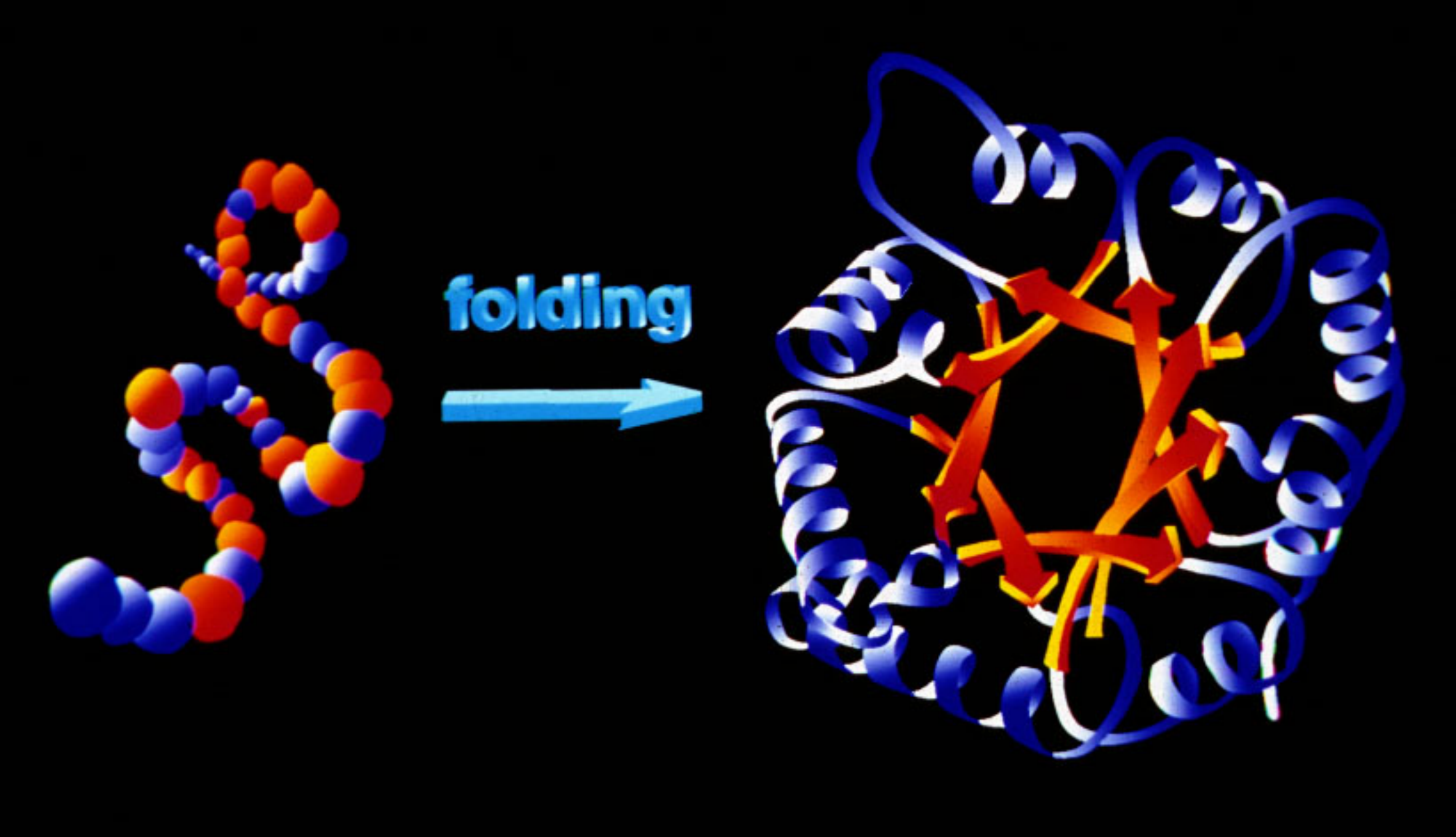
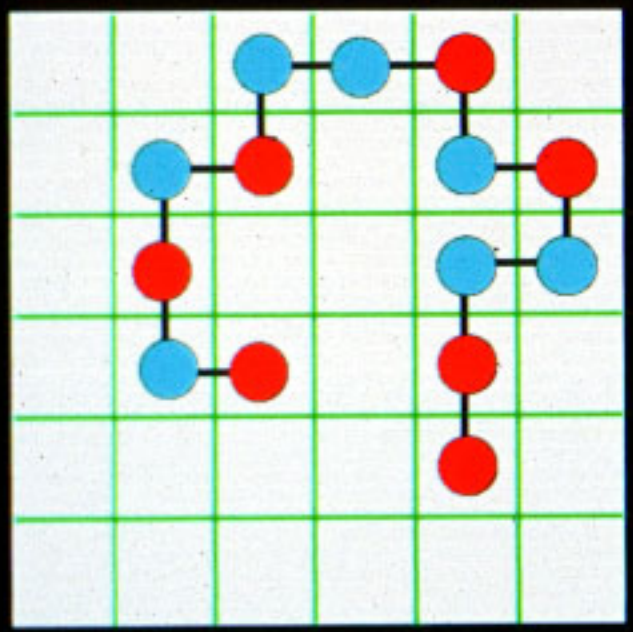

folding

HP is Simplest Folding Code

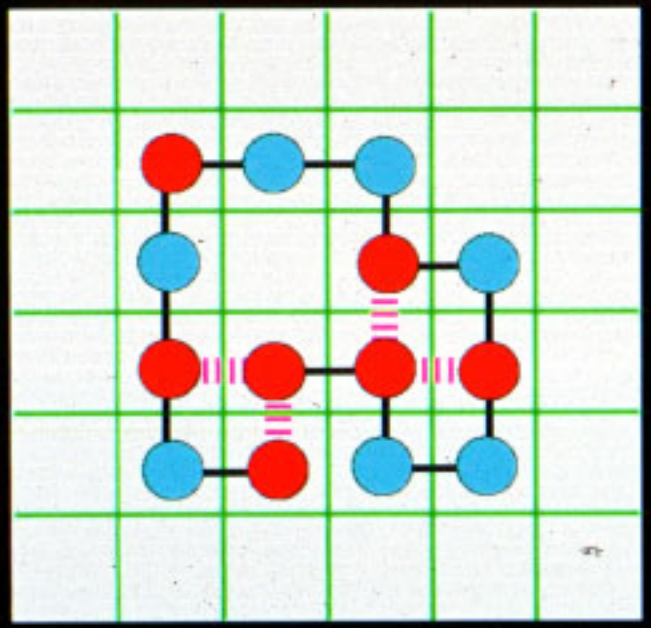
h = #HH contacts

H

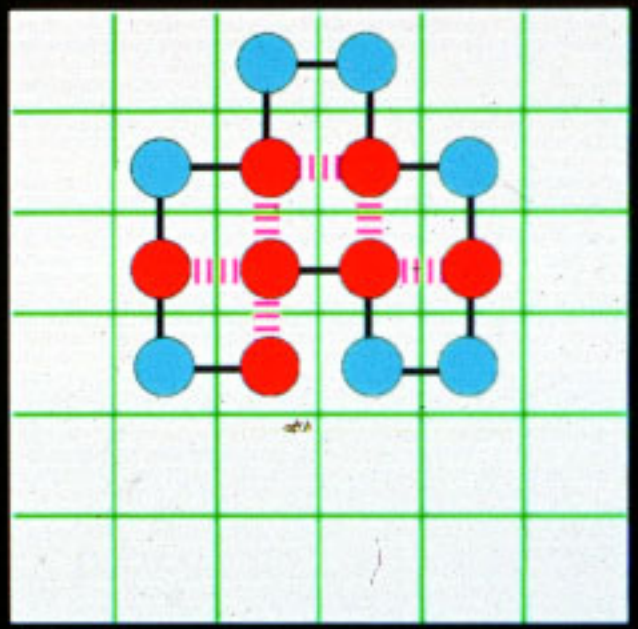
P



h=0

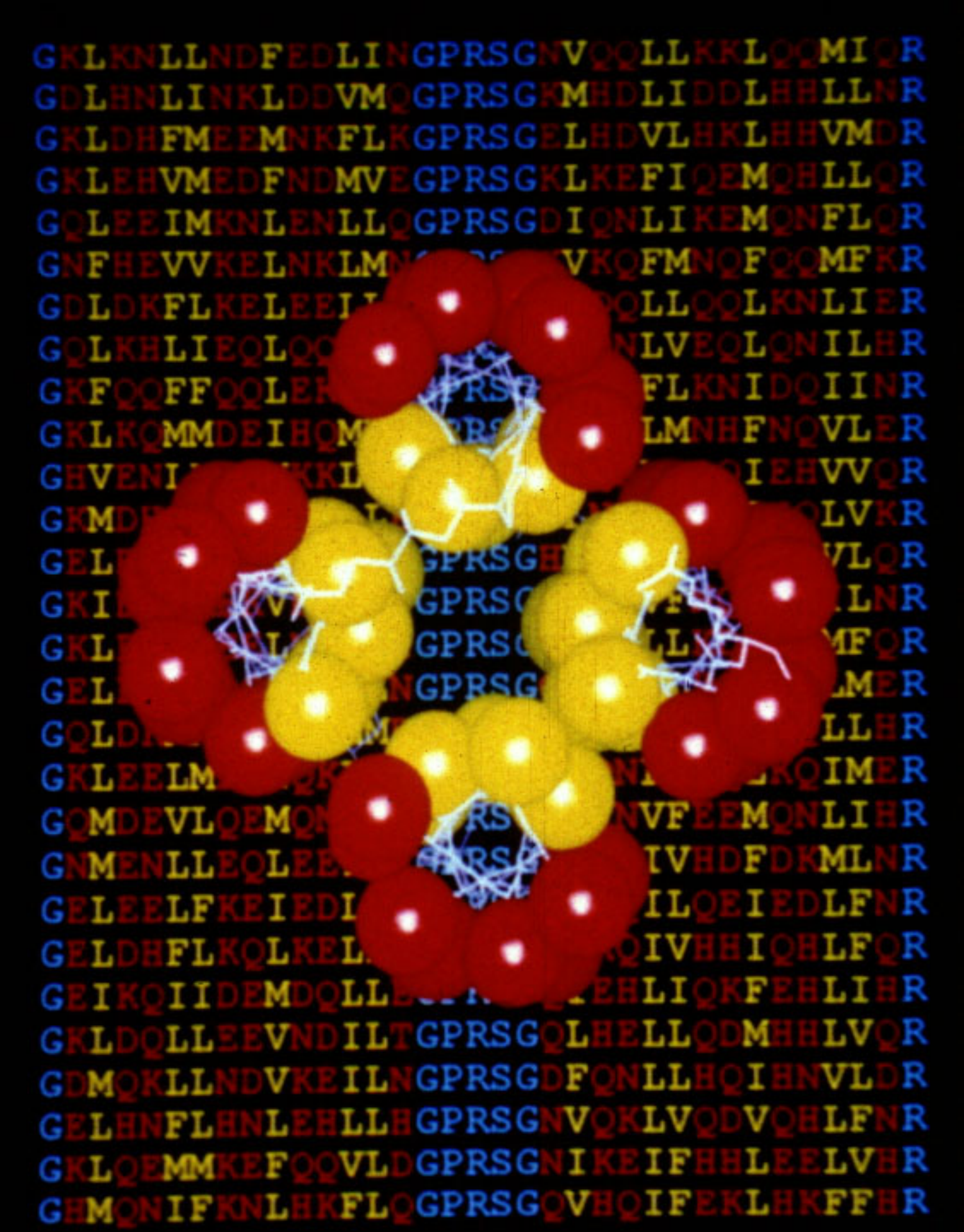


h=4



h=6

Lau & Dill Macromol 22 3986 (1989)

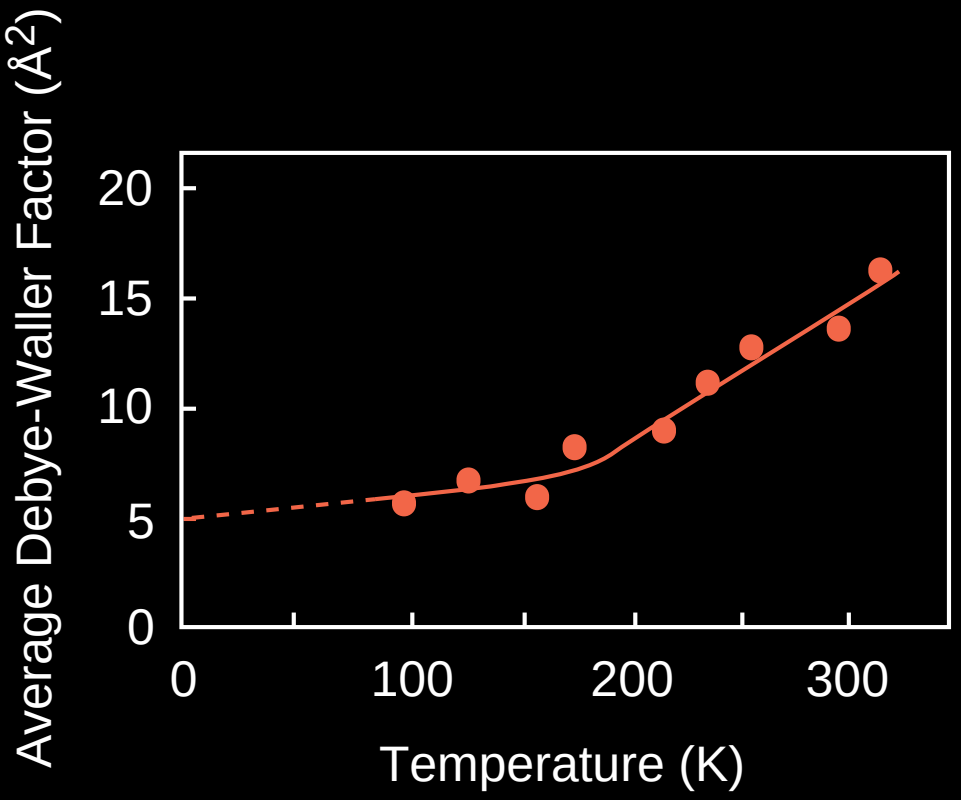


S. Kamtekar *et al.*

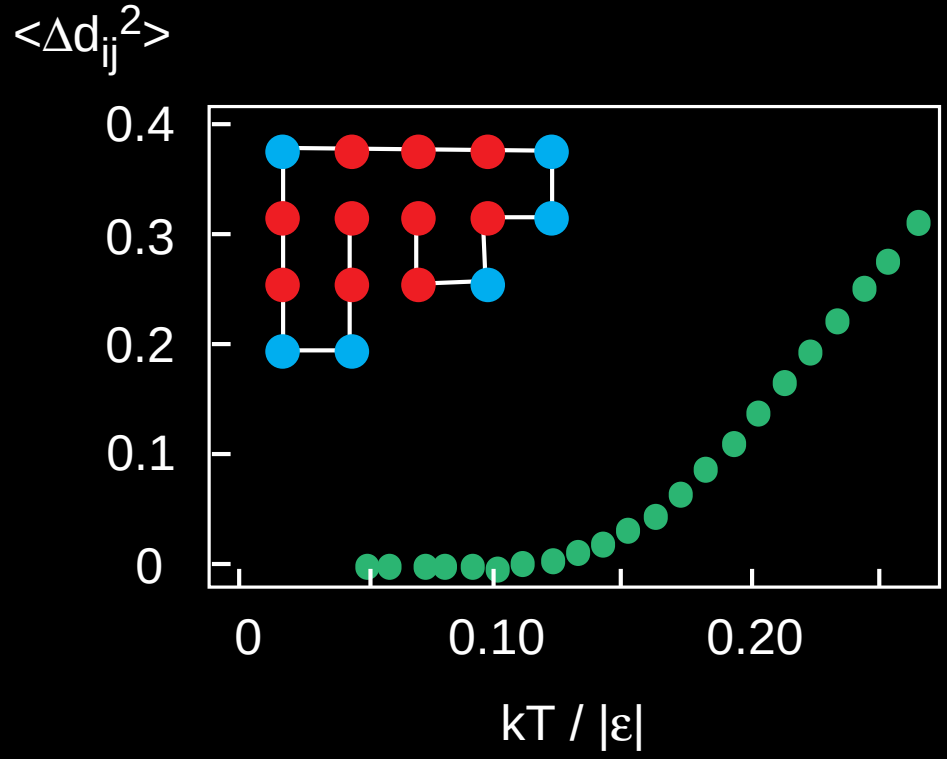
Science 262 1680 (1993)

Fluctuations in Crystals: a "Glass Transition?"

Experiments



Simulation



Tilton *et al.* Biochem 31 2469 (1992)

- Karen Tang

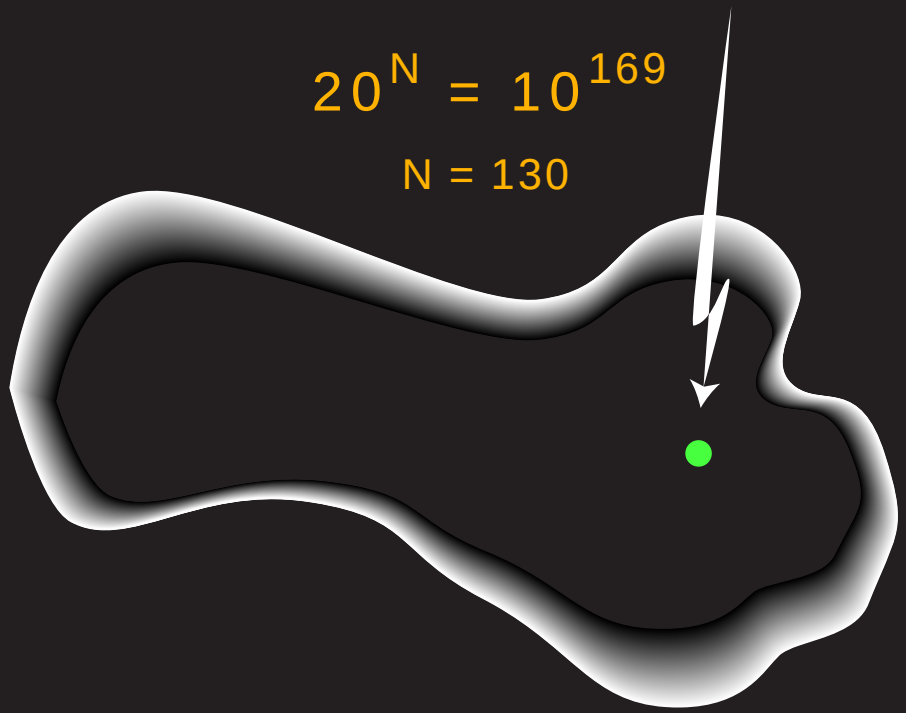
Sequence Space is dense with folded molecules.

