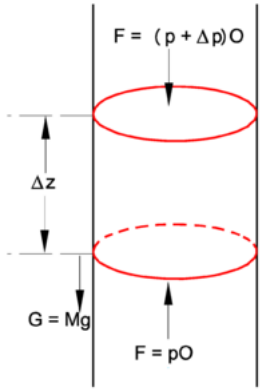
Protein folding

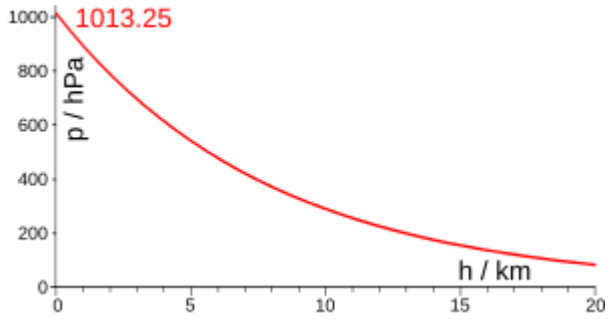
Outline

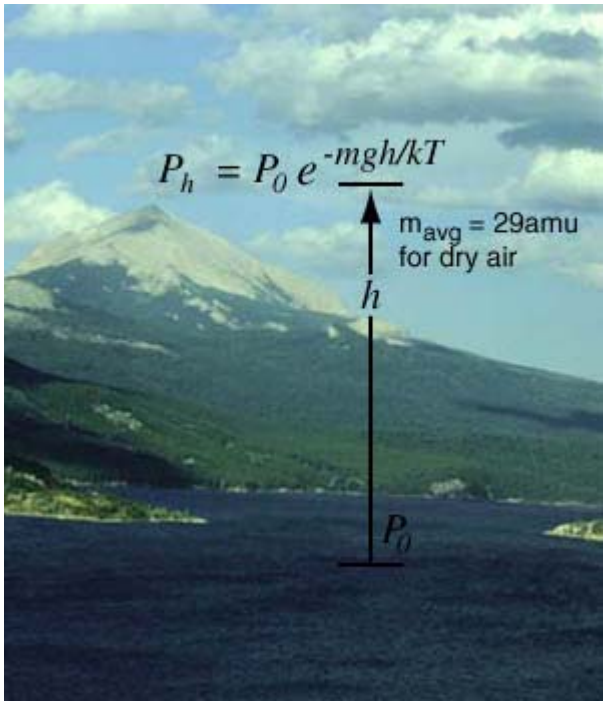
- Entropy and its role in protein stability
- Protein folding

Boltzmann distribution

Barometric equation: change of pressure *P* with height *h*.







Probability *p* of molecule of mass *m* to be found at the height *h* has the same exponential dependence:

$$p \sim \exp(- mgh / kT)$$

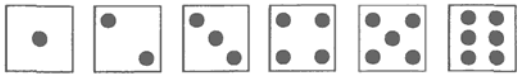
More generally, probability p_i of a system to be found in the microstate having energy E_i:

$$p_i \sim \exp(- E_i / kT)$$

$$p_i = \exp(- E_i / kT) / Z$$

Toy example: dice

One die has six microstates:



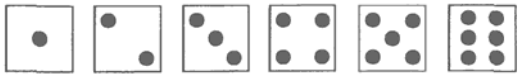
$p_i = 1/6,\ i = 1 \ldots 6$

$p_i \sim \exp(- E_i / kT)$

$E_i = kT \ln 6 + \text{const}$

Where entropy comes from?

One die has six microstates:



$p_i = 1/6,\ i = 1 \ldots 6$

$p_i \sim exp(- E_i / kT)$

$E_i = kT\ ln\ 6 + const$

Two dice have 36 microstates with equal probabilities p_i (= equal energies E_i). If we are interested in the sum of outcomes, the microstates form 11 macrostates.

Consider a macrostate j consisting of M_j microstates:

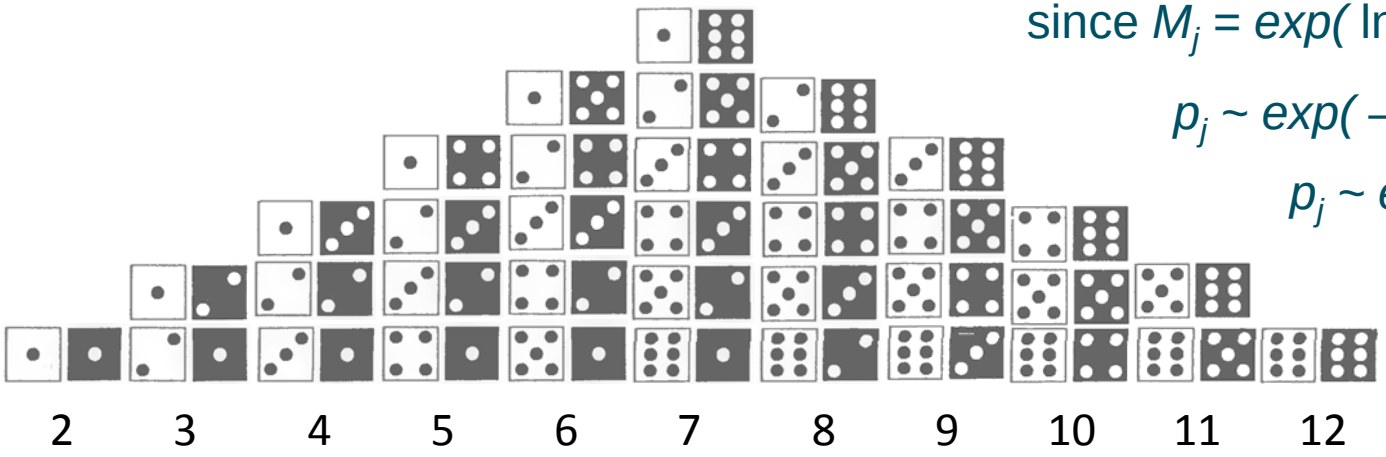
$p_j = M_j\, p_i \sim M_j\, x\, exp(- E_i / kT)$

since $M_j = exp(\ln M_j)$:

$p_j \sim exp(- [E_i - T\, k \ln M_i] / kT)$

$p_j \sim exp(- [E_i - TS_j] / kT)$

$p_j \sim exp(- F_j / kT)$



Entropy: $S_j = k \ln M_j$

Free energy: $F_j = E_i - TS_j$

Entropy and free energy

So, entropy arises due to clustering microstates into macrostates: if we want (or have to) consider macrostates instead of microstates, we should just write entropy.

