PROTEIN PHYSICS

LECTURES 19-21

- *In vivo* folding
- *In vitro* folding: spontaneously
- Levinthal paradox: spontaneously - how?
- Protein folding intermediates
- Two-state folding
- Transition state and protein folding nucleus
- Folding rate theory: solution of Levinthal's paradox

BASIC FACTS:

- ***In vivo*** (in the cell):

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

- Biosynthesis + Folding < 10 – 20 min.

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

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

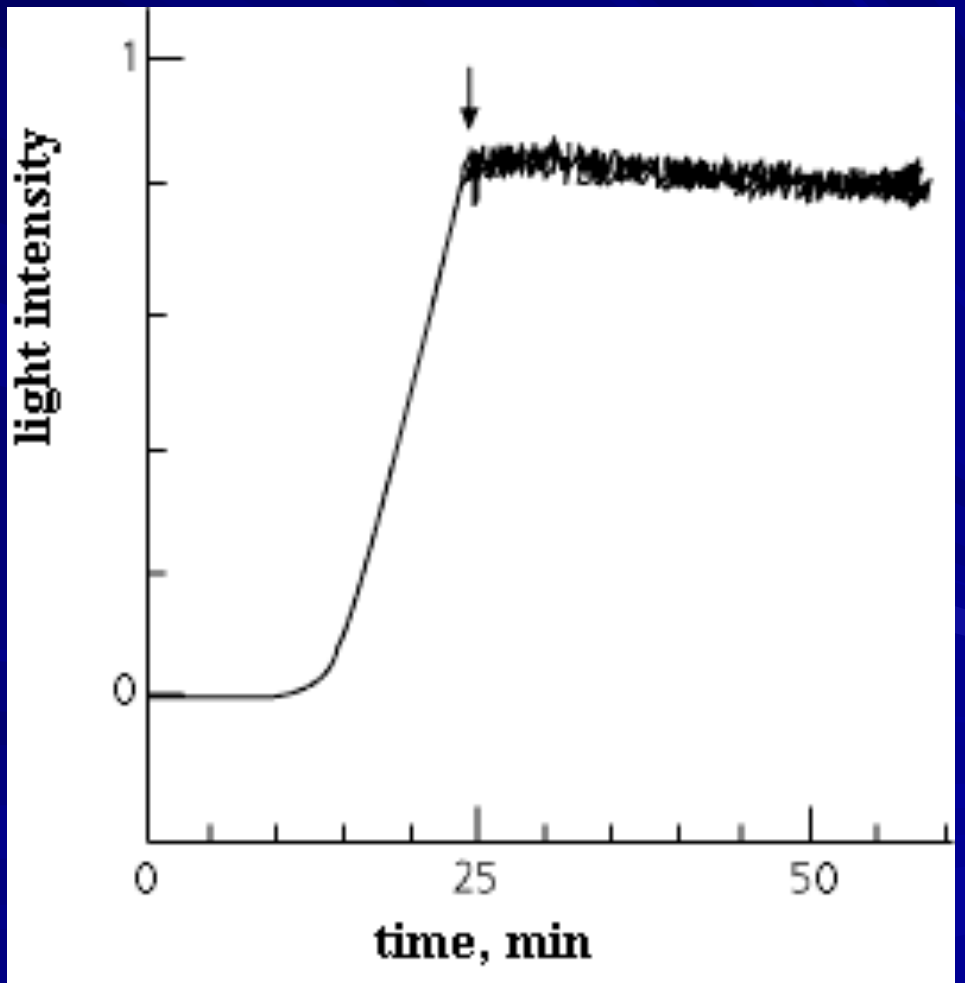
- The main obstacle for *in vivo* folding experiments: nascent protein is small, ribosome (+ …) is large.

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

The main obstacle for *in vivo* folding experiments:
nascent protein is small, ribosome (+ …) is large.
However, one can follow some "rare" protein activity, and use a "minimal" cell-free system

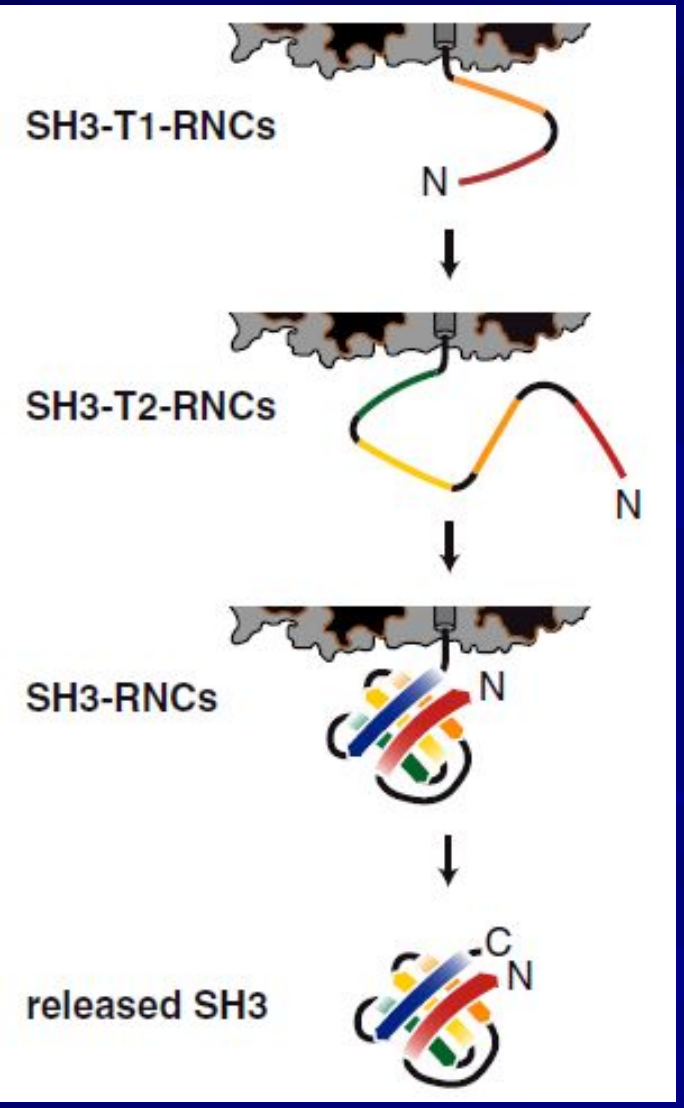
Luciferase activity

(Kolb, Makeev, Spirin, 1994)



Protein folding *in vivo* (at ribosome – at least for small proteins)

≈ as *in vitro*



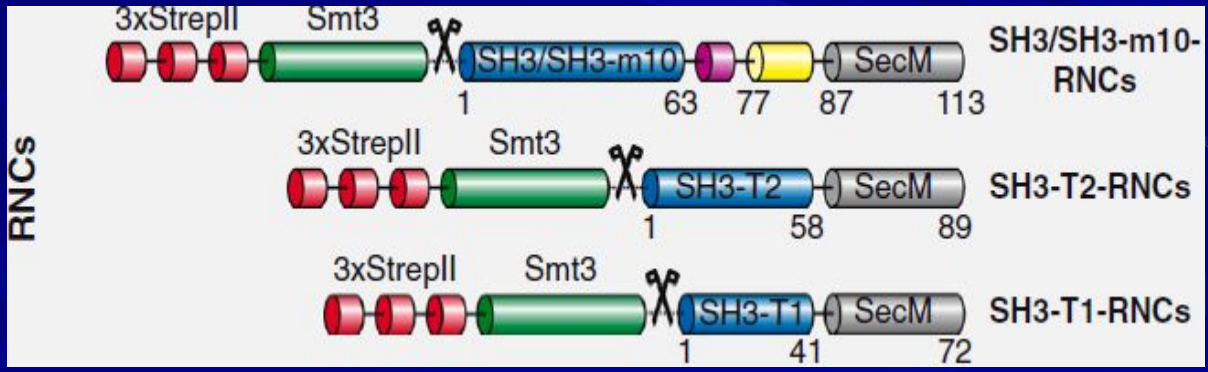
^{15}N, ^{13}C NMR:

Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy.

Eichmann C, Preissler S, Riek R, Deuerling E.

PNAS **107**, 9111 (2010):

«Polypeptides [at a ribosome] remain unstructured during elongation but fold into a compact, native-like structure when the entire sequence is available.»



Protein folding *in vivo* (at ribosome)

^{15}N, ^{13}C NMR:

Monitoring cotranslational protein folding in mammalian cells at codon resolution.

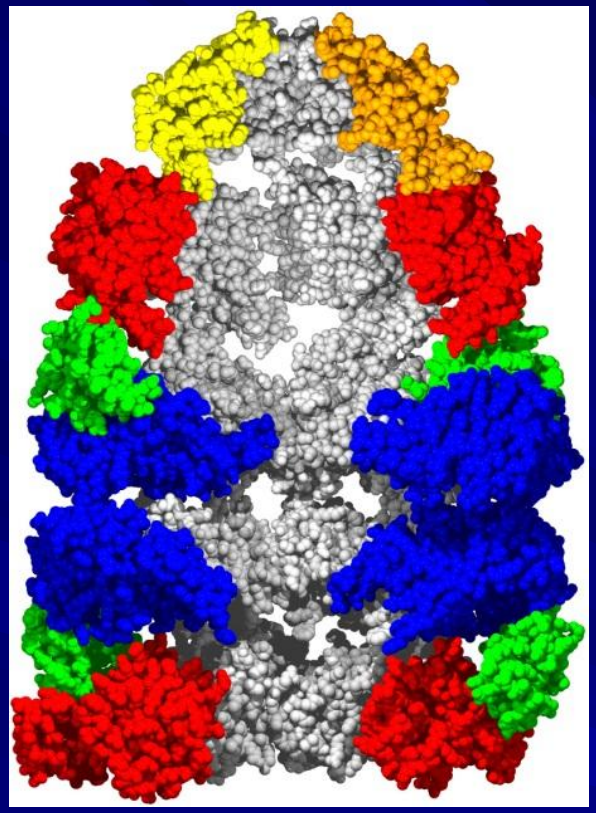
Han Y., David A., Liu B., Magadán J,G., Bennink J.R., Yewdell J.W., Qian S.-B.

PNAS **109**, 12467 (2012):

«…folding immediately after the emergence of the full domain sequence.»

«… displaying two epitopes simultaneously when the full sequence is available.»

Chaperone



GroEL/ES

Folding:

inside or outside

GroEL/ES?

- OUTSIDE

«Active action»? -- NO

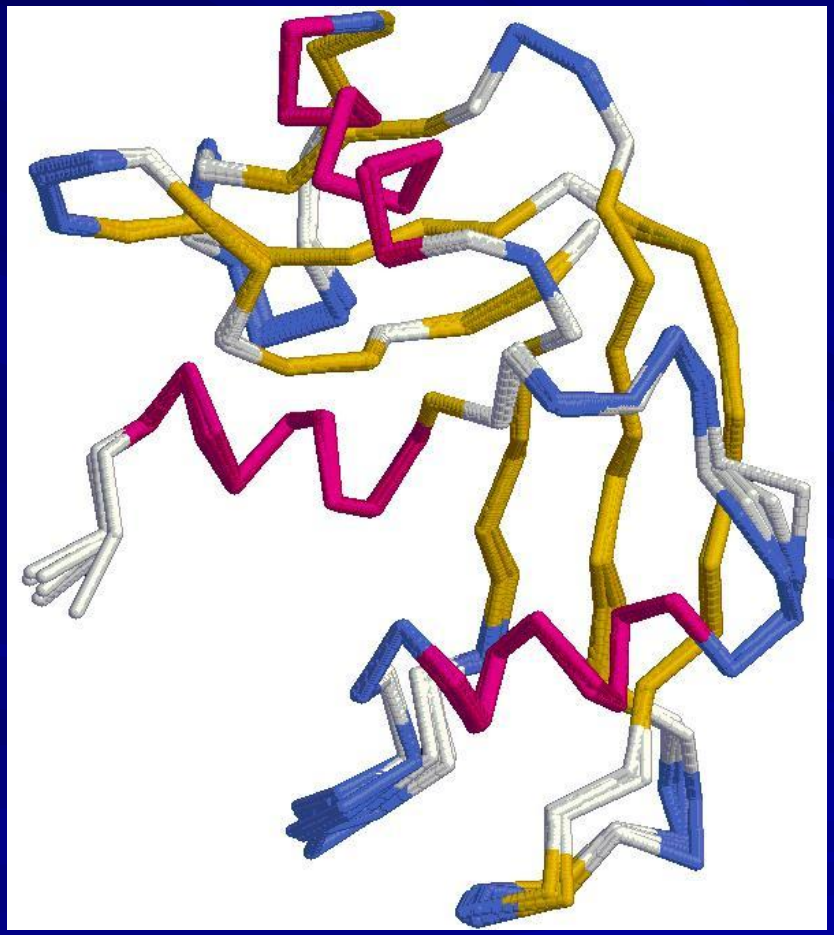
«Anfinsen cage»?

Ellis R.J. 2003
Curr. Biol. 13:R881-3

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

“ambidextrous chaperone activity“
(Weinstock, Jacobsen, Kay, 2014,
PNAS 111(32):11679-84)

PROTEIN CHAIN
CAN FORM ITS UNIQUE 3D STRUCTURE
SPONTANEOUSLY *IN VITRO*
(Anfinsen, 1961: Nobel Prize, 1972)



BASIC FACTS:

· ***In vitro*** (in physico-chemical experiment):

- Unfolded globular **protein is capable of renaturation** (if it is not too large and not too modified chemically after the biosynthesis), i.e., its 3D structure is capable of spontaneous folding [Anfinsen, 1961].

- Chemically synthesized protein chain achieves its correct 3D structure [Merrifield, 1969].

- The main obstacle for *in vitro* folding is aggregation.

Conclusion: Protein structure is determined by its amino acid sequence;
cell machinery is not more than an “incubator” for protein folding.

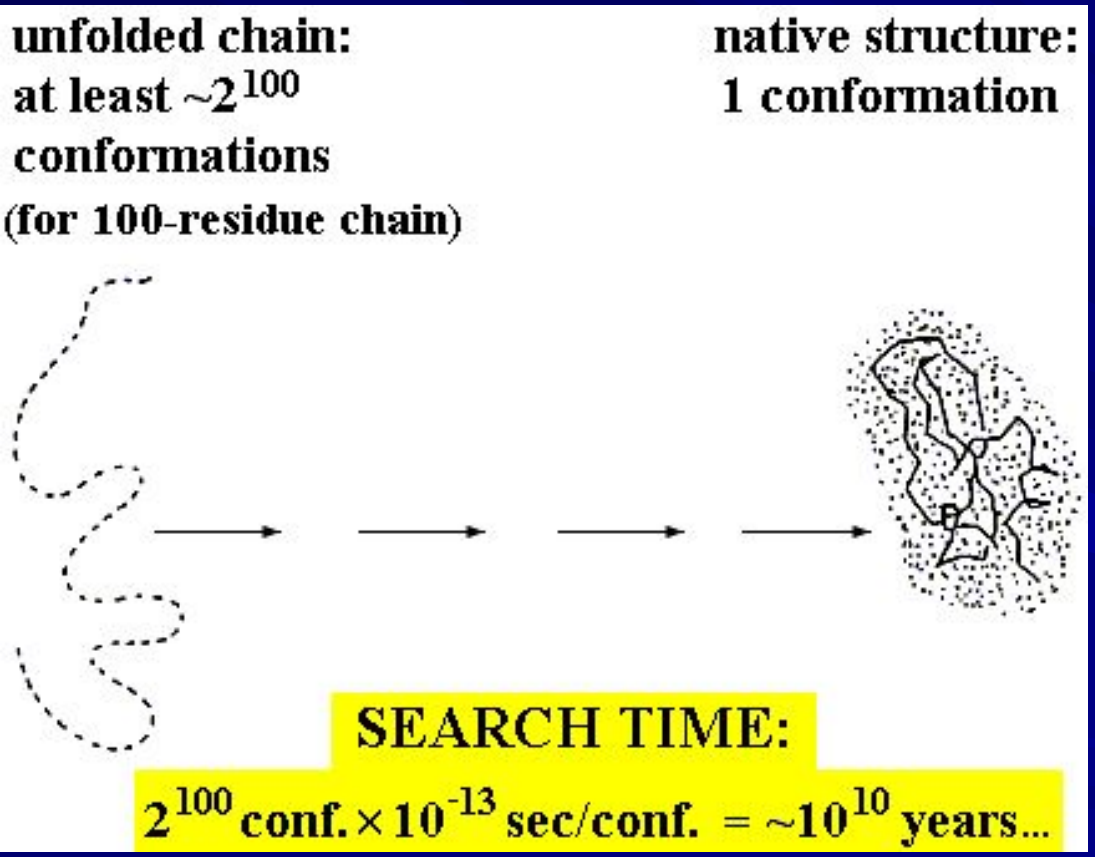
Christian Boehmer **Anfinsen,** Jr.
(1916 –1995)
Nobel Prize 1972

Robert Bruce **Merrifield**
(1921 – 2006)
Nobel Prize 1988

Cyrus **Levinthal**
(1922 –1990)

HOW DOES PROTEIN FOLD?
and even more:
How CAN protein fold spontaneously?