How many sequences must be explored to find:

The Lysozyme
Sequence ?

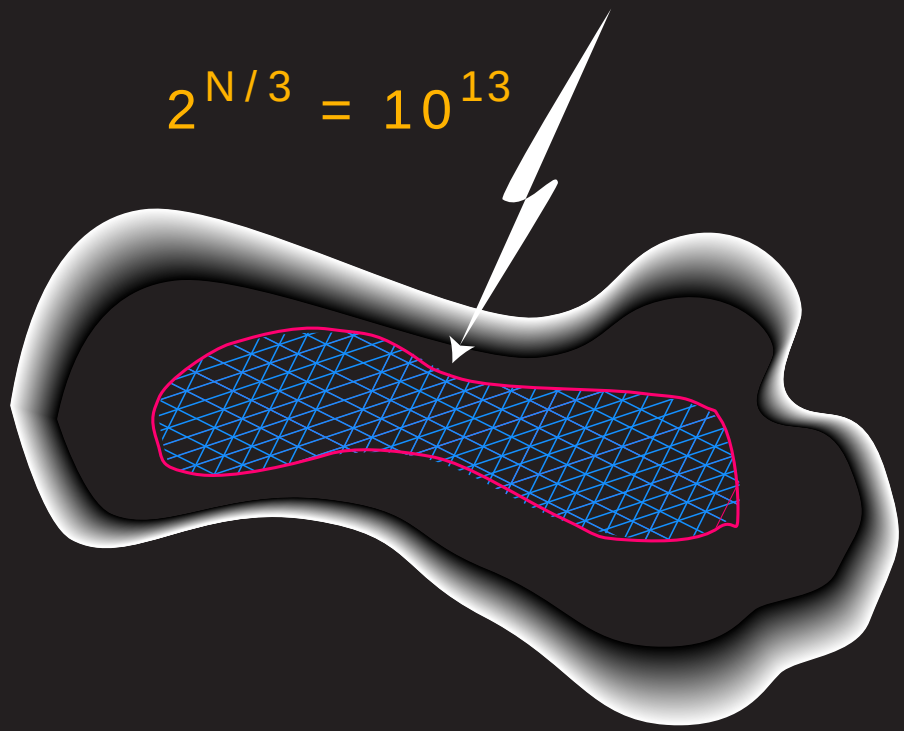
$20^N = 10^{169}$

N = 130

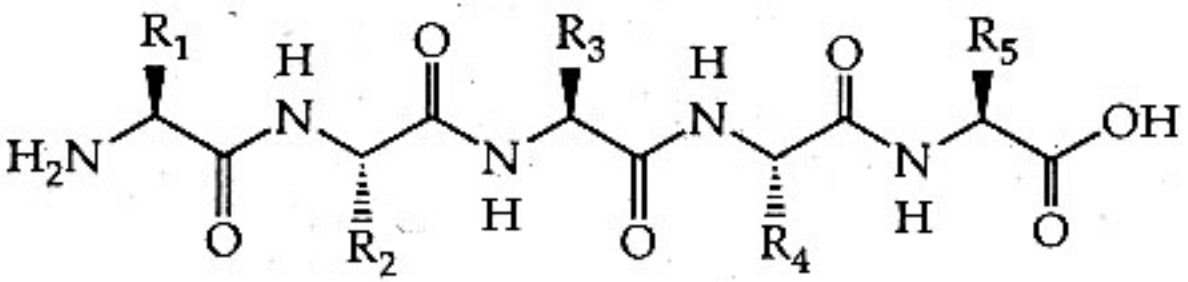


The Lysozyme
Structure ?

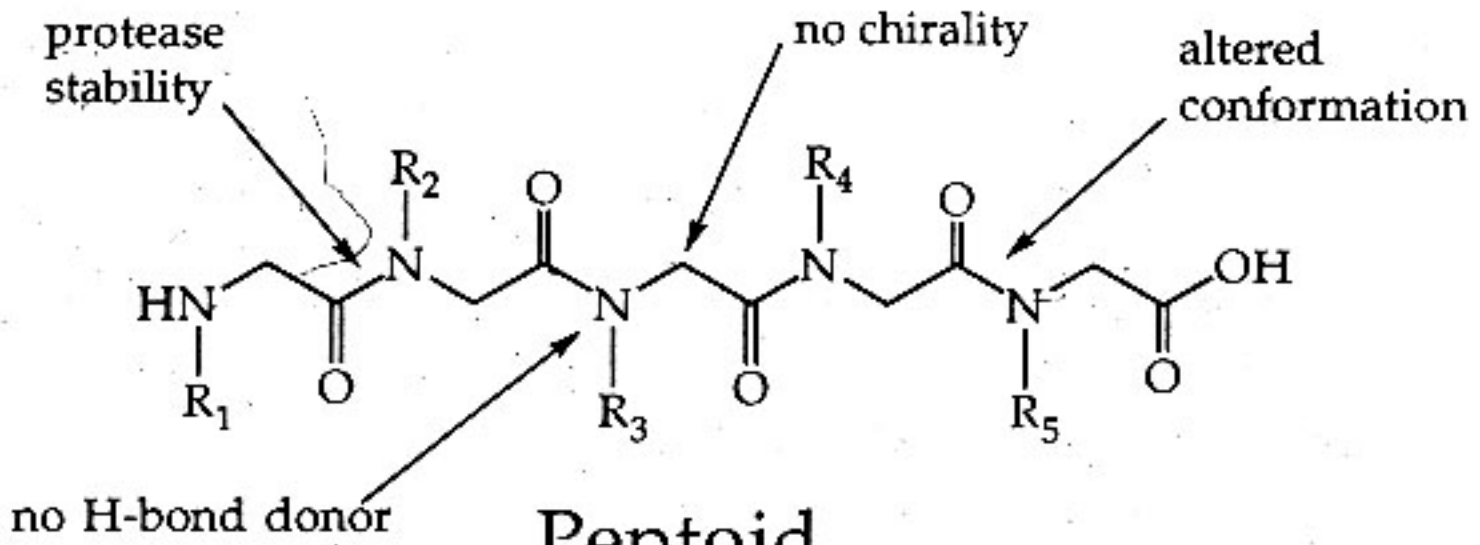
$2^{N/3} = 10^{13}$



Oligo N-Substituted Glycines



Peptide

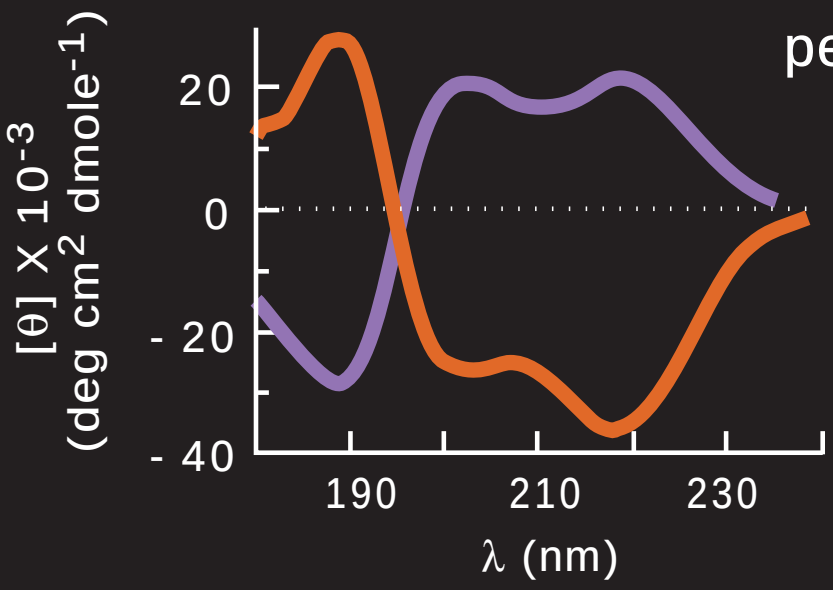


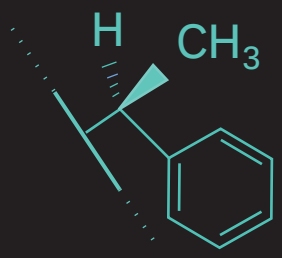
Peptoid

R. Simon *et al.*
PNAS '92 (Chiron Corp.)

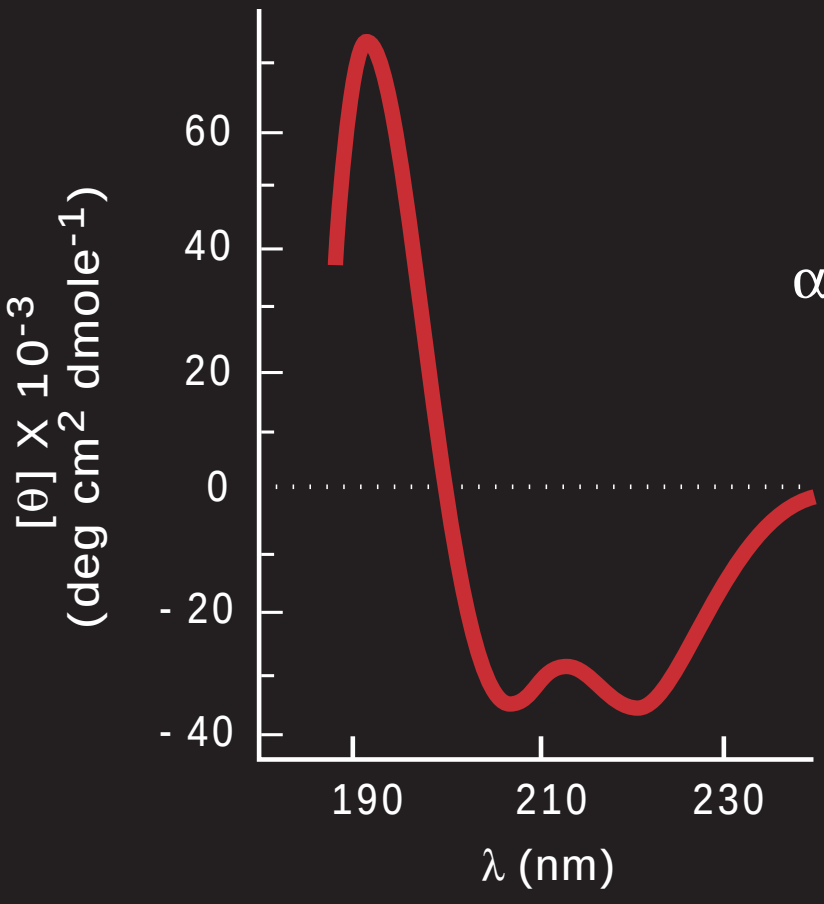
Chirality in Side Chains Can Replace Chirality in Backbone

CD Spectrum looks helix-like. No hydrogen bonds.





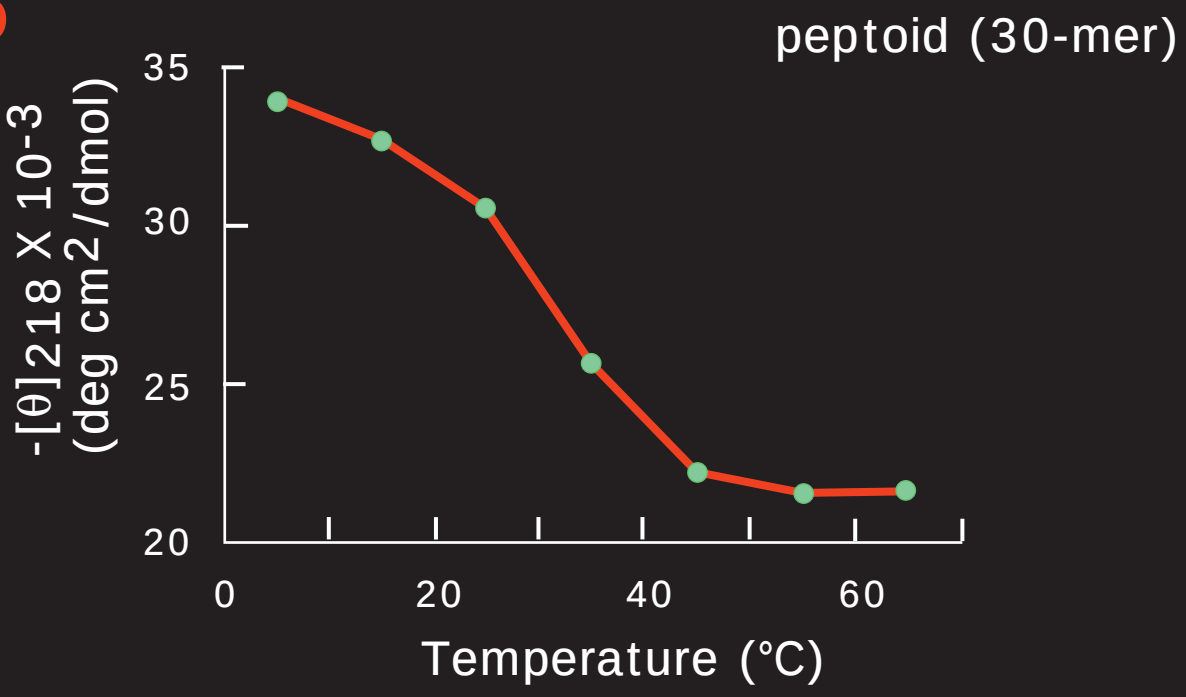
R, S enantiomers.
phenylethylglycine



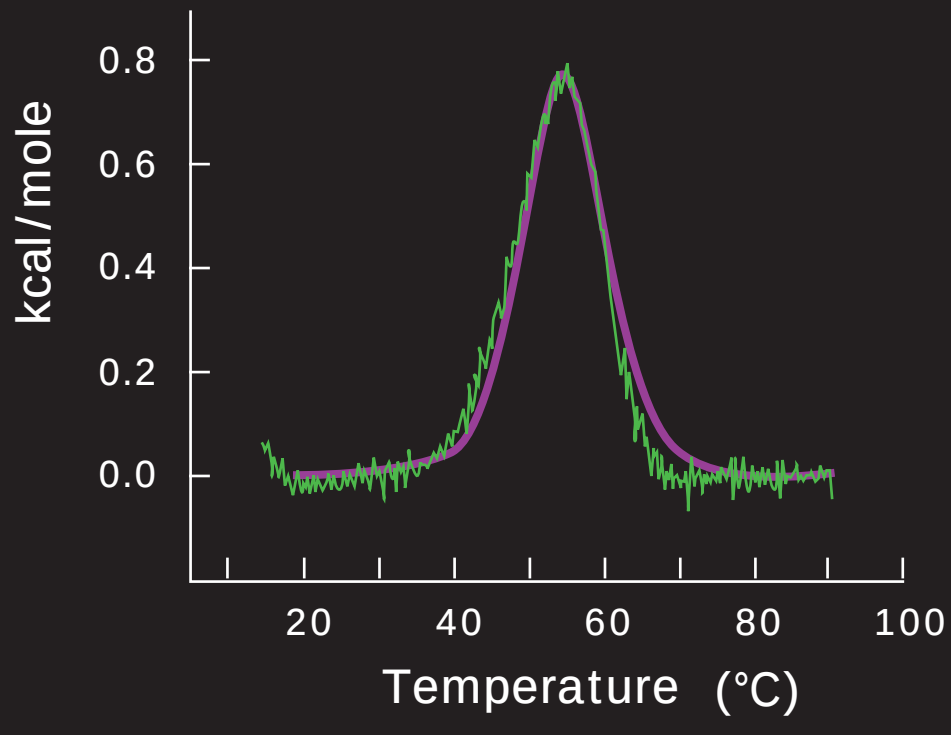
• Kent Kirshenbaum, Anneliese Barron, Richard Goldsmith, Philippe Armand, Erin Bradley, Kiet Truong, Ken Dill, Fred Cohen, and Ron Zuckerman

Peptoid 2° structure is stable, cooperative, despite lack of hydrogen bonds

CD



DSC



PROTEIN FOLDING

- Why is it Fast?
- How can experimental kinetics help speed up computational searching?