Entropy:

$$S = k \ln M$$

Free energy:

$$F = E - TS$$

Free energy and probabilities are coupled:

$$p_j \sim exp(- F_j / kT)$$

Entropy examples (1/2)

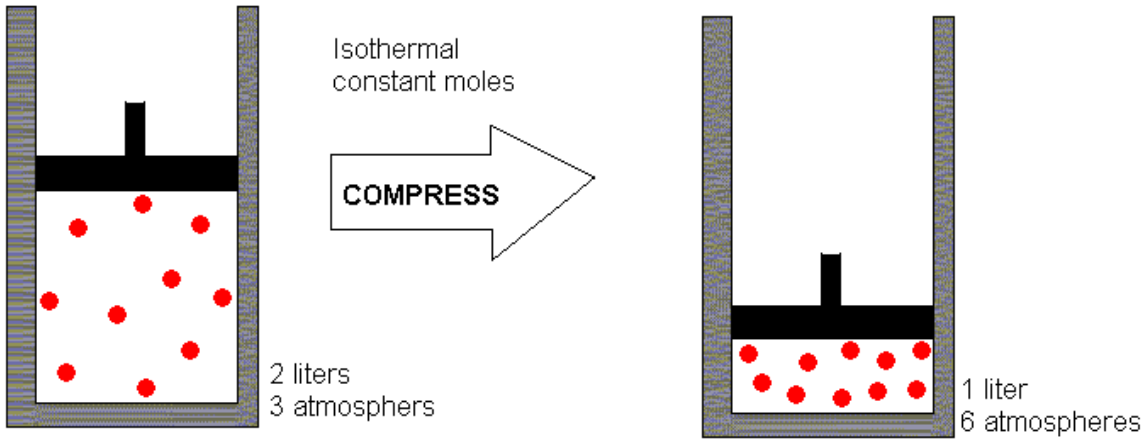
How a polymer looks like in neutral solvent? Why is it elastic?

The most stable (i.e., most probable) macrostate has the linear size in each dimension $r = r_0 n^{1/2}$, where r_0 is the length of a single fragment, and n the number of fragments. Proof by induction.

The stability/elasticity is purely entropic. The unfolded proteins look similar.

Entropy examples (2/2)

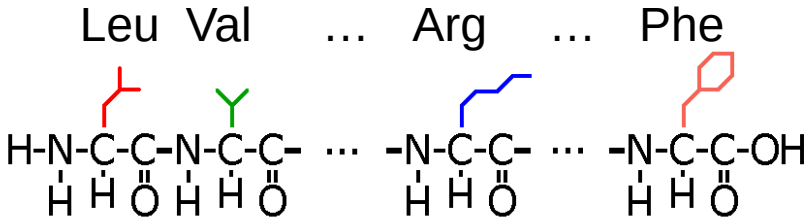
Isothermic compression of ideal gas



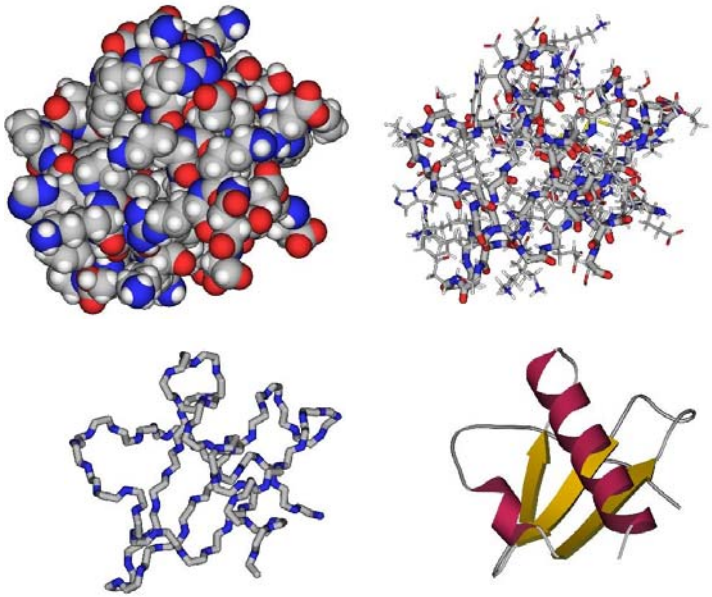
The energy of the ideal gas before and after the compression is the same (because temperature is the same): $U = 3/2 * \nu RT$, ν is the number of moles. The reason for elasticity is that we transfer the gas from the macrostate with high entropy to that with low entropy. This effect is purely entropic.

Role of entropy in protein interactions

Primary structure



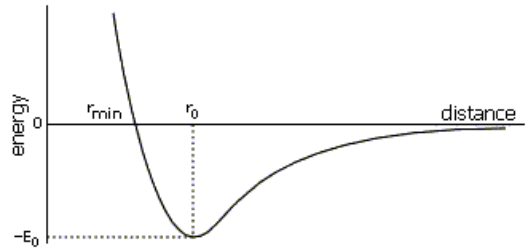
Tertiary structure



We will not consider covalent bonds (except for coordinate bonds and disulfide bonds) because they are the same in the native and unfolded proteins and cancel out.

Van der Waals interactions

Are noncovalent interactions formed nonspecifically by all proteins. They are presented in both the folded and unfolded state of the protein. In the folded state many interactions are between amino acid residues (and with water molecules at the surface). In the unfolded state the interactions are mostly between amino acid residues and the surrounding water molecules.



Lennard-Jones potential:

$$U_{LJ}(r) = E_0[(r_0/r)^{12} - 2(r_0/r)^6]$$

Free energy:

In water: ~ 0.0 to –0.05 kcal/M

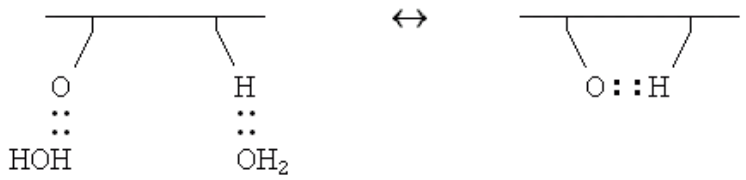
In vacuum: ~ –0.2 to –0.5 kcal/M

Entropy does not play a role

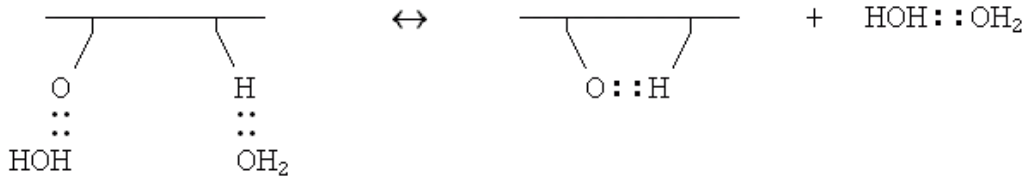
Hydrogen bonds

Are noncovalent interactions occurred in all proteins of electrostatic nature. They are found between directed H- containing dipoles (-OH or –NH) and partially negatively charged O or N atoms.

Formation in vacuum:



Formation in water:



Free energy:

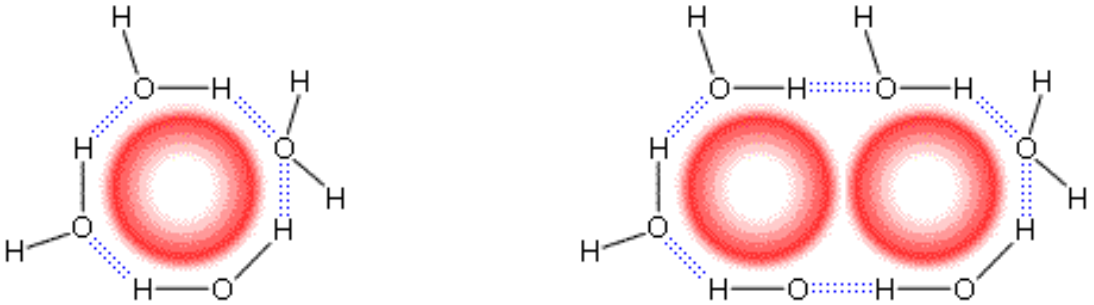
In water: ~ –1.5 kcal/M

In vacuum : ~ –5 kcal/M

Formation (in water): entropic nature

Hydrophobic interactions

Are noncovalent non-specific interactions between hydrophobic groups found in most proteins. The origin is that water molecules want to save hydrogen bonds and are therefore partly constrained when facing hydrophobic groups.



