

# **Protein Folding and Stability**

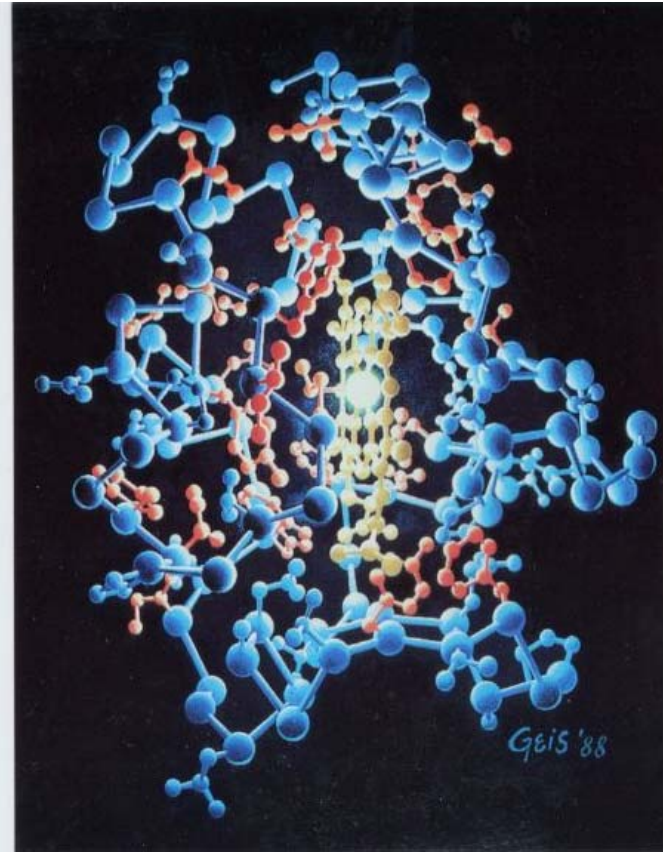
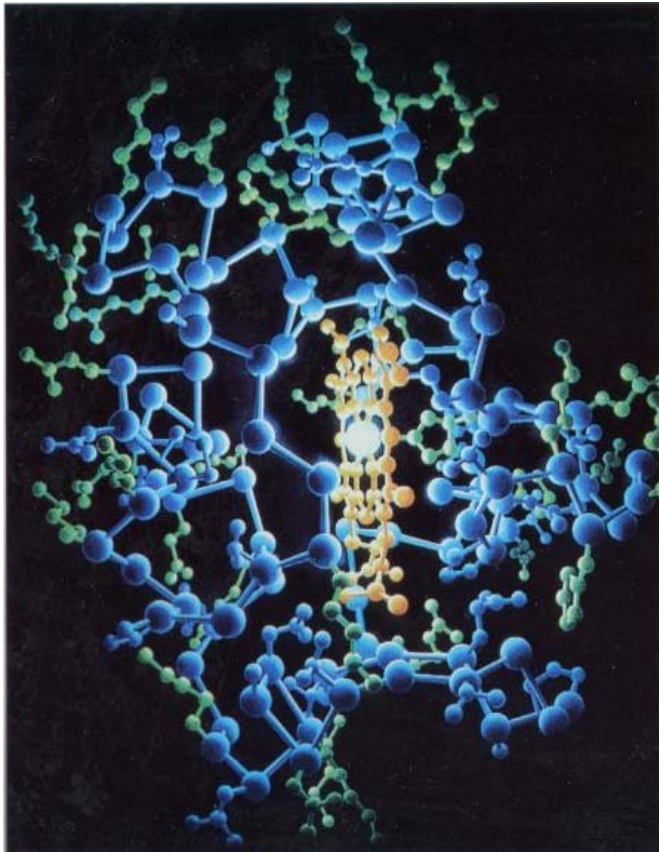
# Protein Folding and Stability

---

- 3D folding of the polypeptide chain involves the interaction of protein regions far apart in the sequence: **long-range, non-sequential contacts**. There are also adjacent interactions in secondary structures
- Structures are stabilized through many **weak interactions** (see table)
- Main source of stability: **Hydrophobic Interactions**
- Main destabilizing force: **Configurational Entropy** of Unfolded State
- Structures are marginally stable
- Protein stability depends on the **balance** between the **favorable entropy change** of folding due to **hydrophobic interactions**, and the **unfavorable entropy change** of folding due to the **loss of configurational freedom**.

# Protein Folding and Stability

<u>Interaction</u>	<u>Effect</u>	<u>Magnitude</u>	<u>Mechanism</u>
Electrostatic (Classical)	destabilizes	Small: $\sim(\text{net charge})^2$	Increases charge density of folded state
Ion Pairing	stabilizes	Intermediate: (15 Kcal/mol)	Confers structural specificity to folded state
Config. Entropy:	destabilizes	Large	Opposes folding into compact state
H-bonding:	stabilizes	Intermediate (short range)	Confers structural specificity to int. architecture
van der Waals	stabilizes	Small	Helps packing
Hydrophobic	stabilizes	Large	Drives formation of compact globular structures
Intrinsic Propensities	stabilizes	Small (important)	Helps attainment of 2° structure