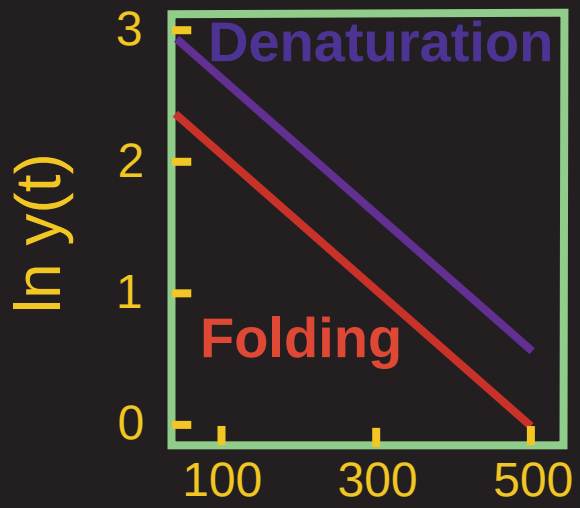
Protein Folding is Fast

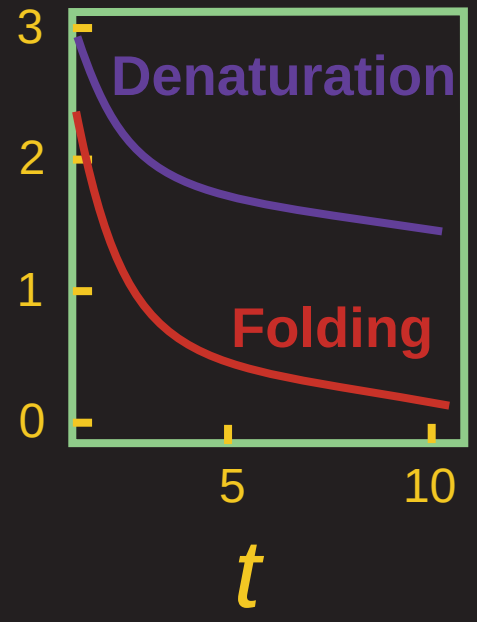
Amino acid unit transition rate $\cong$ 10-100 psec
Number of unit transitions $\cong 10^2 - 10^3$

If no mistakes, 1-100 nsec to fold
Fastest known folding rate* $\cong$ 10 μsec

$\Rightarrow$ Protein makes only $10^{2-}\, 10^5$ mistakes (thermal motion, uphill steps, re-tries of earlier conformations, kinetic traps, etc)

* RE Burton, TG Oas et al. ***Nat Struc Biol* 4**:305 (1997)





Folding Models

2 State

D → N

Multi-State

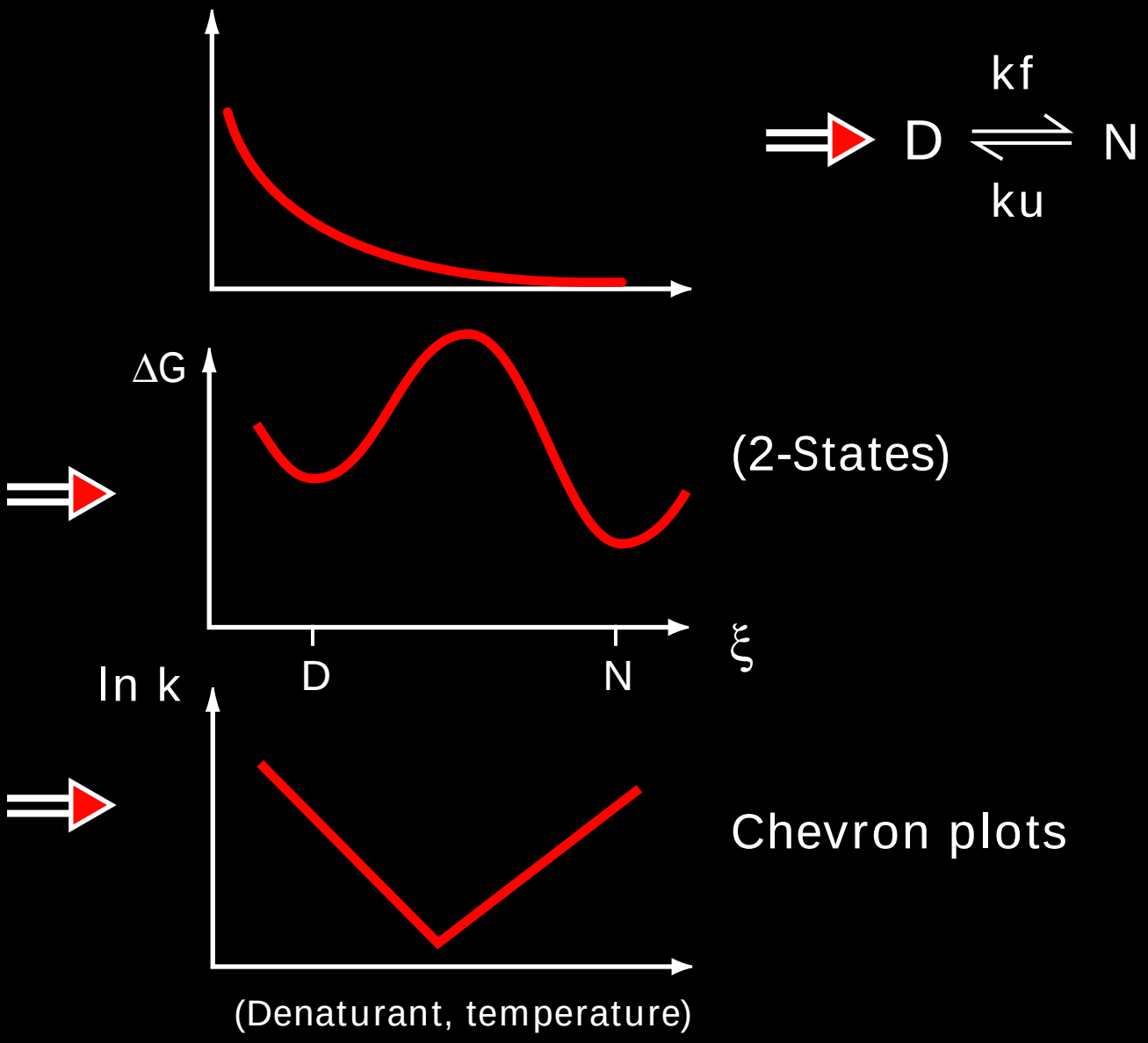
D → I → N "on - path" intermediate

D → N "off - path" intermediate

↓

I

Protein Folding Kinetics
As seen from Transition-State Theory

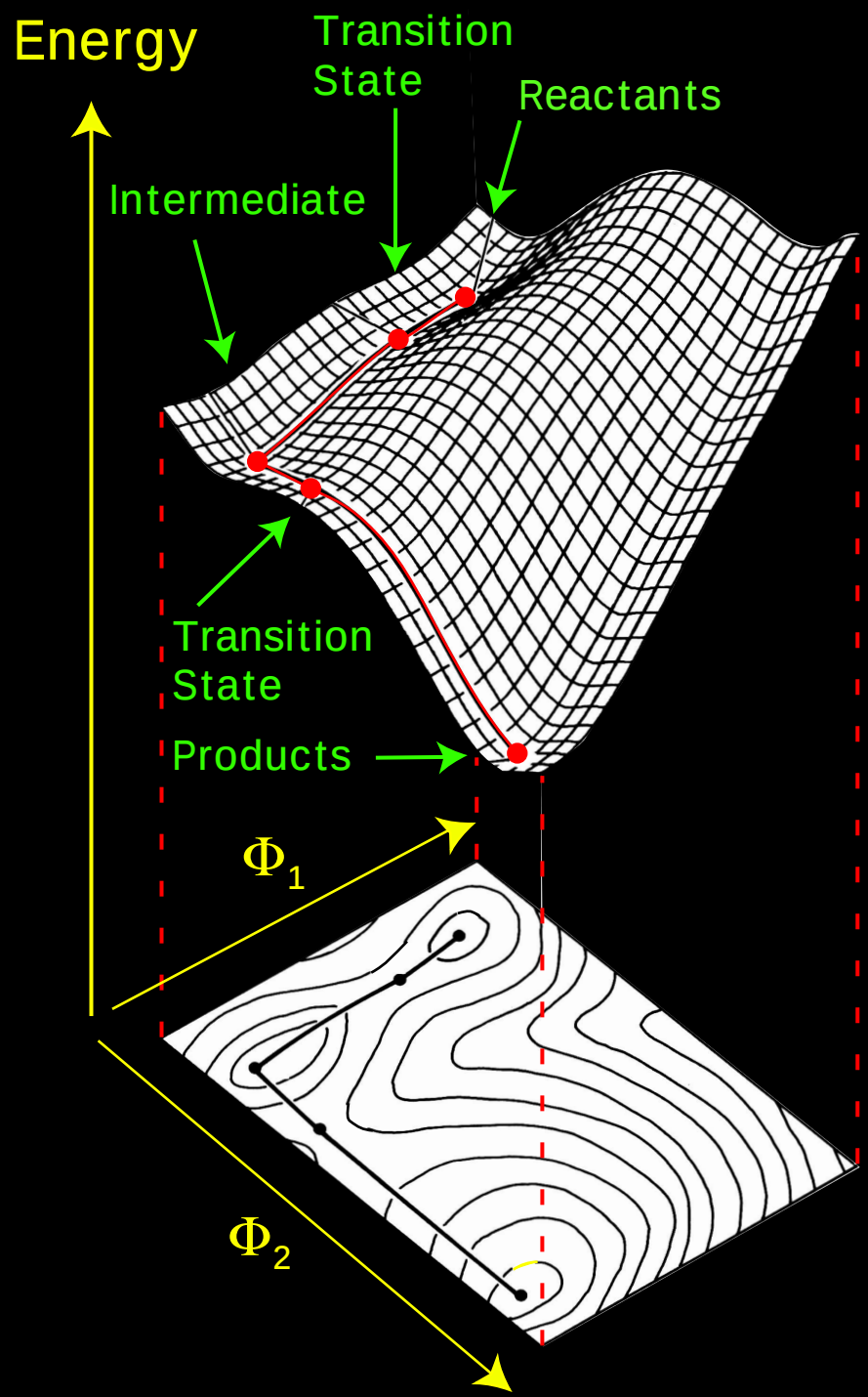


Classical Rate Theory

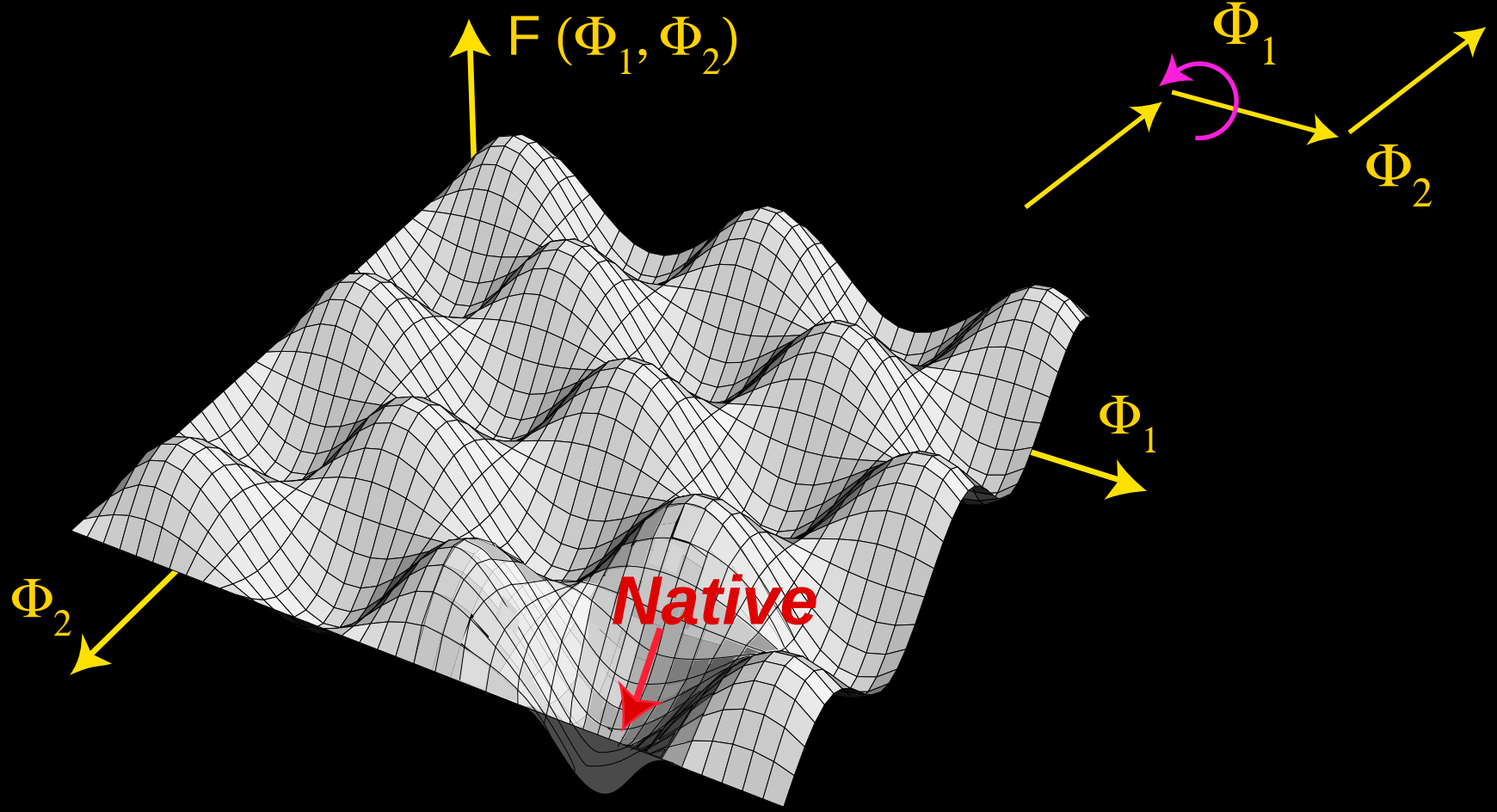
$$\text{rate} = \left(\frac{kT}{h}\right) e^{-\Delta G^{\ddagger}/kT}$$

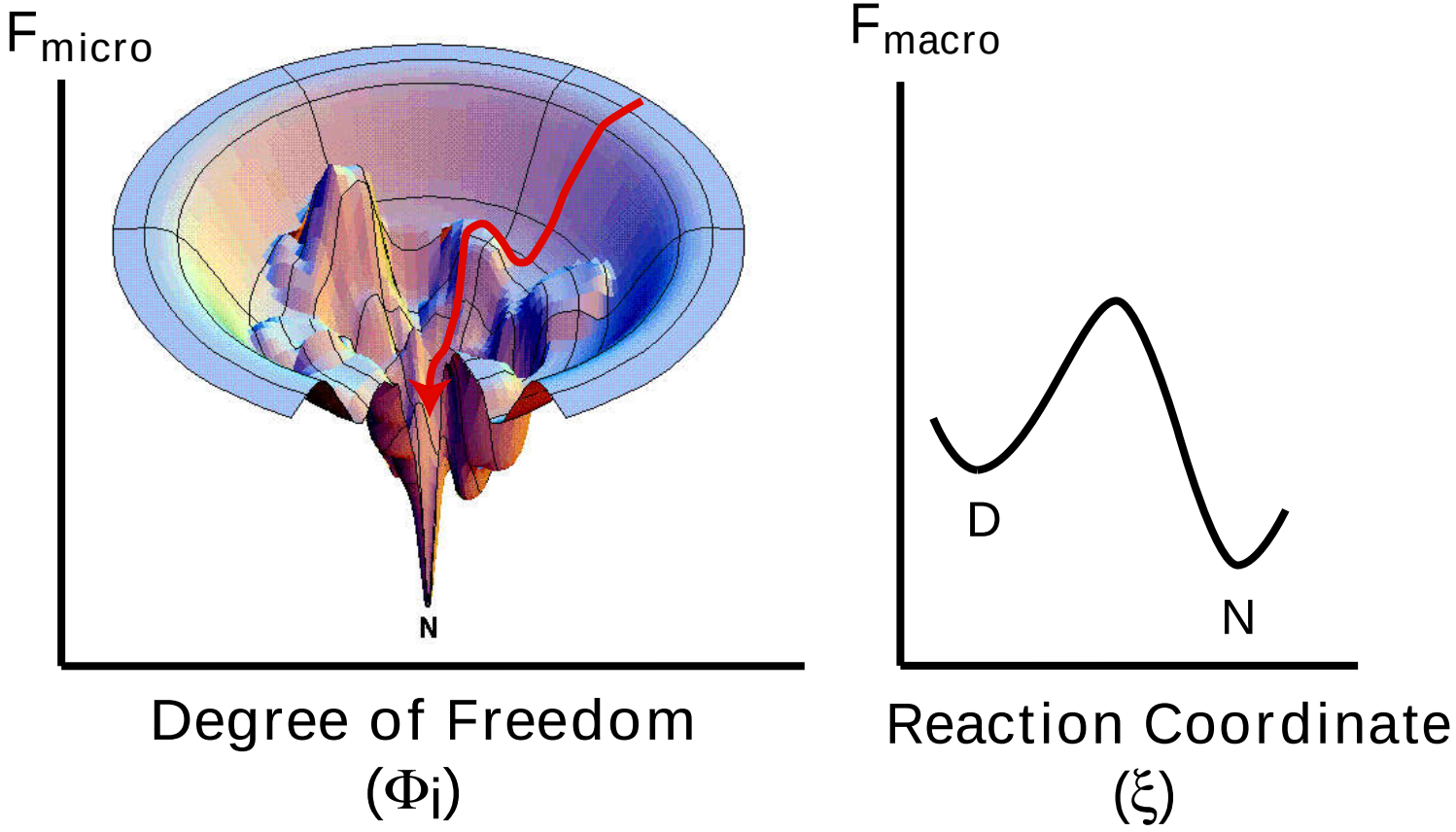
- Explains that chemical processes are slow relative to speed limit, $kT/h = 0.2$ psec, because of energy barriers (activated states)

- Each macrostate (Reactant, I, TS, Product) can be identified with a microstate (a particular molecular structure).