Levinthal paradox (1968):

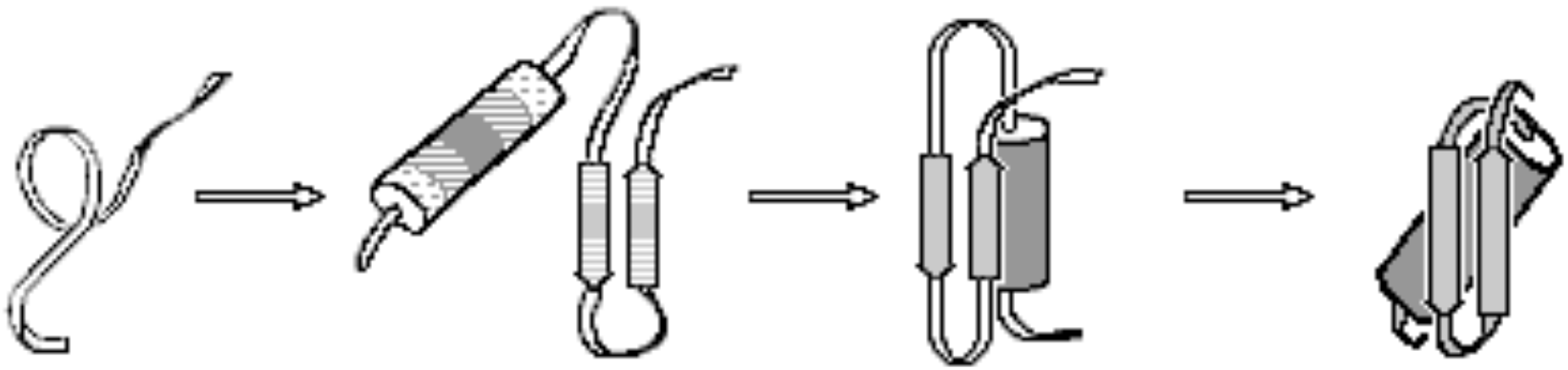


Native protein structure reversibly refolds from various starts, i.e., it is thermodynamically stable.

But how can protein chain find this unique structure - within seconds - among zillions alternatives?

SPECIAL PATHWAYS?? FOLDING INTERMEDIATES??

“Framework model” of stepwise folding (Ptitsyn, 1973)



Unfolded chain

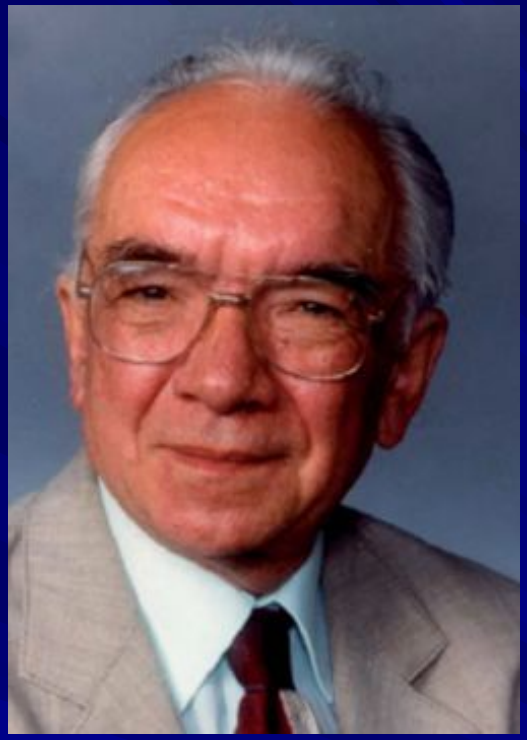
Secondary structure fluctuating around its native position

Native-like secondary structure and folding pattern

Native 3D structure

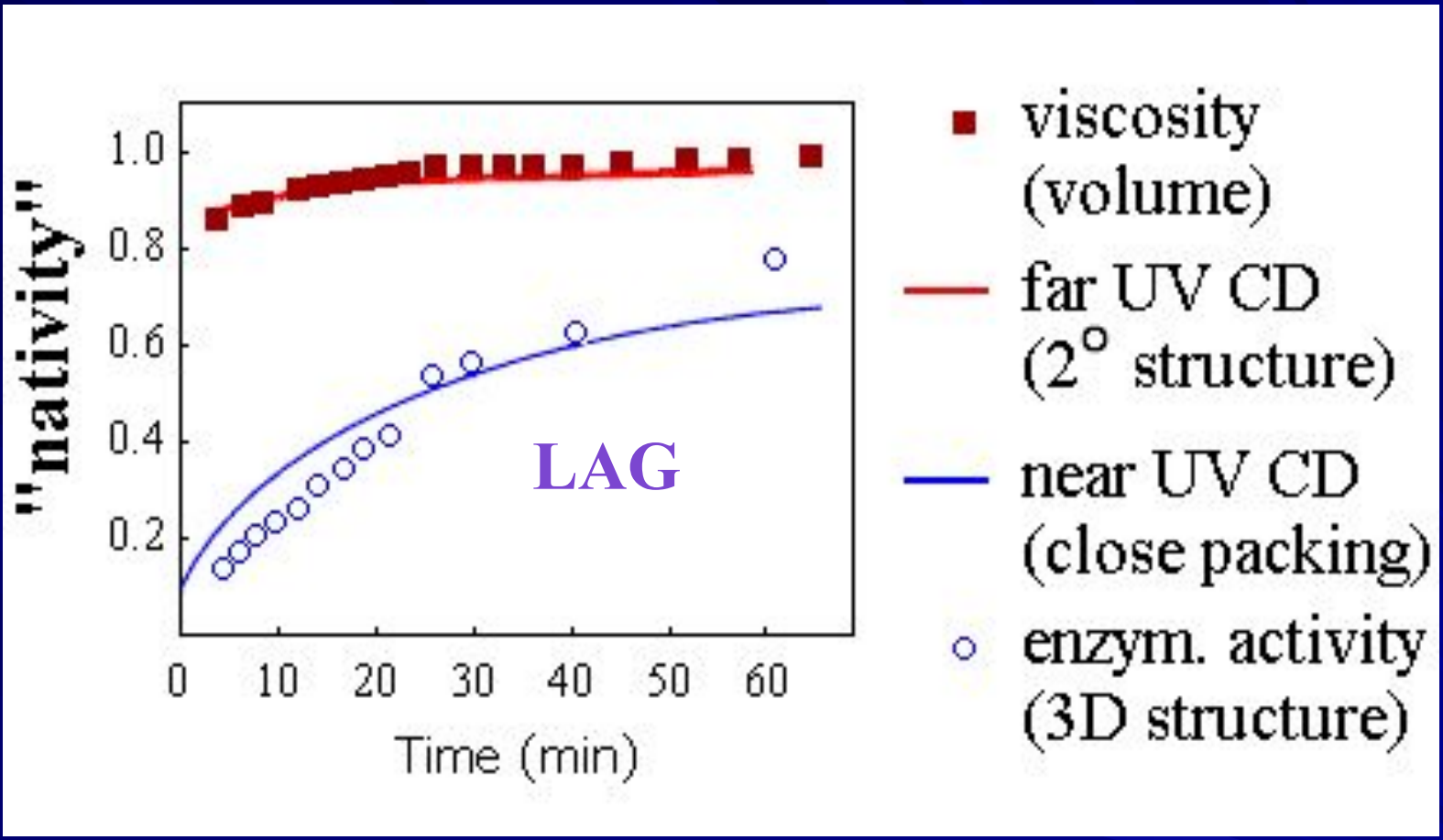
Now: Pre-molten globule

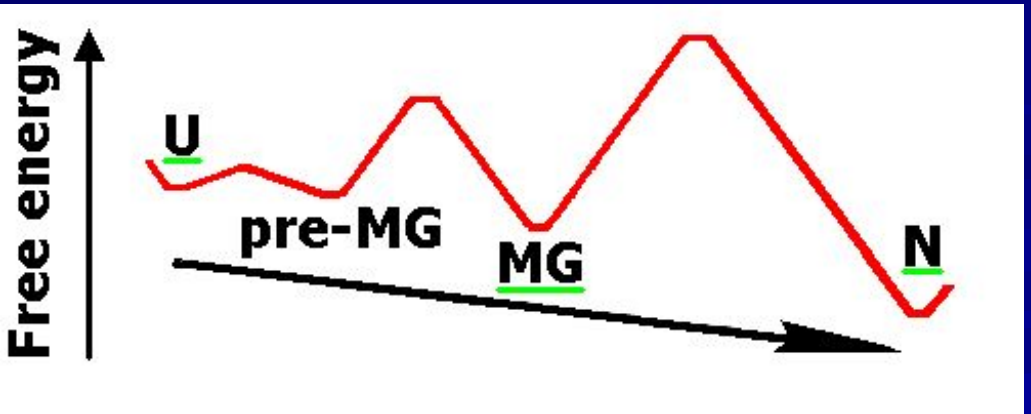
Now: Molten globule



Oleg Borisovich
Ptitsyn
(1929-99)

Kinetic intermediate (molten globule) in protein folding





(Doldikh,…, Ptitsyn, 1984)

Multi-state folding

Found: metastable ("accumulating", "directly observable") folding intermediates.

The idea was: intermediates will help to trace the folding pathway,
- like intermediates in a biochemical reaction trace its pathway.

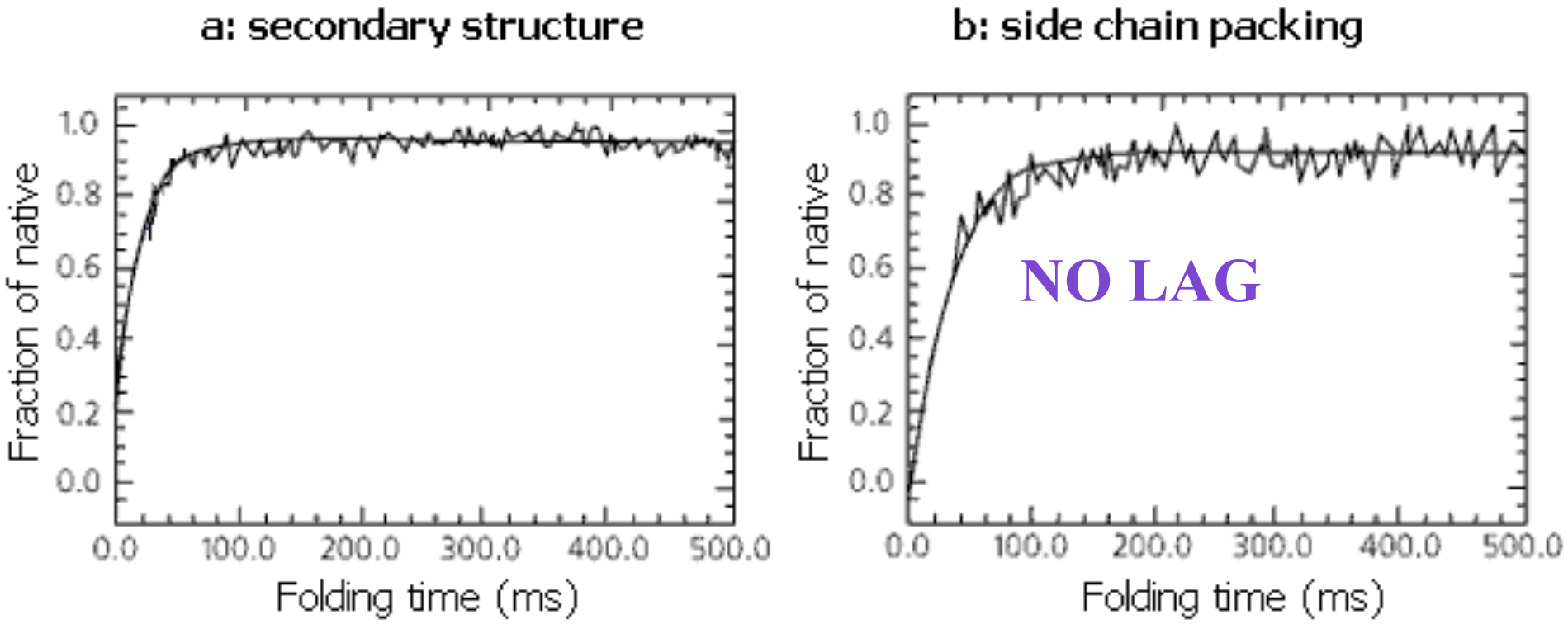
This was a "chemical logic".

However, although protein folding intermediates (like MG) were found for many proteins, the main question as to how the protein chain can find its native structure among zillions of alternatives remained unanswered.

A progress in the understanding was achieved when studies involved small proteins (of 50 - 100 residues).

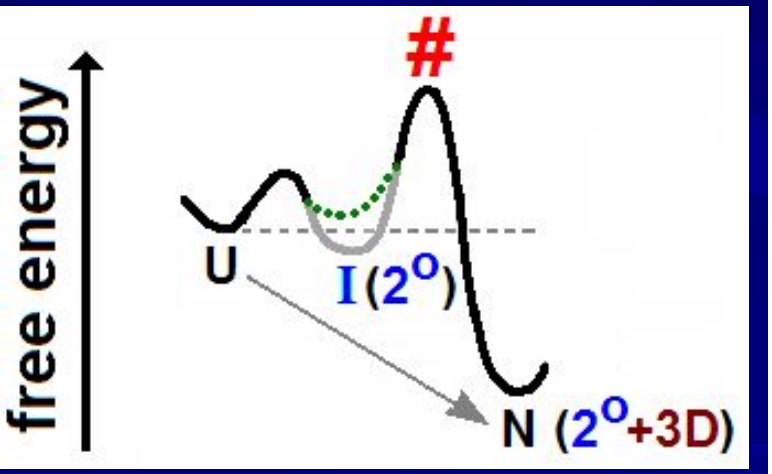
Many of them are "two-state folders": they fold *in vitro* without any observable (accumulating) intermediates, and have only two observable states: the native fold and the denatured coil.