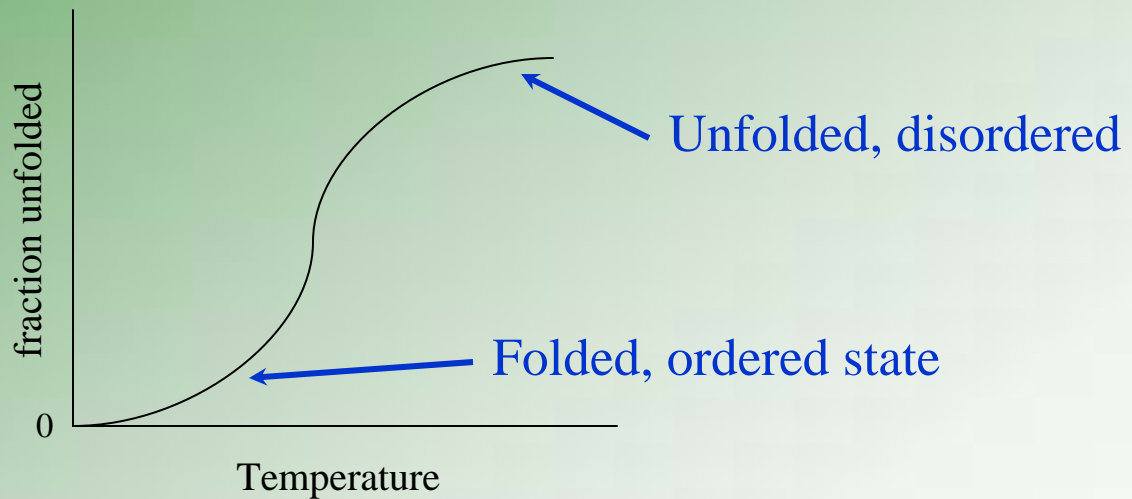


# Protein Folding and Stability

---

## Thermodynamics of Protein Folding:

The tertiary protein structure can be destabilized by temperature (thermal denaturation)



This process resembles the order/disorder transition observed in much simpler systems

An example...

# Protein Folding and Stability



At 0°C, the system is at equilibrium:

$$\Delta G^{\circ} = \Delta H^{\circ} - T_{\text{transition}} \Delta S^{\circ} \equiv 0$$

$$\Delta H^{\circ} = T_{\text{transition}} \Delta S^{\circ}$$

Interactions  
favoring order  
and Ice form

These are  
balanced

Interactions  
favoring disorder  
and liquid form

At -10°C, however, the ice is favored, so:



$$\Delta G^{\circ} = \Delta H^{\circ} - T \Delta S^{\circ}$$

$$\Delta G^{\circ} = 1343 \frac{\text{cal}}{\text{mol}} - 1292 \frac{\text{cal}}{\text{mol}} = + 51 \frac{\text{cal}}{\text{mol}}$$

# Protein Folding and Stability

---

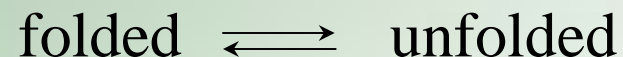
At +10°C, liquid water is favored:

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

$$\Delta G^\circ = 1529 \frac{\text{cal}}{\text{mol}} - 1583 \frac{\text{cal}}{\text{mol}} = -54 \frac{\text{cal}}{\text{mol}}$$

So, whether ice or liquid water is favored, is directed by the **balance between enthalpic and entropic factors**

We seek the same type of description for the equilibrium system:

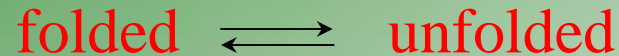


**We want to understand what makes a protein fold and unfold**

# Factors Affecting Protein Stability

---

For the process:



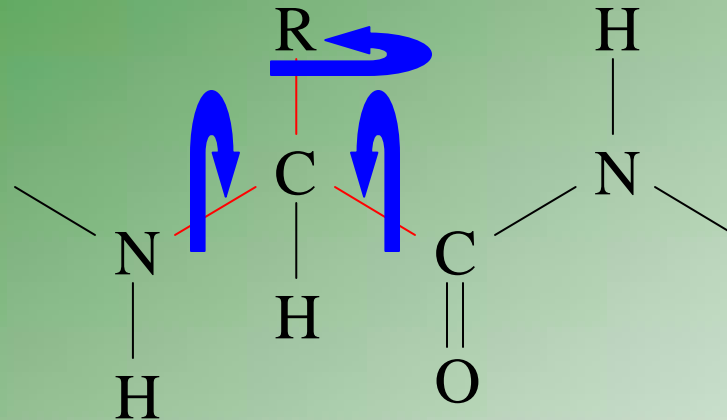
## 1. Configurational Entropy:

Upon unfolding, the polypeptide chain goes from a single (or nearly single) configuration to become a “random coil” with many accessible configurations → to a state of high entropy

Let us consider the polypeptide  
back bone...

# Protein Folding and Stability

---



We can consider **3 degrees of freedom** per aa residue in the chain. Assume that each degree of freedom can adopt 2 conformations. Then, the number of accessible configurations per residue is:

$$2 \times 2 \times 2 = 2^3$$

(each with equal probability)

# Protein Folding and Stability

---

The entropy gained upon unfolding would be:

$$\Delta S^{\circ}_{\text{config.}} = R \ln \frac{\text{Unfolded Configurations}}{\text{Folded Configurations}}$$

$$\Delta S^{\circ}_{\text{config.}} = R \ln \frac{2^3}{1}$$

$$\Delta S^{\circ}_{\text{config.}} = 1.98 \frac{\text{cal}}{\text{mol residue K}} \times \ln 2^3$$

$$\Delta S^{\circ}_{\text{config.}} = +4.1 \frac{\text{cal}}{\text{mol residue K}}$$

The corresponding configurational free energy is:

$$\Delta G_{\text{config}} = -T\Delta S_{\text{config}}$$

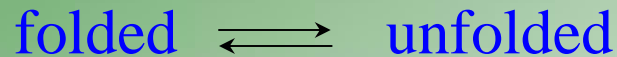
$$\Delta G_{\text{config}} = -298\text{K} \times 4.1 \frac{\text{cal}}{\text{mol residue K}}$$

$$\Delta G_{\text{config}} = -1.2 \frac{\text{Kcal}}{\text{mol residue}}$$

# Protein Folding and Stability

---

The configurational free energy is favorable for the process:



Thus, this free energy term is **highly destabilizing** for proteins

- What, then, accounts for the fact that proteins are usually stable at room temperature?
- Some factors must be able to offset the entropic disorder gained upon unfolding
- Also, **why do proteins unfold at certain temperatures?**