Free energy (for methyl groups):

In water: ~ –0.3 kcal/M

In vacuum: ~ NO

Hydrophobic interactions are of entropic nature

Electrostatic interactions

Are noncovalent specific interactions occurred in most proteins. They are found between positively charged Lys^+, Arg^+, or His^+ and negatively charged Asp^- or Glu^-.

Free energy:

At the protein/water interface: ~ –2 kcal/M

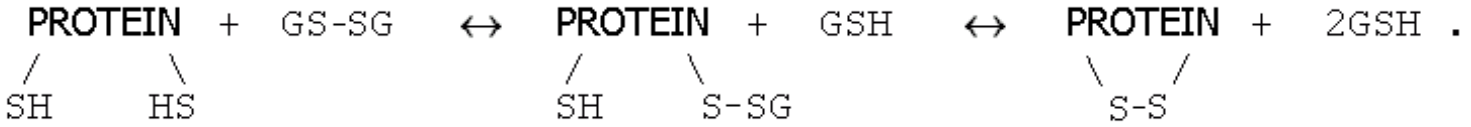
At the protein/vacuum interface: ~ –40 kcal/M

Inside the protein (water or vacuum): ~ –25 kcal/M

Exchange (in water): mostly entropic nature

S-S bonds

Disulfide bonds are covalent bonds formed very specifically between two cysteine residues. They are found mostly in secreted proteins. Inside the cell a special enzyme and glutathiones make the formation of disulfide bonds reversible. Outside the cell disulfide bonds are fixed.



Free energy:

Inside the cell: ~ 0 kcal/M

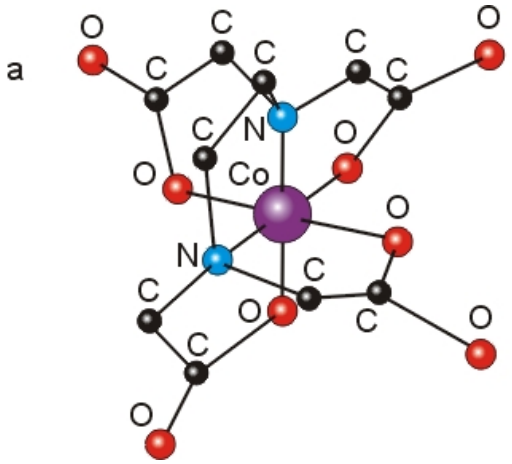
Outside the cell: very high, ~ –100 kcal/M, as for usual covalent bond

Exchange (in water): entropic nature

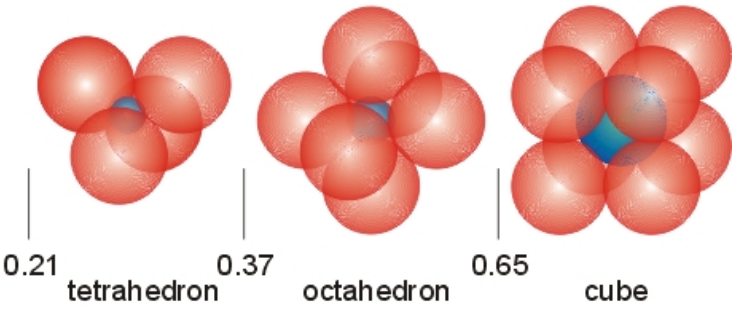
Coordinate bonds

Coordinate bonds are covalent bonds formed very specifically by protein with metal ions (Zn++, Fe+++, …). They are found in metal-binding proteins.

Protein



Water



Free energy:

In water: ~ –6 kcal/M and higher

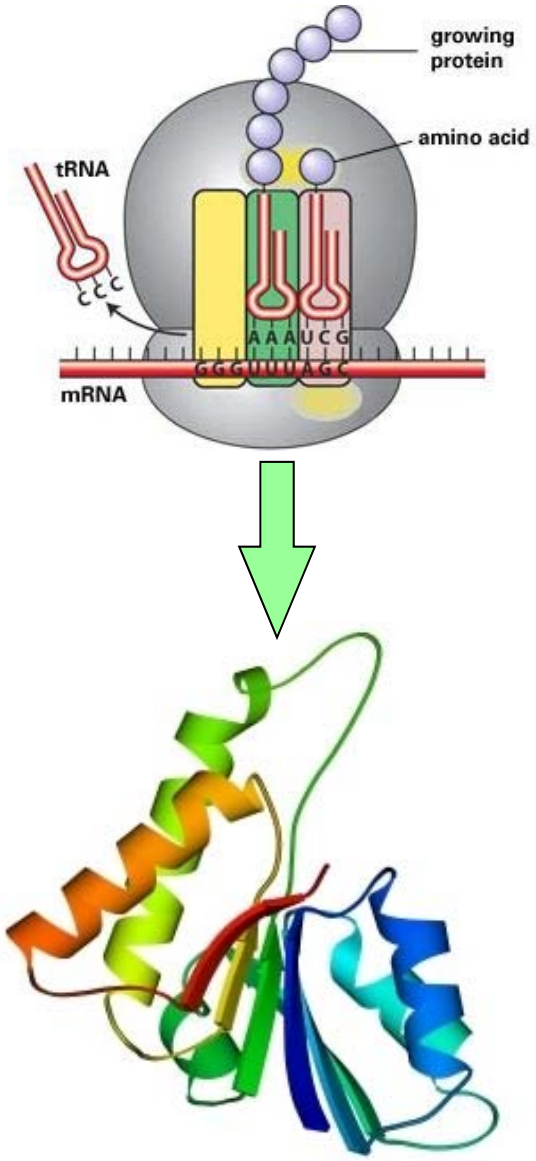
In vacuum: very high, ~ –100 kcal/M, as for usual covalent bond

Exchange (in water): entropic nature

Outline

- Entropy and its role in protein stability
- Protein folding

Biosynthesis and protein folding



Protein folding *in vivo*

- RNA-encoded protein chain is synthesized at a ribosome.

- Folding of large (multi-domain) protein: during the biosynthesis.

- Folding is aided by special proteins “chaperons” and enzymes like disulfide isomerase.

- ^{15}N, ^{13}C NMR: Polypeptides remain unstructured during elongation but fold into a compact, native-like structure when the entire sequence is available.