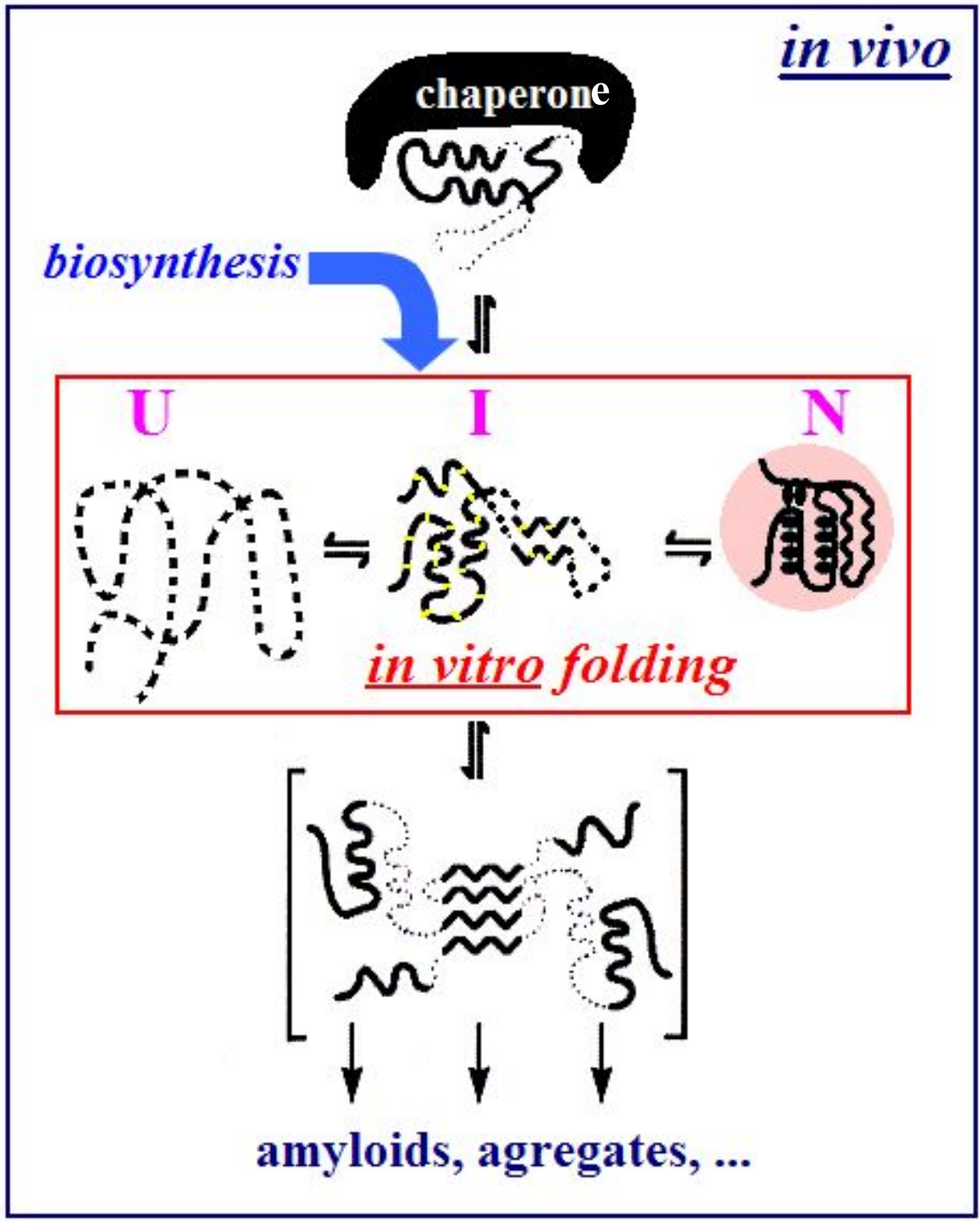
"Two-state" folding: without any observable (accumulating in experiment) intermediates



The "two-state folders" fold rapidly: not only much faster than larger proteins (not a surprise), but as fast as small proteins having folding intermediates (that were expected to accelerate folding…)

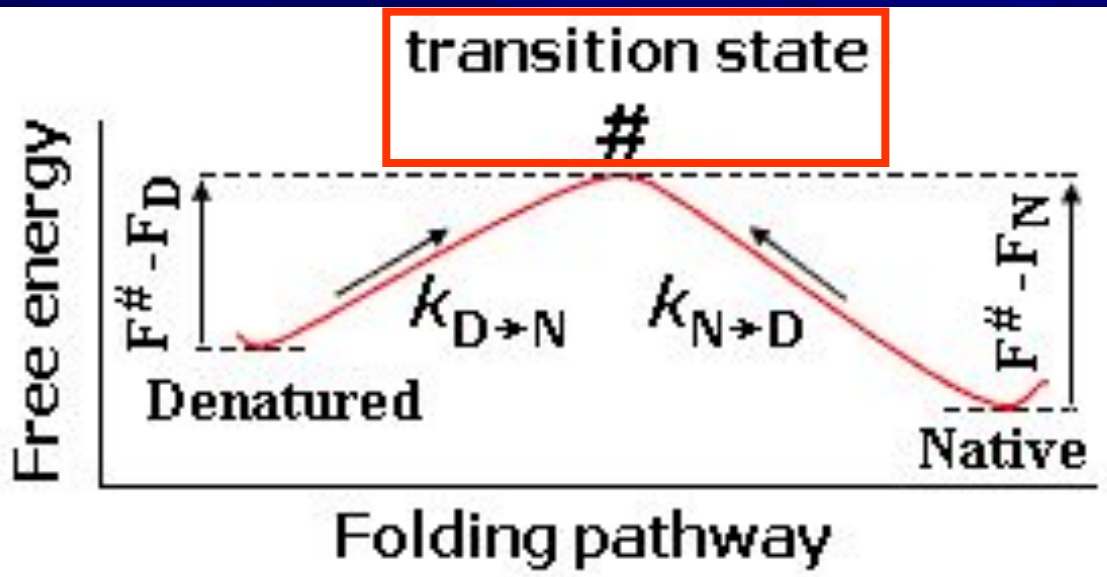
PROTEIN FOLDING: current picture





What to study in the "two-state" folding where there are no intermediates to single out and investigate?

Answer: just here one has the best opportunity to study the *transition state*, the bottleneck of folding.



"detailed balance": the same pathways for D→N and N→D

$$k_{D\to N} = k_0 \exp[-(F^{\#}-F_D)/RT], \quad k_{N\to D} = k_0 \exp[-(F^{\#}-F_N)/RT]$$

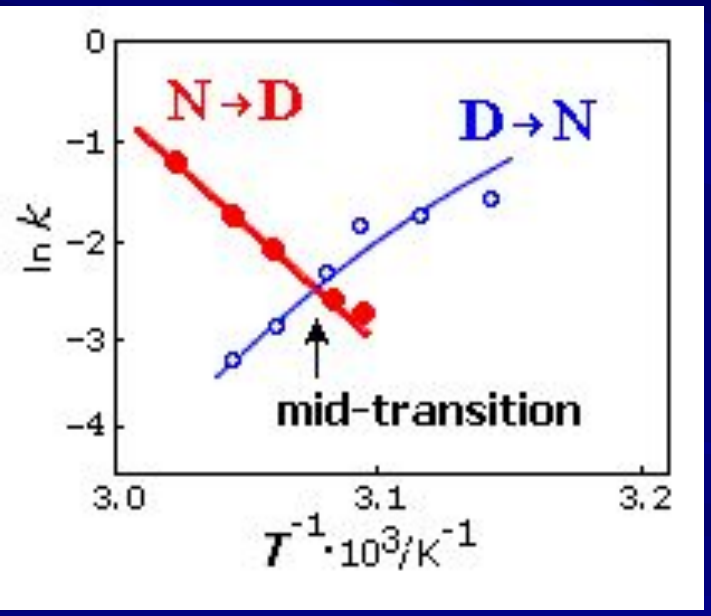
EXPERIMENT GIVES: $k_{\text{apparent}(D\leftrightarrow N)} = k_{D\to N} + k_{N\to D}$

to equilibrium

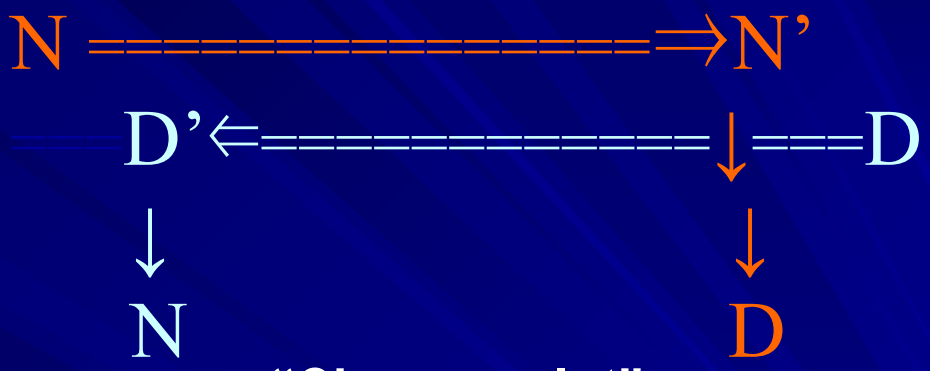
$$-\ln(k_{D\to N}/k_{N\to D}) = (F_N - F_D)/RT$$

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

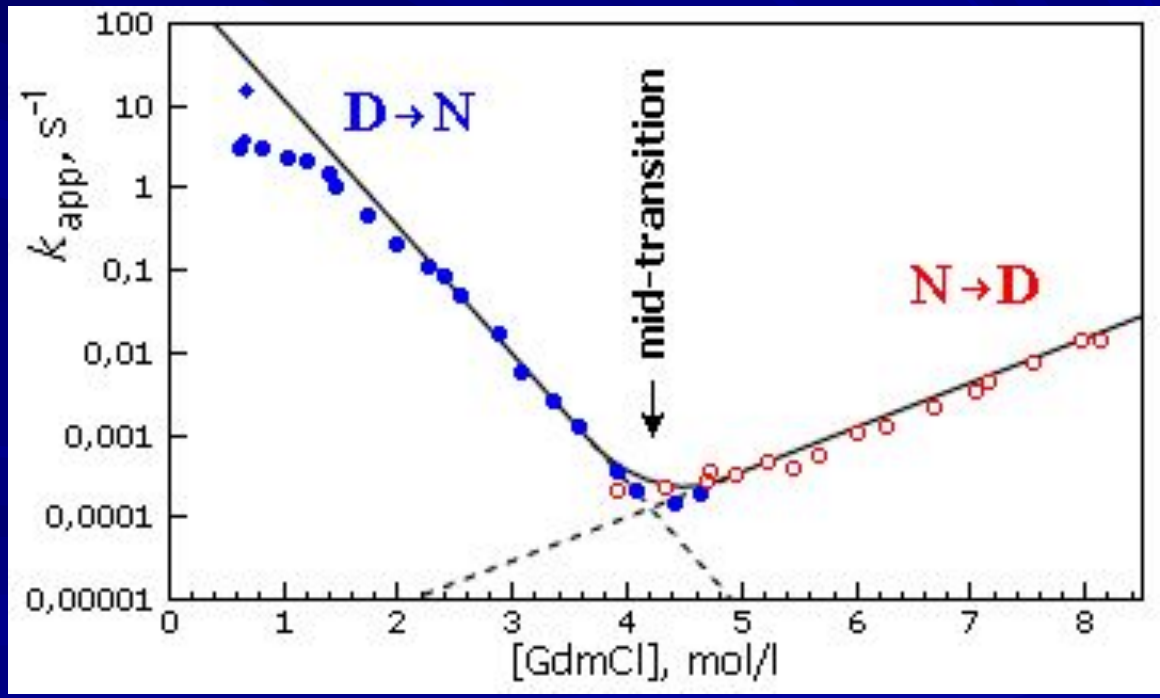
(a)



(b)



"Chevron plot"

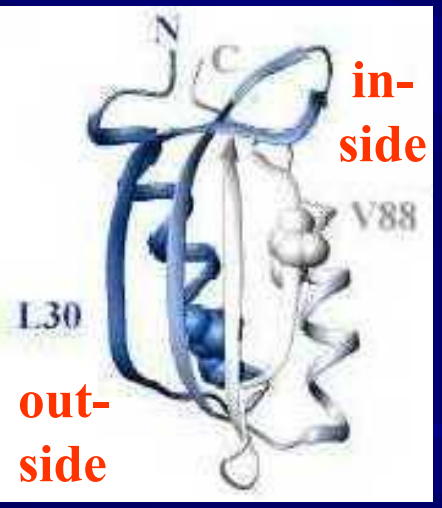


Sir Alan Roy Fersht, 1943

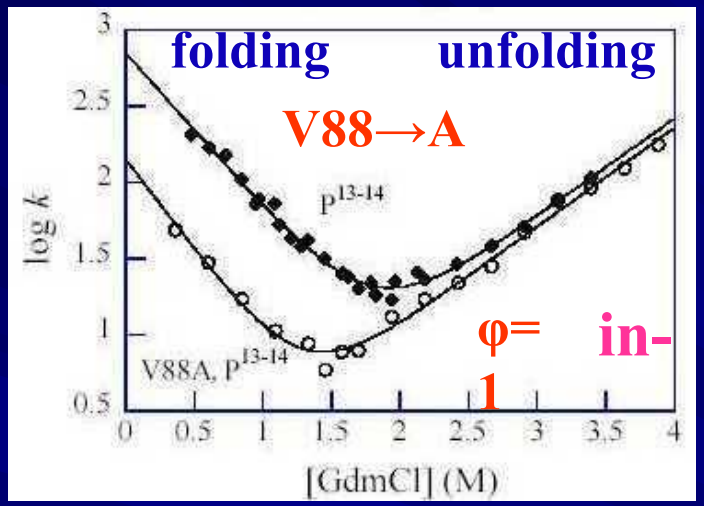
Protein engineering

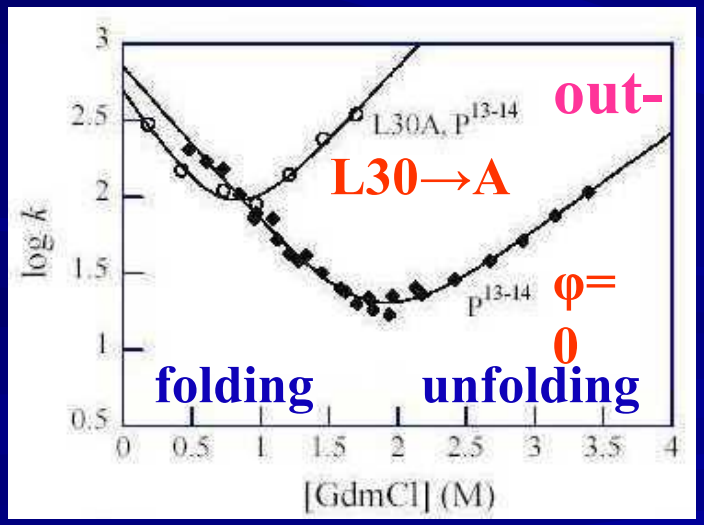
Folding nucleus

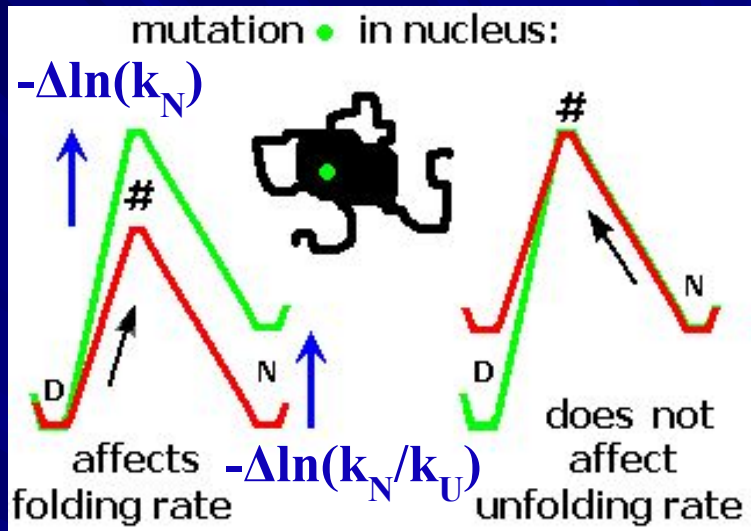
Folding nucleus: Site-directed mutations show residues belonging and not-belonging to the "nucleus", the native-like part of transition state (Fersht, 1989)