# Protein Folding and Stability

## 2. The Hydrophobic Effect:

In the mid 1950s **Kauzman** and **Tanford** began to investigate what really happened when a protein unfolds and exposes many aliphatic, and aromatic side-chains to the solvent. They carried out a series of thermodynamic experiments to transfer hydrophobic substances from organic solvents to water. Here are the data:

<u>System</u>	<u><math>\Delta G^\circ_{\text{transfer}}</math></u>	<u><math>\Delta H^\circ</math></u>	<u><math>\Delta S^\circ</math></u>
CH <sub>4</sub> in benzene → CH <sub>4</sub> in water	+2600 cal	-2800 cal	-18 cal/mol
CH <sub>4</sub> in ether → CH <sub>4</sub> in water	+3300 cal	-2400 cal	-19 cal/mol
C <sub>2</sub> H <sub>4</sub> in benzene → C <sub>2</sub> H <sub>4</sub> in water	+2920 cal	-1610 cal	-15 cal/mol
C <sub>2</sub> H <sub>6</sub> in benzene → C <sub>2</sub> H <sub>6</sub> in water	+3800 cal	-2200 cal	-20 cal/mol

# Protein Folding and Stability

---

- There is nothing surprising about the free energies being greater than 0 – hydrophobic substances do not like water!
- Interestingly, however, the **enthalpy of transfer** is **favorable**. This result was unexpected.
- Also they found that the **entropic change** of the transfer, is what **makes the whole reaction unfavorable**. Namely, the solvation of these compounds in water led to **a reduction in the entropy of the transfer reaction**.

# Protein Folding and Stability

---

**Conclusion:** Hydrophobic groups must be able to form reasonably good hydrogen bonds with water (hence  $\Delta H < 0$ ), but it must be that the process required **CAVITATION** in the solvent (water) which must **arrange and organize around** the hydrophobic substances, leading to a negative  $\Delta S$  (of solvent, and not of the dissolved molecule itself)

# Protein Folding and Stability

---

Tanford and Kauzman then proposed that a major driving factor in protein folding is the **solvent disordering** that ensues **when the protein folds** sequestering in its interior (away from water) all the hydrophobic side-chains

# Protein Folding and Stability

---

- This is the so-called “**hydrophobic effect.**” It nicely explains why it is that all folded proteins have a “greasy” core and expose their polar residues to the solvent, by placing them on the surface.
- In fact, here is some data on aa side-chains:

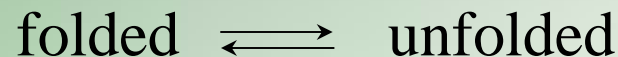
<u>Residue</u>	<u>R</u>	<u><math>\Delta G_{\text{transfer}}</math></u>
Ala	-CH <sub>3</sub>	+730 cal/mol
Val	-CH(CH <sub>3</sub> ) <sub>2</sub>	+1690 cal/mol
Leu	-CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	+2400 cal/mol
Ile	-CH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	+2970 cal/mol
Ser	-CH <sub>2</sub> OH	+40 cal/mol

# Protein Folding and Stability

Thus, for a protein with 100 aa of which 20 are Leu, 20 Phe, and 10 Tyr:

$$\begin{aligned} 20 \text{ Leu} &= 20 \times 2.4 \frac{\text{Kcal}}{\text{mol}} = + 48 \frac{\text{Kcal}}{\text{mol}} \\ 20 \text{ Phe} &= 20 \times 2.6 \frac{\text{Kcal}}{\text{mol}} = + 52 \frac{\text{Kcal}}{\text{mol}} \\ 10 \text{ Tyr} &= 10 \times 2.9 \frac{\text{Kcal}}{\text{mol}} = + 29 \frac{\text{Kcal}}{\text{mol}} \\ &\quad \underline{\quad \quad \quad} \\ &\quad \quad \quad + 129 \frac{\text{Kcal}}{\text{mol}} \end{aligned}$$

Therefore, for the process...



$$\Delta G_{\text{overall}}^{\circ} = \Delta G_{\text{config}}^{\circ} + \Delta G_{\text{hydrophobic}}^{\circ}$$

$$\Delta G_{\text{overall}}^{\circ} = -120 \frac{\text{Kcal}}{\text{mol}} + 129 \frac{\text{Kcal}}{\text{mol}}$$

$$\Delta G_{\text{overall}}^{\circ} = +9 \frac{\text{Kcal}}{\text{mol}} \quad \text{i.e. protein remains stable (marginally)}$$

# Protein Folding and Stability

---

Thus, the surprising fact is that the **factor responsible for the stability of proteins is the entropic cost of organizing the solvent around hydrophobic residues**, when the protein denatures.

**And again... water, the solvent, is the silent but all important factor**

Also... proteins are marginally stable: their favorable free energy of folding, being the **small difference between two large factors of opposite sign**.

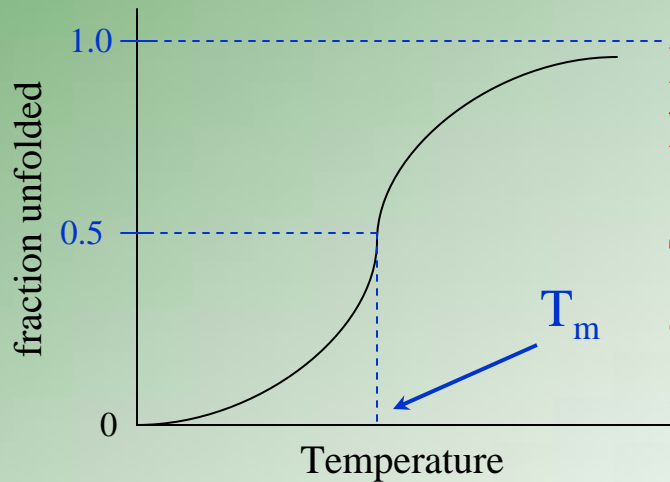
**Let us examine thermal denaturation...**

# Protein Folding and Stability

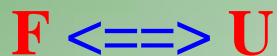
## Thermal Denaturation:

From the previous discussion, it is not surprising that relatively small changes in ionic strength, temperature, etc., lead to denaturation of a protein (they are only marginally stable). Let's look at thermal denaturation:

The 'melting temperature' or  $T_m$  is the temp at which 50% of the molecules are unfolded.



