A theoretical physicist looks at protein structure

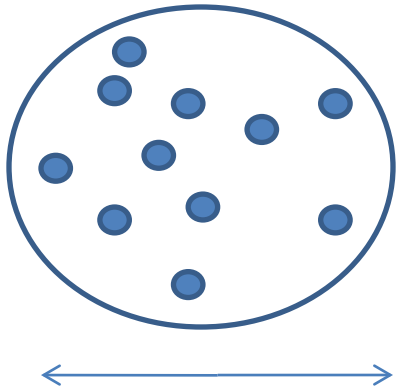
Kerson Huang

Physics Department, MIT

Institute of Advanced Studies, NTU

2013

Consider
N monomers formed under some rule



Flory exponent $R \propto N^{\nu}$ (for large *R* and *N*)

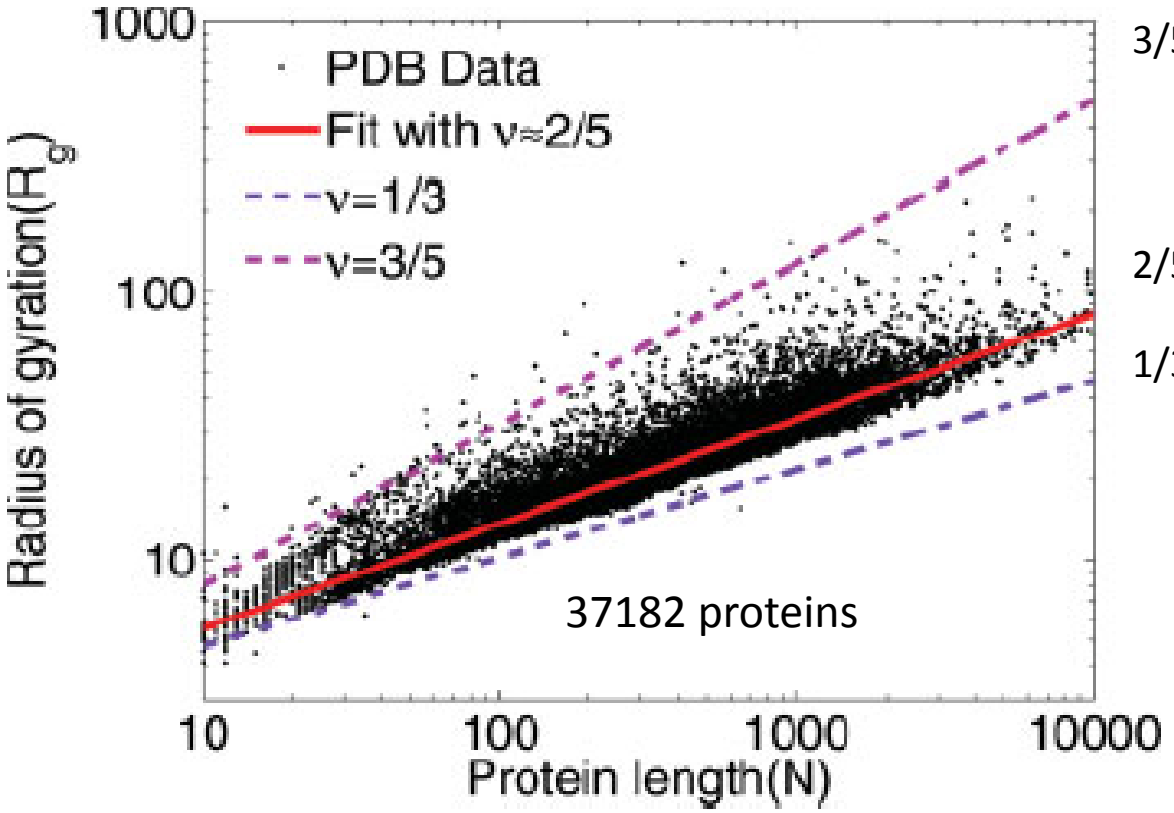
X-ray structure factor

$$g(k) \propto k^{1/\nu}$$

$1/\nu$ is a measure of the compactness of the structure.

Protein data analysis

L. Hong & J. Lei, *J. Polymer Sci.* B**47**,207 (2009)



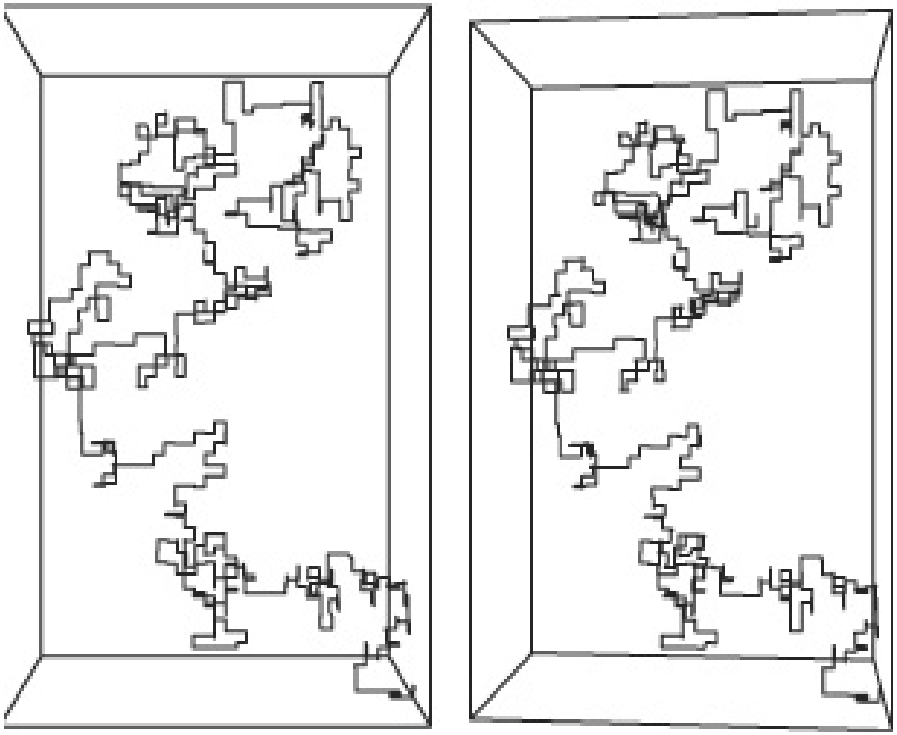
3/5 (unfolded state)

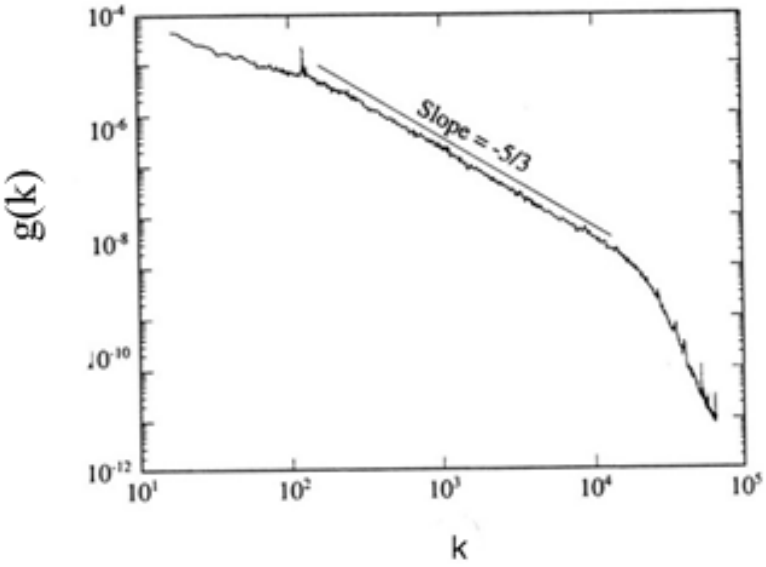
2/5 (native state)

1/3 (solid)

SAW (self-avoiding walk)

Model of polymer chain





Power spectrum of turbulence

$\nu = 3/5$

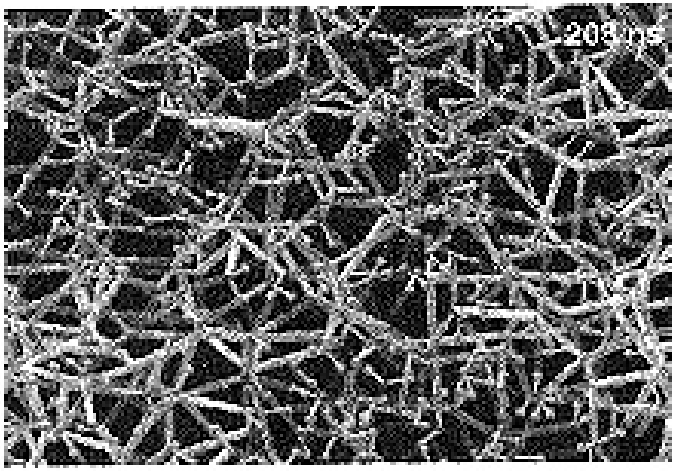
$g(k) \propto k^{-5/3}$ (Kolomogorov 5/3 law in turbulence)

SAW can simulate
- unfoleded protein chain
- vortex lines in turbulent fluid.