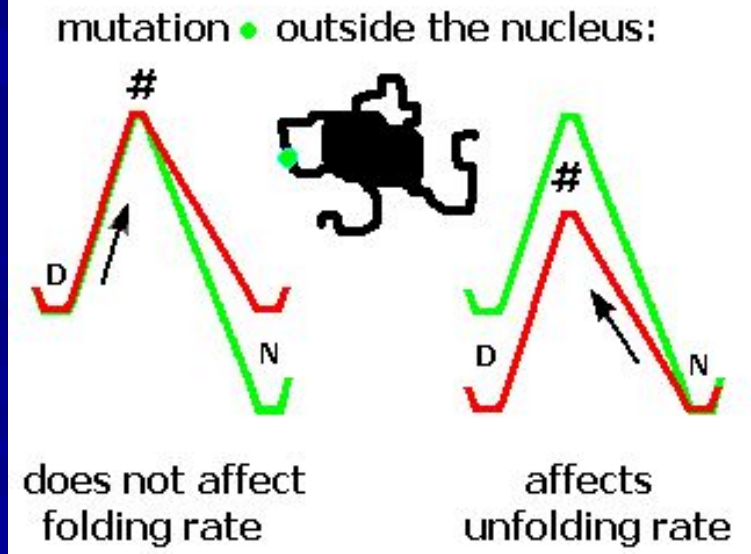


$$\varphi = \frac{\Delta \ln(k_N)}{\Delta \ln(k_N/k_U)}$$



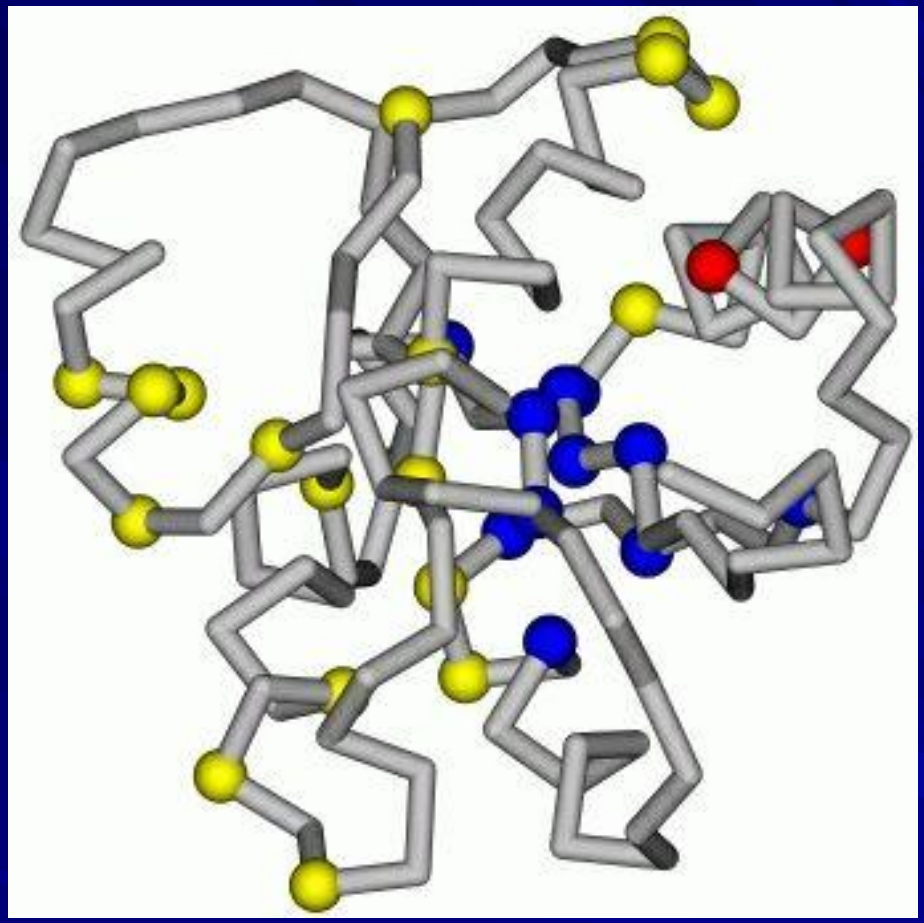






Folding nucleus in CheY protein

(Lopez-Hernandes & Serrano, 1996)



n In nucleus

n Outside

“difficult”

Folding nucleus is often shifted to some side of protein globule and does not coincide with its hydrophobic core; folding nucleus is NOT a molten globule

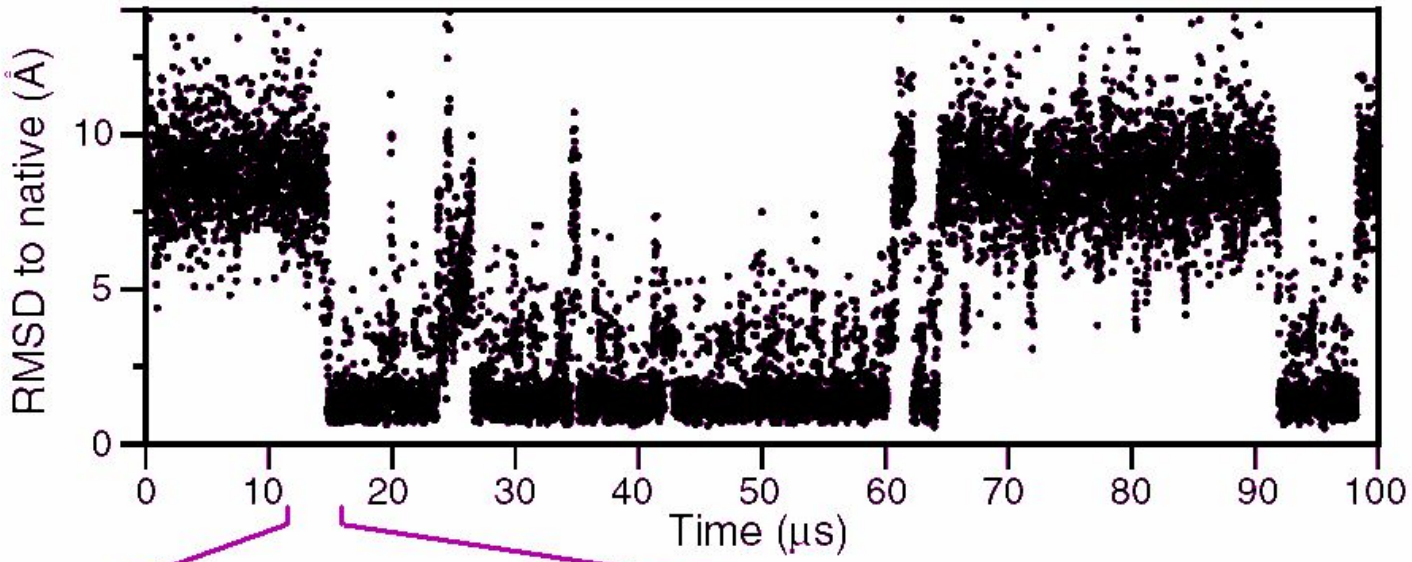
“Hot point” in protein physics: advanced MD simulations

David E. Shaw

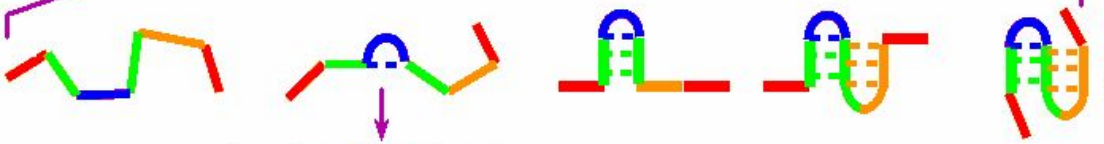
“D. E. Shaw Research”
US$ 3.5 billion
Supercomputer “Anton”

FiP35 protein: simulation of folding

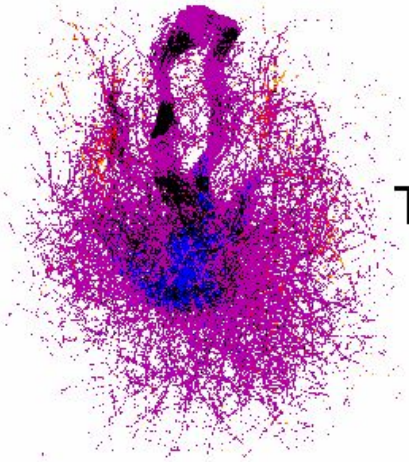
D.E. Shaw et al., Oct. 2010, *Science* **330**, 341



folding pathway



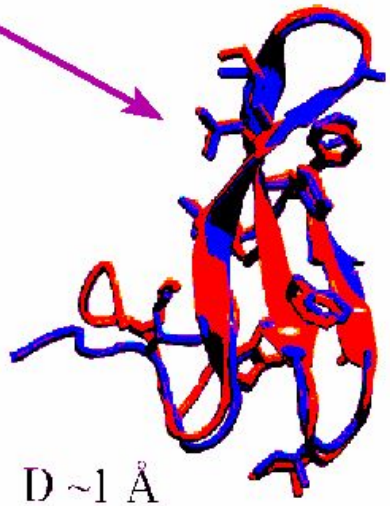
Coil



TS ensamble

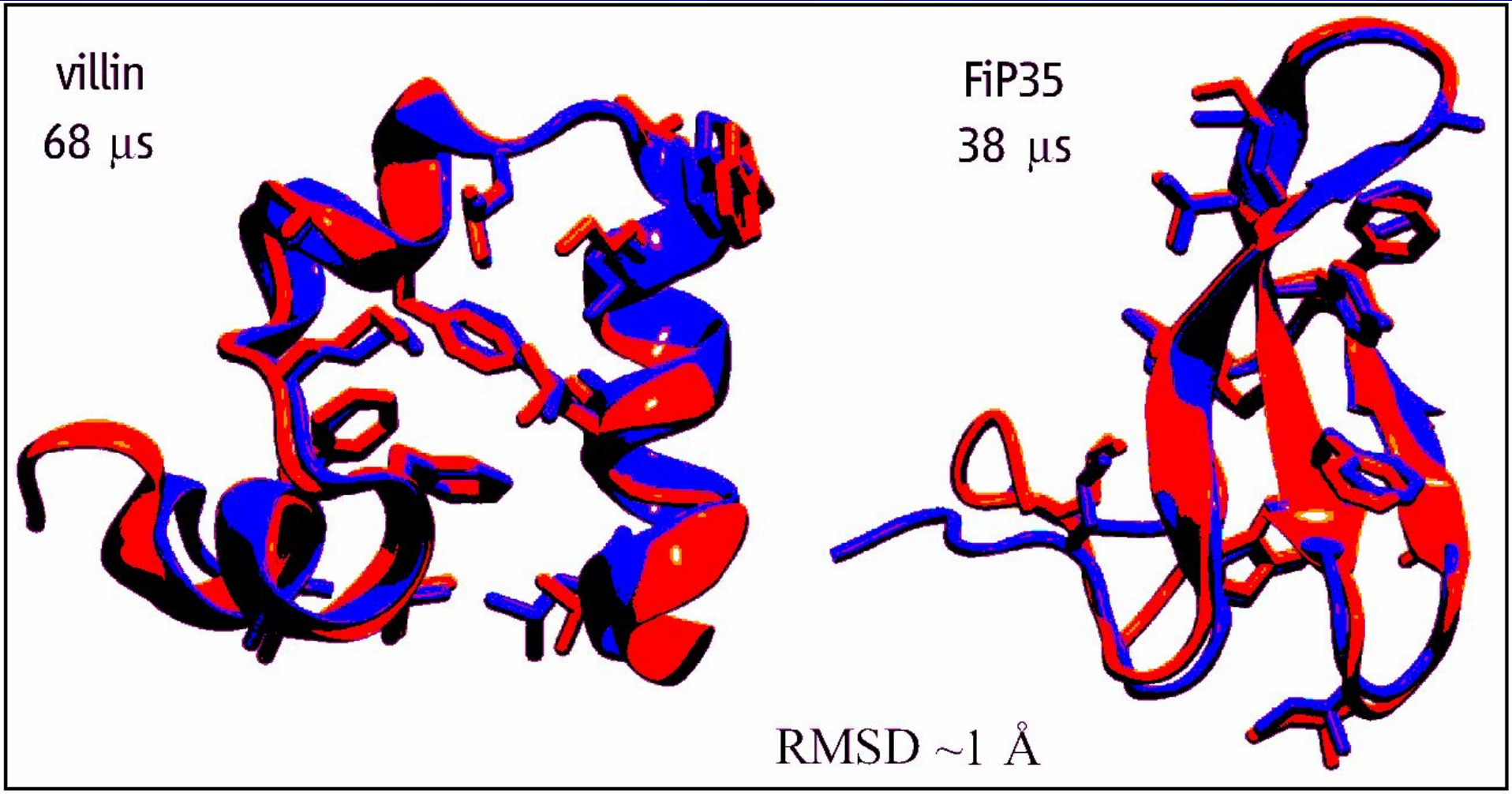
phase separation

Native



D ~1 Å

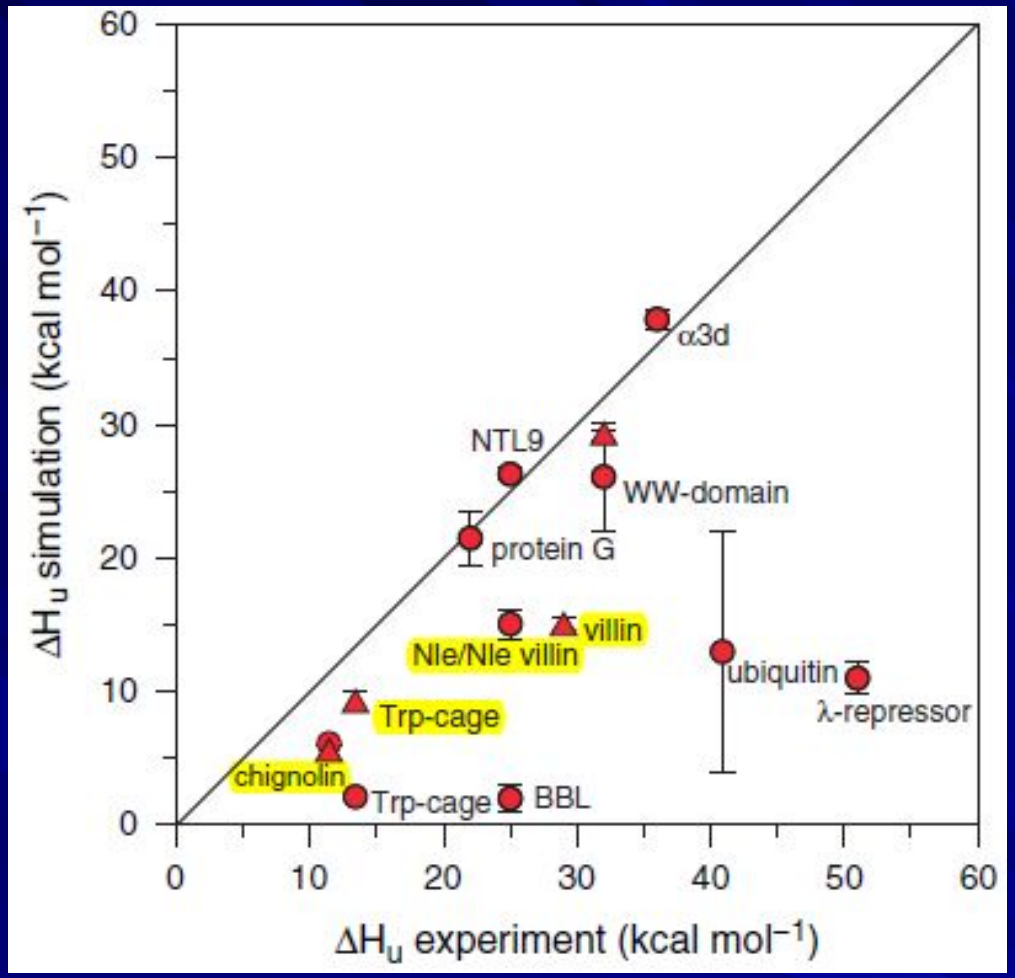
“A priory” computed 3D folds of small proteins



modified version of the Amber ff99SB force field:

K. Lindorff-Larsen *et al.*, *Proteins* **78**, 1950 (2010)

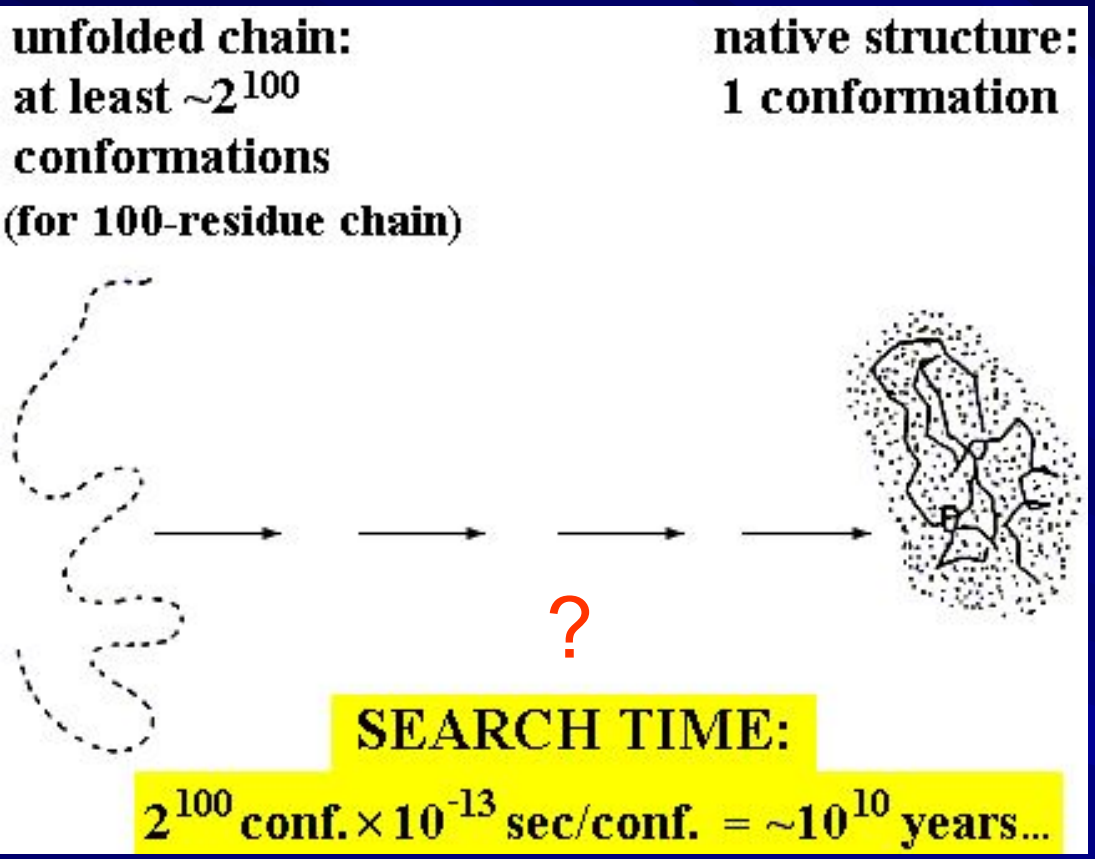
D.E. Shaw et al., Oct. 2010, *Science* **330**, 341



BUT: unfolding enthalpies are predicted VERY BADLY!

S. Piana, J.L. Klepeis, D.E Shaw
Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations
Current Opinion in Structural Biology 2014, 24:98–105

Back to Levinthal paradox

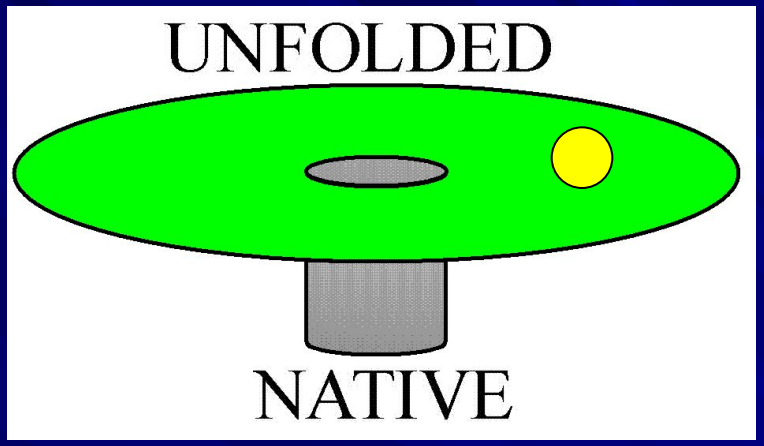


Native protein structure reversibly refolds from various starts, i.e., it is thermodynamically stable.

But how can protein chain find this unique structure - within seconds - among zillions alternatives?

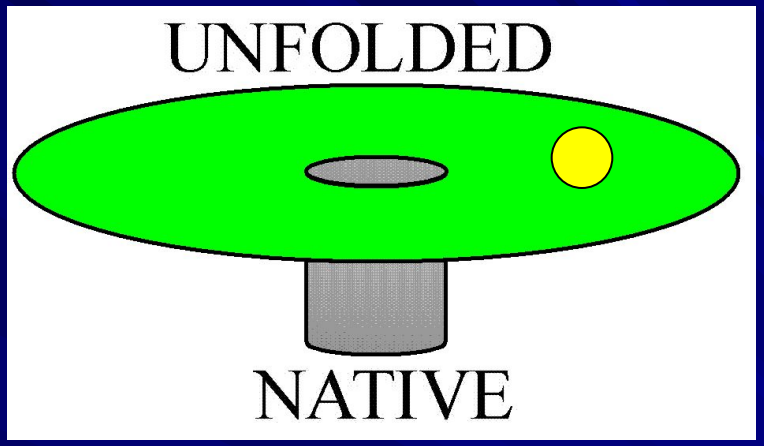
However, the same problem – how to find one configuration among zillions – is met by crystallization and other 1-st order phase transitions.

Is “Levinthal paradox” a paradox at all?



***L*-dimensional “Golf course”**

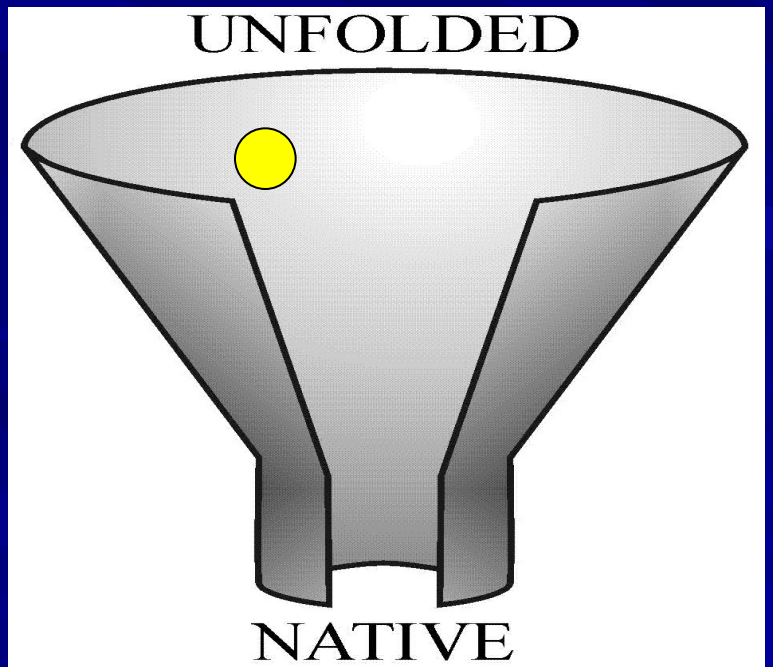
Is “Levinthal paradox” a paradox at all?



***L*-dimensional “Golf course”**

…any tilt of energy surface solves this “paradox”… (?)

“Funnel”: entropy_by_energy compensation

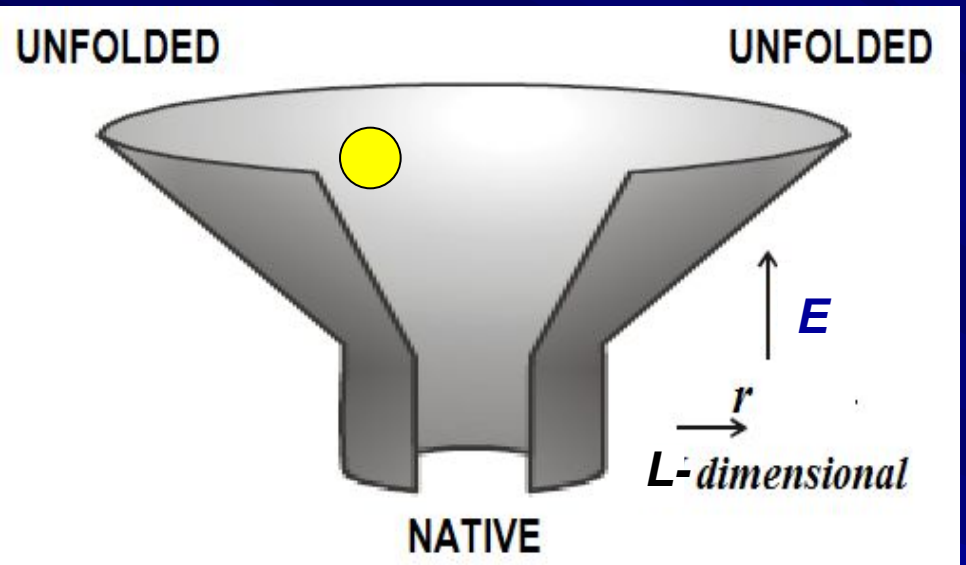


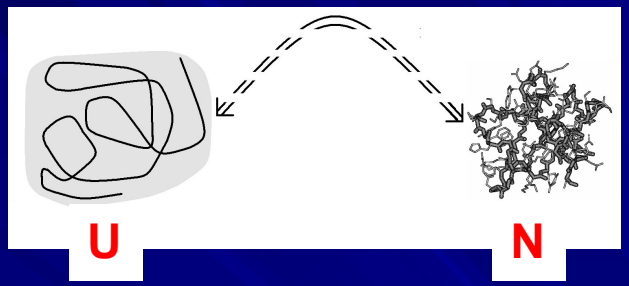
Simple *L*-dimensional “funnel” (without phase separation)

Zwanzig, 1992;
Bicout & Szabo, 2000

L-dimensional "folding funnel"?

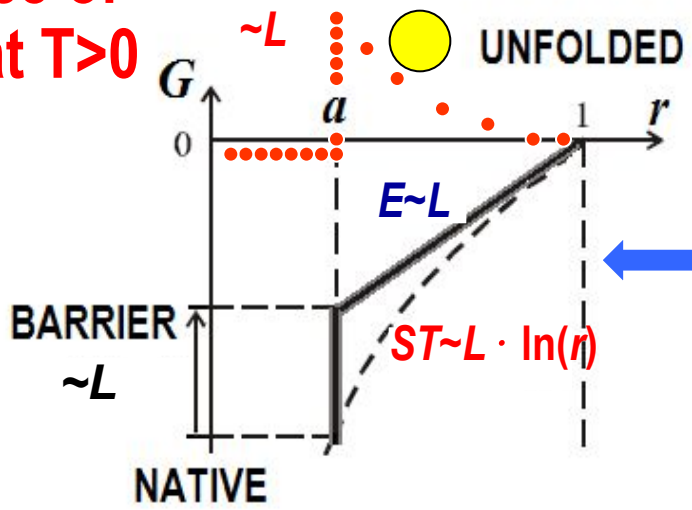
Sly simplicity of a "folding funnel" (without phase separation)





All-or-none transition for 1-domain proteins
(in thermodynamics: Privalov,1974; in kinetics: Segava, Sugihara,1984)

Resistance of entropy at T>0



- NO *simultaneous* explanation to
(I) "all-or-none" transition
(II) folding within non-astron. time at mid-transition

Funnel helps, but ONLY when T is much lower than $T_{mid\text{-}tr.}$!!

A special pathway?

Phillips (1965) hypothesis:

folding nucleus is formed by the N-end of the nascent protein chain, and the remaining chain wraps around it.

for single-domain proteins: **NO:**

Goldenberg & Creighton, 1983:

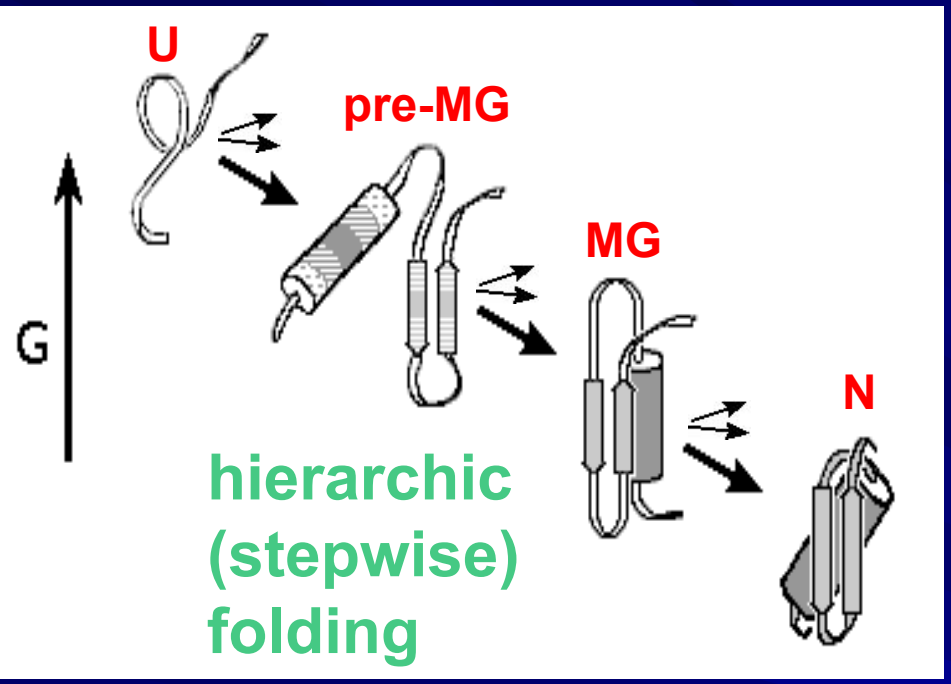
circular permutants:
N-end has no special role in the *in vitro* folding.