Protein melting is a highly cooperative process. Steepness of the curve at  $T_m$  is a measure of the cooperativity: it implies a 2-state process.



At  $T_m$ ,  $K_{eq} = 1$

$$\Delta G_{\text{overall}}^{\circ} = 0 = \Delta H_{\text{overall}}^{\circ} - T_m \Delta S_{\text{overall}}^{\circ}$$

$$T_m = \frac{\Delta H_{\text{overall}}^{\circ}}{\Delta S_{\text{overall}}^{\circ}} = \frac{\Delta H_{\text{Folded} \rightarrow \text{Unfolded}}^{\circ}}{\Delta S_{\text{config}}^{\circ} + \Delta S_{\text{hydrophobic}}^{\circ}}$$

# Protein Folding and Stability

---

Most **small globular proteins** are characterized by a highly cooperative transition from native to denatured state, as revealed by the ratio of their *van't Hoff* to experimental calorimetric enthalpies of denaturation:

$$\frac{\Delta H_{VH}}{\Delta H_{calorimetric}} = 1.05 \pm 0.03$$

This is not the case for multi-domain proteins. Papain, vgr., give values of 1.8 for this ratio.

Experimentally, it is found that protein denaturation is accompanied by a large change in heat capacity at constant pressure:  $\Delta C_p$

# Protein Folding and Stability

---

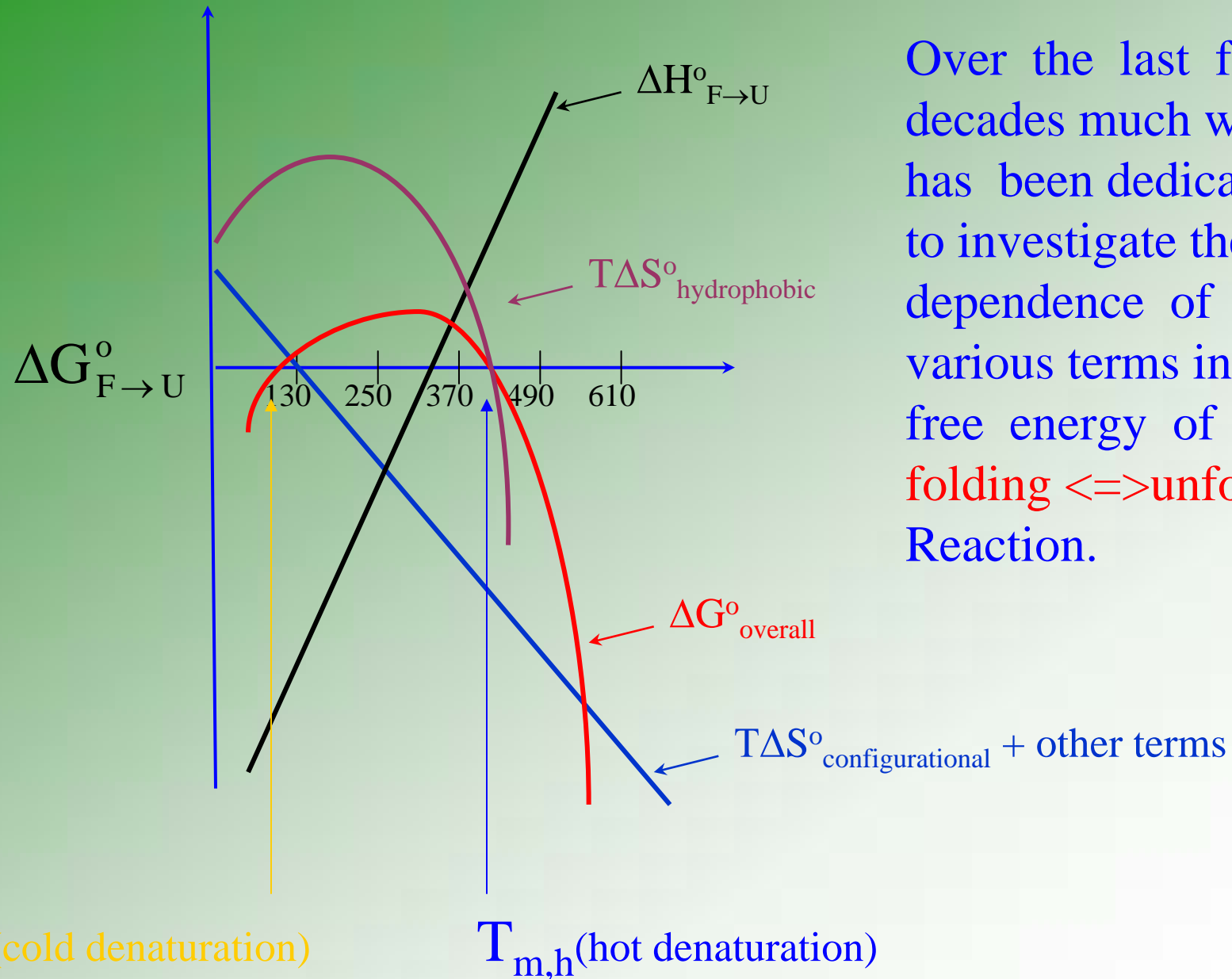
This large change in heat capacity is thought to be dominated by the amount of **non-polar surfaces** exposed to the solvent, and the **tendency of water to order itself around these groups**.

The finite heat capacity change also indicates a **T-dependent enthalpy and entropy of unfolding**, so that the the free energy of unfolding can be written as:

$$\Delta G_{unfold} = \Delta H^* - T\Delta S^* + \Delta C_p \left[ (T - T^*) - T \ln \frac{T}{T^*} \right]$$

This equation predicts that the **stability of proteins attains an optimal value** at certain temperature.