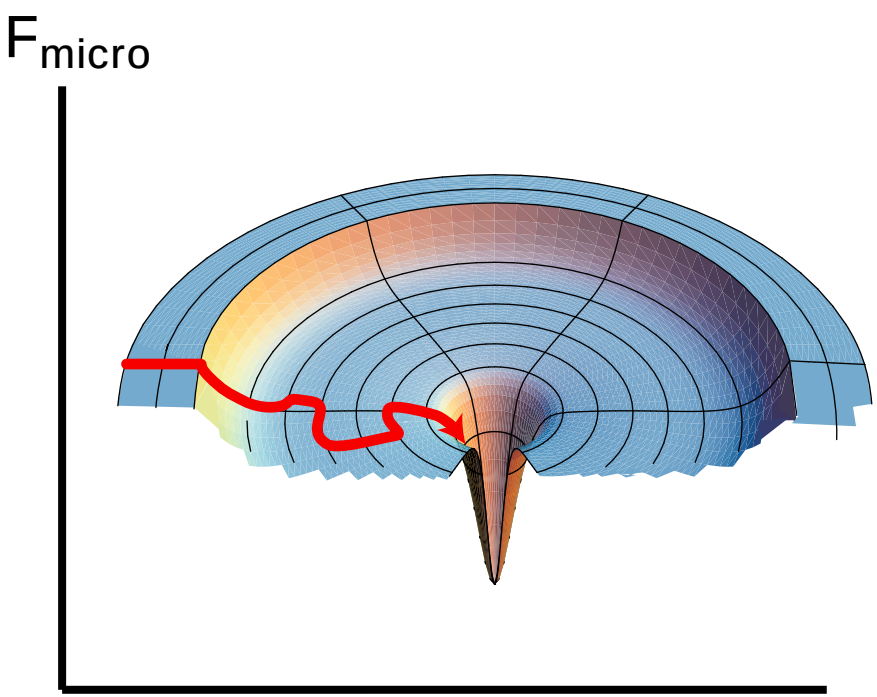

Energy
Transition State
Reactants
Intermediate
Transition State
Products
Φ_1
Φ_2

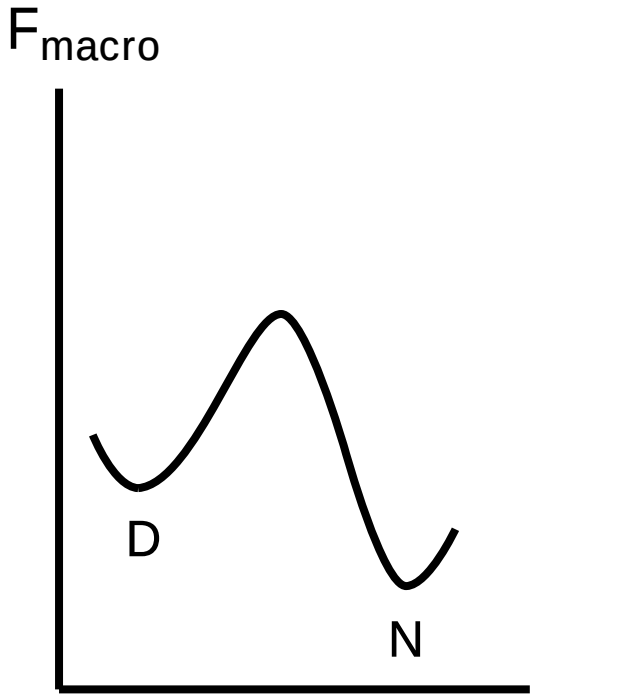
Energy Landscape





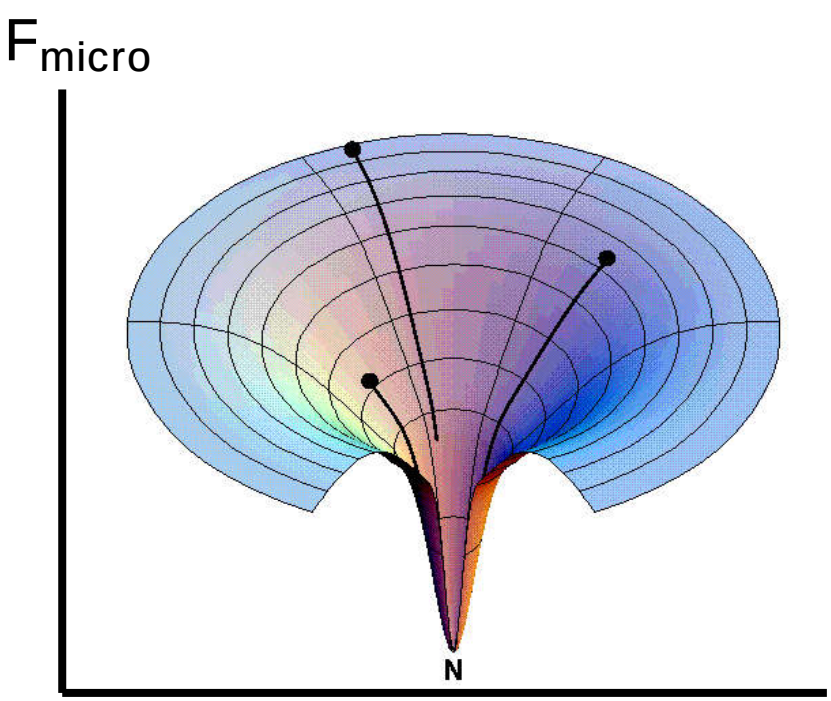
Energy Barrier

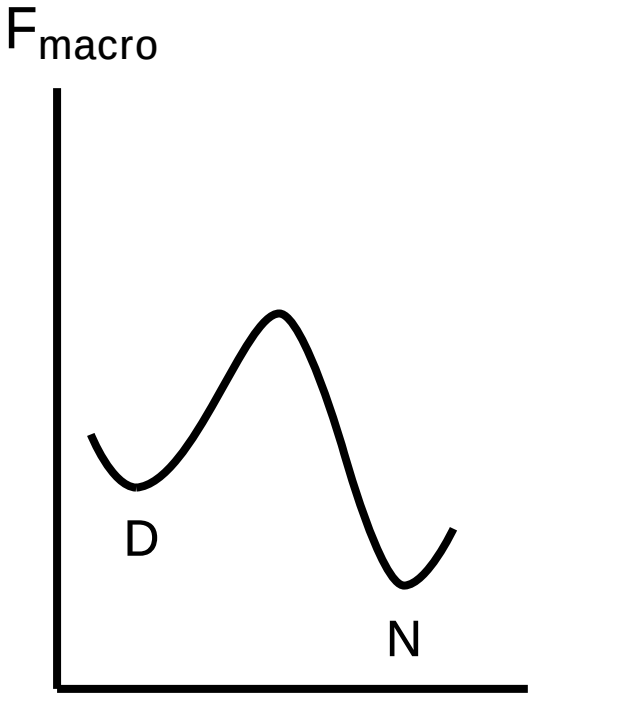




Conformational Entropy Barrier

What is the Barrier?





How to Find the Transition State

$$\frac{d\mathbf{C}}{dt} = \mathbf{AC} \qquad \mathbf{A} = \begin{bmatrix} k_{11}\ k_{12}\ k_{13}.... & & \\ k_{21} & & \\ k_{22} & & \\ & & k_{NN} \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix}$$

