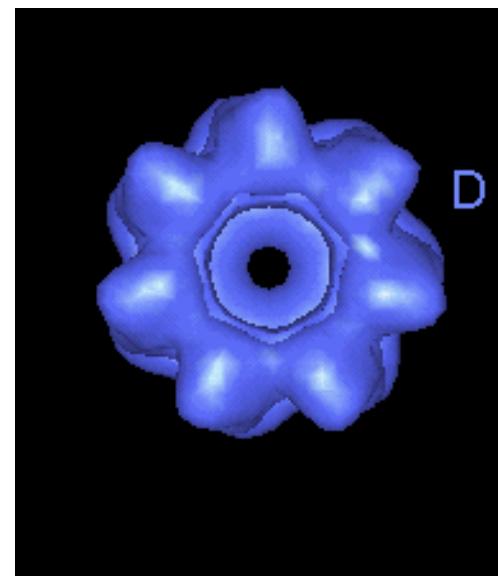
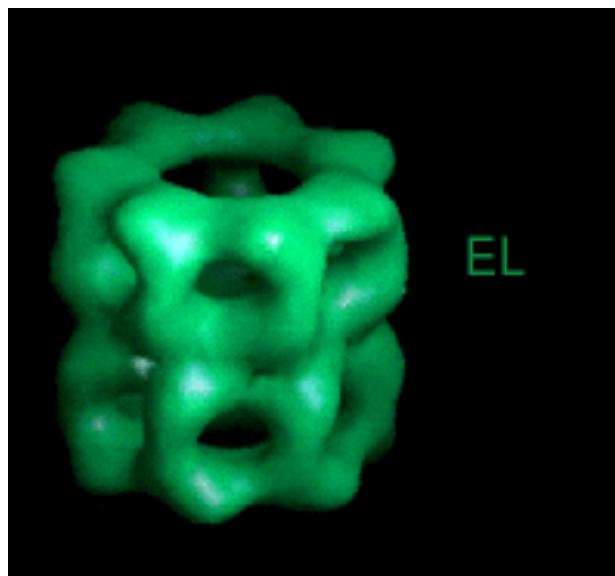
Substantial changes in the shape of the hole and in the outer surface  
modifications in the contacts between domains



## ***In vivo:* chaperone proteins GroEL**

Then: second chaperone protein, GroES, binds to one end of the GroEL cylinder, producing a hollow structure → state D.

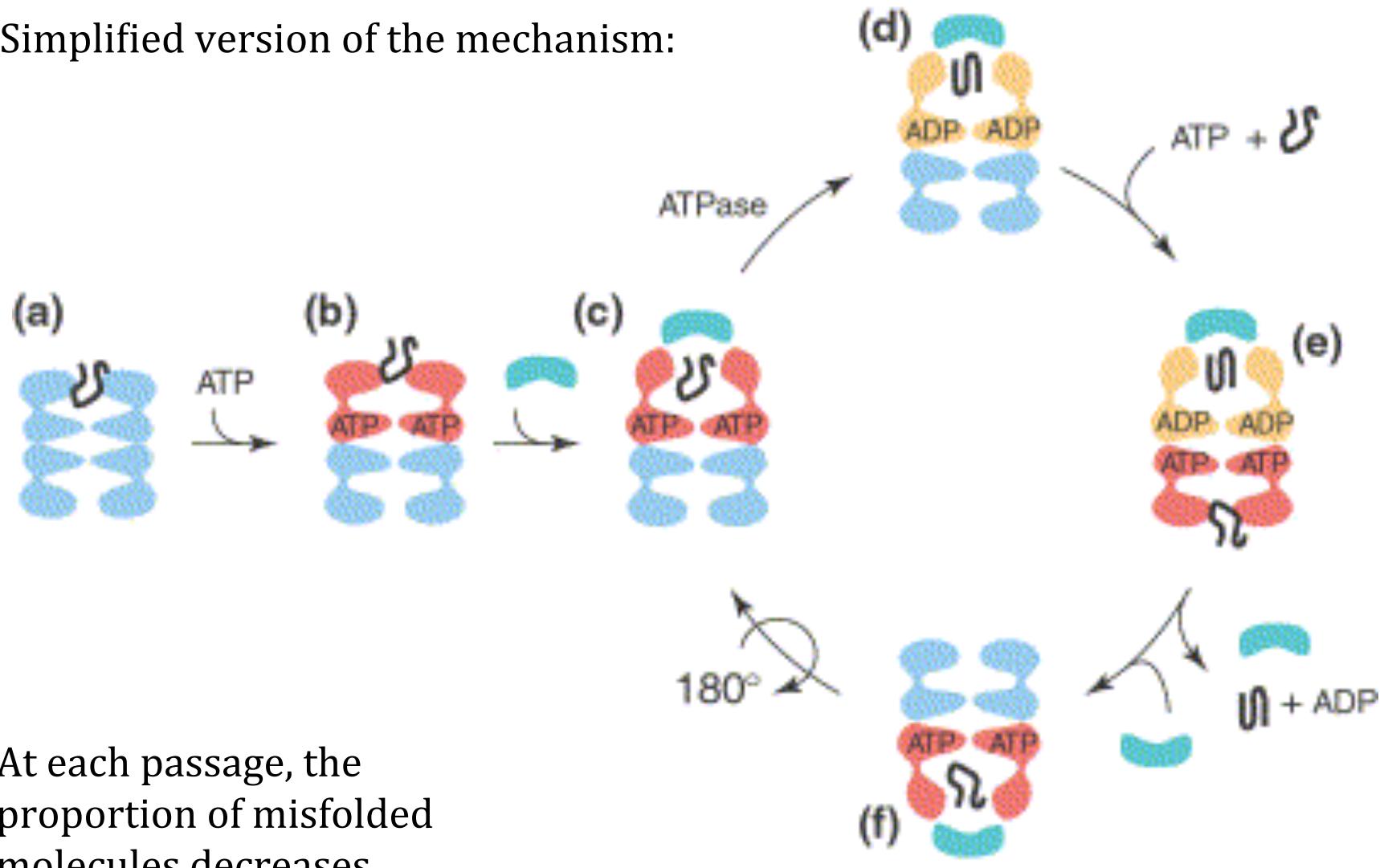
- The combination of ATP and GroES causes the expansion of the cavity of GroEL
- The conformational change also changes the residues on the inner surface of the cavity : hydrophobic → hydrophilic
- The protein folds in this in vitro-like environment
- The protein is then expelled



=> GroEL gives the misfolded proteins a 2<sup>nd</sup>, 3<sup>rd</sup> chance to get correctly folded

## *In vivo:* chaperone proteins GroEL + GroES

Simplified version of the mechanism:



At each passage, the proportion of misfolded molecules decreases.

## Protein folding in the cell - *in vivo*

Sequential process in which members of different families of chaperones bind to a protein during its biosynthesis and folding. Each chaperone interacts with a particular conformer of the protein.

When a protein leaves one chaperone protein, it is transferred to the next chaperone, and thus progresses along the folding process.

Multidomain proteins need chaperone proteins more crucially than small proteins to reach their native structure.

So: the structure is basically encoded in the sequence, but the protein may need help to reach it, - at least *in vivo*.

And yet, some proteins do not make it... => conformational diseases (see next part !)