# Protein Folding and Stability



$T_{m,c}$  (cold denaturation)

$T_{m,h}$  (hot denaturation)

# Protein Folding and Stability

---

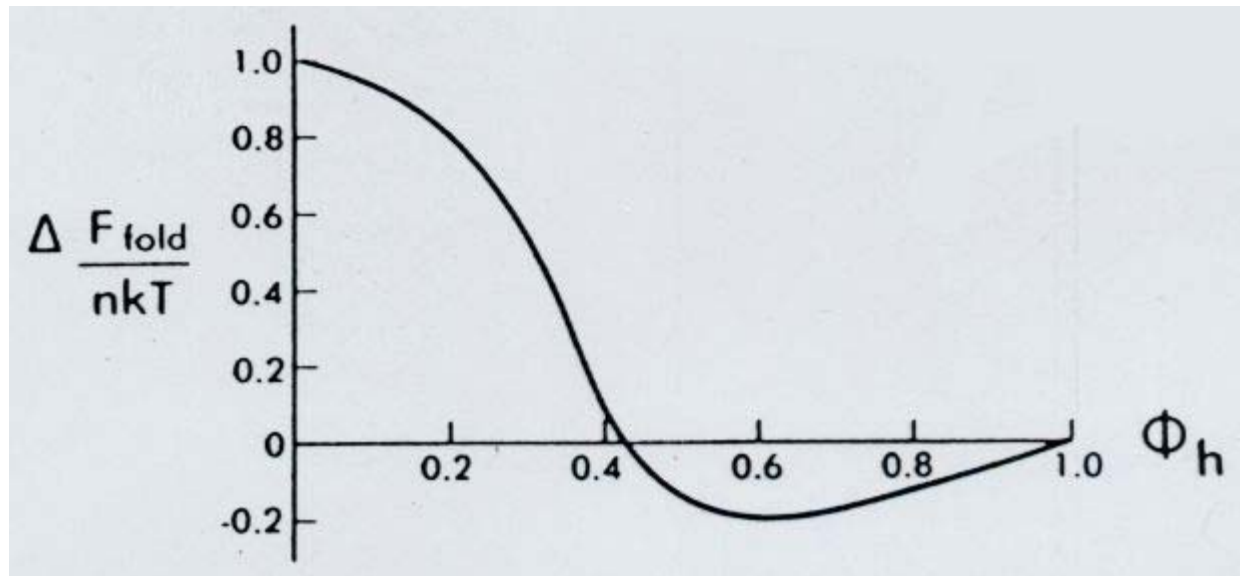
We have seen that the general architectural pattern of proteins is:

Hydrophobic residues → core

Polar or hydrophilic residues → exposed to the solvent

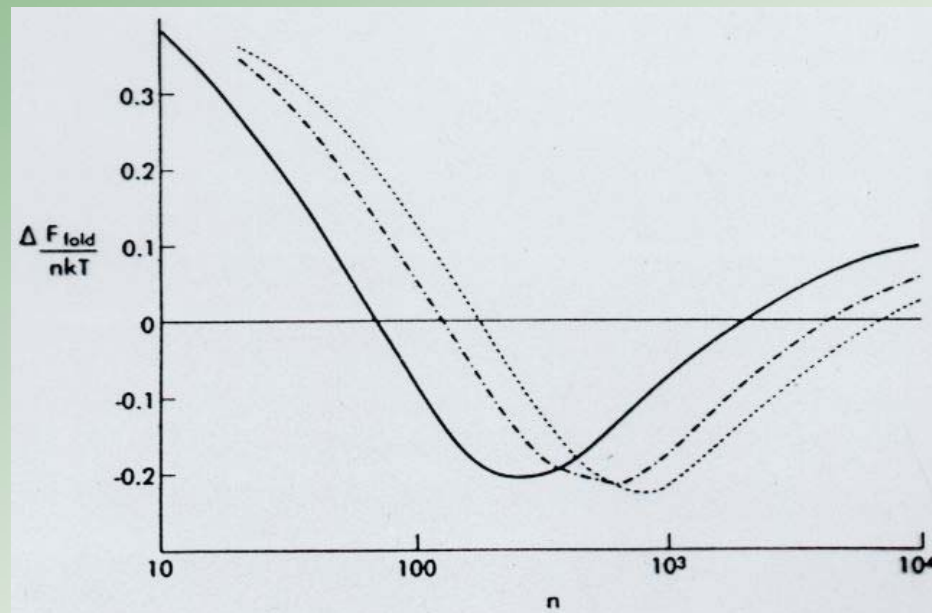
This pattern has several important consequences

a) For any given size, the protein requires a minimum fraction of apolar residues to attain stable folding (see fig. 2)