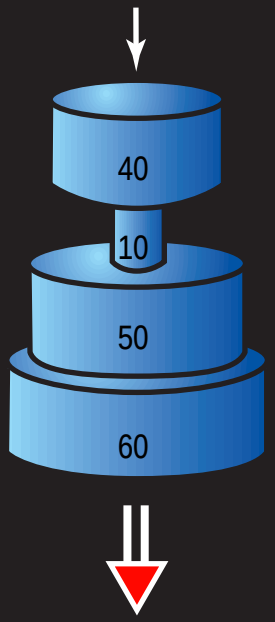
- Diagonalize, get slowest eigenvalue

- Corresponding eigenvector is TS ensemble

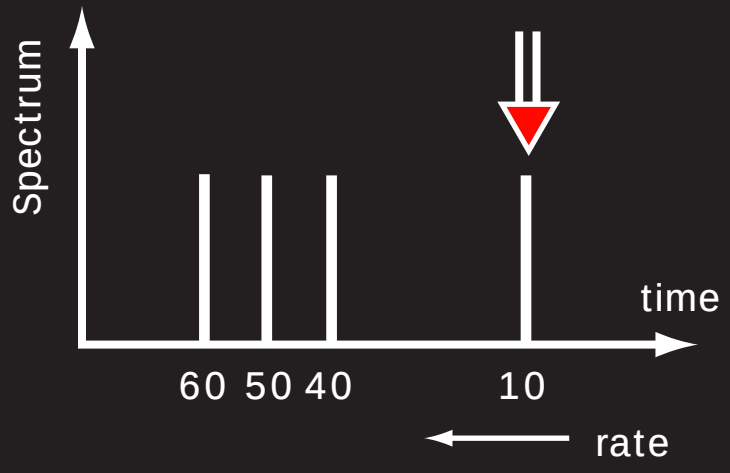
Model: 2D exact short-chain Go model

Macroscopic rate is a collective property, not a property of a single bottleneck

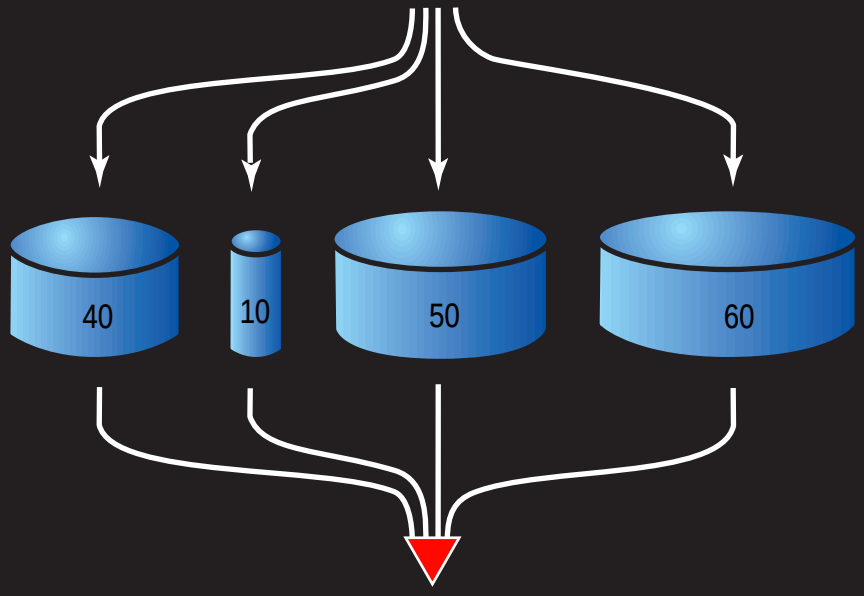
Series Process



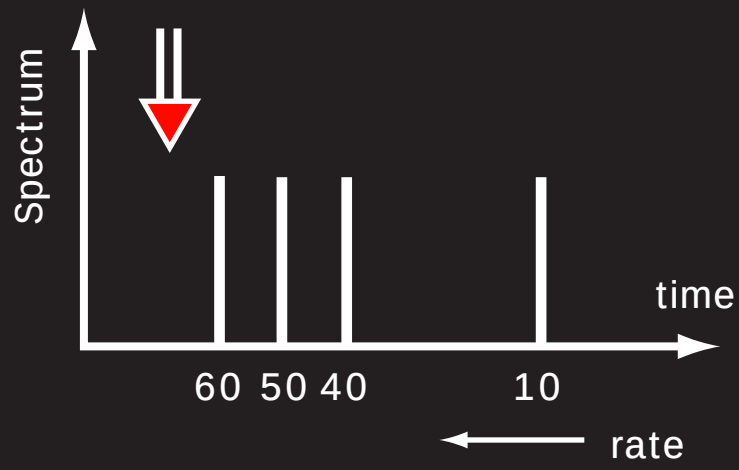
macroscopic rate ≡ bottleneck rate



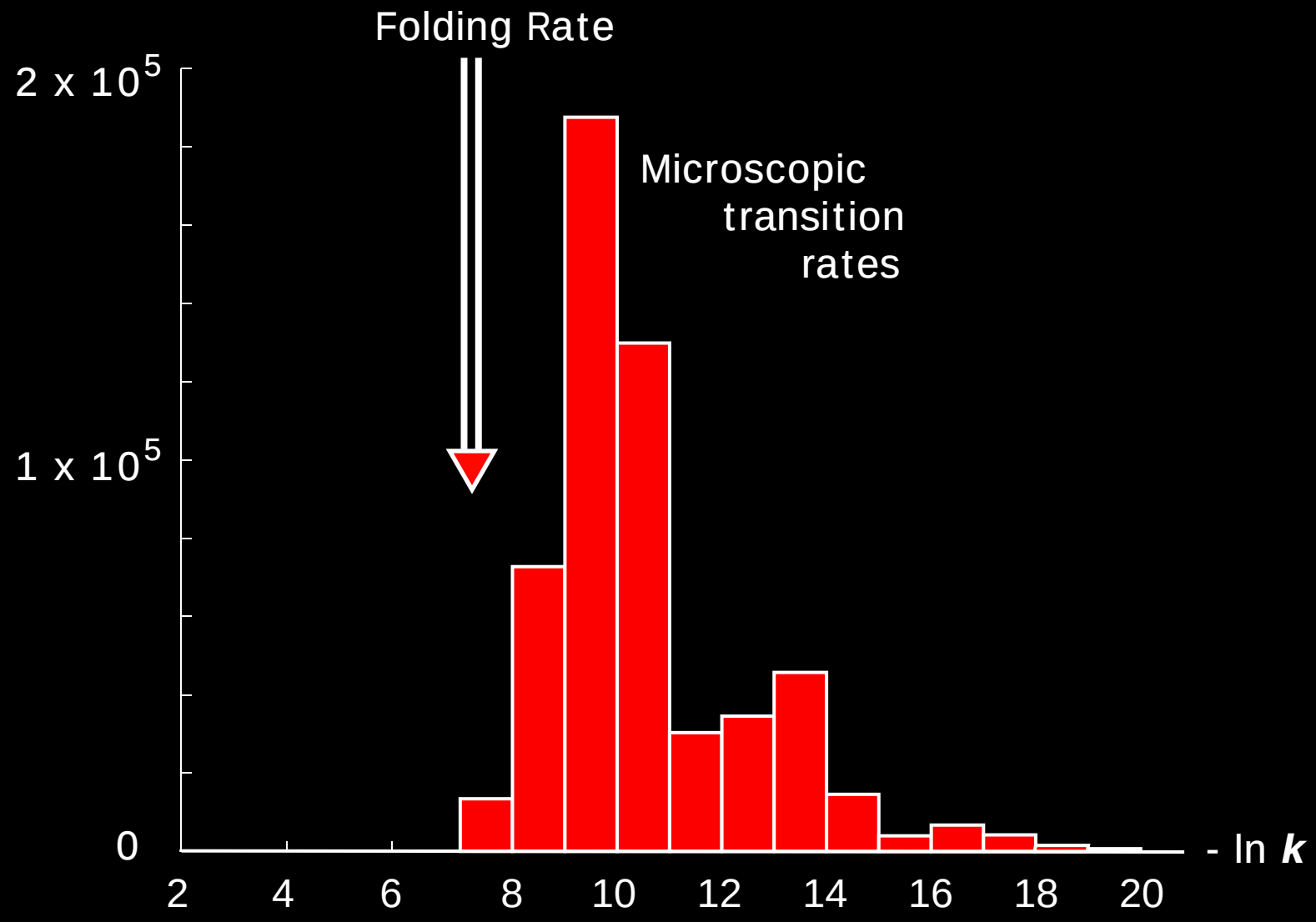
Parallel Process



macroscopic rate > μrates

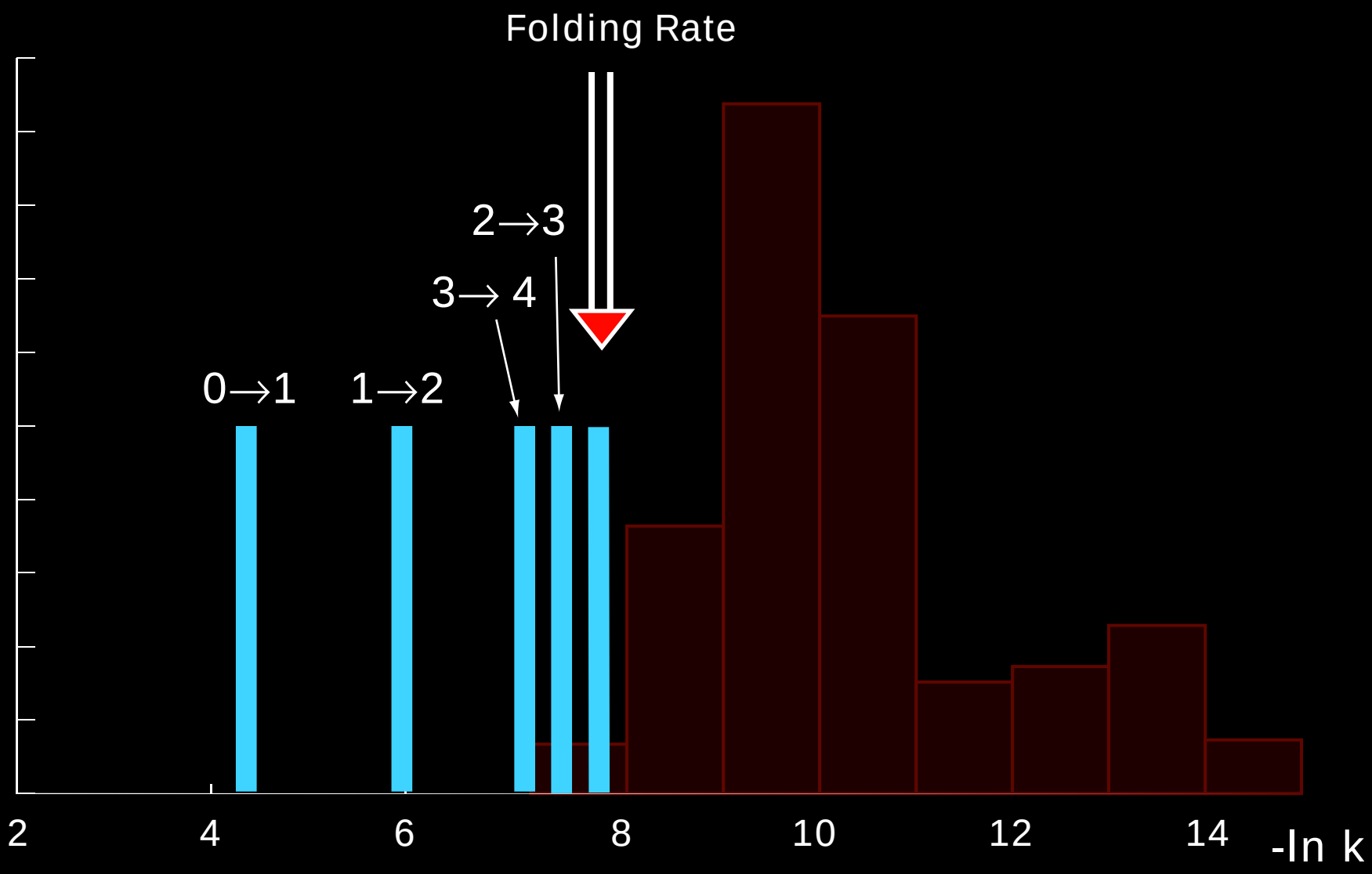


Folding is Faster than Microscopic Transition Rates



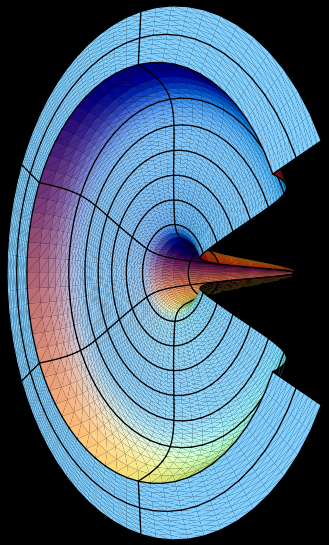
jack schonbrun

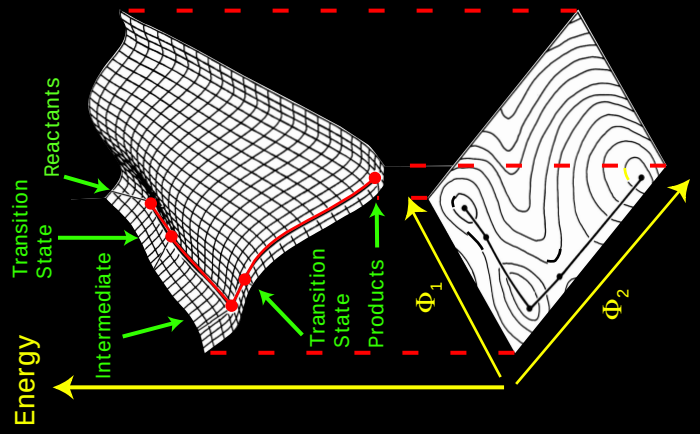
Transitions Between Energy Levels Are Fast

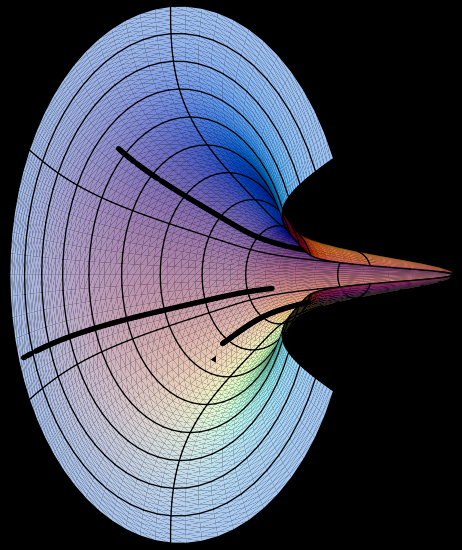


- jack schonbrun

Types of Rate Limiting Processes:



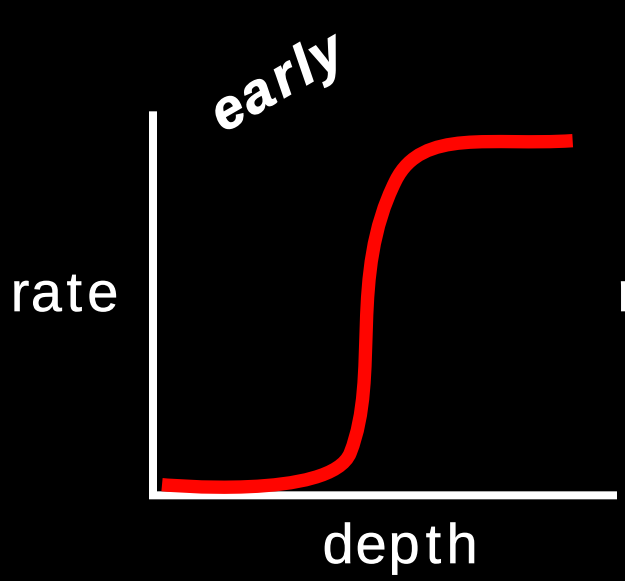


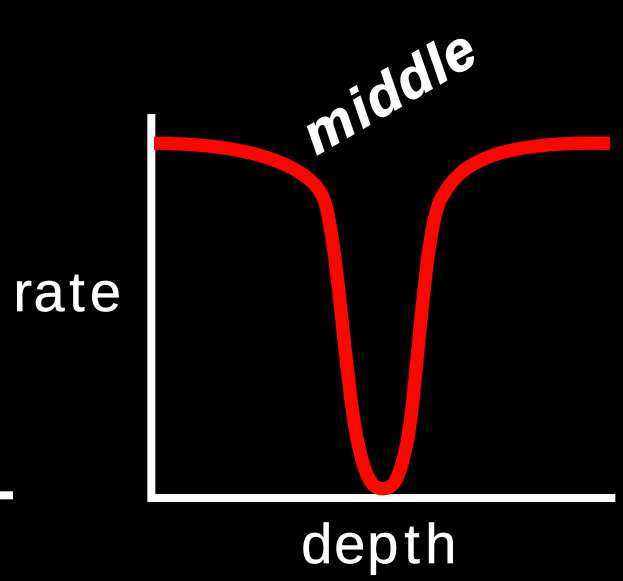


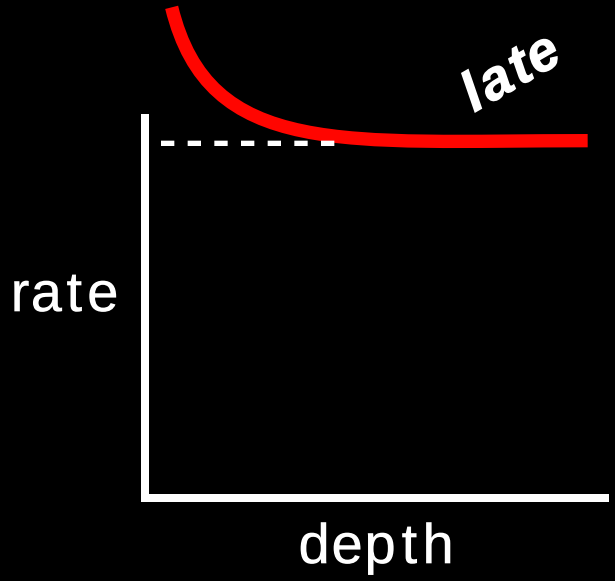
Random search

Specific microscopic barrier

Funnelling the trajectories

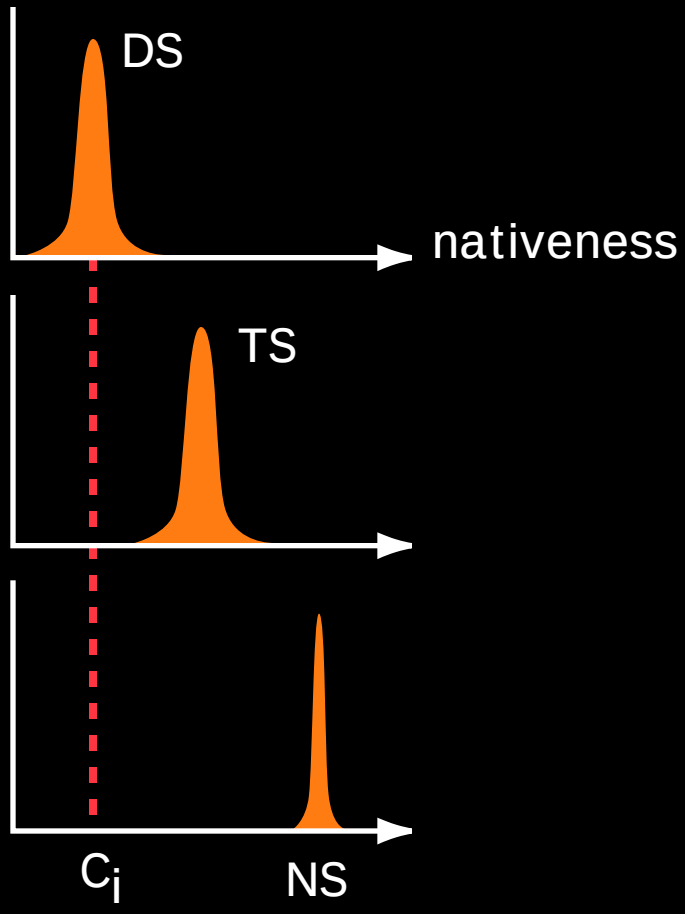






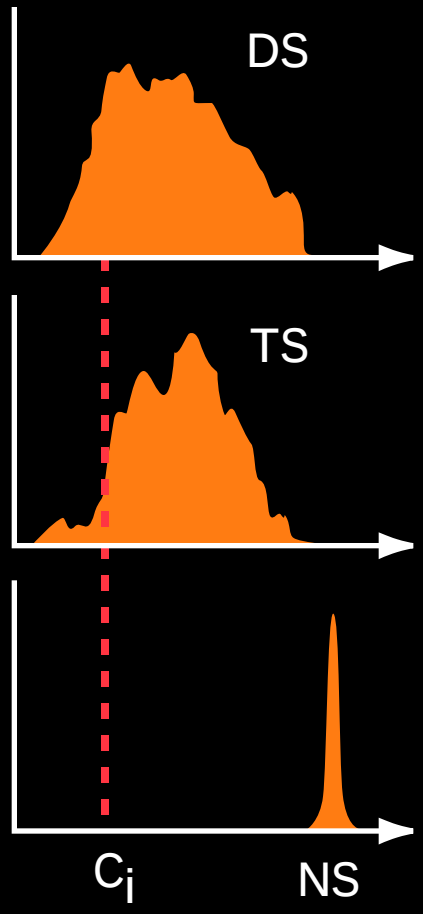
Which Microconformations are Transition State Macroconformations?

Series Model



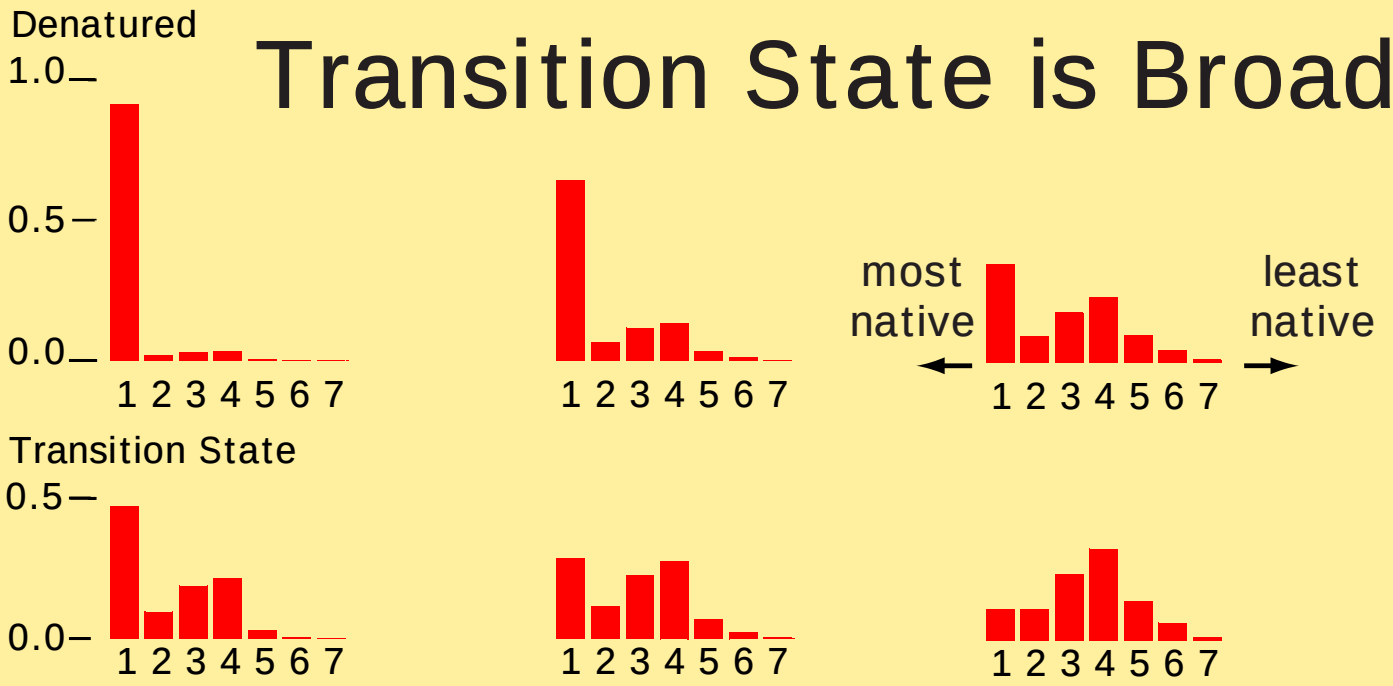
localized ensembles:
Conformation C_i is either D or TS

Parallel Model

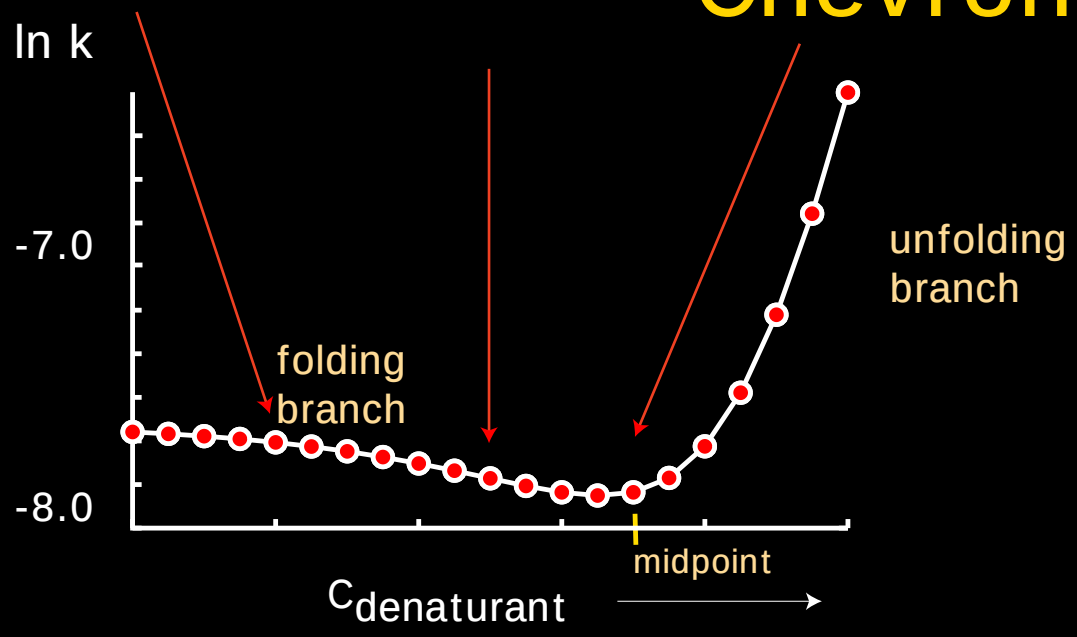


delocalized ensembles:
Conformation C_i can be in both D or TS

Transition State is Broad



Chevron Plot



D → TS → N or

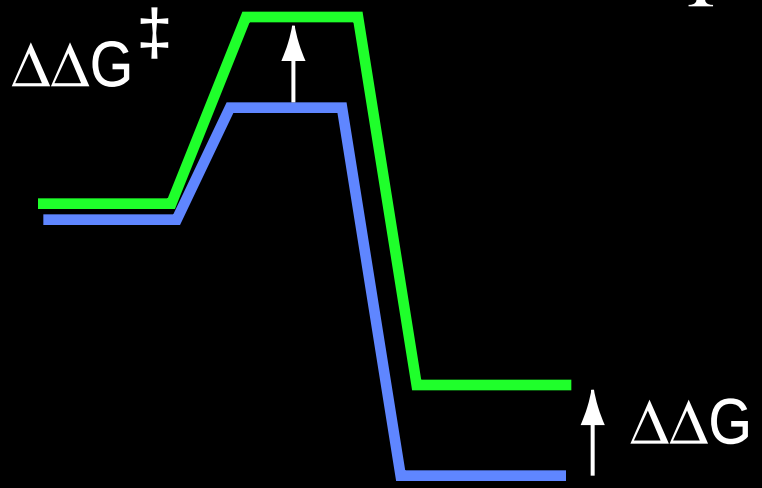
TS → D → N ?