However, for **many**-domain proteins:

Folding from N-end domain, ≈ domain after domain

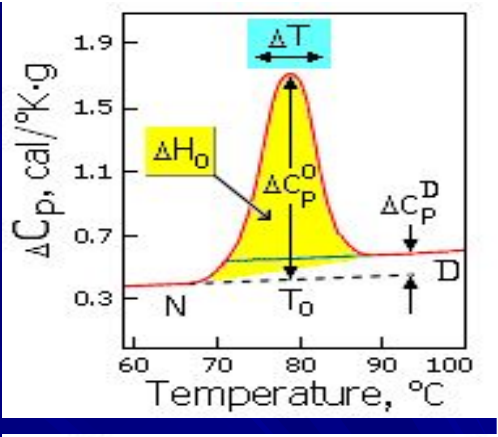
DO NOT CONFUSE N-END DRIVEN FOLDING WITHIN DOMAIN
(which seems to be absent)
and
N-DOMAIN DRIVEN FOLDING IN **MANY**-DOMAIN PROTEIN
(which is observed indeed)

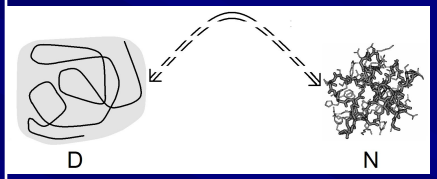
Sly simplicity of hierarchic folding
as applied to resolve the Levinthal paradox



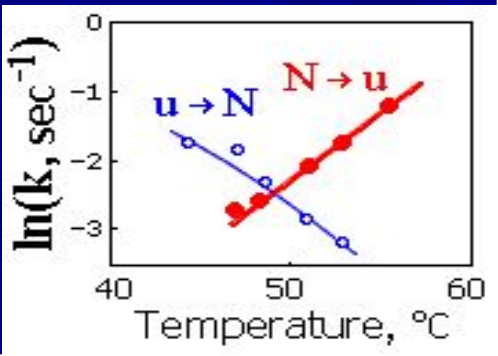
All-or-none transition:

In thermo-dynamics





In kinetics

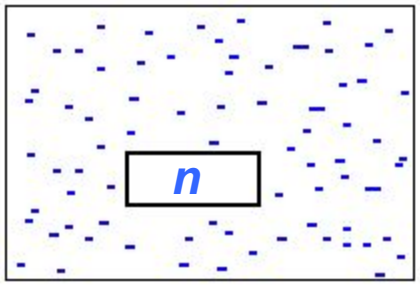


Folding intermediates must become more and more stable for hierarchic folding. This cannot provide *a simultaneous* explanation to

(i) folding within non-astronomical time;

ii) "all-or-none" transition, i.e., co-existence of only native and denatured molecules in visible amount;

ii) the same 3D structure resulting from different pathways



1-st order phase transition: rate of nucleation

Crystallization, classic theory

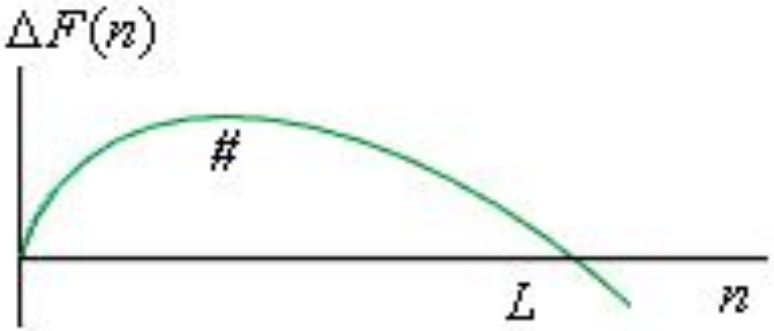
$$INITIATION\ TIME \sim \tau \times \exp(+\Delta F^{\#}/RT)$$

$\Delta F^{\#}$: free energy of critical nucleus

τ : time of 1 step (~ps ÷ ns)

CONSECUTIVE REACTIONS:

TRANSITION TIME ≅ SUM OF TIMES ≈ Max. barrier TIME



$\Delta F(n) = \Delta\mu n + Bn^{2/3}$: free energy of n-particle nucleus. New phase: $\Delta\mu < 0$.

Critical nucleus [from $d(\Delta F)/dn = 0$]: Max. ΔF: $\Delta F^{\#} = (4/27)B^3/(\Delta\mu)^2$

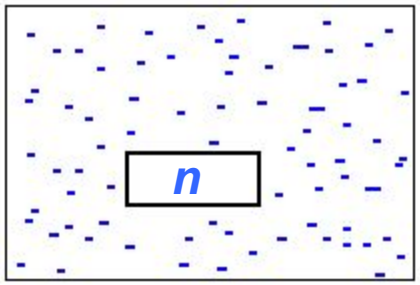
$$n^{\#} = (8/27)B^3/(-\Delta\mu)^3$$

Min. size L of stable new phase [from $\Delta F(L) = 0$]: $L = B^3/(-\Delta\mu)^3$

Then $\Delta F^{\#} = (4/27)BL^{2/3}$,

$$n^{\#} = (8/27)L$$

$$INITIATION\ TIME \sim \tau \times \exp(const \times L^{2/3})$$



1-st order phase transition: rate of nucleation

Crystallization, classic theory

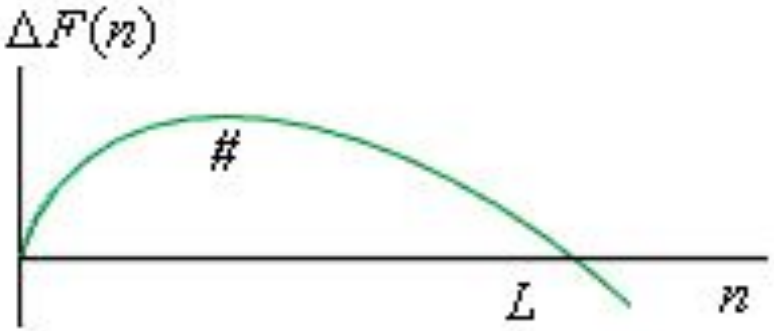
$$INITIATION\ TIME \sim \tau \times \exp(+\Delta F^{\#}/RT)$$

$\Delta F^{\#}$: free energy of critical nucleus

τ : time of 1 step (~ps ÷ ns)

CONSECUTIVE REACTIONS:

TRANSITION TIME ≅ SUM OF TIMES ≈ Max. barrier TIME



$\Delta\mu \approx -\Delta T \cdot (H_m/T_m)$ $\quad B \sim H_m$

$\Delta F(n) = \Delta\mu n + Bn^{2/3}$: free energy of n-particle nucleus. New phase: $\Delta\mu < 0$.

Critical nucleus [from $d(\Delta F)/dn = 0$]: Max. ΔF: $\Delta F^{\#} = (4/27)B^3/(\Delta\mu)^2$

$n^{\#} = (8/27)B^3/(-\Delta\mu)^3$

Min. size L of stable new phase [from $\Delta F(L) = 0$]: $L = B^3/(-\Delta\mu)^3 \approx (T_m/\Delta T)^3$

Then $\Delta F^{\#} = (4/27)BL^{2/3}$,

$n^{\#} = (8/27)L$

ALL → ∞ at ΔT → 0

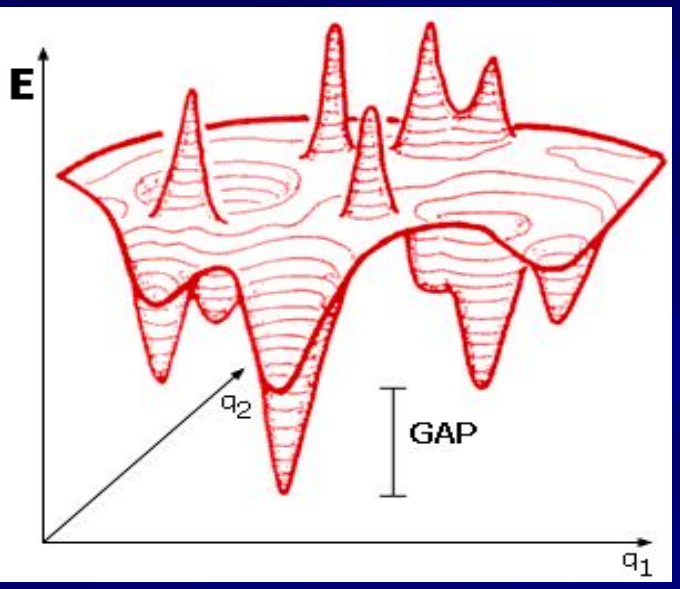
$$INITIATION\ TIME \sim \tau \times \exp(const \times L^{2/3})$$

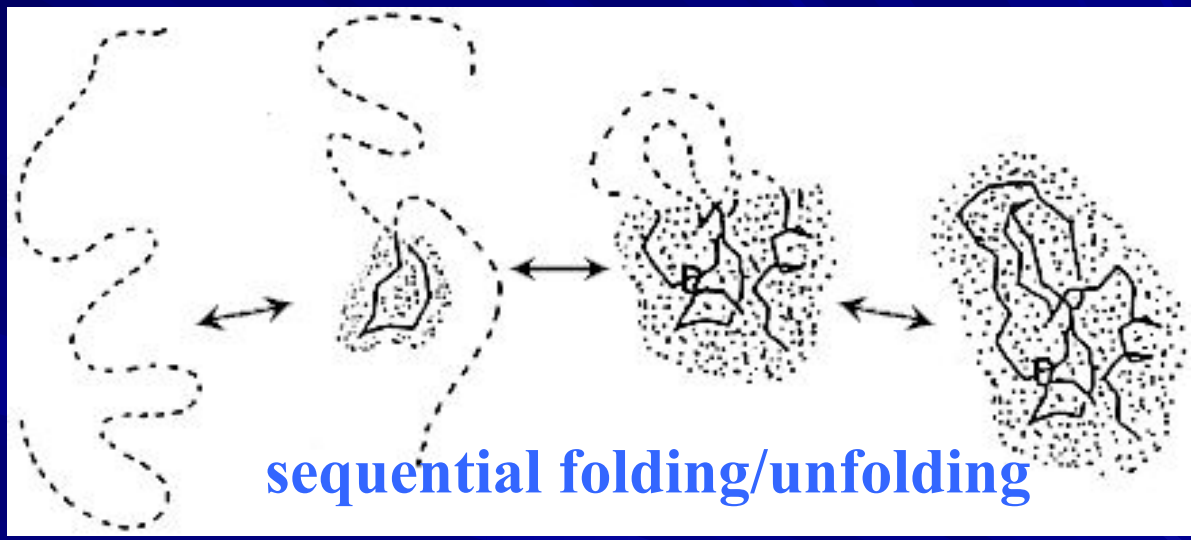
For macroscopic bodies ↓

ACTUALLY: hysteresis… INITIATION at walls, admixtures, …

For proteins, the microscopic bodies ↓

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





The same pathways: "detailed balance"

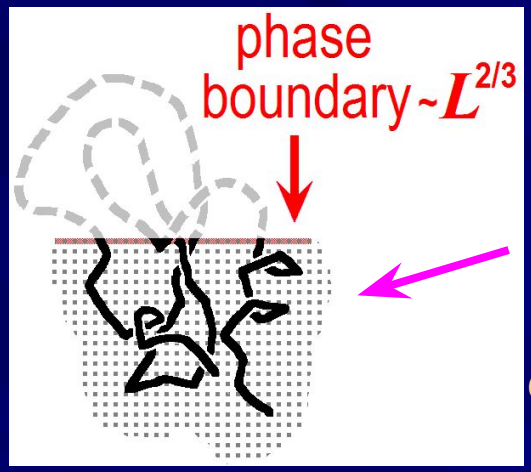
How fast the most stable fold will be achieved?

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

HOW FAST the most stable state is achieved?

free energy barrier →

F (U) loops = F (N)



→ $\Delta F^{\#} \sim L^{2/3}$ · surface_tension

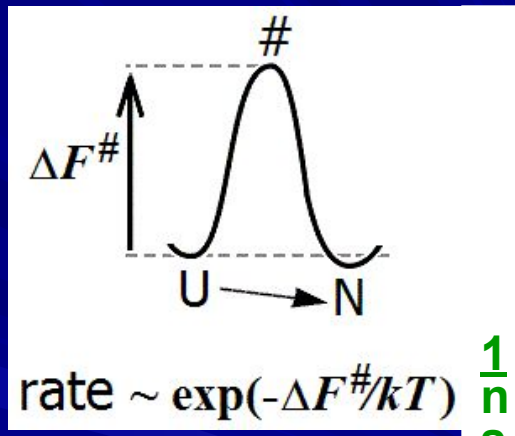
a) micro-; b)

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: $\Delta F^{\#} \le L^{2/3} \cdot {}^1/_2[{}^3/_2 RT \cdot \ln(L^{1/3})]$

$\cdot + L/(\sim 100)$



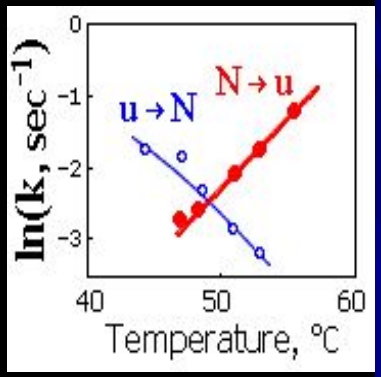
[Flory] [knots]