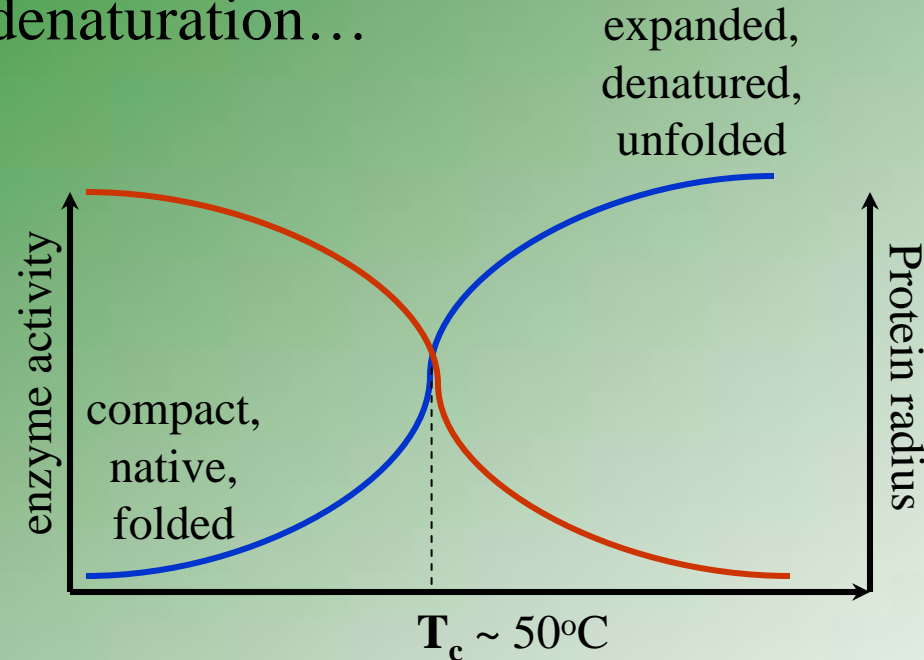
# Protein Folding and Stability

- b) At higher fractions of apolar residues the protein is first stabilized and then destabilized
- c) For a constant fraction of apolar residues, there should be an optimal range of protein sizes that confer the highest stability to the globular structure (see fig. below)
- d) As the size of the protein increases, structure tends to **segregate** in connected globular domains. Larger proteins  $\rightarrow$  larger subunits



# Protein Folding and Stability

Back to thermal denaturation...



No covalent bonds are broken or formed in this heating process

$\therefore$  Activity depends on a structure maintained by weaker forces: 3D native structure

Working hypothesis: Somehow sequence codifies for the 3D structure

# Protein Folding and Stability

---

## Questions:

1. **Is there a folding code?** Can the sequence specify the 3D structure?
2. **Is the code linear?** Do some partial sequences fold in the same manner no matter where they are placed in the sequence?
3. **Is there a “unit” of folding information?** Is there a particular set of aa sequences responsible for folding?
4. **What is the contribution of a single aa residue to folding?**

Let us examine some facts...!

# Protein Folding and Stability

---

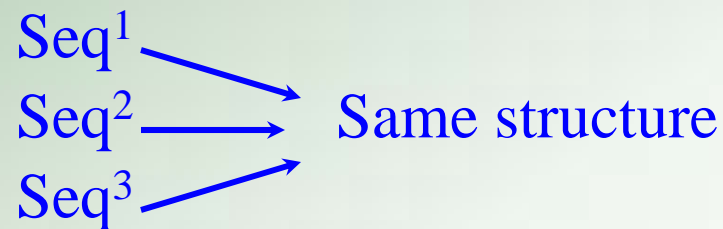
## Facts:

1. X-ray data shows: small identical aa sequences (3-4 aa) have different structure in different proteins

This fact argues that it is the sequence “in the context of the rest of the sequence” that is important

2. Practically identical structures are observed for proteins with very different sequences

This fact argues that functionality matters, and not the sequence. The code is said to be degenerate:



# Protein Folding and Stability

---

## More Facts:

3. Only a minority of residues in the protein are important in determining the structure:

(i) RNase mutants have many aa replacements, but have the same X-ray structure

(ii) Certain  $\alpha$ -helices in a protein can be completely replaced by a run of Ala and still maintain the  $\alpha$ -helical structure (Brian Mathews, U. of Oregon)

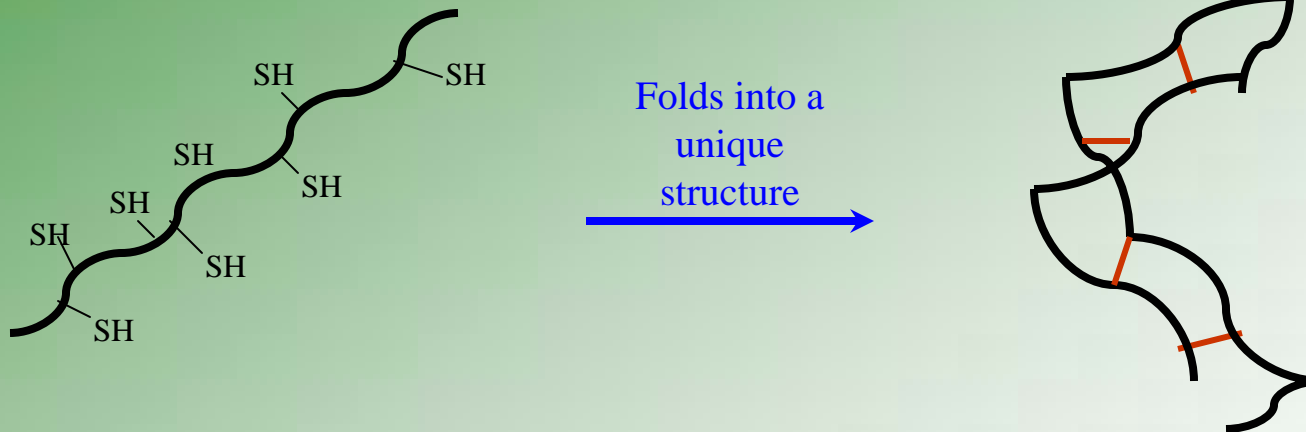
# Protein Folding and Stability

## Ribonuclease Studies:

RNase A: endonuclease (124 aa)

: there are 8 cys residues  $\rightarrow$  4 S-S bonds

After biosynthesis...