Φ Value Analysis *

$\Delta G = -RT \ln K$

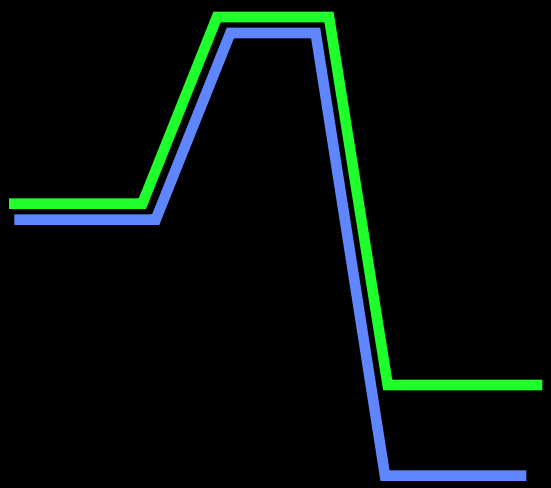
$\Delta G^{\ddagger} = -RT \ln k_f$

$$\Phi = \frac{\Delta\Delta G^{\ddagger}}{\Delta\Delta G}$$



$\Phi = 1$

At mutation site:
TS has Native-like structure

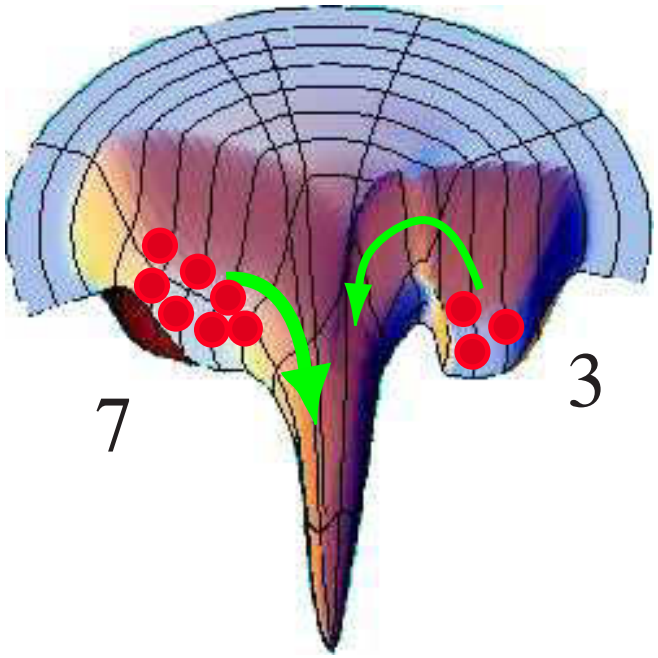


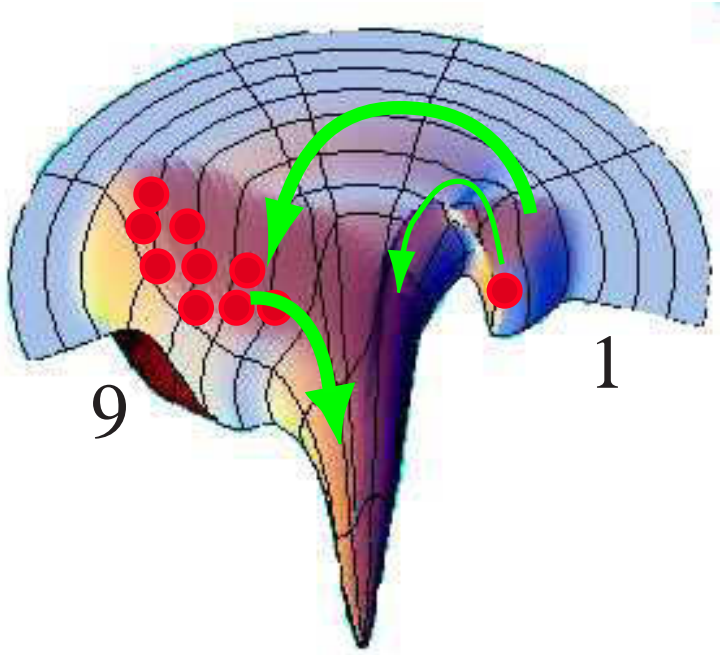
$\Phi = 0$

TS has Denatured-like structure

* A Fersht, Structure and Mechanism in Protein Science. Freeman (1999)

Negative Φ values come from Redirected Flow in Parallel Processes





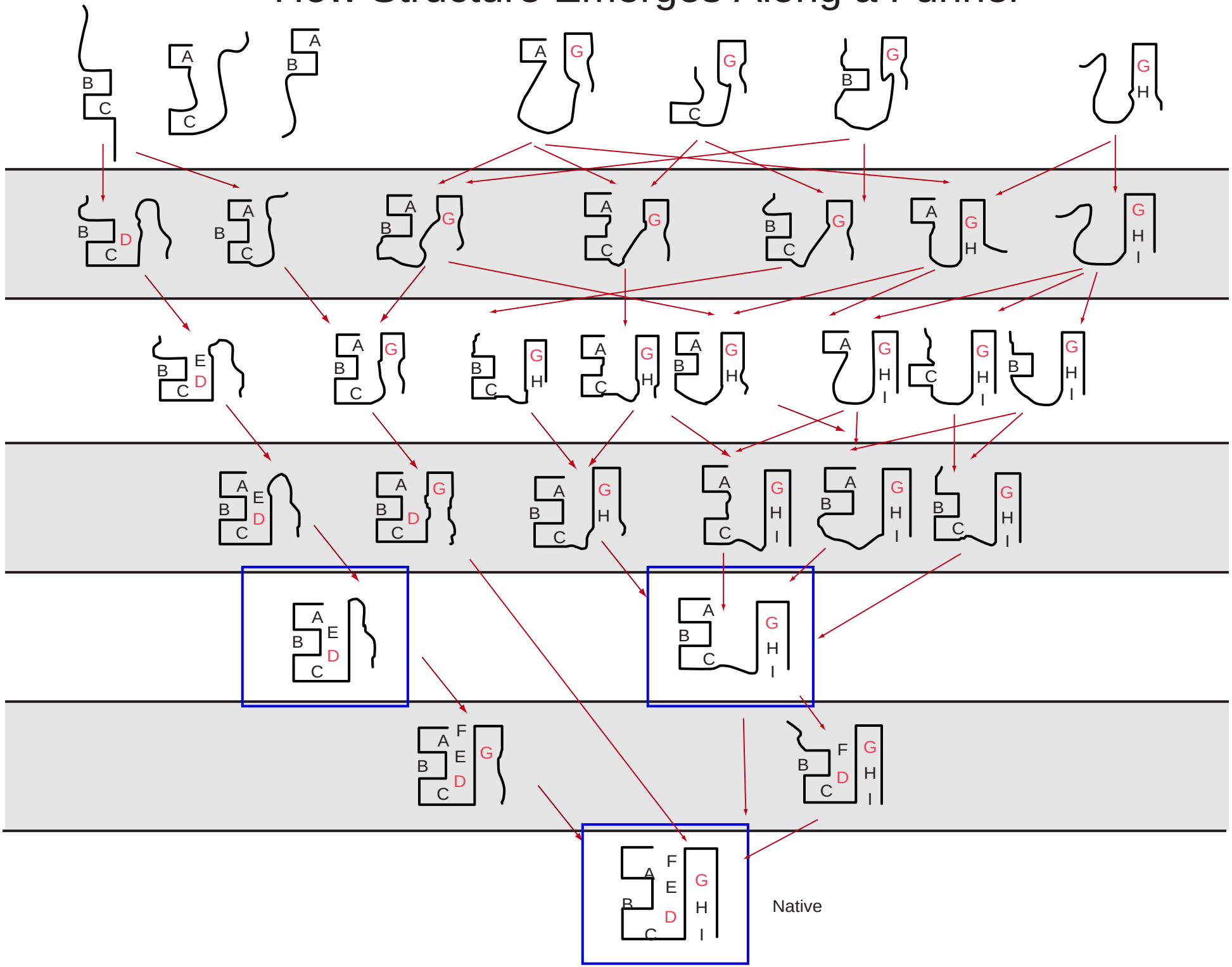
$$\text{Rate} = \left(\frac{7}{10}\right)1 + \left(\frac{3}{10}\right)(0.1) = 0.73$$

$$\text{Rate} = \left(\frac{9}{10}\right)1 + \left(\frac{1}{10}\right)(0.1) = 0.91$$

Destabilization leads to higher folding rates

- Banu Ozkan

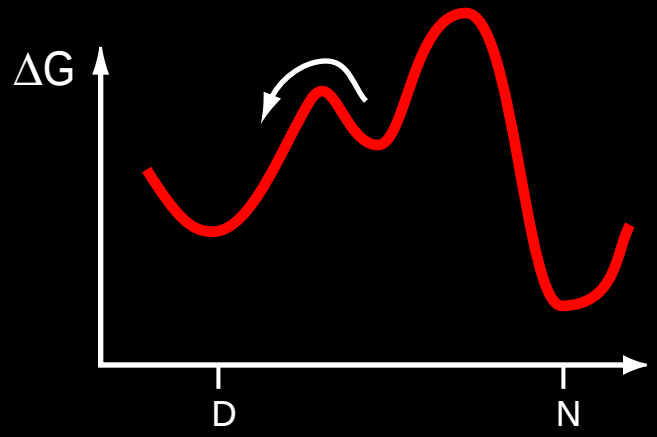
How Structure Emerges Along a Funnel



Mechanism of Chaperone Action

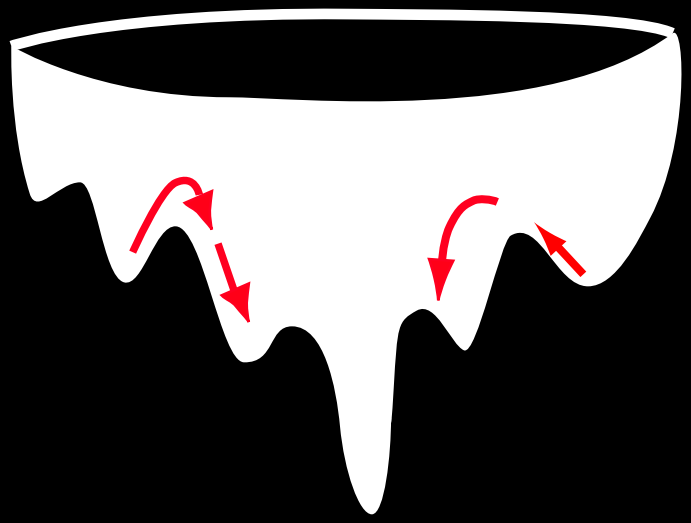
Series model dilemma:

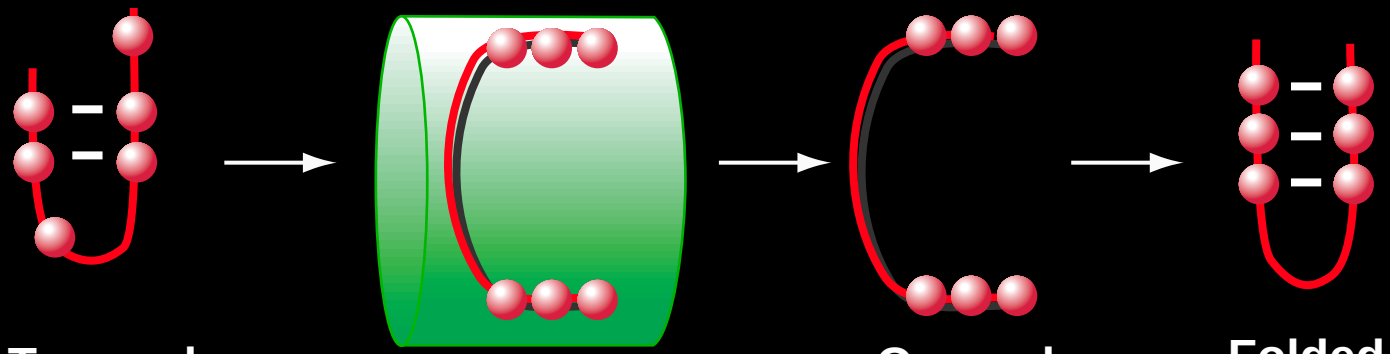
- How to recognize specific TS?
- Unfolding can't help a protein fold



Parallel model solution

- Unfolding a protein can help it fold





- Lorimer, Thirumalai, 1996. - Chan, Dill, 1996

Summary--
2-state Kinetics can come from:

Pathways
single rxn coord,
bottleneck step,
macro-rate < slowest micro-rate,
macrostates correspond to microstates

OR

Funnels
multiple routes,
early acceleration,
macro-rate > fast micro-rates
macrostates are ensembles

- Kinetics is a collective property of landscapes.
  Not a property of a single trajectory.

- In 2-state folding, what is the barrier?
  The whole folding process, not just collapse.

- Transition States are broad.
  They overlap with Denatured States.

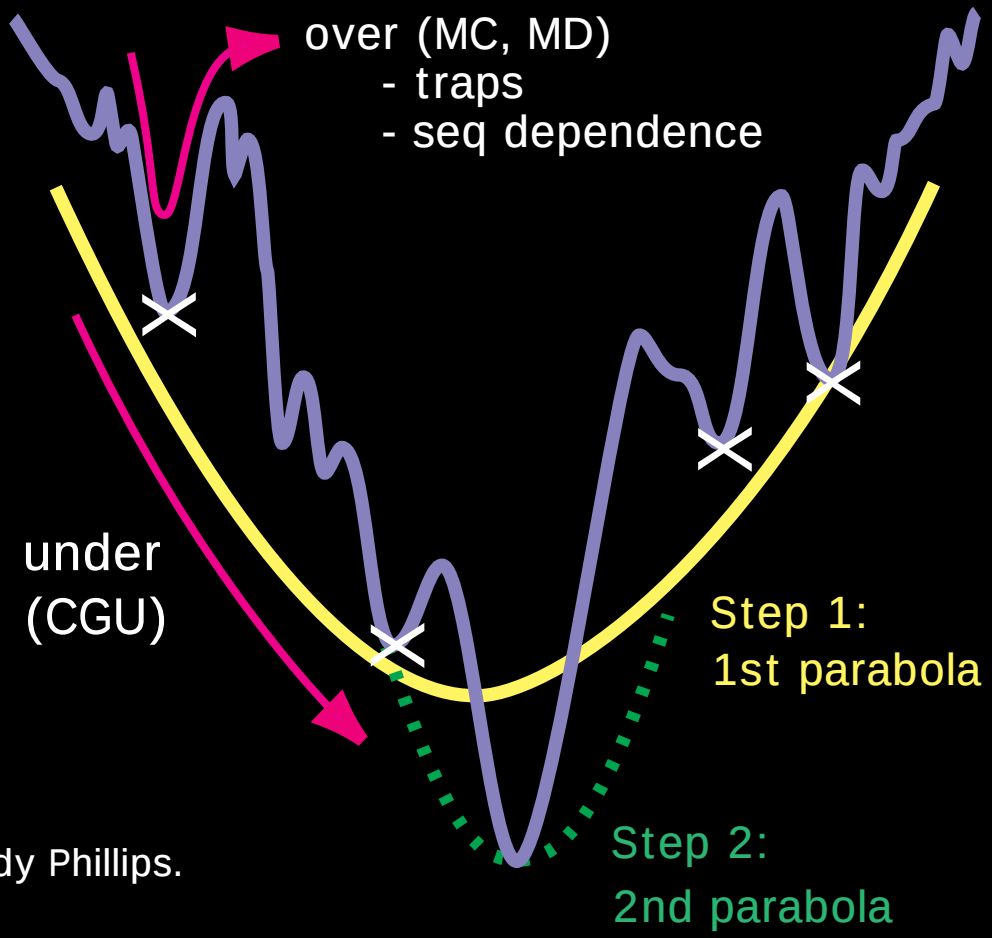
- Nonclassical Φ values are evidence for parallel steps.