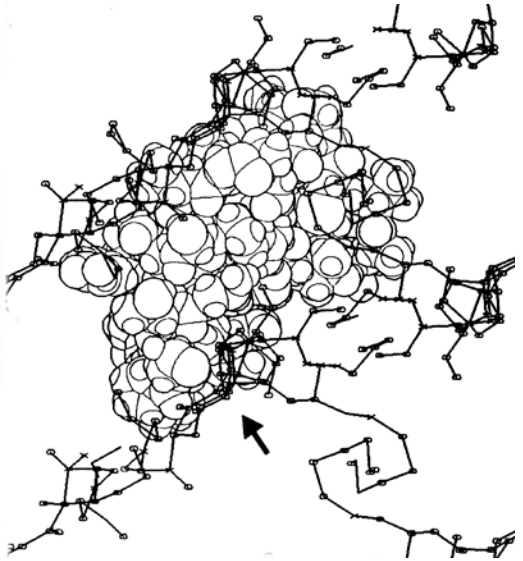
The Flory exponent has universality

Protein folding

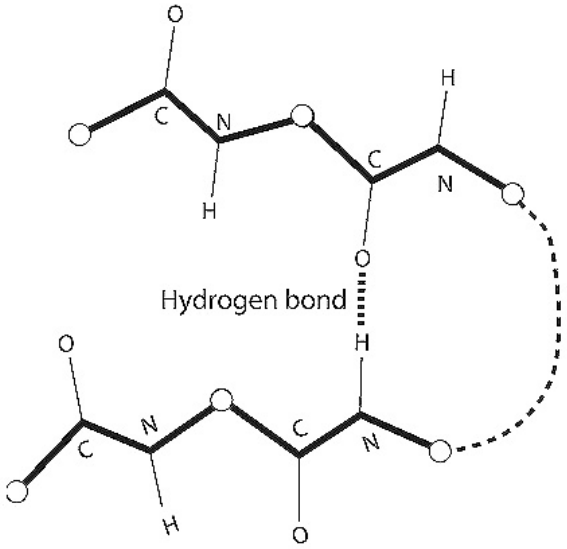
- hydrophobic forces
- hydrogen bonding



Water network (H bonds)



Protein with hydrophobic residues disrupts water network



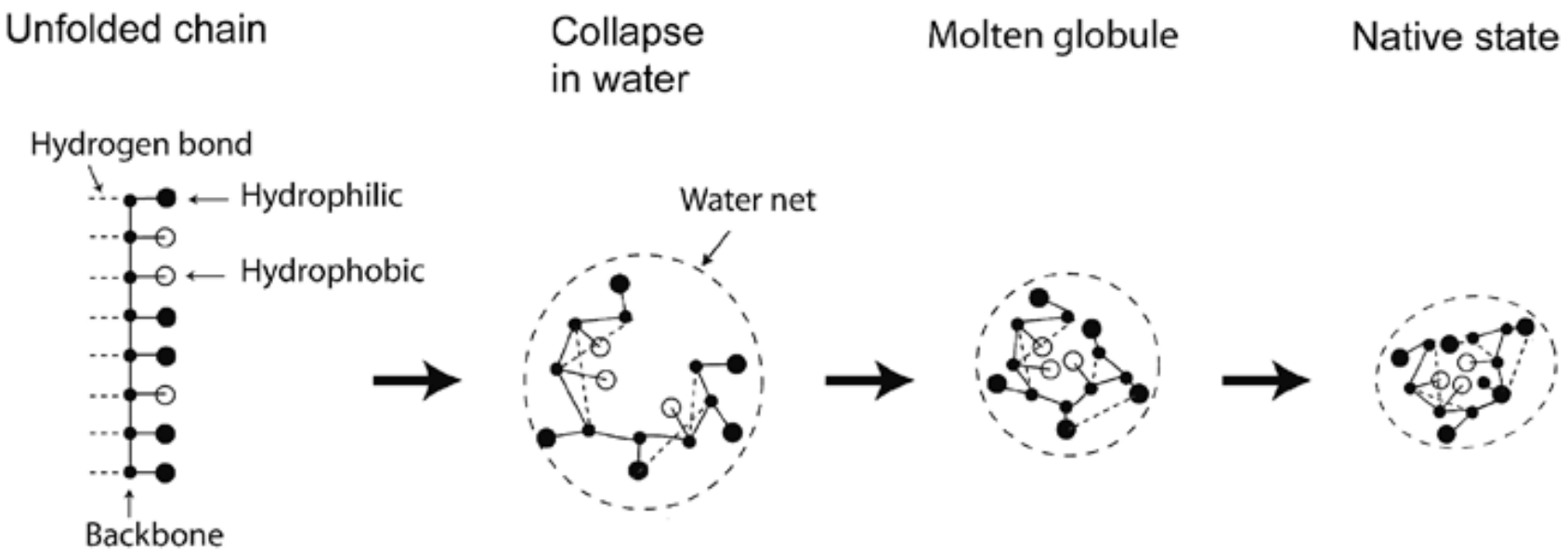
Internal H bonds

- Distance O-H = 2 A
- C=O and N-H antiparallel

Protein folding

Main forces:

- hydrophobic forces
- hydrogen bonding



$\nu = 3/5$ $\nu = ?$ $\nu = 2/5$

Extension of Flory theory leads to the following generalization

L. Hong & J. Lei, J. Polymer Sci. B47,207 (2009)

$$\nu = \frac{d+2}{(D+2)d}$$

$$D = \text{spatial dimension}$$

$$d = \text{fractal dimension}$$

Put D=3

Molten globule is as compact as native state:

Expected exponent = 2/5

Pre-globule (V.N. Uversky, Protein Sci. 11, 739 (2002):

Exponent = 0.411 +- 0.016 = 3/7

Expected stages of protein folding:

Unfolded	→	Pre-globule	→	Molten globule	→	Native state
$\nu = 3/5$		3/7		2/5		2/5

We shall verify this by computer simulation of protein folding

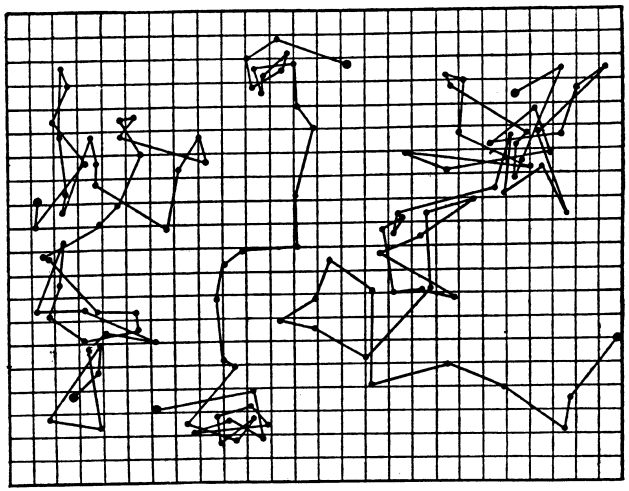
Simulation of protein folding:
Brownian motion of molecular chain in water

First consider
Brownian motion of one particle --- random walk.
Governed by Langevin equation:

Random force

Dissipation

$$m\ddot{x} = F(t) - \gamma\dot{x}$$



Perrin's famous drawings

This equation can be solved
- analytically
- through computer simulation of random walk

Conditioned random walk --- in potential G(x)

MonteCarlo: Simulation of thermalization

E = Energy = -dG/dx
Start with initial state
Generate random update

- If E decreases, accept it.
- If E increases, accept with probability

Canonical ensemble

Back to protein folding

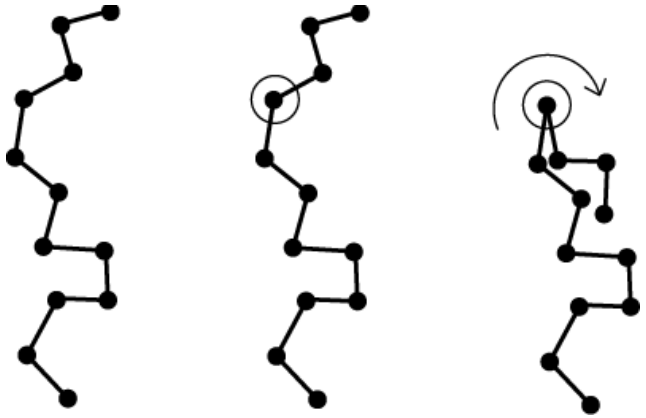
Unfolded states: SAW (self-avoiding walk)

To generate one SAW from another, use **pivot algorithm**:

Start with initial SAW chain

- **@** Choose arbitrary pivot
- Rotate end portion about pivot (by changing torsion angles)
- If no overlap, accept update
- If overlap, go to **@**

This will generate a uniform ergodic ensemble of SAW.