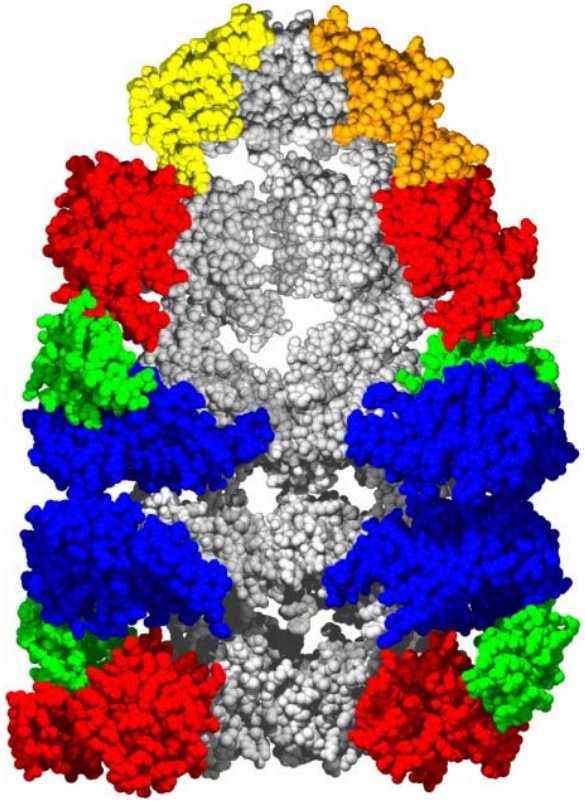
Chaperones

Folding: inside or outside

GroEL/ES? - OUTSIDE

GrEL/ES only decreases protein concentration of not-yet-folded protein in solution
(Marchenkov & Semisotnov, 2009, *Int. J. Mol. Sci.*, 10: 2066-83)

GrEL/ES does not accelerate folding
(Marchenko et al., PNAS, 2015)

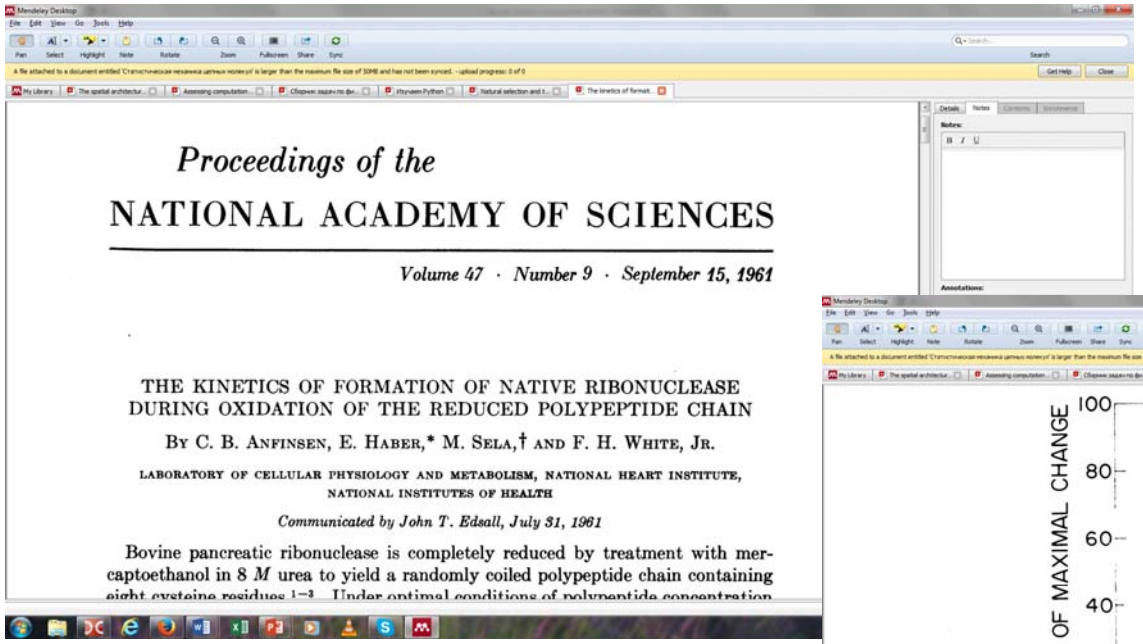


GroEL/ES

Protein folding *in vitro*

Water is a bad solvent for protein → Urea 8M is a good solvent for protein

Protein is folded ← Protein is unfolded



Proceedings of the

NATIONAL ACADEMY OF SCIENCES

Volume 47 · Number 9 · September 15, 1961

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

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

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

Communicated by John T. Edsall, July 31, 1961

Bovine pancreatic ribonuclease is completely reduced by treatment with mercaptoethanol in 8 *M* urea to yield a randomly coiled polypeptide chain containing eight cysteine residues [1–3]. Under optimal conditions of polypeptide concentration

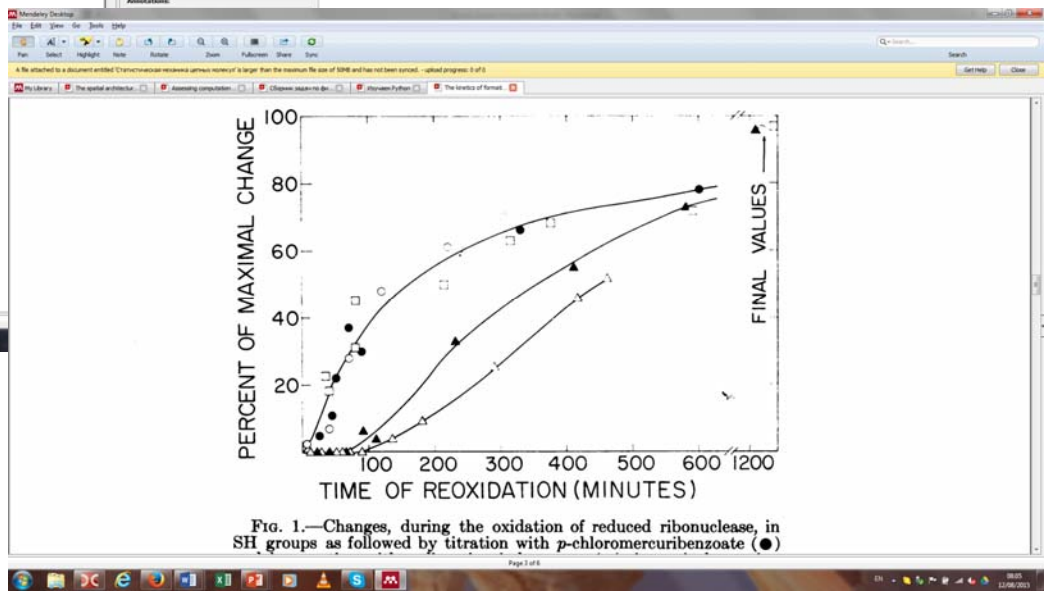
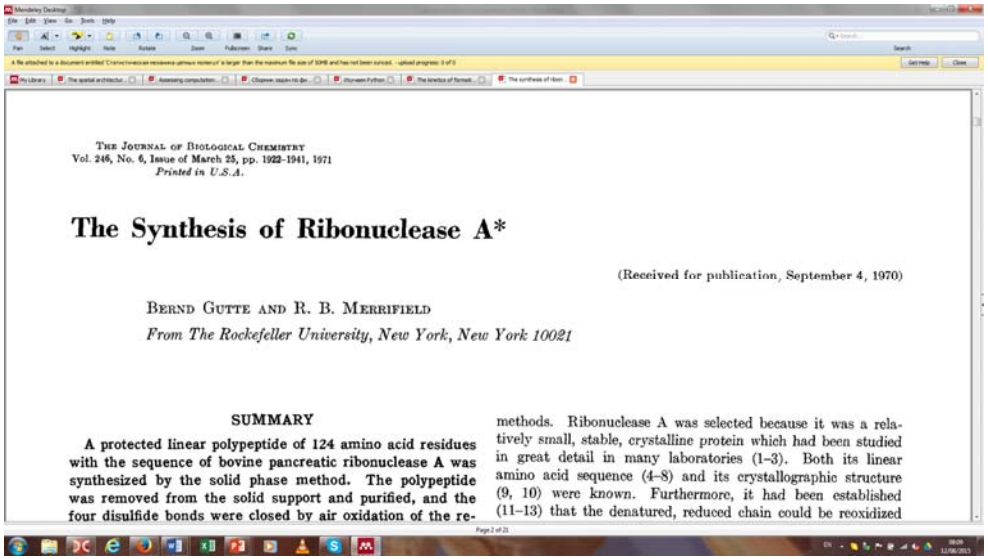


FIG. 1.—Changes, during the oxidation of reduced ribonuclease, in SH groups as followed by titration with *p*-chloromercuribenzoate (●)

Renaturation after chemical synthesis



THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 246, No. 6, Issue of March 25, pp. 1922-1941, 1971
Printed in U.S.A.