- Terminology that applies to series processes,
  but not necessarily to parallel processes:
  (before, after)
  (backward, forward),
  (productive, unproductive (intermediates))

Thanks to:

Jack Schonbrun
Banu Ozkan
Ivet Bahar
Hue Sun Chan
Ben Rosen
Andy Phillips
Ken Foreman
NIH, NSF

A Fast Search Strategy from Landscape-ology



- CGU: Ben Rosen, Andy Phillips.
J Comp Bio '97

CGU:

- Finds Global Minimum
  - Depends on n, not sequence
    - Search time ~ n^4

SA Gets Trapped High on the Landscape Relative to CGU

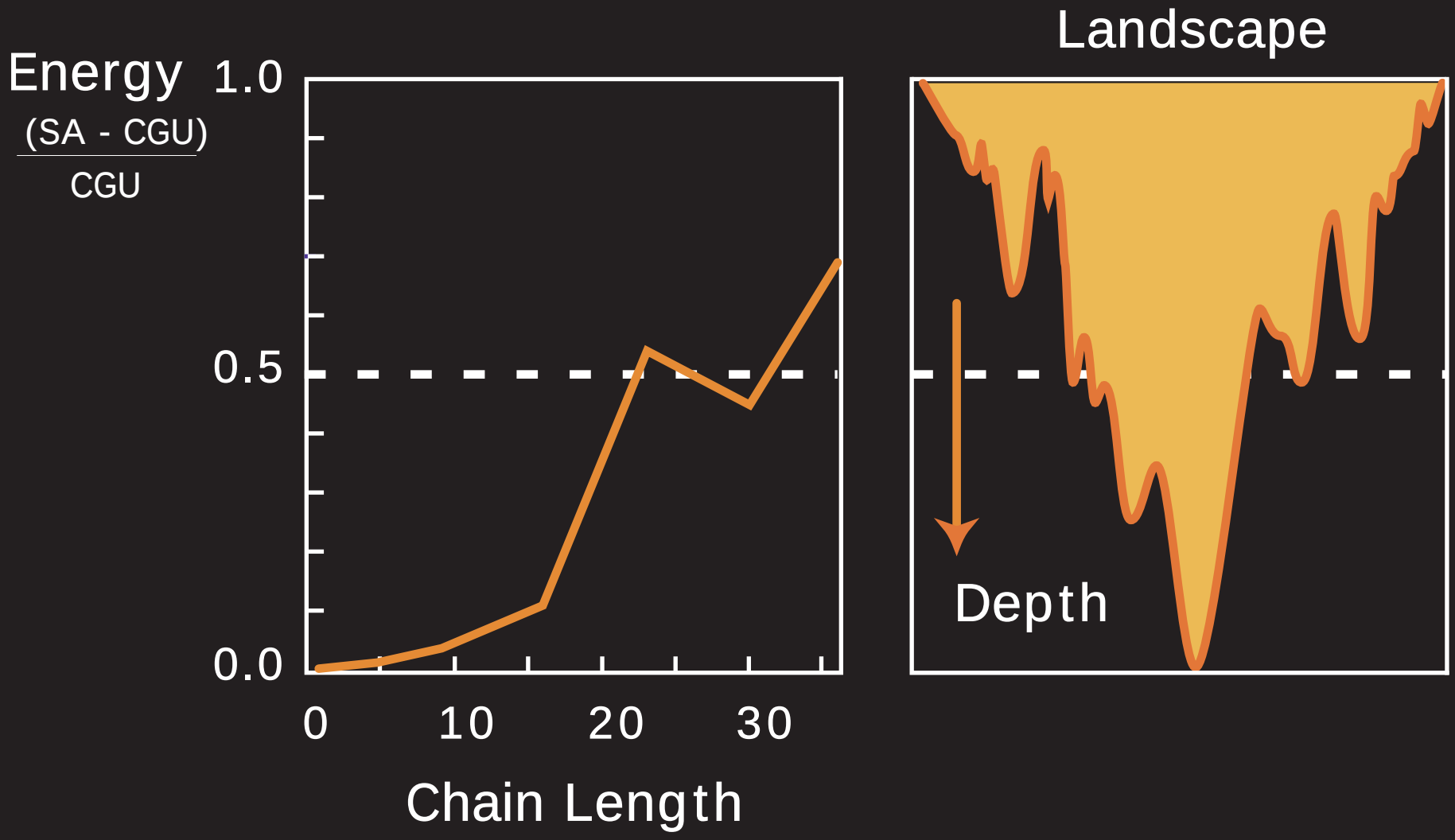


Table 2

ϕ Values for folding and difference energies for intermediate and transition state

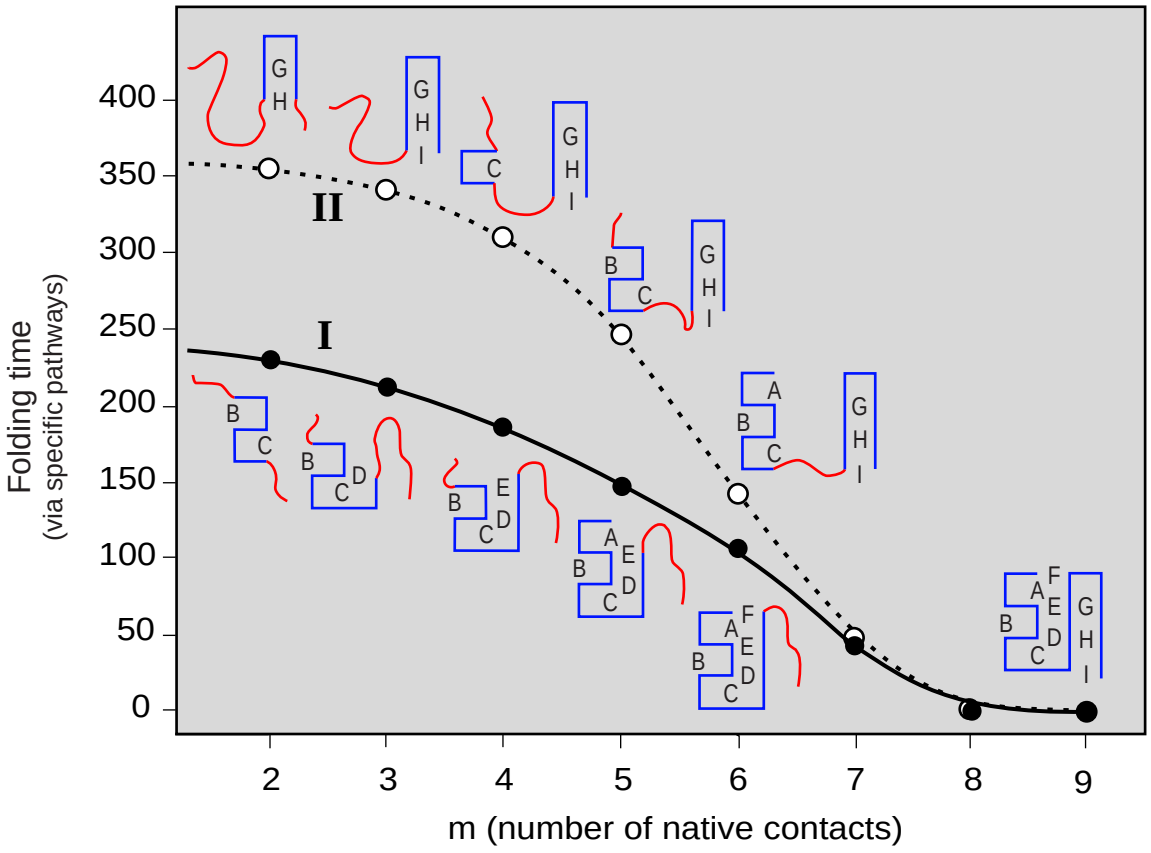
Residue and mutation	$\Delta\Delta G_{U-I}$ (kcal mol^{-1})†	$\Delta\Delta G_{U-\ddagger}$ (kcal mol^{-1})†	$\Delta\Delta G_{U-F}$ (kcal mol^{-1})†	ϕ_I	$\phi_\ddagger$
IV4	0·00	0·00	0·67	0·0	0·0
IA4	0·00	0·07	1·50	0·0	0·0
VA4	0·00	0·08	0·83	0·0	0·1
NA5	0·17	0·19	2·06	0·0	0·1
TA6‡	0·69	0·68	2·30	0·3	0·3
VT10	0·70	1·03	2·58	0·3	0·4
VA10	1·0	1·27	3·63	0·3	0·3
YA13§	1·54	1·88	3·71	0·4	0·5
YA13/YA17§	1·51	1·99	4·64	0·3	0·4
LA14‡	2·37	2·84	4·55	0·5	0·6
TS16‡	1·28	1·46	1·68	0·8	0·9
YA17§	1·02	1·28	2·26	0·5	0·6
HQ18‡	1·18	1·21	1·42	0·8	0·9
NA23	−0·05	−0·09	2·50	0·0	0·0
IV25	−0·27	−0·28	1·18	−0·2	−0·2
TA26‡	0·02	−0·04	2·00	0·0	0·0
EG29	−0·24	−0·25	1·90	−0·1	−0·1
VT36	−0·07	−0·04	1·15	0·0	0·0
VA36	−0·34	−0·12	1·34	−0·2	−0·1
ND41	−0·04	−0·06	2·51	0·0	0·0
VT45	−0·20	−0·16	2·44	−0·1	−0·1
VA45	−0·53	−0·32	1·83	−0·3	−0·2
IV51	−0·28	−0·29	1·85	−0·1	−0·1
DN54	−0·49	−0·53	2·42	−0·2	−0·2
DA54	−0·47	−0·51	3·10	−0·2	−0·2
IT55	0·28	0·42	1·00	0·3	0·4
IA55	0·68	0·76	1·28	0·5	0·6
NA58	1·93	2·04	2·17	0·9	0·9
KR62	0·32	0·37	0·43	0·8	0·9
IV76	−0·10	0·02	0·88	0·0	0·0
IA76	0·45	0·91	2·04	0·2	0·5
VA76	0·55	0·89	1·16	0·5	0·8
NA77	−0·02	−0·03	1·88	0·0	0·0
YF78§	0·16	0·15	1·41	0·1	0·1
NA84	0·37	0·32	2·24	0·2	0·1
IV88§	0·95	1·31	1·40	0·7	0·9
IA88§	2·48	3·97	4·16	0·6	1·0
VA88	1·58	2·71	2·76	0·6	1·0
LV89	0·12	0·03	0·03	0·5	0·1
LT89	1·56	2·99	2·90	0·5	1·0
VT89	1·39	2·91	2·55	0·5	1·1
SA91	1·09	1·80	1·93	0·7	0·9
SA92	1·74	2·61	2·74	0·6	1·0
IV96§	0·56	0·55	0·95	0·6	0·6
IA96§	2·26	2·86	3·32	0·7	0·9
VA96	1·69	2·31	2·37	0·7	1·0
TV105	0·69	1·11	2·25	0·3	0·5
IV109	0·14	0·13	0·82	0·2	0·2
IA109	0·91	1·34	2·22	0·4	0·6
VA109	0·77	1·21	1·40	0·6	0·9

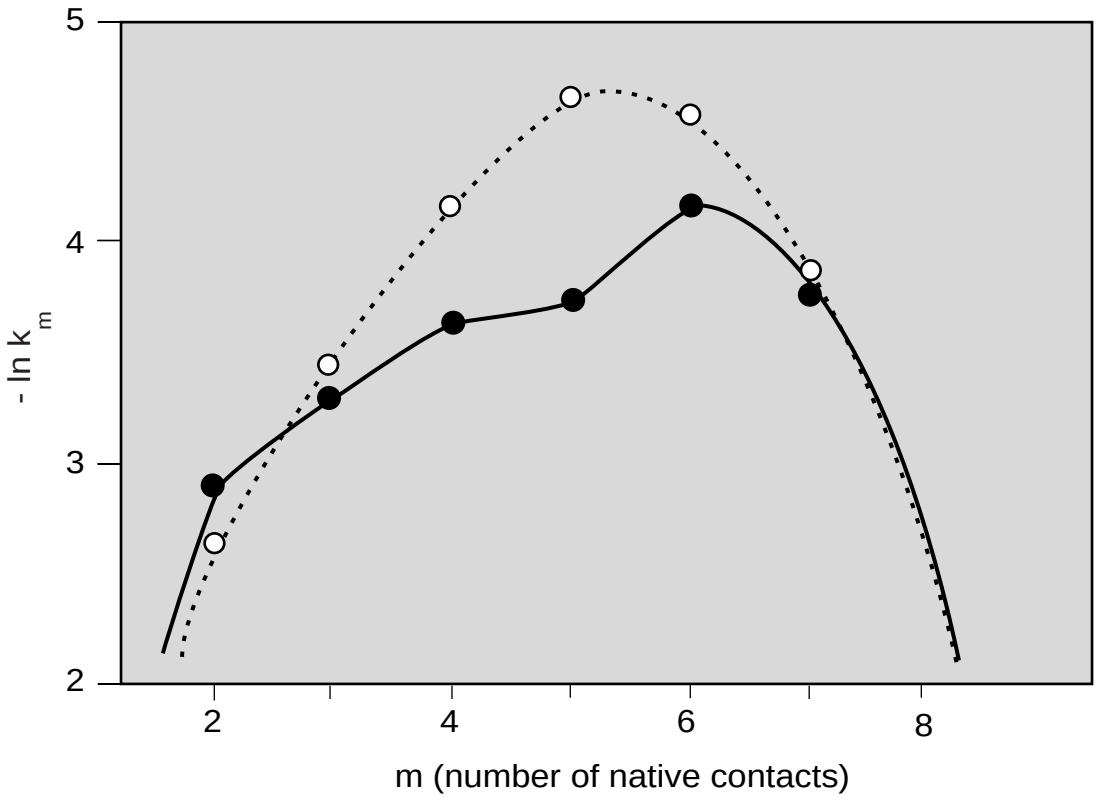
$\Delta\Delta G_{U-I}$, $\Delta\Delta G_{I-\ddagger}$ and $\Delta\Delta G_{U-F}$ are the difference energies defined in the text and paper I of the series, and ϕ_I and $\phi_\ddagger$ are the ϕ values for folding in water. The values of $\Delta\Delta G_{U-F}$ have been modified from paper II of this series (Table 3) by using the data from Table 7 of that paper to correct for the effect of ~4 M-urea.

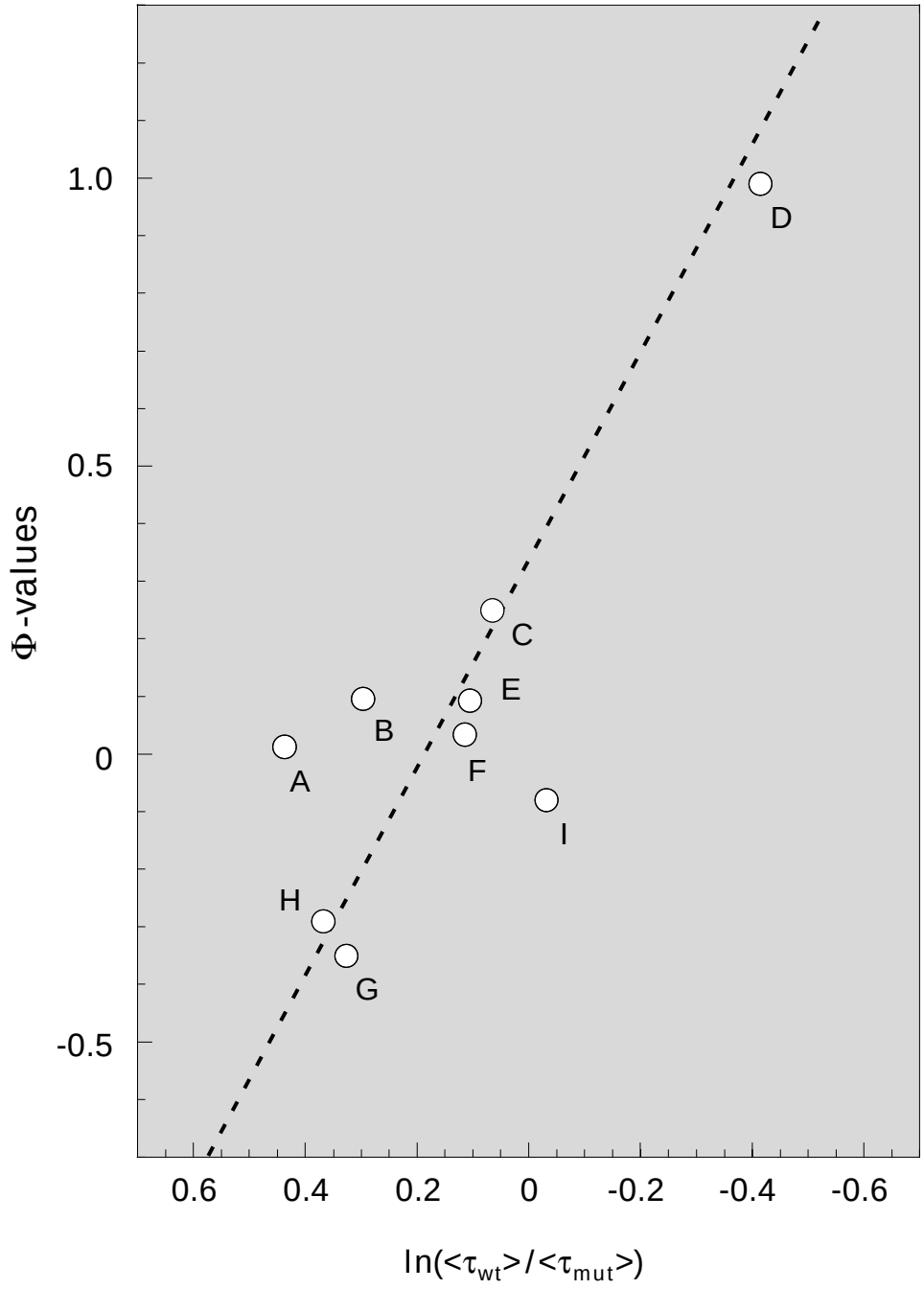
†The energies are measured with a standard error of ±0·15 kcal mol^{-1}.

‡Taken from Matouschek *et al.* (1990).

§Taken from Horovitz *et al.* (1991).

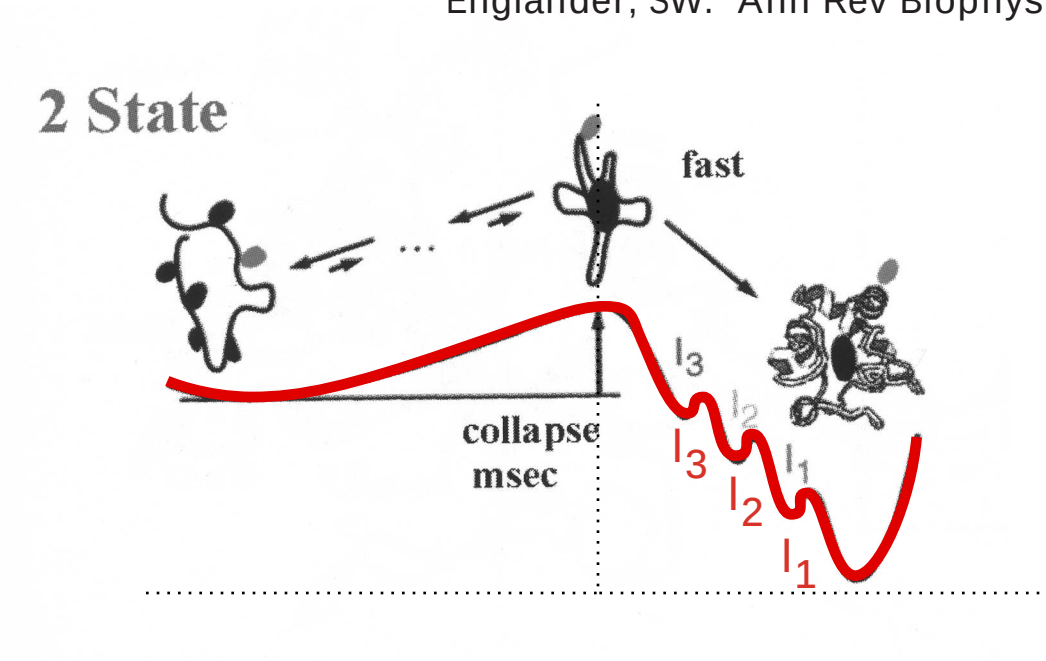

Folding time
(via specific pathways)
400
350
300
250
200
150
100
50
0
II
I
2
3
4
5
6
7
8
9
m (number of native contacts)


5
4
3
2
- ln k m
2
4
6
8
m (number of native contacts)


1.0
0.5
0
-0.5
Φ-values
D
C
E
B
F
A
I
H
G
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
ln(<τwt>/<τmut>)

Pathway Model: Collapse comes first, then detailed structure.