Protein folding: CSAW (Conditioned self-avoiding walk)

E = Energy of a configuration

- Start with initial chain.
- Generate SAW by pivot algorithm.
- If E decreases, accept it.
- If E increases, accept with probability

This will generate a canonical ensemble of folded states.

Model depends entirely on choice of the configuration energy E.

Simplest CSAW model:

Consider only
- hydrophobic forces
- hydrogen bonding

$K_1 =$ Contact no. of hydrophobic residues

$K_2 =$ No. of hydrogen bonds

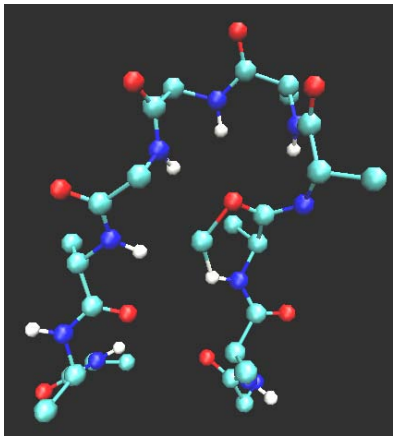
Contact no. = No. of nearest neighbors,
not counting those along the chain.

Folding of Chignolin by CSAW

Jinzhi Lei

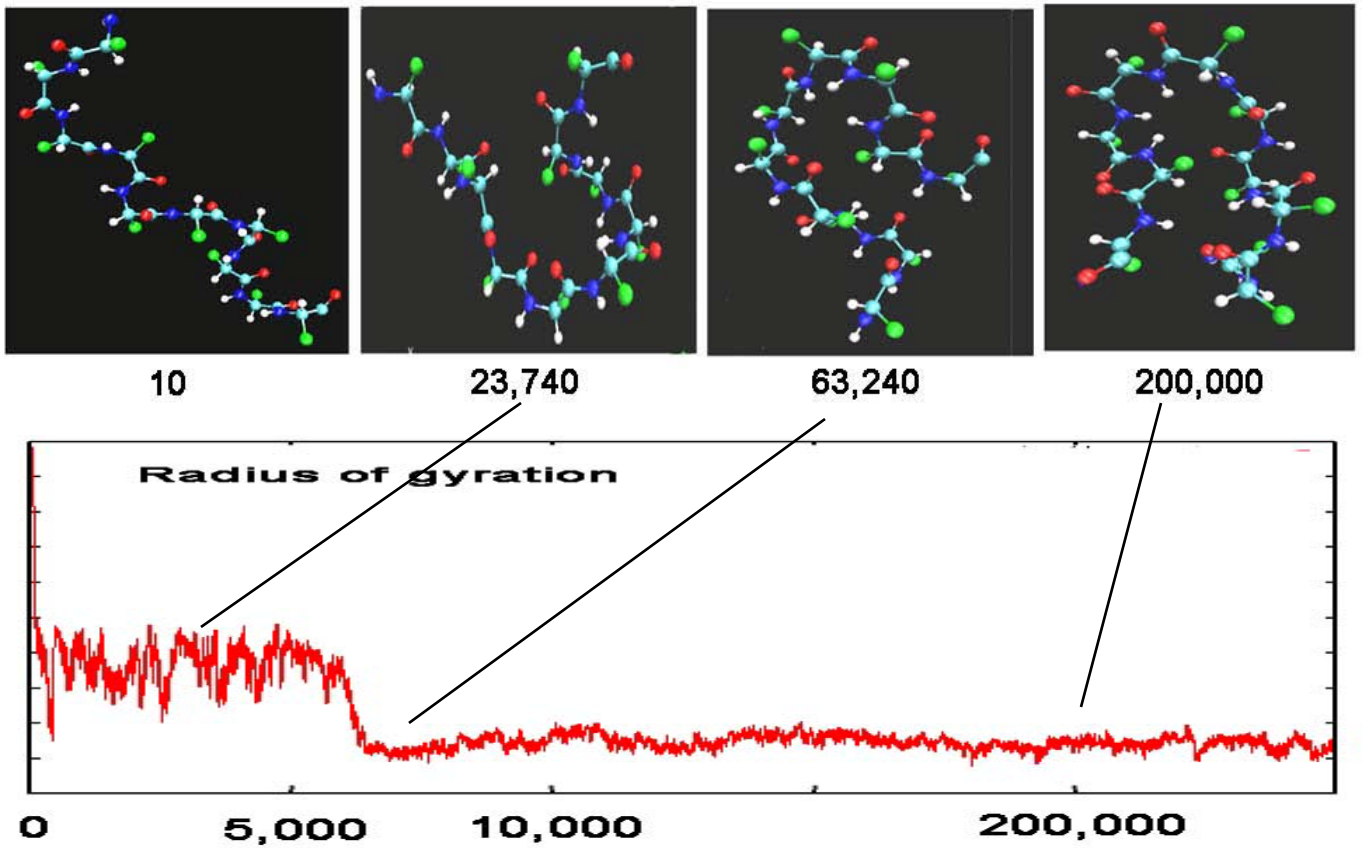
250,000 MC steps

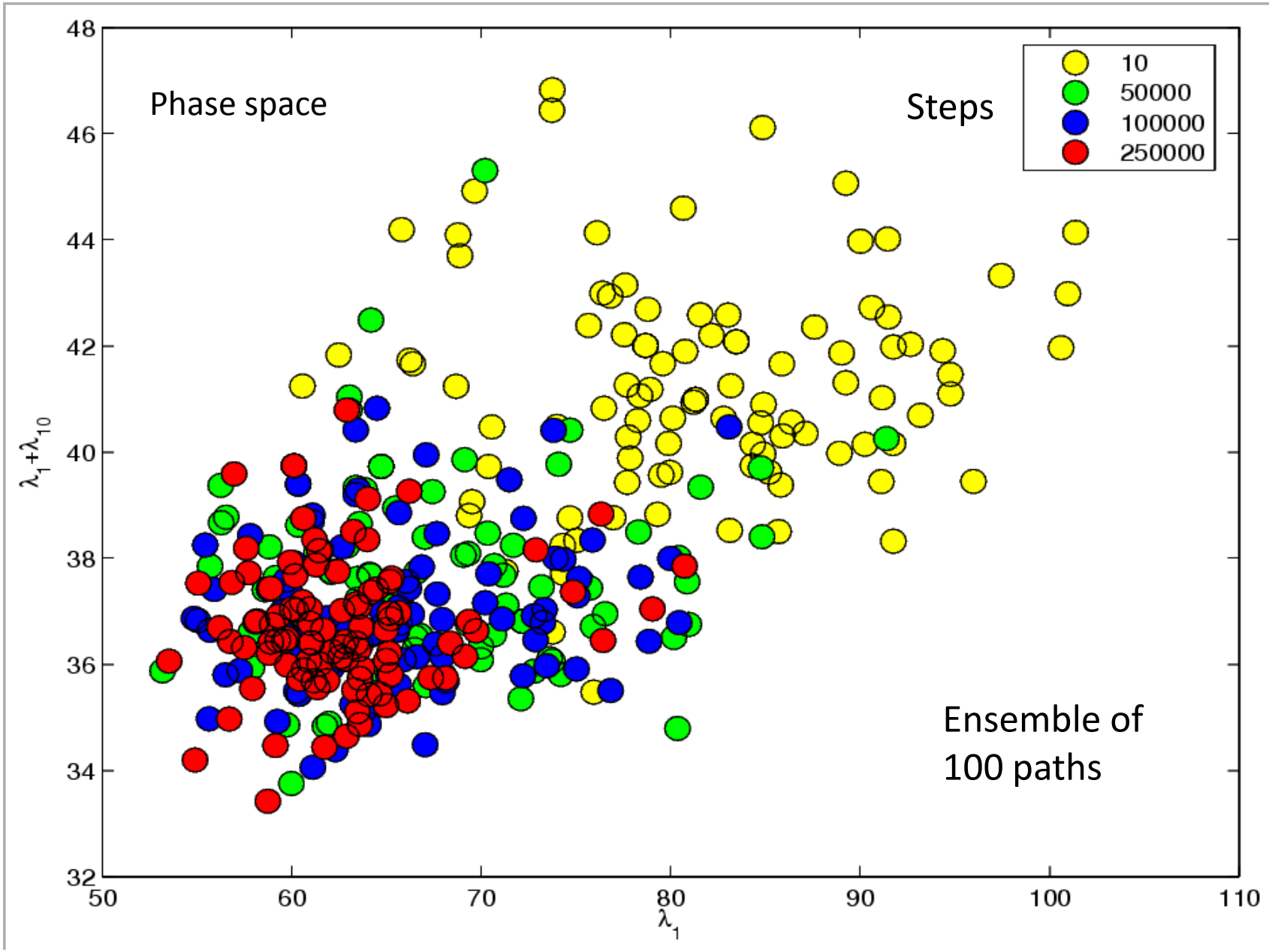
10 minutes on work station

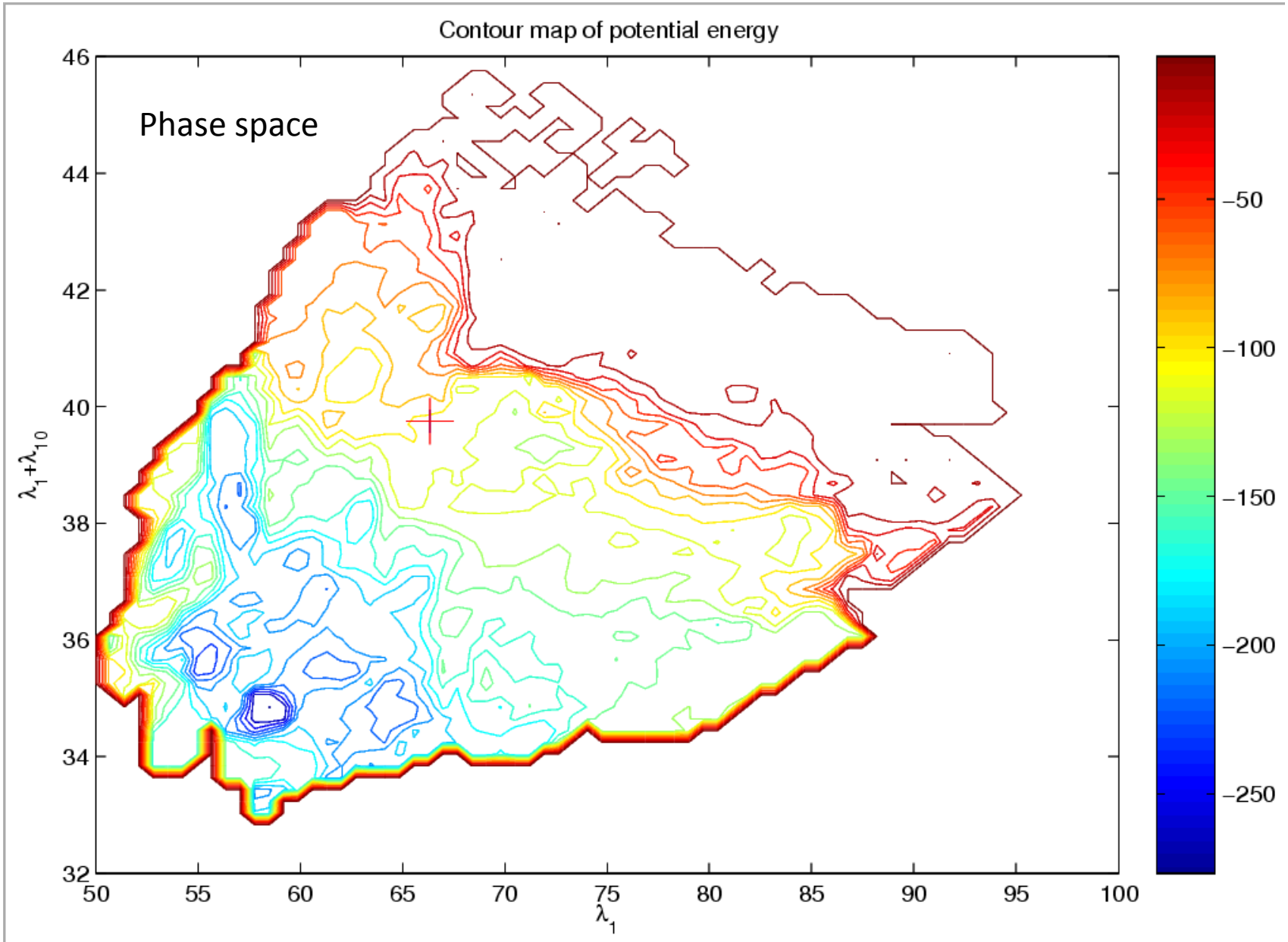


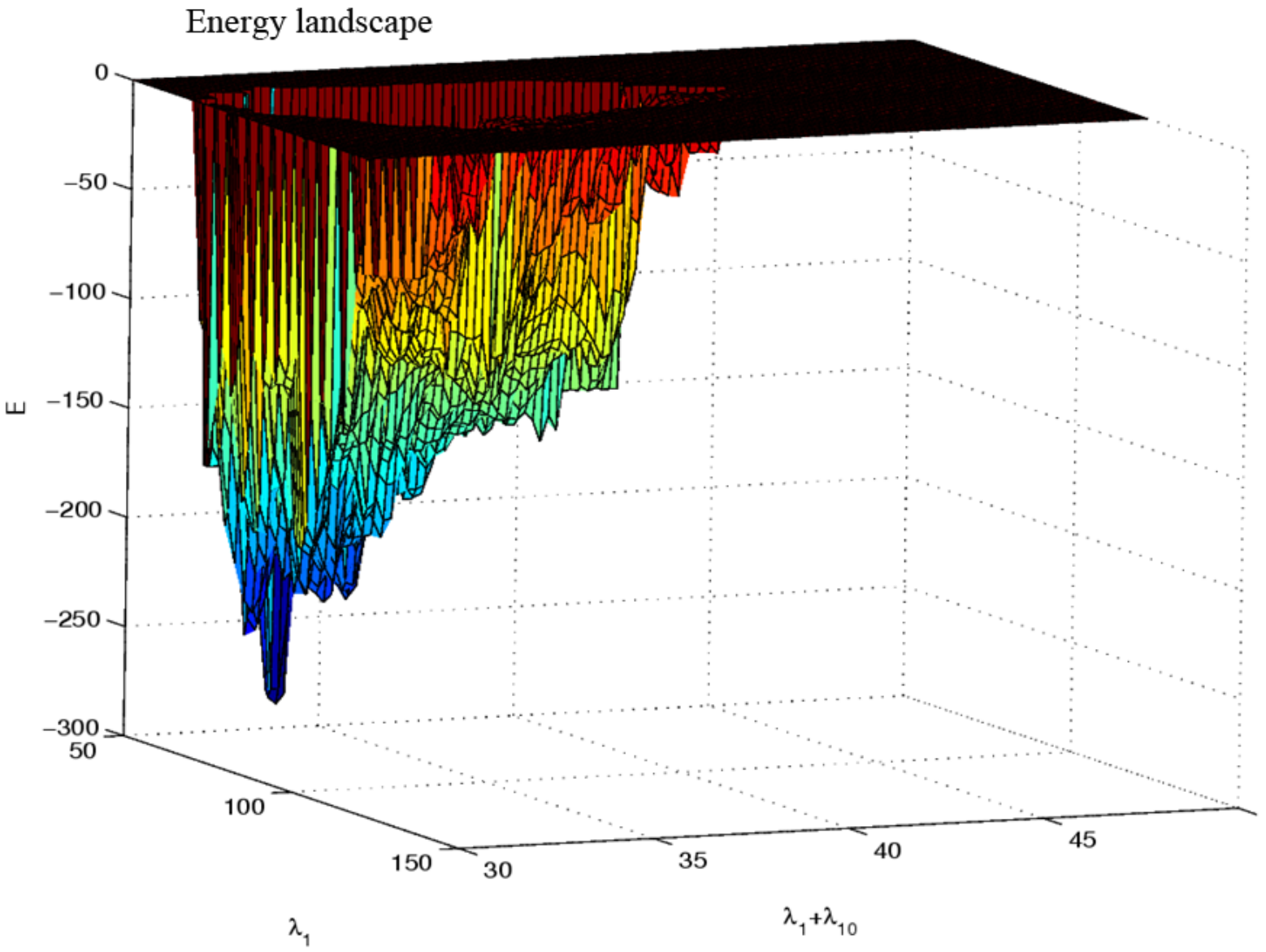
Chignolin:

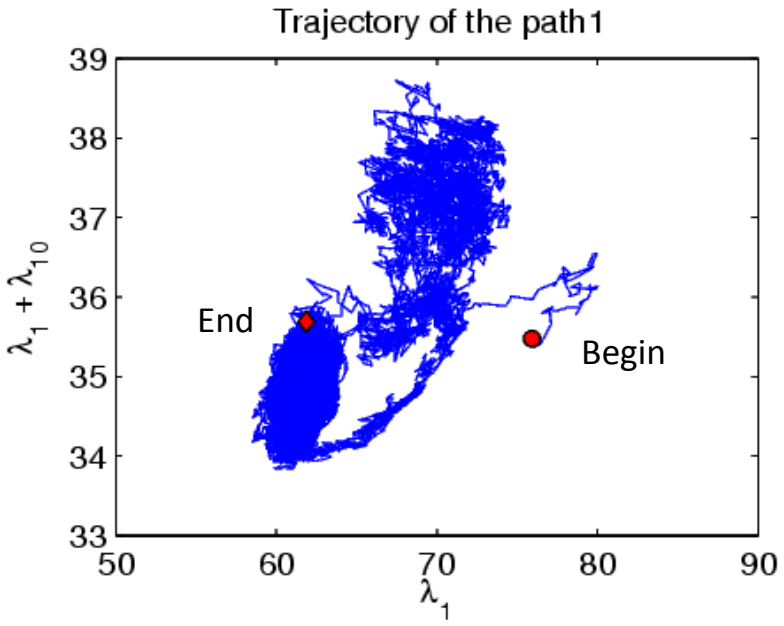
- Synthetic protein
- 10 amino acids
- Native state: beta hairpin
- Honda et al, Tsukuba, 2004

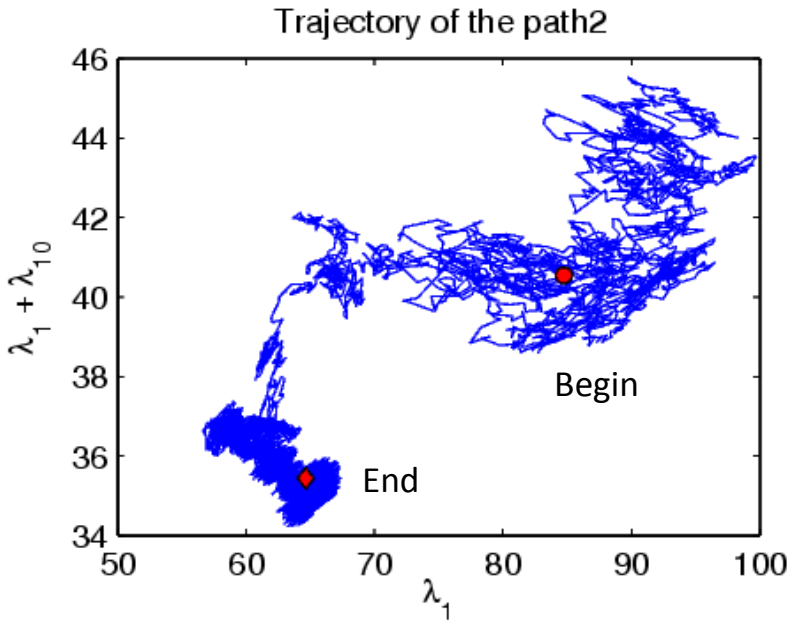


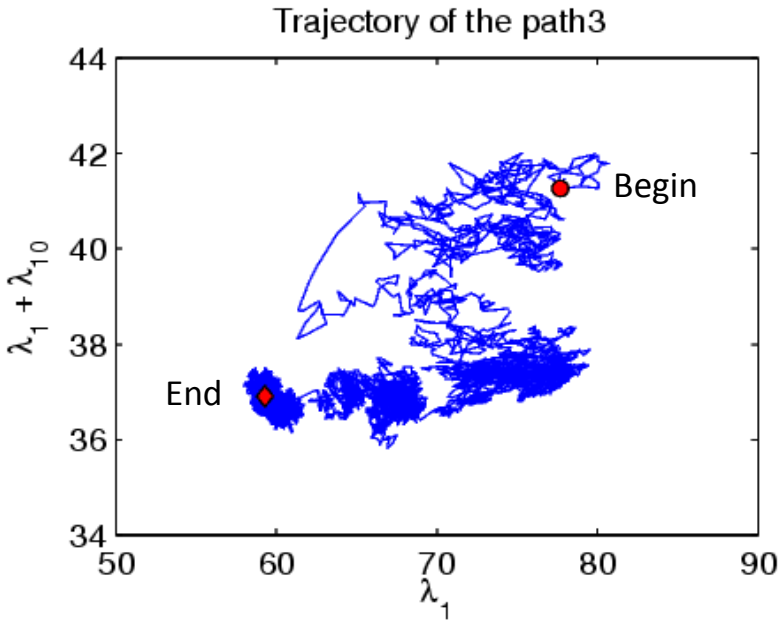

Phase space
Steps
10
50000
100000
250000
Ensemble of
100 paths
$\lambda_1+\lambda_{10}$
λ_1
48
46
44
42
40
38
36
34
32
50
60
70
80
90
100
110

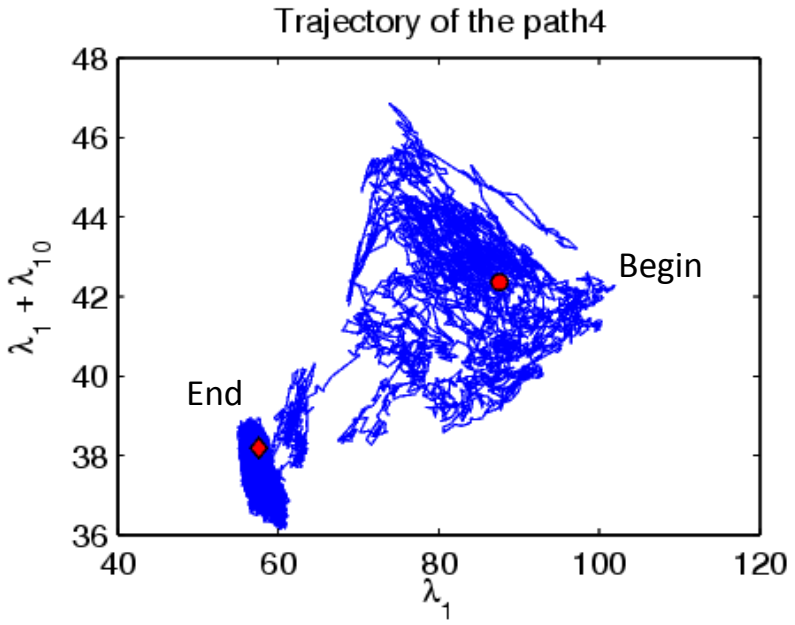

Contour map of potential energy
Phase space
46
44
42
40
38
36
34
32
50
55
60
65
70
75
80
85
90
95
100
λ_1
$\lambda_1+\lambda_{10}$
−50
−100
−150
−200
−250


Energy landscape
0
−50
−100
−150
−200
−250
−300
E
50
100
150
λ1
30
35
40
45
λ1+λ10