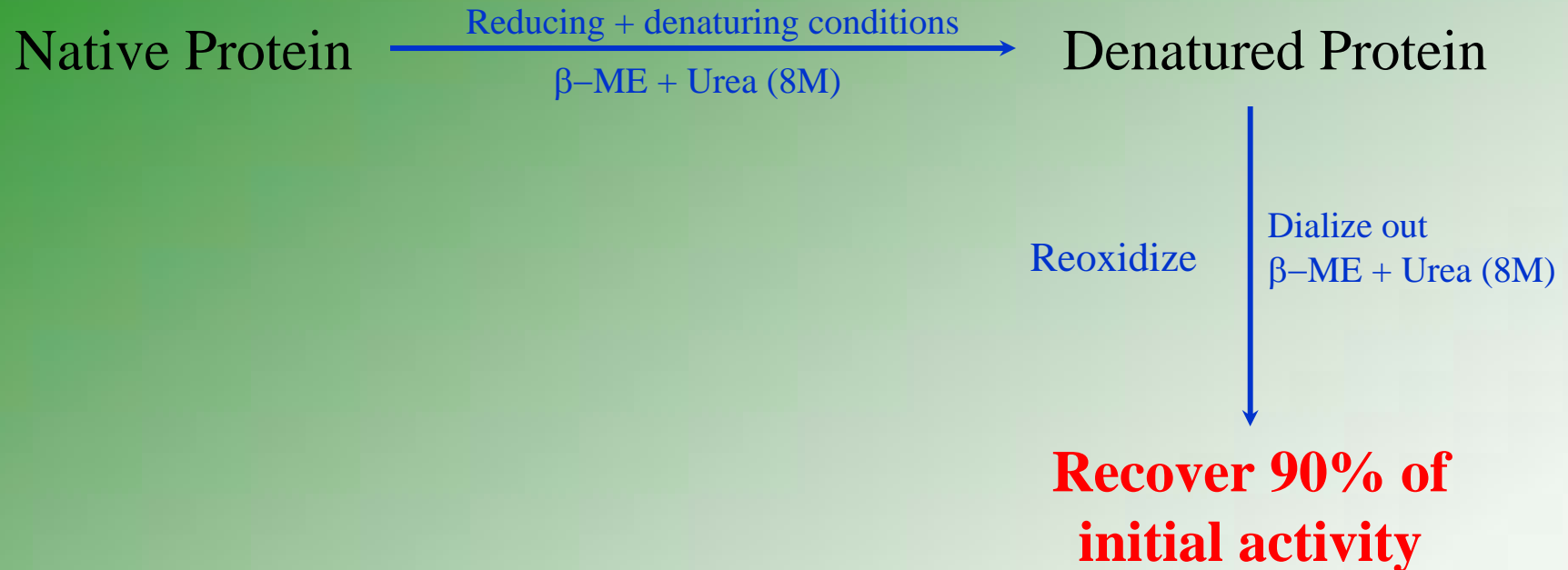


We can do two experiments on this native protein...

# Protein Folding and Stability

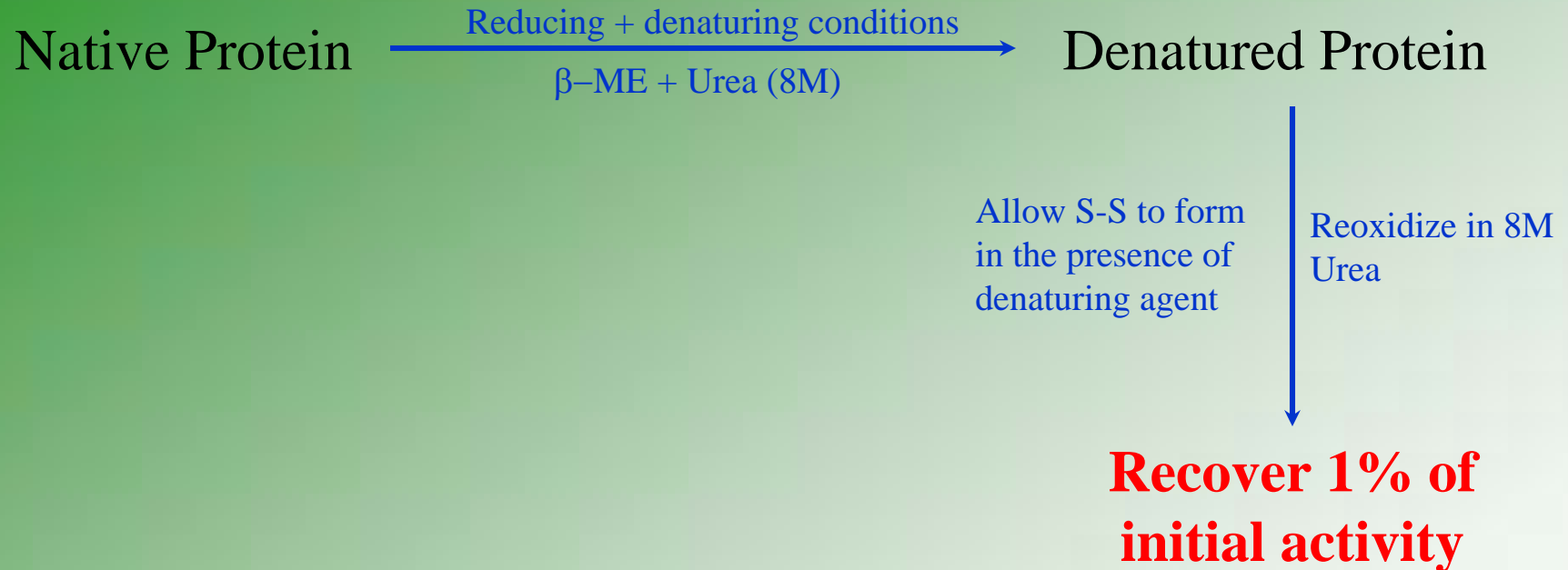
---

1.



# Protein Folding and Stability

2.



Why does this occur?

# Protein Folding and Stability

---

Anfinsen and co-workers showed:

This result is what you would expect from purely random S-S formation:

How many possible sets of 4 S-S bonds can be formed from 8 SH groups?