The Synthesis of Ribonuclease A*

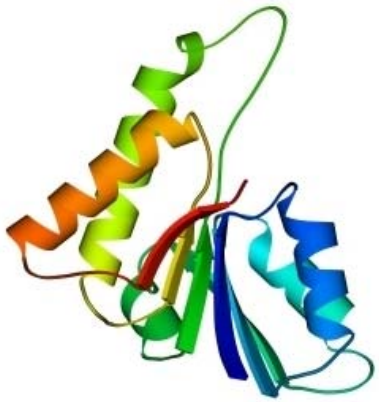
(Received for publication, September 4, 1970)

BERND GUTTE AND R. B. MERRIFIELD
From The Rockefeller University, New York, New York 10021

SUMMARY

A protected linear polypeptide of 124 amino acid residues with the sequence of bovine pancreatic ribonuclease A was synthesized by the solid phase method. The polypeptide was removed from the solid support and purified, and the four disulfide bonds were closed by air oxidation of the re-

methods. Ribonuclease A was selected because it was a relatively small, stable, crystalline protein which had been studied in great detail in many laboratories (1–3). Both its linear amino acid sequence (4–8) and its crystallographic structure (9, 10) were known. Furthermore, it had been established (11–13) that the denatured, reduced chain could be reoxidized



"The synthetic RNase A was indistinguishable from natural bovine pancreatic RNase A by gel filtration …, by chromatography …, and by electrophoresis. Amino acid analyses, peptide maps of tryptic digests, and the Michaelis constant agreed well with those of the natural enzyme."

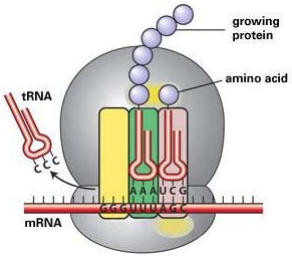
Christian Boehmer
Anfinsen, Jr.
(1916 –1995)
Nobel Prize 1972

Robert Bruce
Merrifield
(1921 – 2006)
Nobel Prize 1988

Native structure is the global minimum

Experiment

Biosynthesis



Renaturation

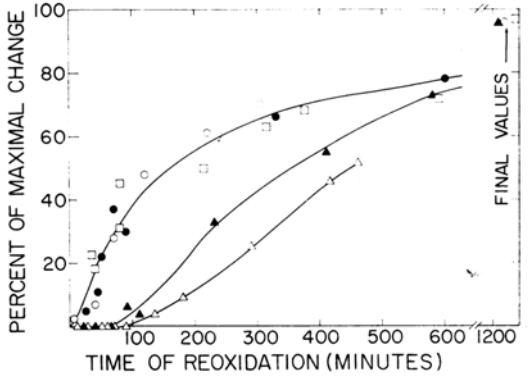
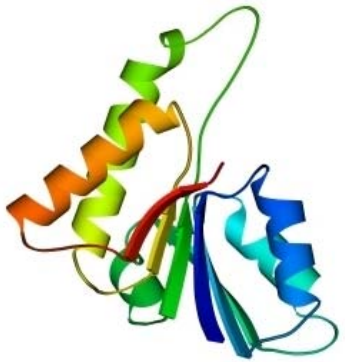


FIG. 1.—Changes, during the oxidation of reduced ribonuclease, in SH groups as followed by titration with *p*-chloromercuribenzoate (●)

Protein
3D structure is a
global minimum
of free energy



Chemical synthesis

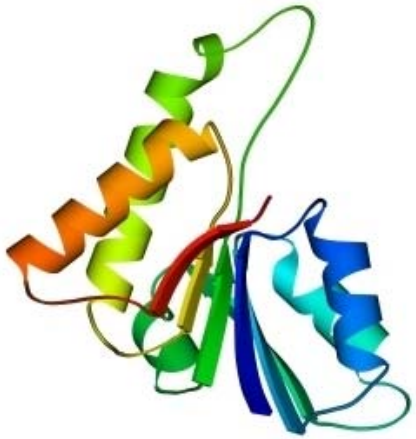
Physics of protein folding

Unfolded structure

$p_u \sim exp(- E_u / kT)$

Native structure

$p_N \sim exp(- E_N / kT)$



Many unfolded conformations

... M_U

$p_U = M_U \, p_u \sim M_U \, x \, exp(- E_u / kT)$

Since $M_U = exp(\ln M_U)$:

$p_U \sim exp(- [E_u - T \, k \ln M_U] / kT)$

$p_U \sim exp(- [E_u - TS] / kT)$

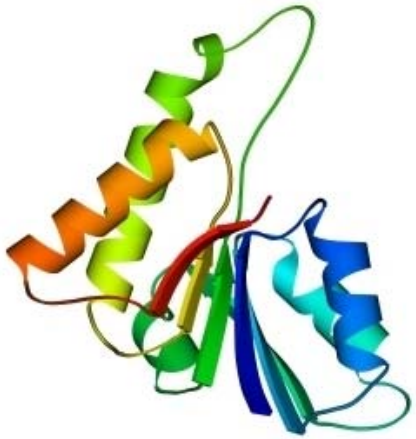
Physics of protein folding

Unfolded structure

$p_u \sim exp(- E_u / kT)$

Native structure

$p_N \sim exp(- E_N / kT)$



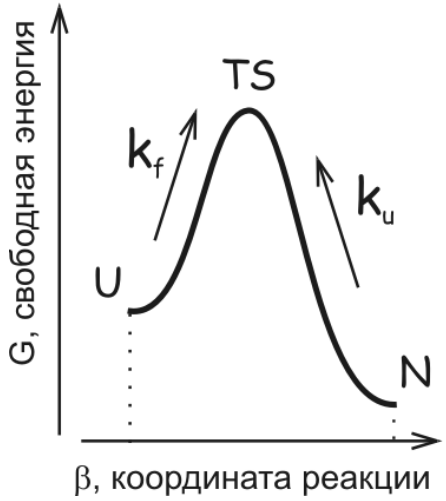
Many unfolded conformations

$p_U = M_U \, p_u \sim M_U \, x \, exp(- E_u / kT)$

Since $M_U = exp(\ln M_U)$:

$p_U \sim exp(- [E_u - T \, k \ln M_U] / kT)$

$p_U \sim exp(- [E_u - TS] / kT)$



Entropy

$S_i = k \ln M_i$

Free energy:

$G_i = E_i - TS_i$

$p \sim exp(- G_i / kT)$

Protein folding

Questions:

1. How do proteins fold?

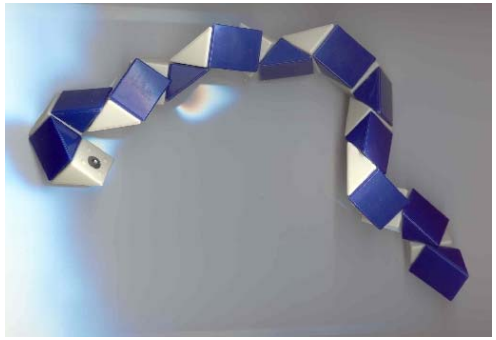
2. Can we predict protein structure from its amino acid sequence?