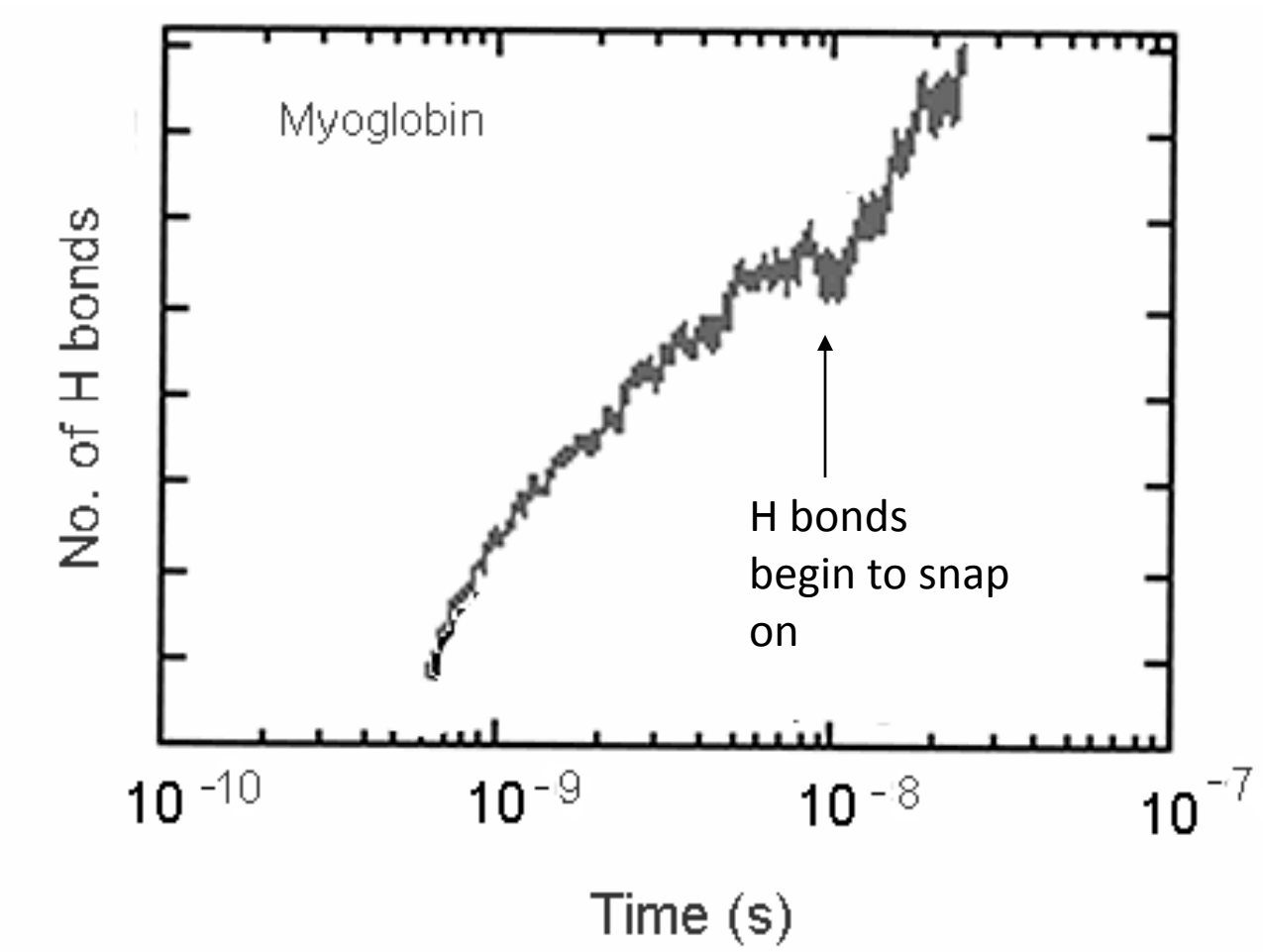

Trajectory of the path1
End
Begin
$\lambda_1 + \lambda_{10}$
λ_1

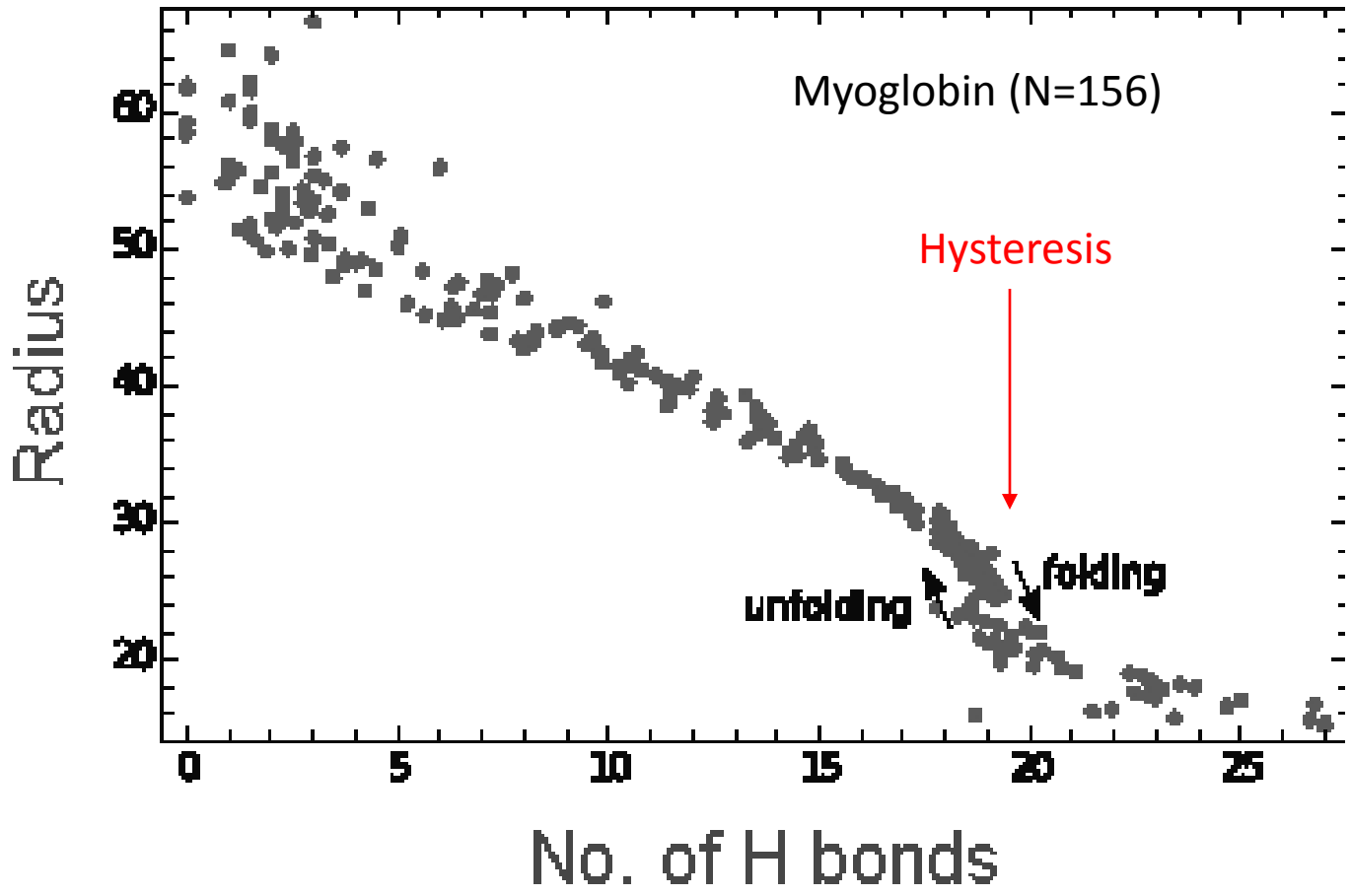

Trajectory of the path2
Begin
End
$\lambda_1 + \lambda_{10}$
λ_1


Trajectory of the path3
Begin
End
$\lambda_1 + \lambda_{10}$
λ_1


Trajectory of the path4
Begin
End
$\lambda_1 + \lambda_{10}$
λ_1

Simulation of Myoglobin (N=156)




Myoglobin (N=156)
Hysteresis
folding
unfolding
Radius
No. of H bonds
60
50
40
30
20
0
5
10
15
20
25

Stages of protein folding

Lei, J, & Huang, K. 2009. Elastic energy of proteins and stages of protein folding. *Europhys. lett.*, 88, 68004.