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

micro loops

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

Nucleus:
not as small,
it comprises
30-60%
of the protein

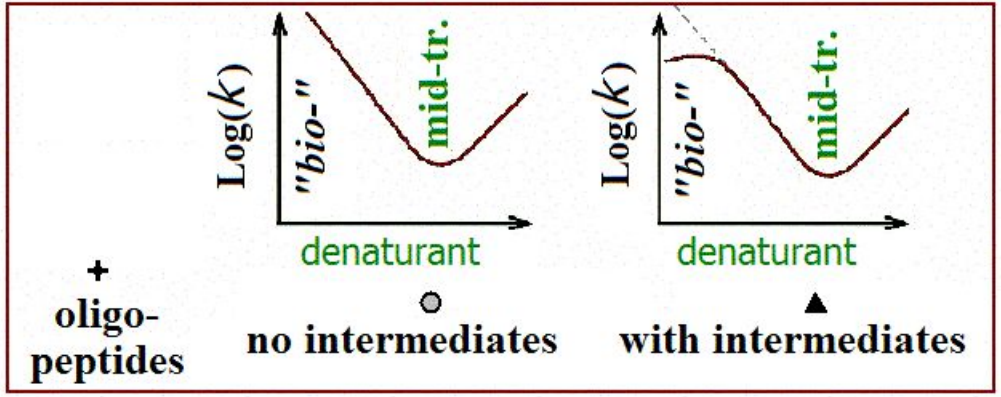


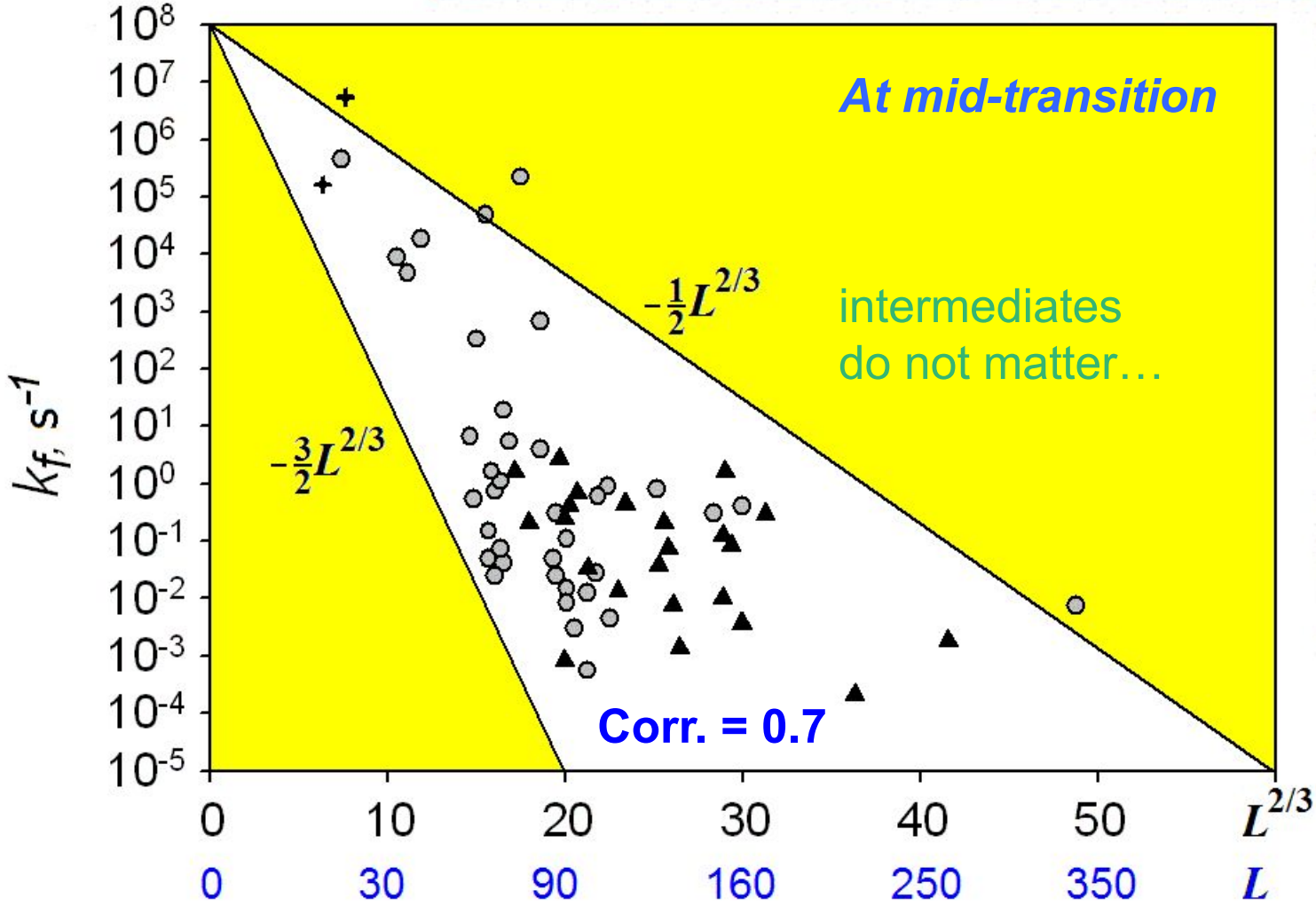
FOLDING RATE:

$$k_f \sim \exp[-(\tfrac{1}{2} \div \tfrac{3}{2})L^{2/3}]\ \text{ns}^{-1}$$

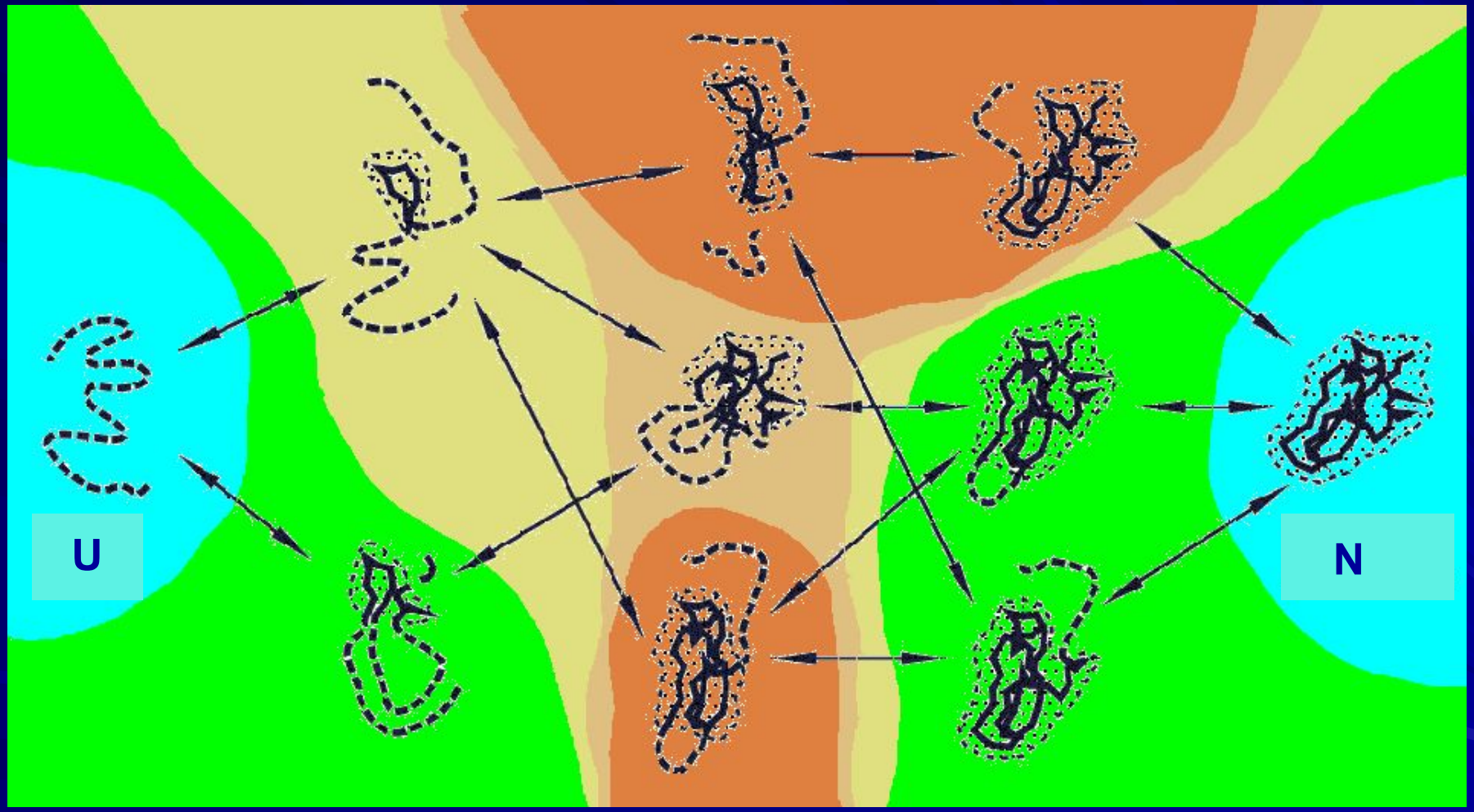
form-factor

loops

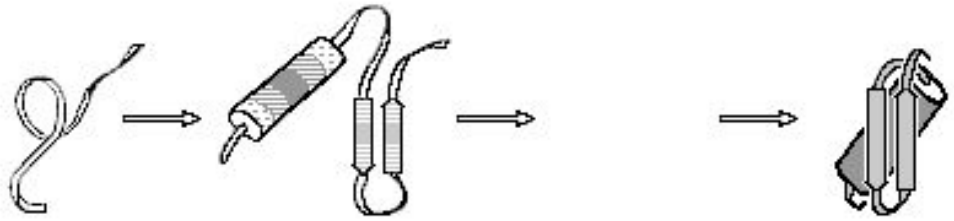




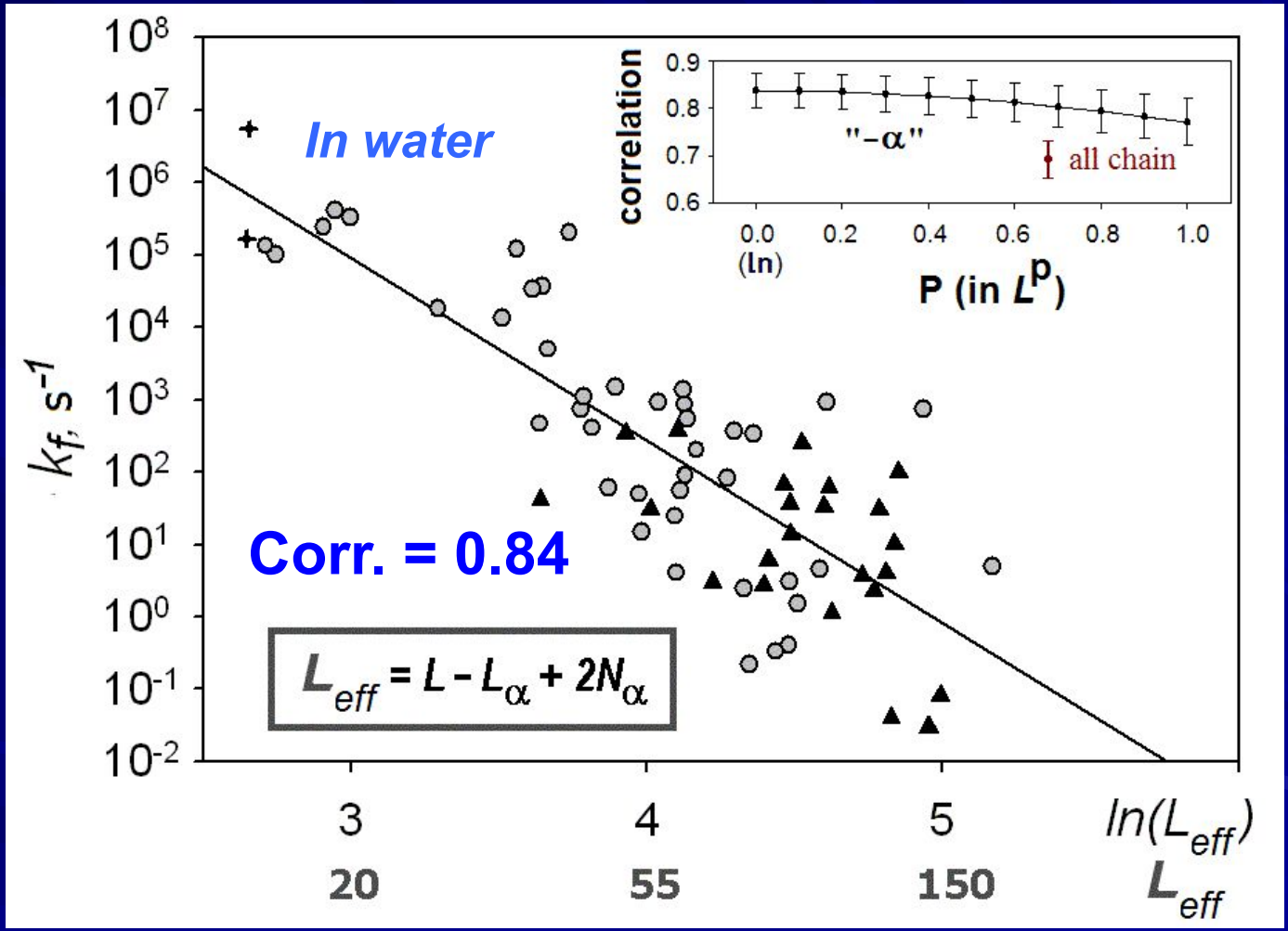
***Any* stable fold is *automatically* a focus of rapid folding pathways. *No "special pathway" is needed*.**



α-helices decrease effective chain length.
THIS HELPS TO FOLD!



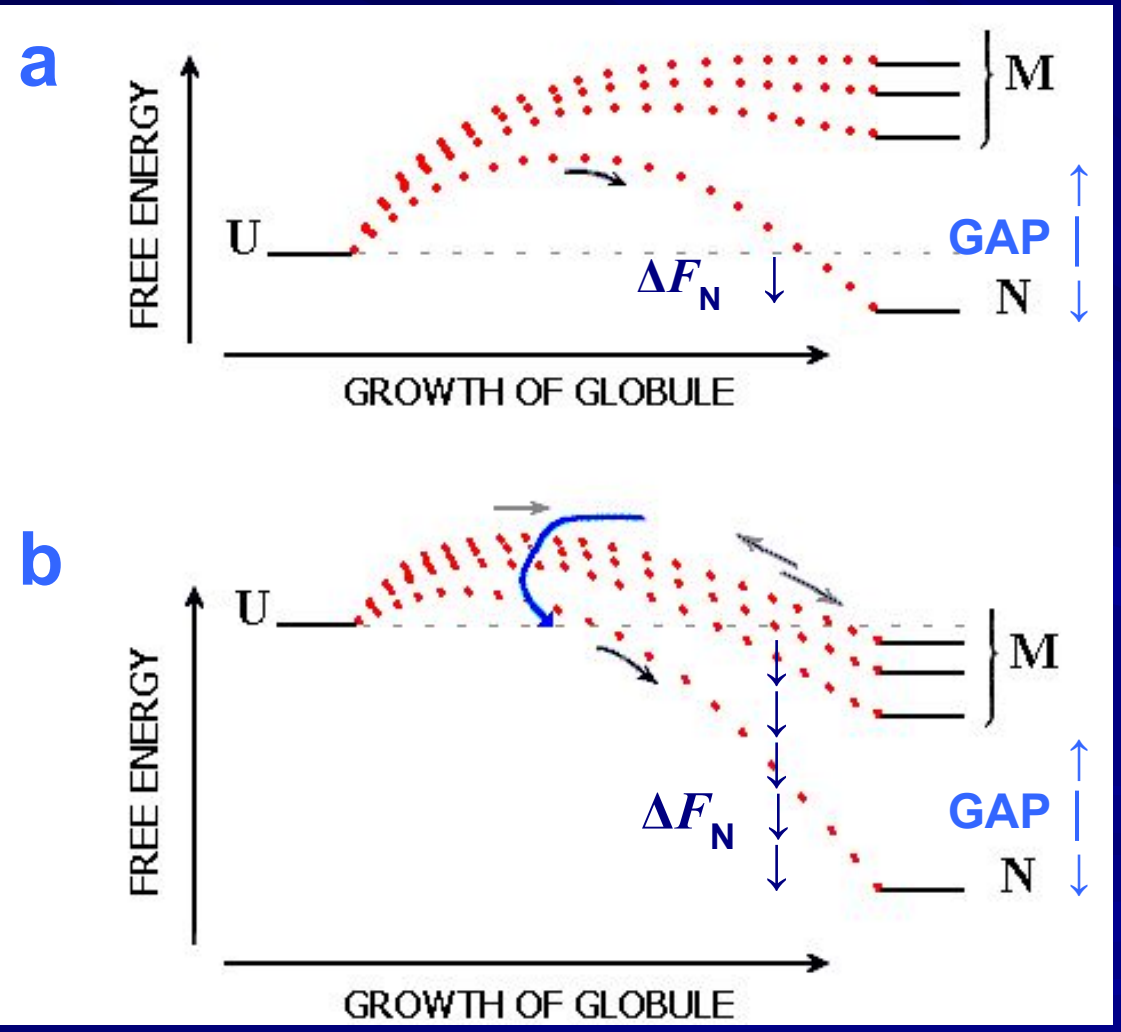
α-HELICES ARE PREDICTED FROM THE AMINO ACID SEQUENCE



Ivankov D.N., Finkelstein A.V. (2004) Prediction of protein folding rates from the amino-acid sequence-predicted secondary structure. - *Proc. Natl. Acad. Sci. USA*, 101:8942-8944.

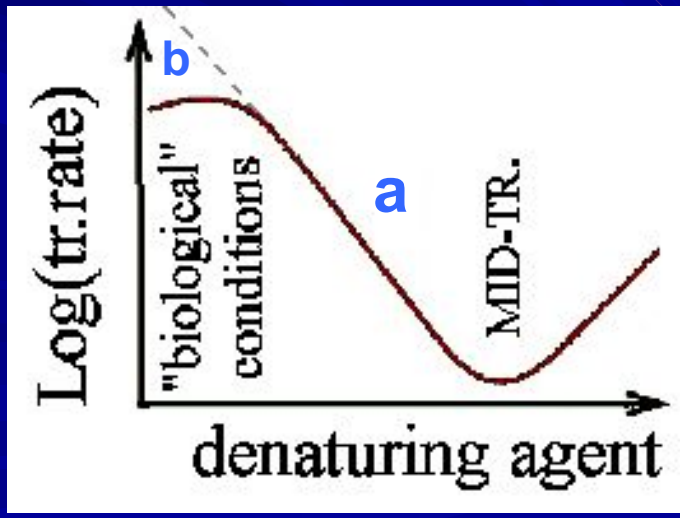
Up to now, a vicinity of mid-transition has been considered.