3. Can we create proteins with new properties?

Finding the global minimum

To check all the variants and choose the best one?

Analogy: Rubik's snake consists of 24 fragments



Sequence space:

Out of $4^{23} \approx 7 \cdot 10^{13}$ variants,

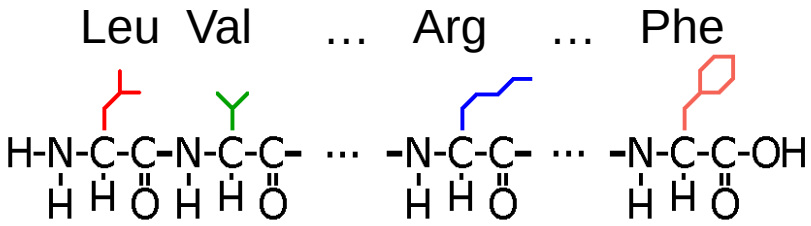
$\approx 10^{13}$ are possible.

One second per variant ≈ 300 000 yrs for all variants

DOES NOT WORK ☹

Finding the global minimum

To check all the variants and choose the best one?



Cyrus **Levinthal**
(1922 –1990)

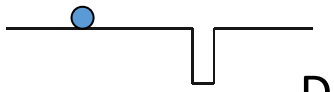
conformations ~ 10 x 10 x … x 10 x … x 10 = 10^L

1ps per conformation => time ~ 1ps x 10^L

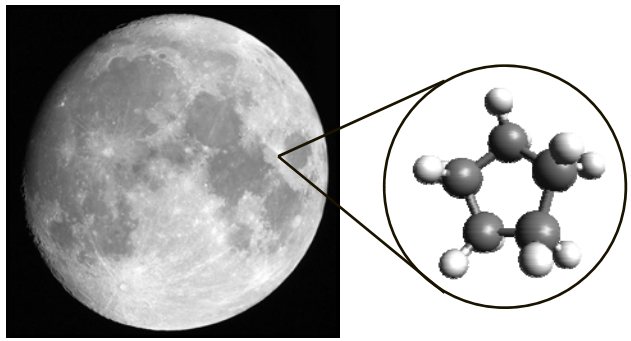
For small domain of L = 100 amino acid residues:

~10^{30} = 1 000 000 000 000 000 000 000 000 000 000 000 000 conformations
time ~ 15 000 000 000 yrs

The same time for random search (golf model):



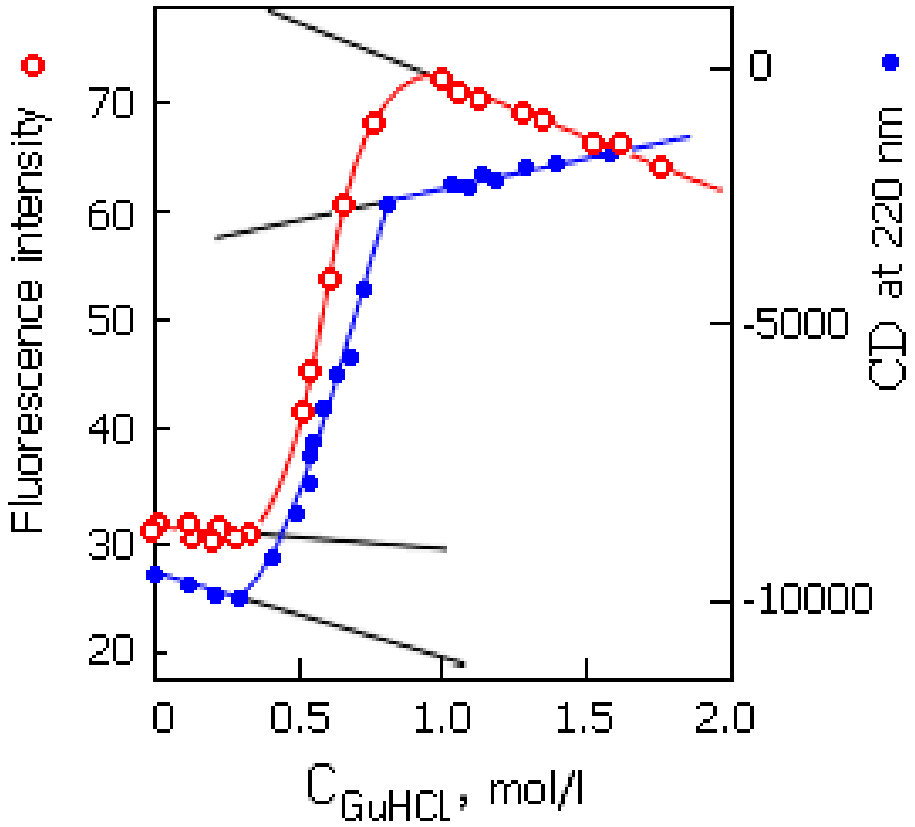
DOES NOT WORK ☹



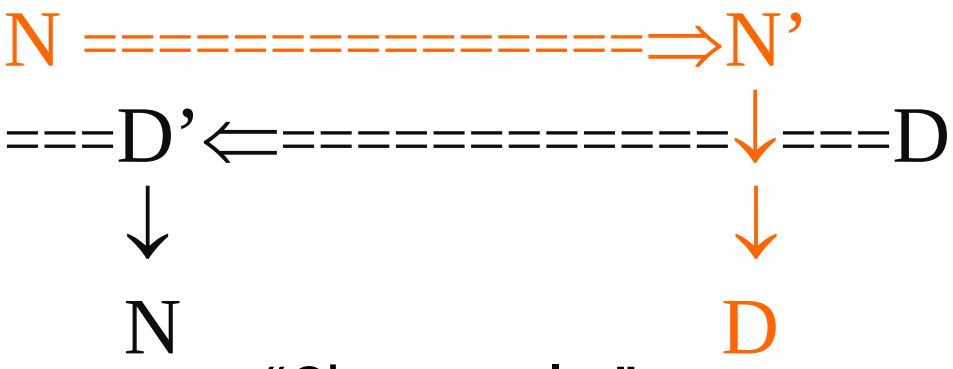
Levinthal: should be special pathway leading to local minimum

Experiments on protein folding

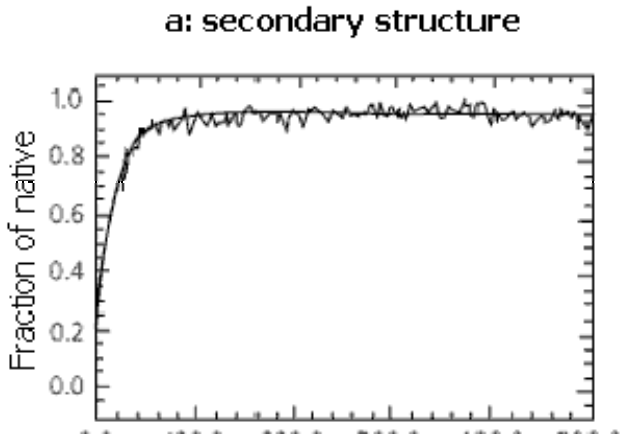
Protein denaturation ***in vitro***: cooperative transition