First SH can form 7 possible S-S (leaves 6 SH)

Second SH can form 5 possible S-S (leaves 4 SH)

Third SH can form 3 possible S-S (leaves 2 SH)

Fourth SH can form 1 possible S-S (leaves no SH)

The total # of sets of 4 S-S bonds is:  $7 \times 5 \times 3 \times 1 = 105$

Only one of these is the correct one, and the probability of it by random chance is:

$$\frac{1}{105} \approx 1\% \quad \text{More importantly...}$$

# Protein Folding and Stability

---

## They concluded:

1. Sequence of 124 aa contains information needed for folding
2. Correct S-S are formed because of interactions of parts distant in the linear sequence

Question... **Is the final native state**

(i) **Kinetically determined** (i.e. the correct combination of S-S forms **faster** than the others)

(ii) **Thermodynamically determined** (i.e. the correct combination of S-S has the **lowest energy**)

# Protein Folding and Stability

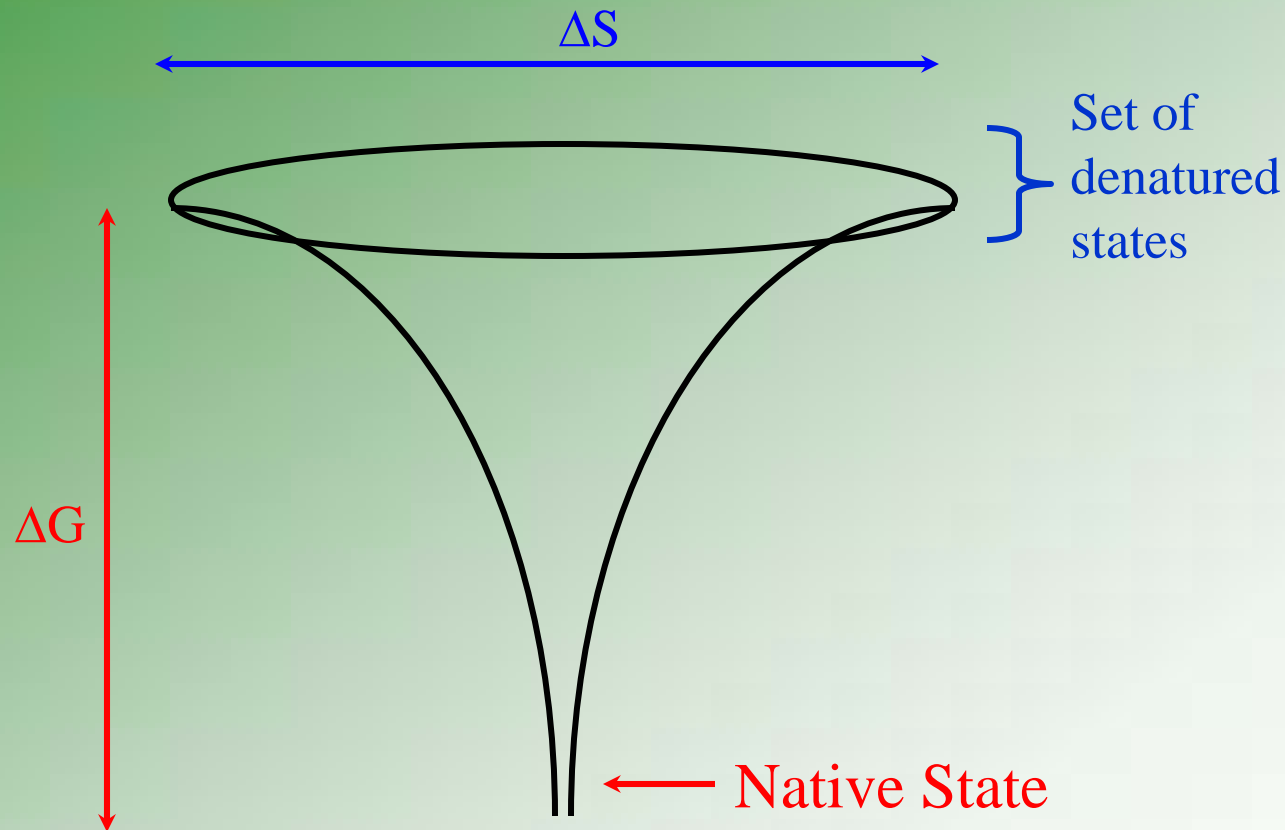
---

The search for the folded state cannot be a random exploration:

- A 100 aa peptide would have to explore  $\sim 100^{100}$  conformations
- Even exploring  $\sim 10^{13}$ /sec would require  $\sim 10^{87}$  sec (the universe is only  $10^{11}$  sec old! – this is called **Levinthal's Paradox**)
- But the idea that proteins follow a single narrow “pathway” is inconsistent with the fact that each molecule in the ensemble would have a different initial configuration ( $\phi_i, \psi_i$ )
- It is possible that the multiplicity of possible folding pathways form a sort of “funnel”...

# Protein Folding and Stability

This funnel can be represented as:



Trade of stabilizing “free energy” of interactions for configurational entropy.

# Protein Folding and Stability

---

There must be an ordered pathway of folding and as native state is approached, structure becomes more stable.

These ideas are the result of the so-called *Principle of Minimal Frustration* itself a re-statement of *Go's Consistency Principle*.

A system is said to be **frustrated** if no configuration of the system can optimize all the interactions. Most protein sequences do not fold spontaneously into a well-defined “native” states.

These principles essentially try to explain the ability of proteins to organize across various length scales (Go) and to explain the apparent existence of a minimal energy state easily reachable.

These principles apply to **naturally** occurring protein sequences which would have evolved the ability to minimize frustration.

# Protein Folding and Stability

---

So, how is the folded state attained then?

The answer to this question is the subject of much controversy.

Some authors have suggested that secondary structures form first and tertiary compaction next.

Studies made on simplified lattice models (Ken Dill, UCSSF) have suggested that the first thing that happens is a hydrophobic collapse that greatly reduces the total number of accessible configurations; thereupon secondary and later tertiary structures organize.

Other proposals involve a hierarchical organization for proteins with pre-formation of secondary structures in the unfolded state.

# Protein Folding and Stability

---

These proposals are not necessarily exclusive of each other

A **hydrophobic collapse** can occur despite the fact that some secondary structure is already organized in the unfolded state.

And indeed this is the most likely and currently the more accepted scenario.

# Protein Folding and Stability

---

# Protein Folding and Stability

---

So: there must be:

Ordered pathway of folding and as native state is approached, structure becomes more stable

Let us examine some **folding pathway facts...**