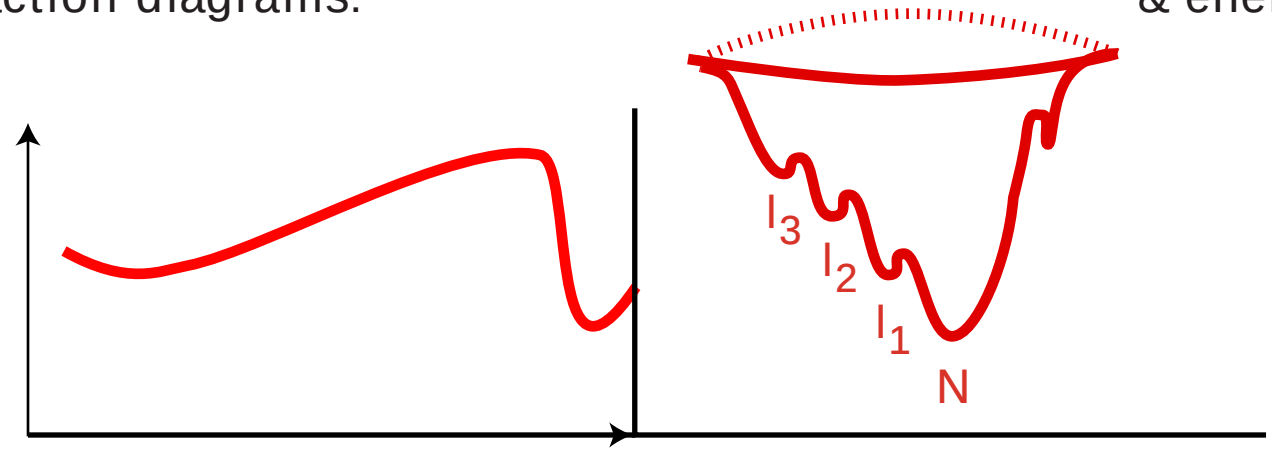
Englander, SW. Ann Rev Biophys Biomol Struc 29:213 (2000)



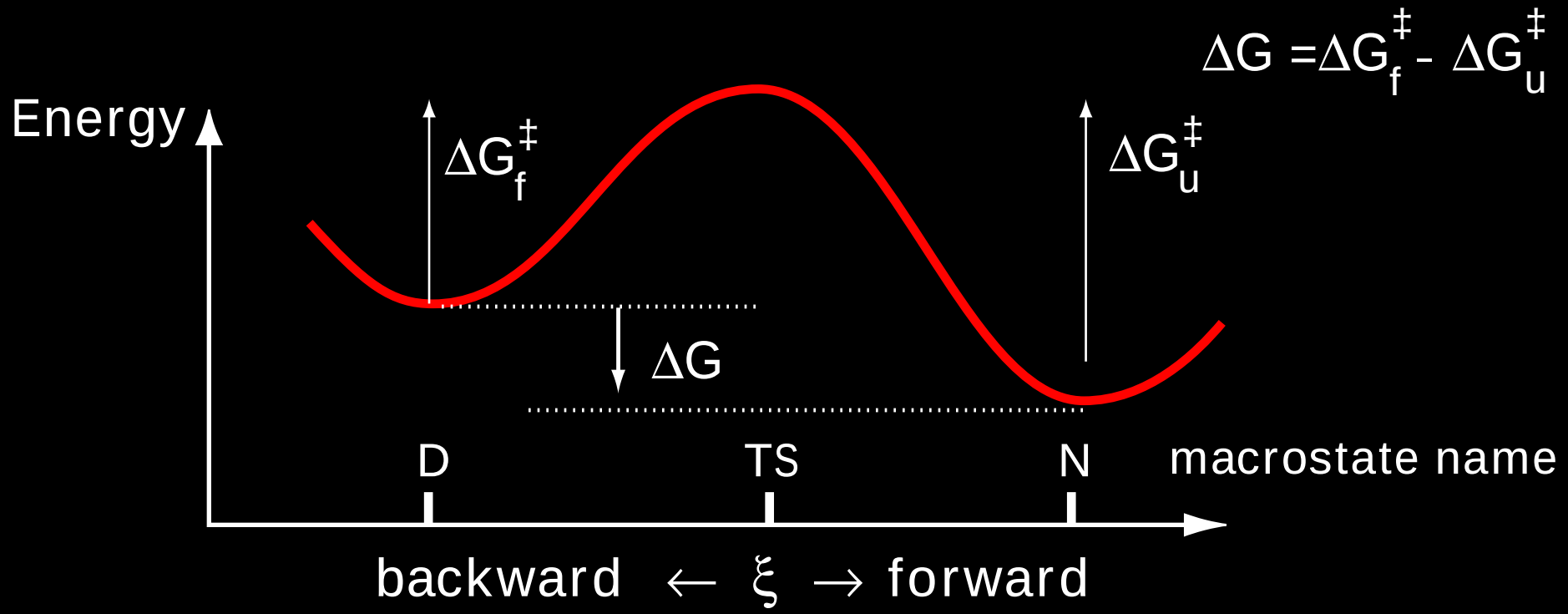
Funnel Model: Collapse and structure formation are simultaneous

This side is about
kinetics,
reaction diagrams:

This side is about
thermodynamics
& energy landscapes:



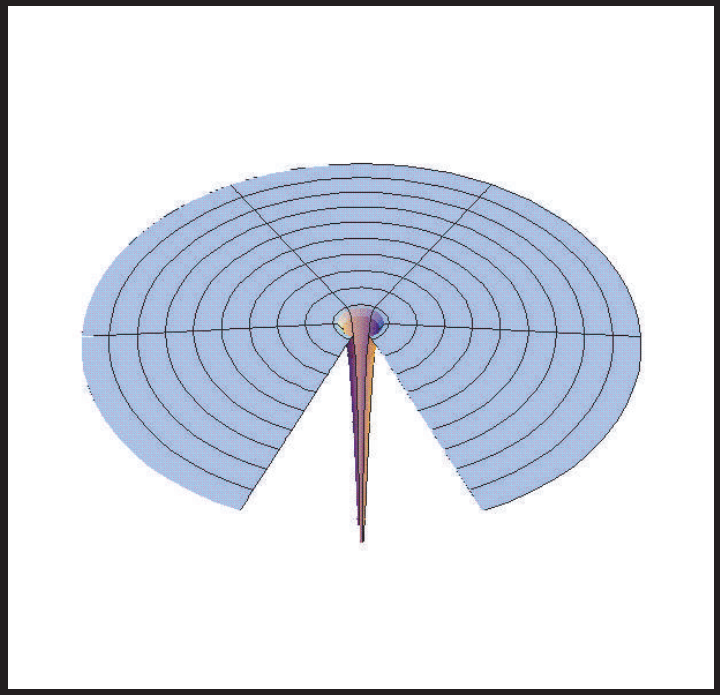
The Classical Transition State

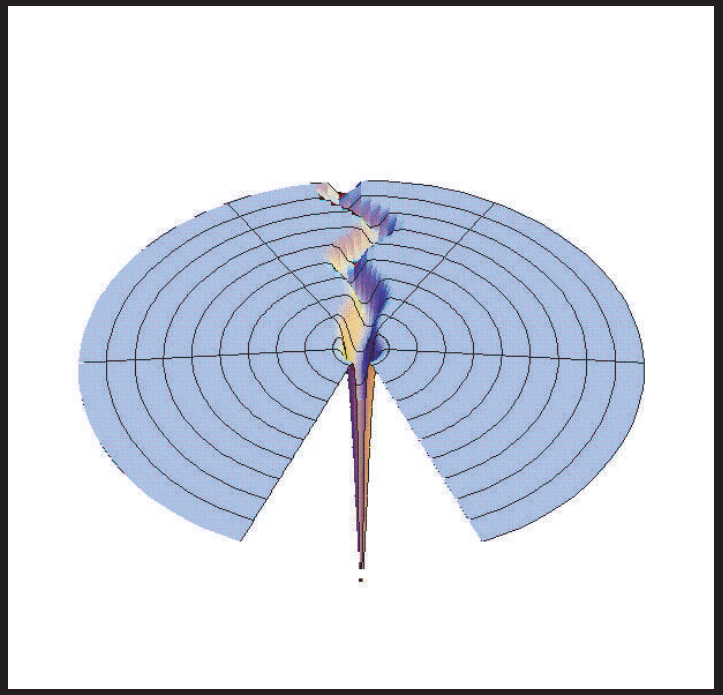


localized ensembles	$(C_1\ C_2\ C_3)$.....	$(C_i\ C_{i+1})$	(C_N)	microstates

- Macrostates are localized ensembles of microstates.
- States are in series and don't overlap.
- Single reaction coordinate. Forward & backward directions.

Levinthal's Paradox

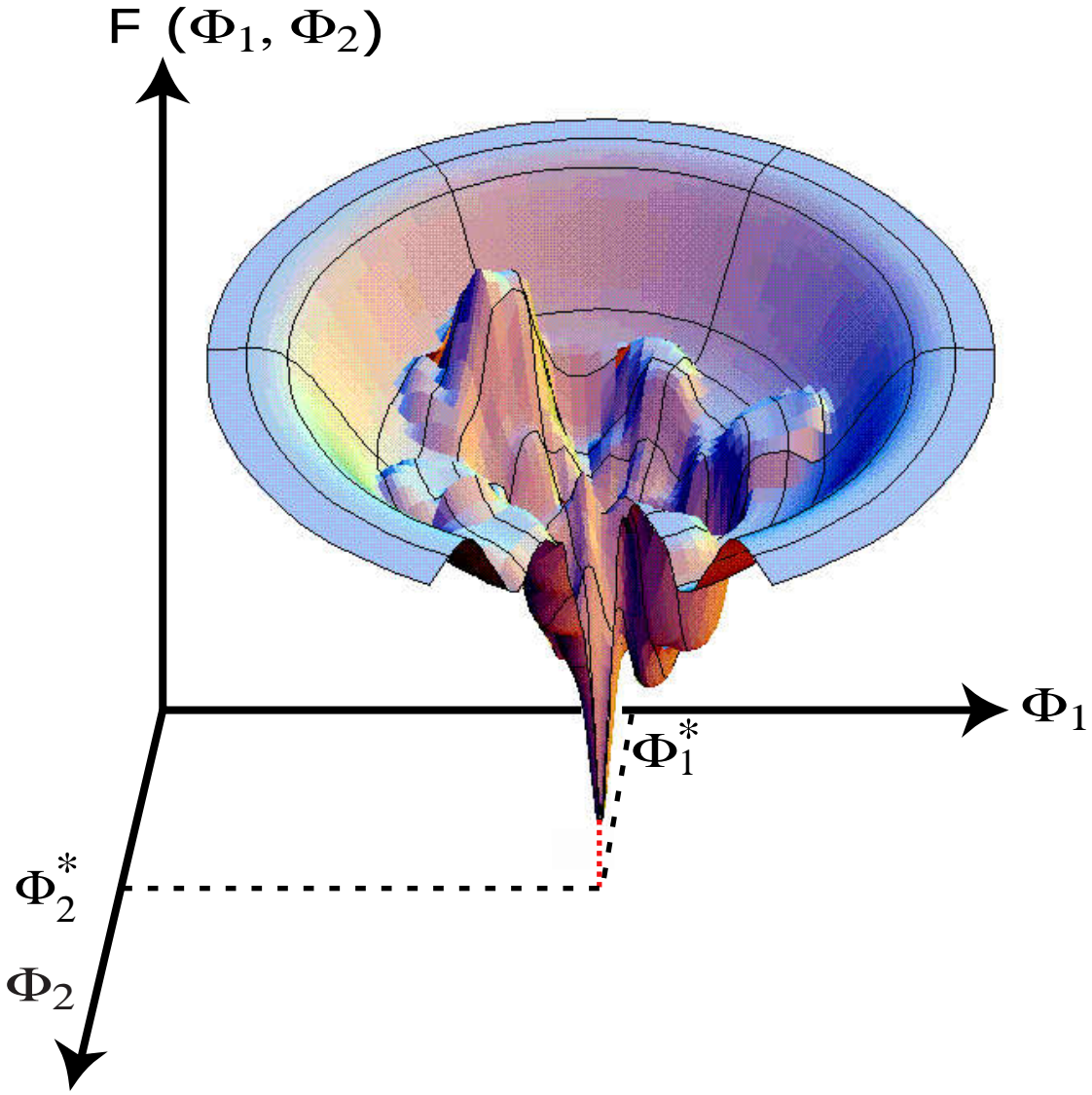




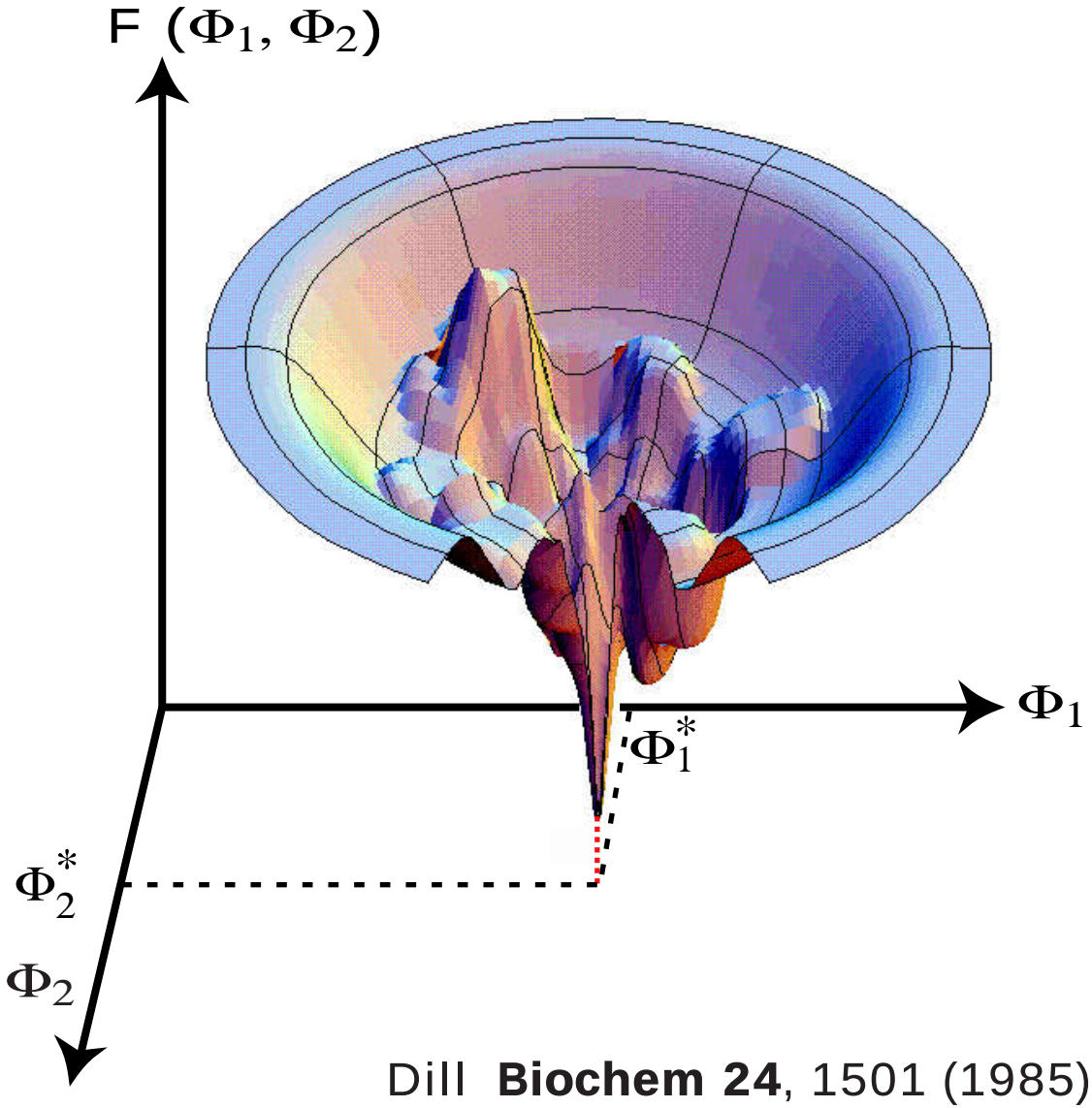
- Proteins find global min (Anfinsen)
- How so fast? (Levinthal)

Hue Sun Chan (Nat Struct Biol 4:10 (1997))

Energy Landscape



Protein Folding Energy Landscape



Dill **Biochem 24**, 1501 (1985)
Bryngelson & Wolynes **PNAS 84**, 7524 (1987)