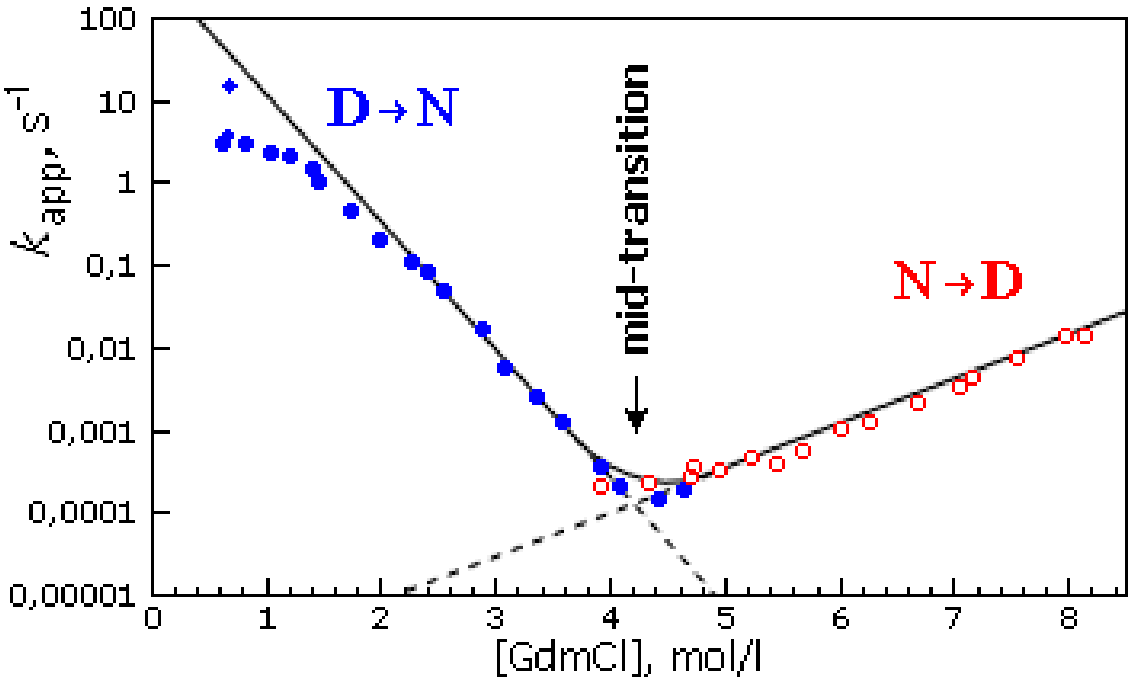


Experiments on protein folding



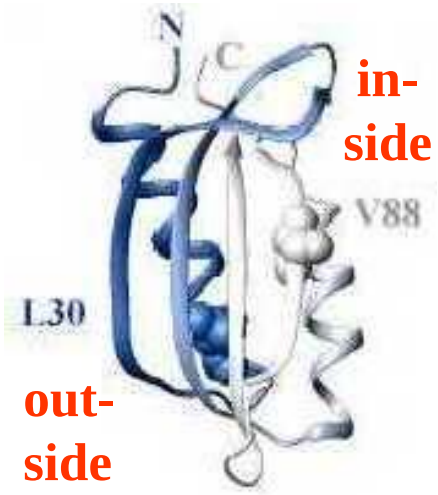
"Chevron plot"



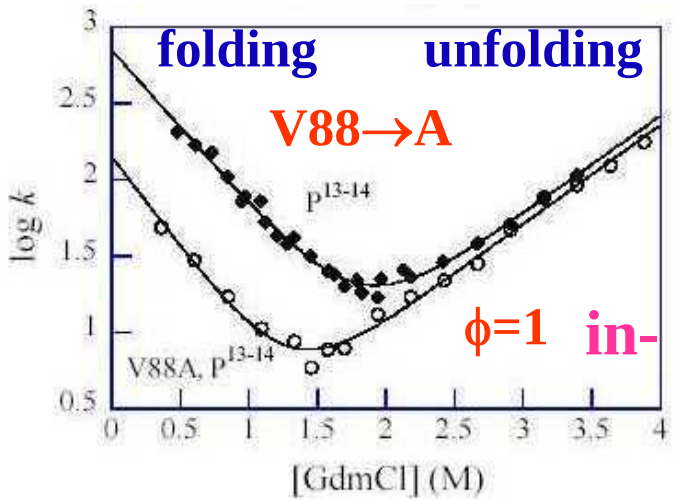


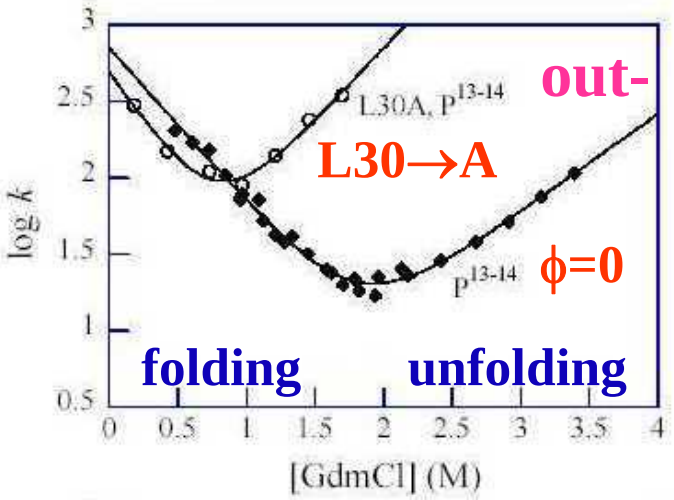
"Chevron plots": ***Reversible folding and unfolding even at mid-transition, where*** $k_{D \to N} = k_{N \to D}$

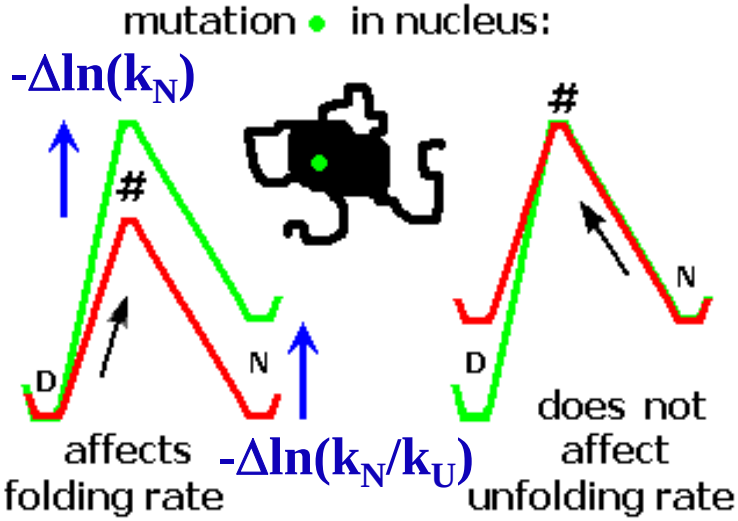
Folding nucleus: Site-directed mutations show residues belonging and not-belonging to the "nucleus", the native-like part of transition state (Fersht, 1989)