When globules become more stable than U:

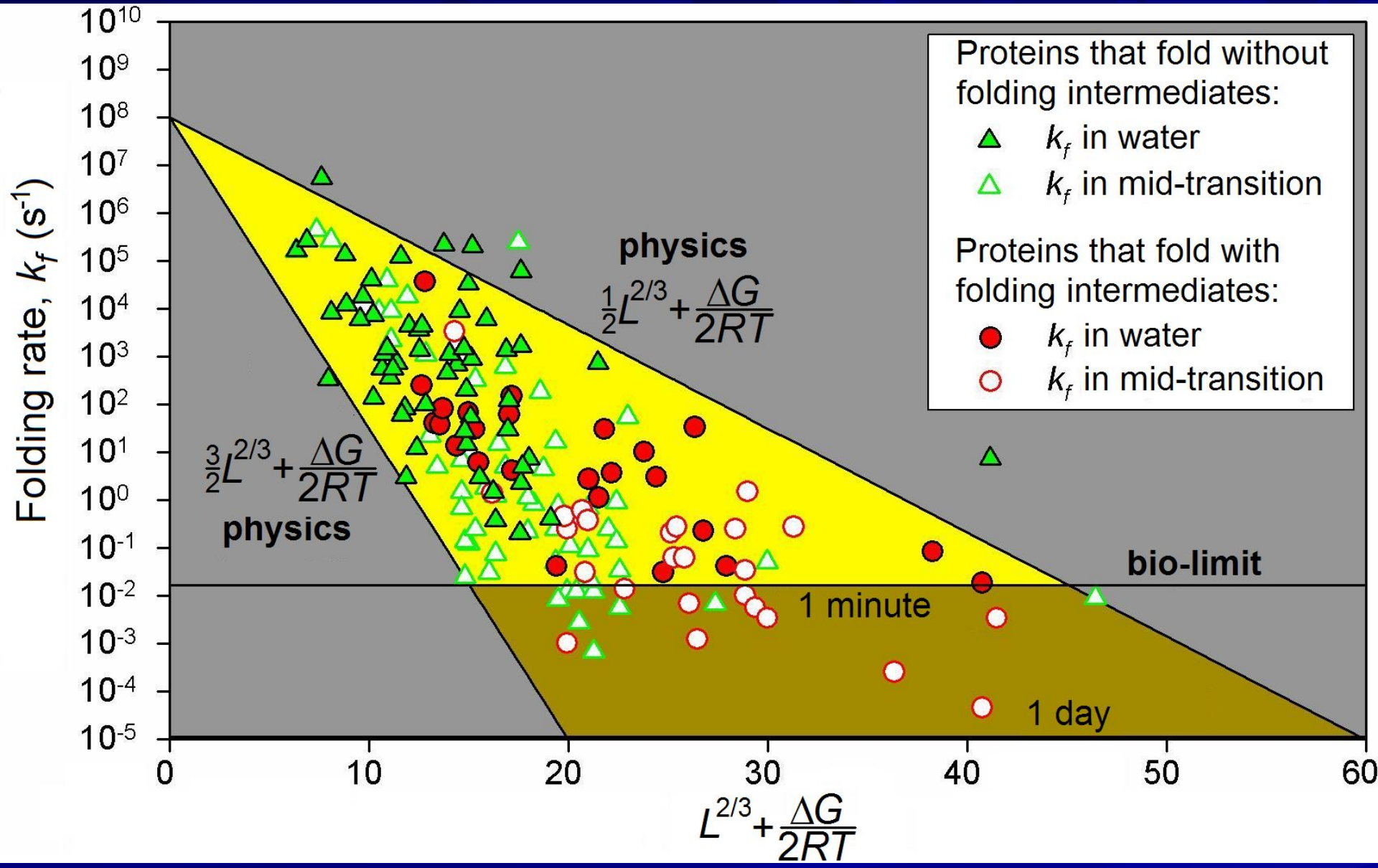


1) Acceleration:
$\Delta \ln k_f \approx -{}^1/_2 \Delta F_N / RT$

2) Large gap → large acceleration due to ΔF_N *before* "rollover" caused by stability of intermediates M at "bio-conditions"

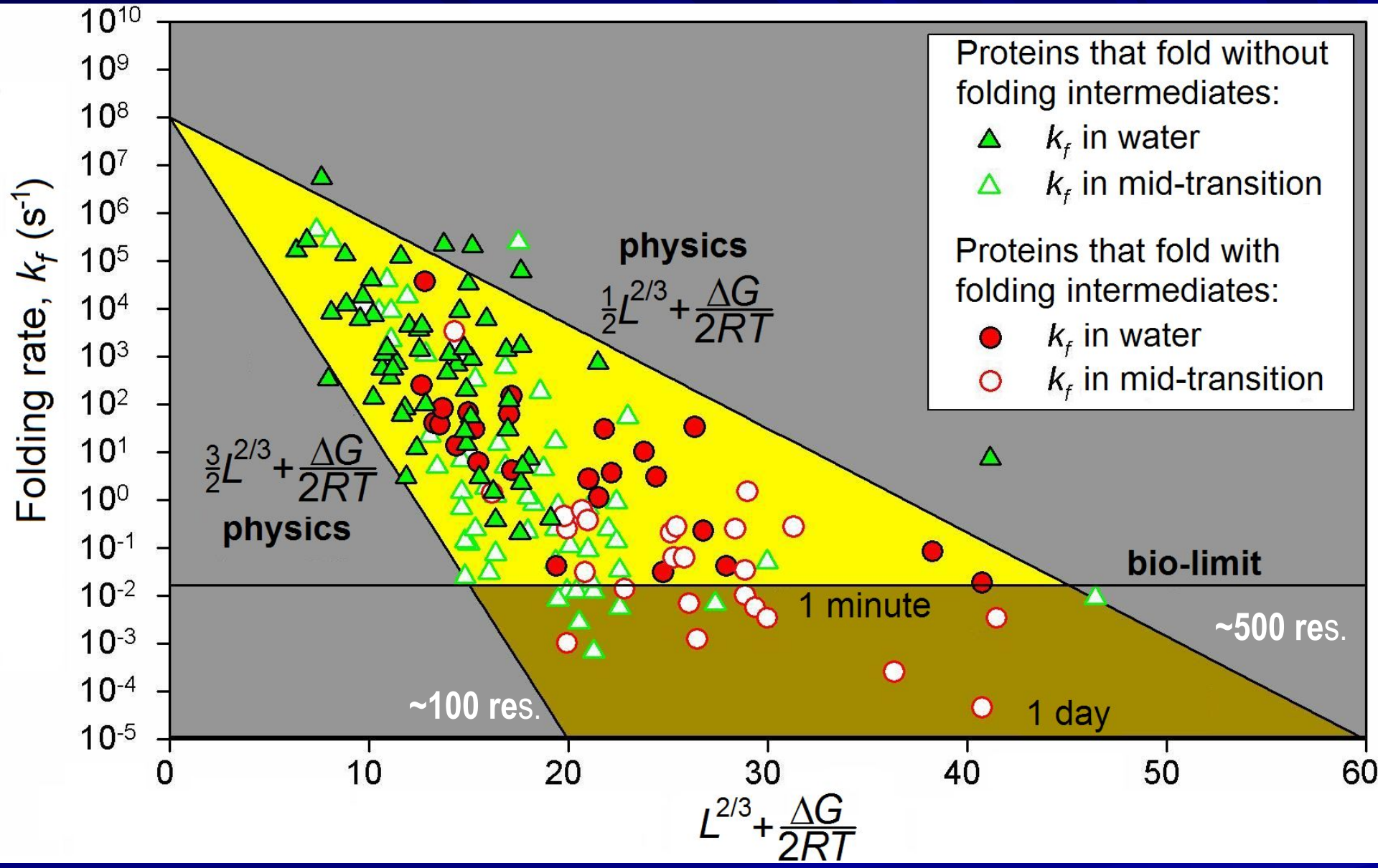


Finkelstein, Badretdinov; *Folding & Design*, 1997, 1998]. Finkelstein; *Les Houches, Session 77*, 2003]



Garbuzynskiy, Ivankov, Bogatyreva, Finkelstein (2013) *PNAS* **110**:147

Finkelstein, Badretdinov; *Folding & Design*, 1997, 1998]. Finkelstein; *Les Houches, Session 77*, 2003]

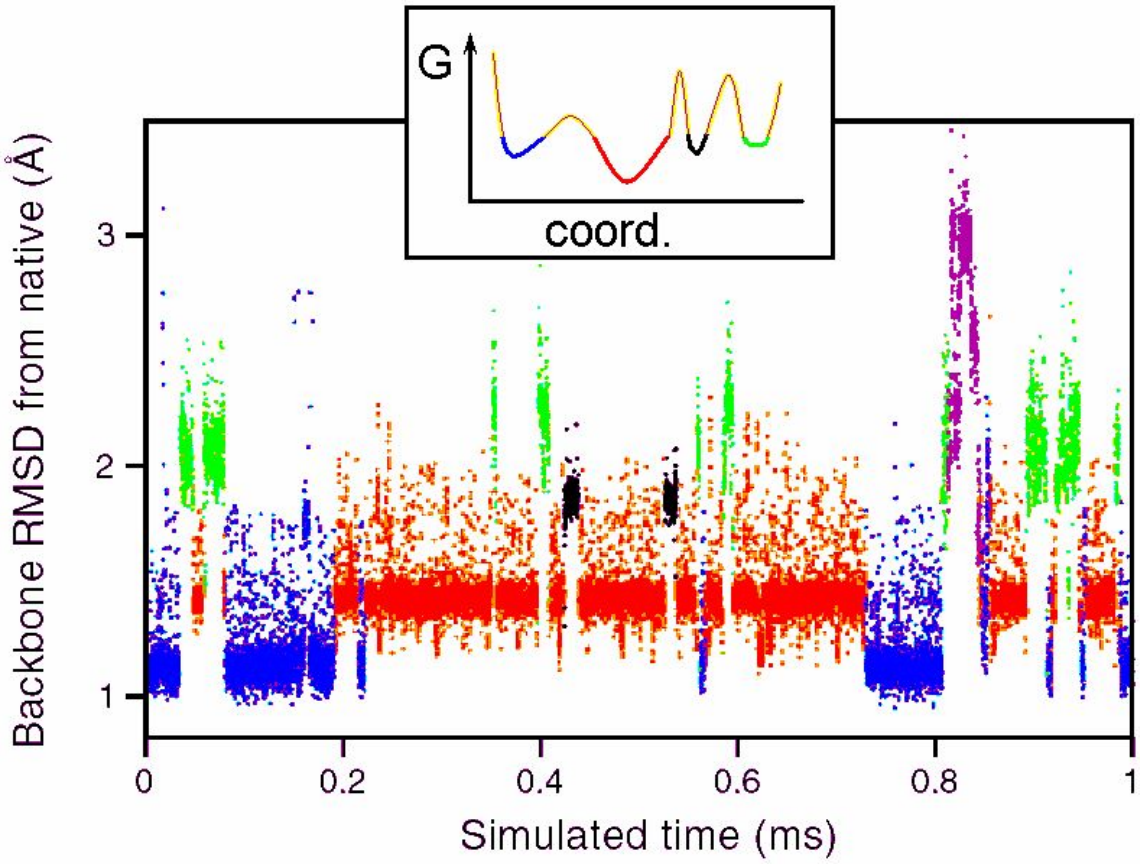


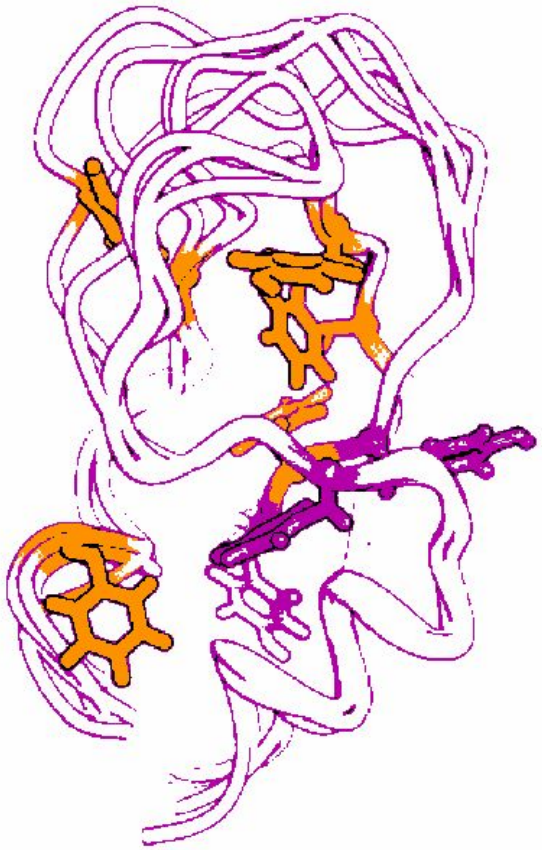
Garbuzynskiy, Ivankov, Bogatyreva, Finkelstein (2013) *PNAS* **110**:147

D.E. Shaw et al., Oct. 2010, *Science* **330**, 341

modified version of the Amber ff99SB force field:
K. Lindorff-Larsen *et al.*, *Proteins* **78**, 1950 (2010)

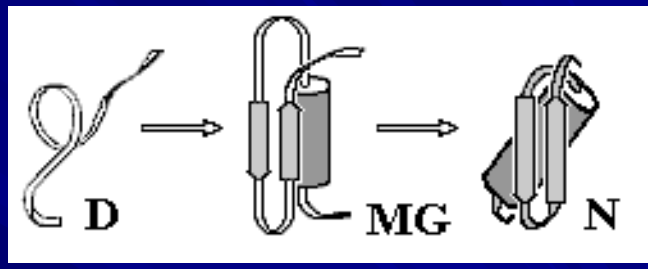
Native-state dynamics of BPTI (modeling)





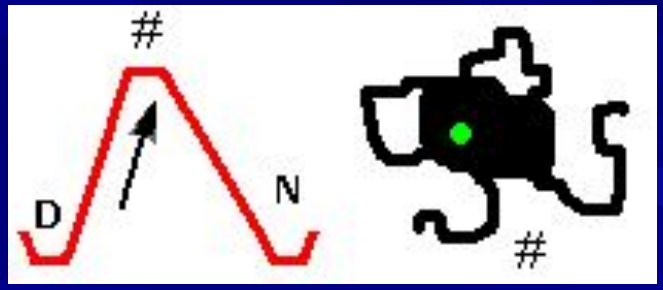
Protein Structures: Kinetic Aspects

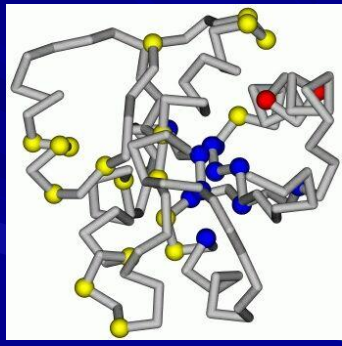
- *In vivo* folding & *in vitro* folding
- Protein folds spontaneously: how can it?



- Protein folding intermediates; MG

- Transition state & folding nucleus





- Protein folding rate theory: solution of Levinthal's paradox

$$t \sim \exp[(\tfrac{1}{2} \div \tfrac{3}{2}) L^{2/3}] \text{ ns}$$