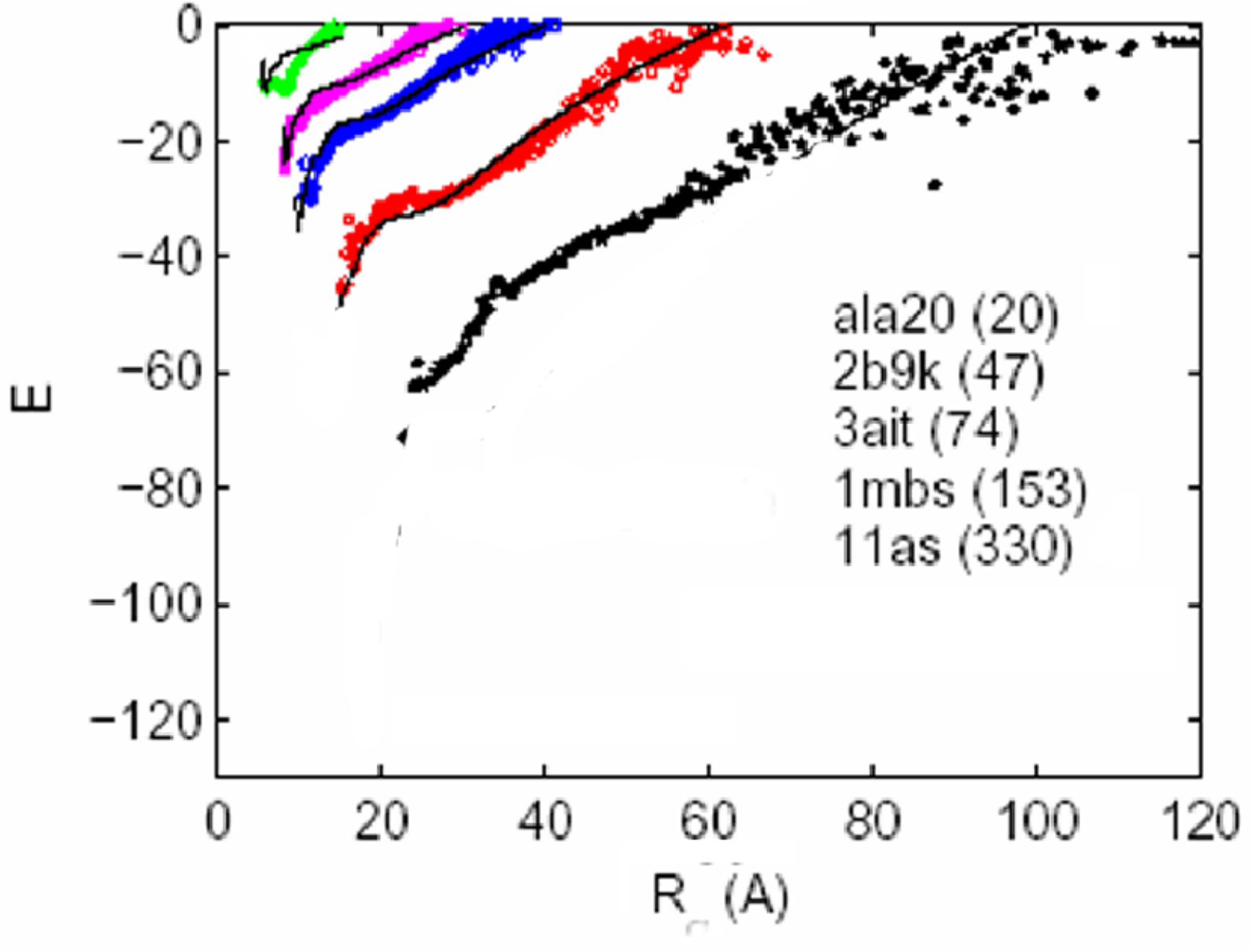
Folding of 5 proteins by CSAW

Protein	N
Ala20	20
2b9K	47
3ait	74
Myoglobin	153
11as	330

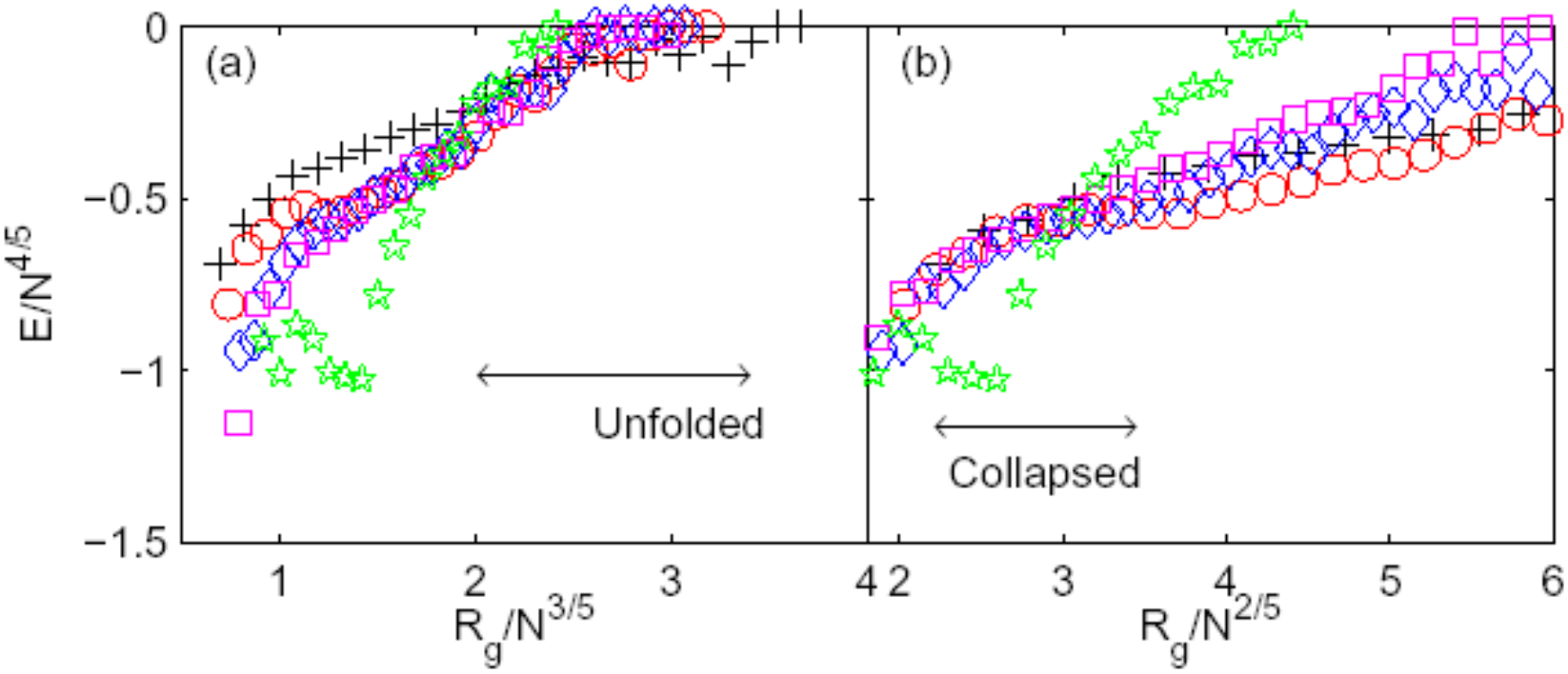
Elastic energy:

E(R,N) = Ensemble average of potential energy, for given value of radius R.

Elastic force = -dE/dR.