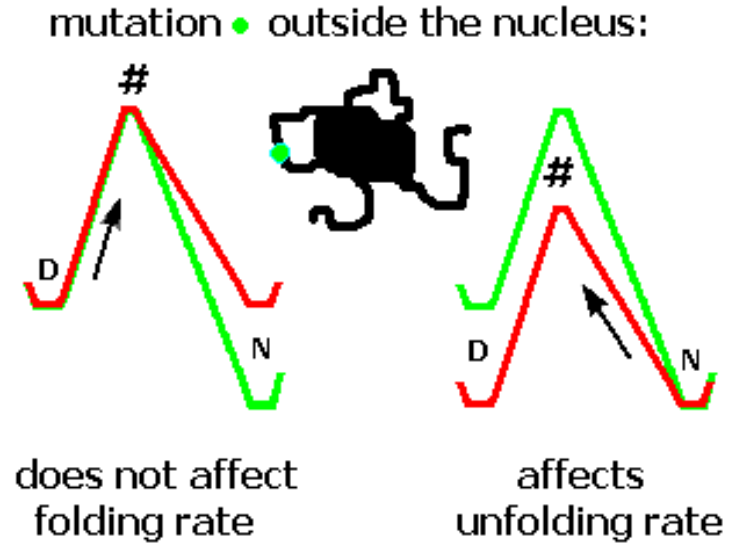


$$\phi = \frac{\Delta \ln(k_N)}{\Delta \ln(k_N/k_U)}$$



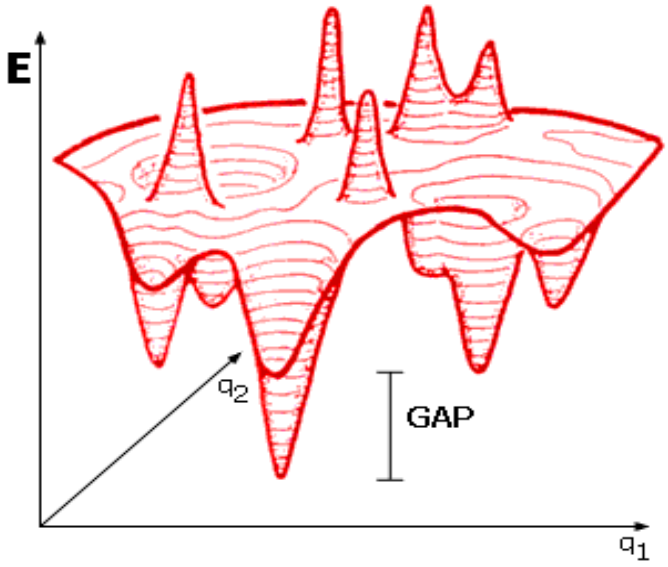


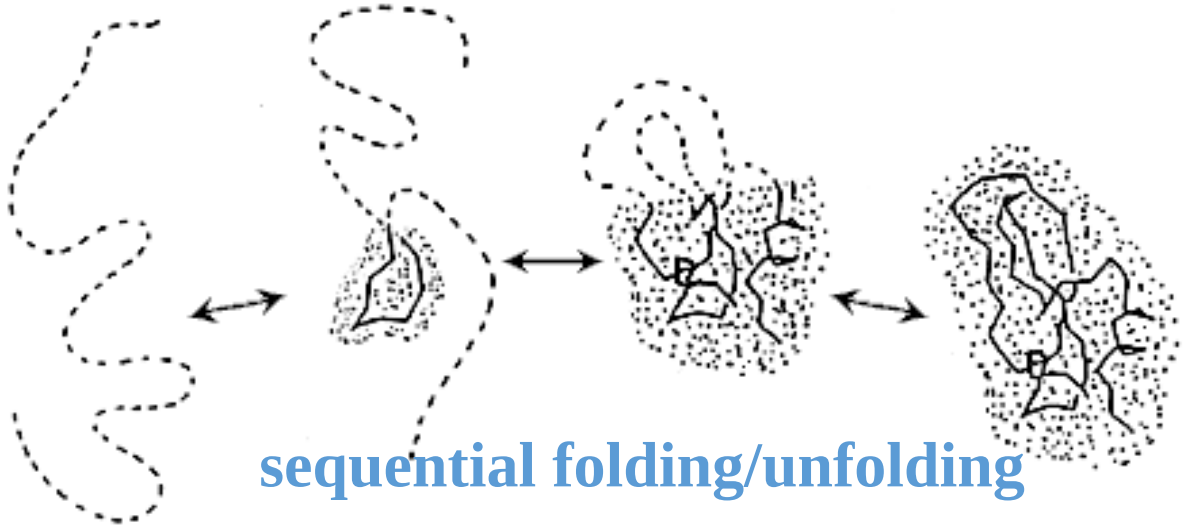




Solution of Levinthal paradox

Let us consider sequential folding (or unfolding) of a chain that has a large energy gap between the most stable fold and the bulk of the other ones; and let us consider its folding close to the thermodynamic mid-transition





sequential folding/unfolding

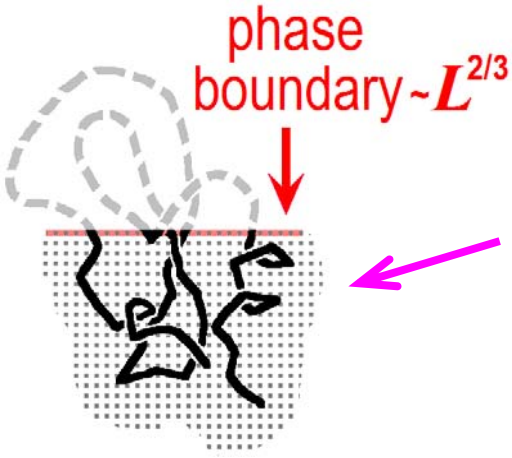
The same pathways: "detailed balance"

How fast the most stable fold will be achieved?

Note. Elementary rearrangement of 1 residue takes 1-10 ns. Thus, 100-residue protein would fold within ms, if there were no free energy barrier at the pathway…

HOW FAST the most stable state is achieved?

free energy barrier →



$\rightarrow \Delta F^{\#} \sim L^{2/3} \cdot$ surface tension

a) micro-; b) loops

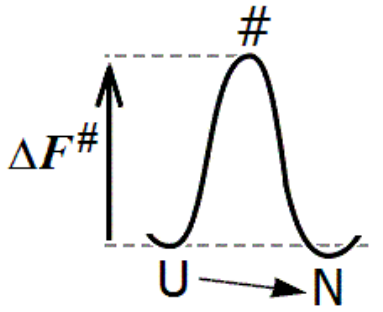
$\max\{\Delta F^{\#}\}$: when

compact folded nucleus: ~1/2 of the chain

micro: $\Delta F^{\#} \approx L^{2/3} \cdot [\varepsilon/_4]$; $\varepsilon \approx 2RT$ [experiment]

loops: $\Delta F^{\#} \leq L^{2/3} \cdot {}^1/_2 [{}^3/_2 RT \cdot \ln(L^{1/3})] \cdot + L/(\sim 100)$

[Flory] [knots]



rate ~ $\exp(-\Delta F^{\#}/kT)$ $\frac{1}{\text{ns}}$

$$\Delta F^{\#}/RT \sim ({}^1/_2 \div {}^3/_2)\, L^{2/3}$$

micro loops

***Any* stable fold is *automatically* a focus of rapid folding pathways:**
"Folding funnel" with phase separation. No "special pathway" is needed.