0
−20
−40
−60
−80
−100
−120
E
0
20
40
60
80
100
120
R (A)
ala20 (20)
2b9k (47)
3ait (74)
1mbs (153)
11as (330)

Replotting with different scalings

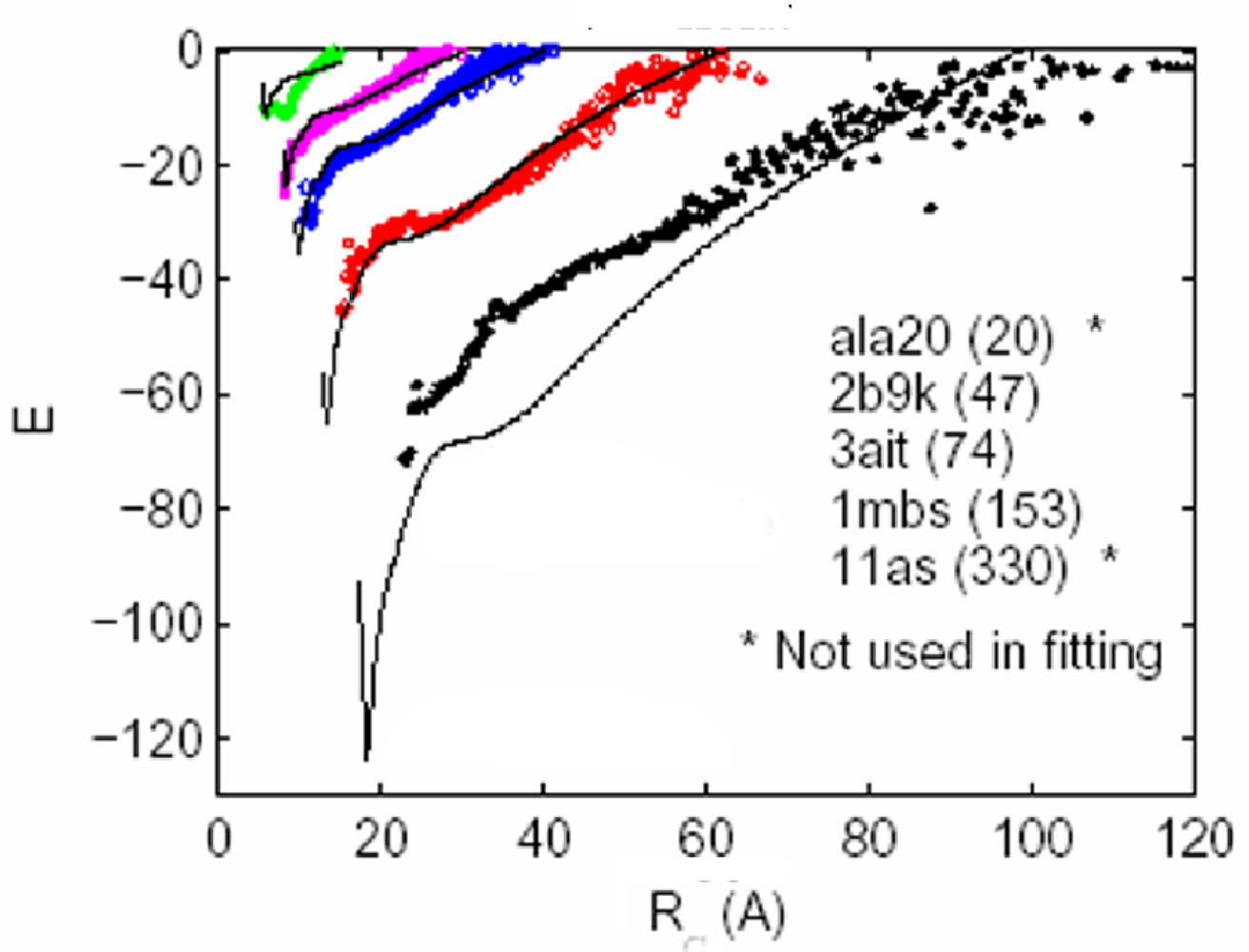


- Unfolded states scale with universal exponent 3/5
- Collapsed states scale with universal exponent 2/5
- Exception: polyalanine (N=20) (green circles)

Proposed universal form

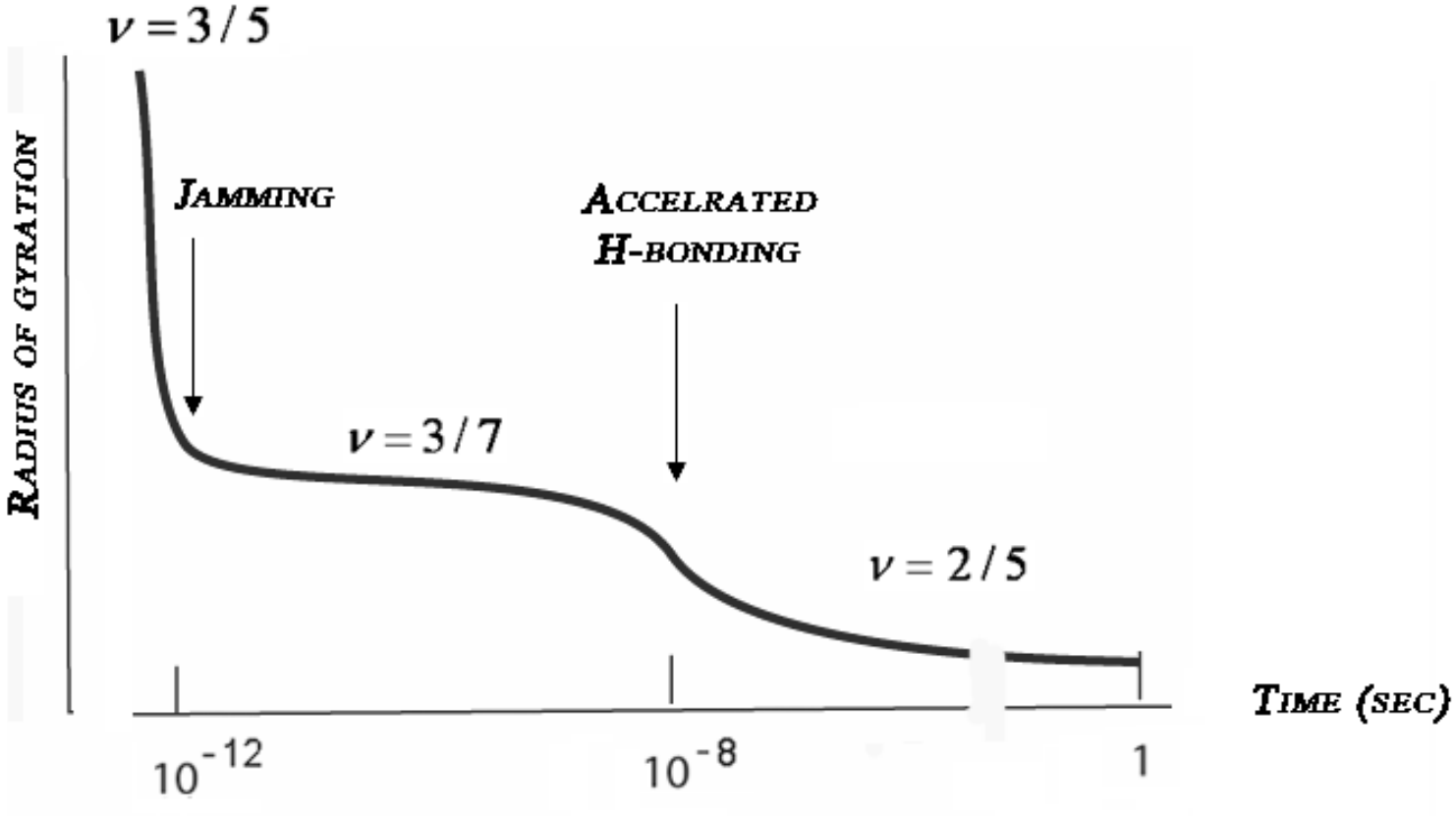
$$E(R_g, N) = aN^{4/5} + b(N R_g)^{1/2} + c(\rho)N^2/R^3$$

$$\rho = \frac{R}{N^{2/5}}$$



Proposed sequence of stages

Proposed stages of protein folding



Unfolded	Pre-globule	Molten globule	Native state
Hydrophobic collapse stopped by jamming	Viscoelastic rearrangement	H bonds provide rigidity.	Side chains locked on

Lifetime of stages depends on protein

References

K. Huang, "CSAW: A dynamical model of protein folding", arXiv:cond-mat/0601244 (2006).

K. Huang, "Conditioned self-avoiding walk (CSAW): stochastic approach to protein folding," *Biophysical Reviews and Letters* (Singapore), **2**, 139-154 (2007).

J. Lei and K. Huang, "Dynamics of alpha helix formation in the CSAW model", *Eur. Phys. J*. E 27, 197 (2008).

J. Lei and K. Huang, "Elastic energy of proteins and the stages of protein folding", *Eur. Phys. Lett*. **88**, 68004 (2009).

H.W. Leong, L.Y. Chew, and K. Huang, "Normal modes and phase transition of the protein chain based on the Hamiltonian formalism", *Phys. Rev*. E **82**, 011915 (2010).

J. Lei and K. Huang, "Protein Folding: A Perspective from Statistical Physics", in *Protein Biochemistry, Synthesis, Structure and Cellular Functions*, E. C. Walters, ed (Nova Publishers Hauppauge, New York, 2010), Chap. 21, pp. 579-604.

K. Huang, "Protein folding stages and universal exponents", *Mod. Phys. Lett*. **B 24**, 1 (2010).