Some Problems in Protein Folding
A prejudiced View

Dr. Ed O'Brien (Cambridge, UK)
Prof. Changbong Hyeon(KIAS)
Prof. Riina Tehver (UMD)
Dr. Govardhan Reddy (UMD)
Dr. Jie Chen (UMD)
Dr. Zhenxing Liu (UMD)

NSF

Is the Protein Folding Problem Solved?
"Many have solved the problem as they have defined it" --- Eaton

Many Facets of Folding

1. Structure Prediction
2. Protein & Enzyme Design
3. Folding Kinetics & Mechanisms
4. Crowding Effects
5. Relation to aggregation
6. Molecular Chaperones

Folding at the center stage

Folding Kinetics

Experiments

- Prot Engg (TSE)
- SAXS/NMR (DSE)
- FAST Folding (T jump; Rapid Mixing)
- SM FRET (Folding/ unfolding)
- LOT/AFM (Force Ramp Force Quench)

Theory

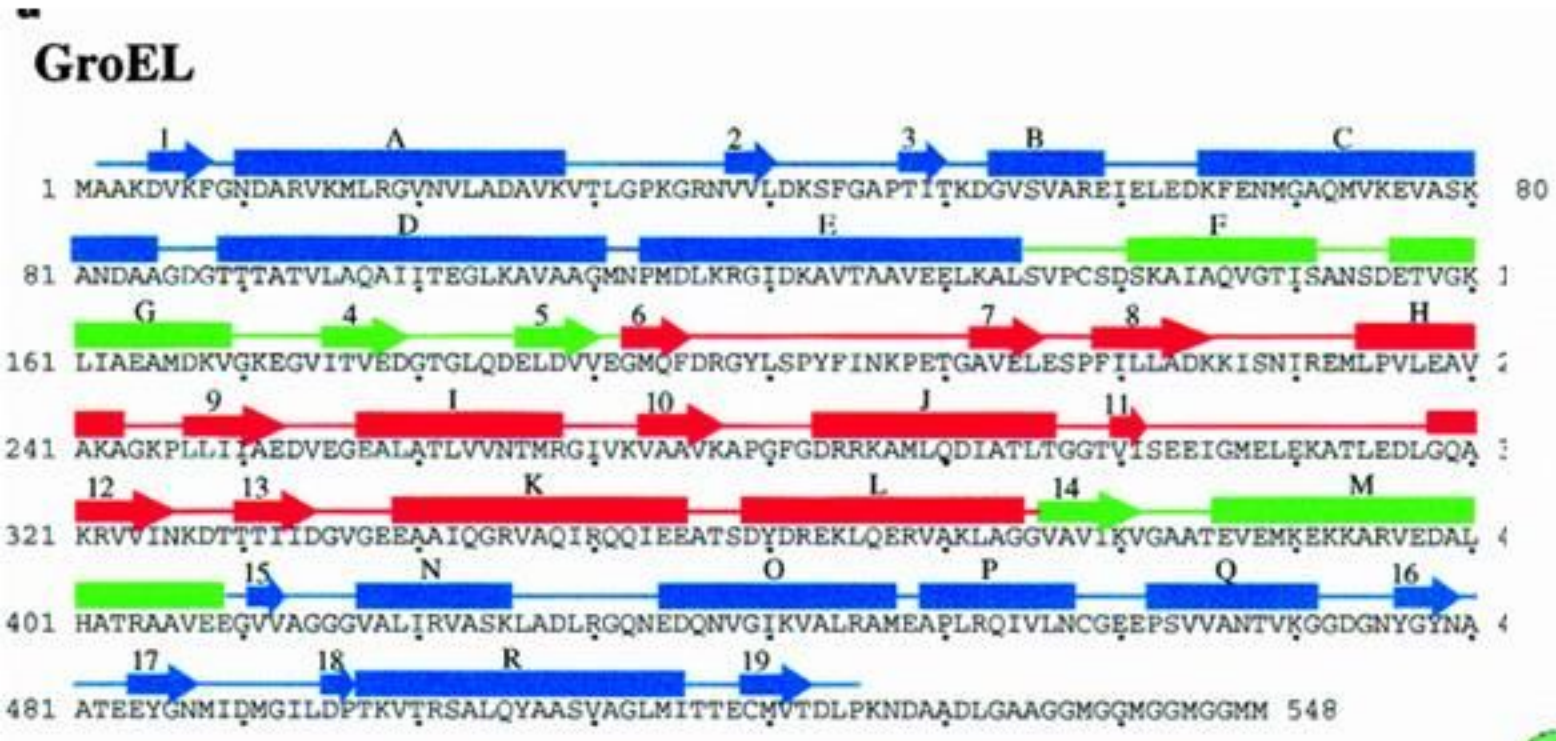
- Statistical Mechanics (Energy Landscape)
- Minimal Models (Lattice/Off-Lattice)
- MD Simulations
- Bioinformatics (Evolutionary Imprint)

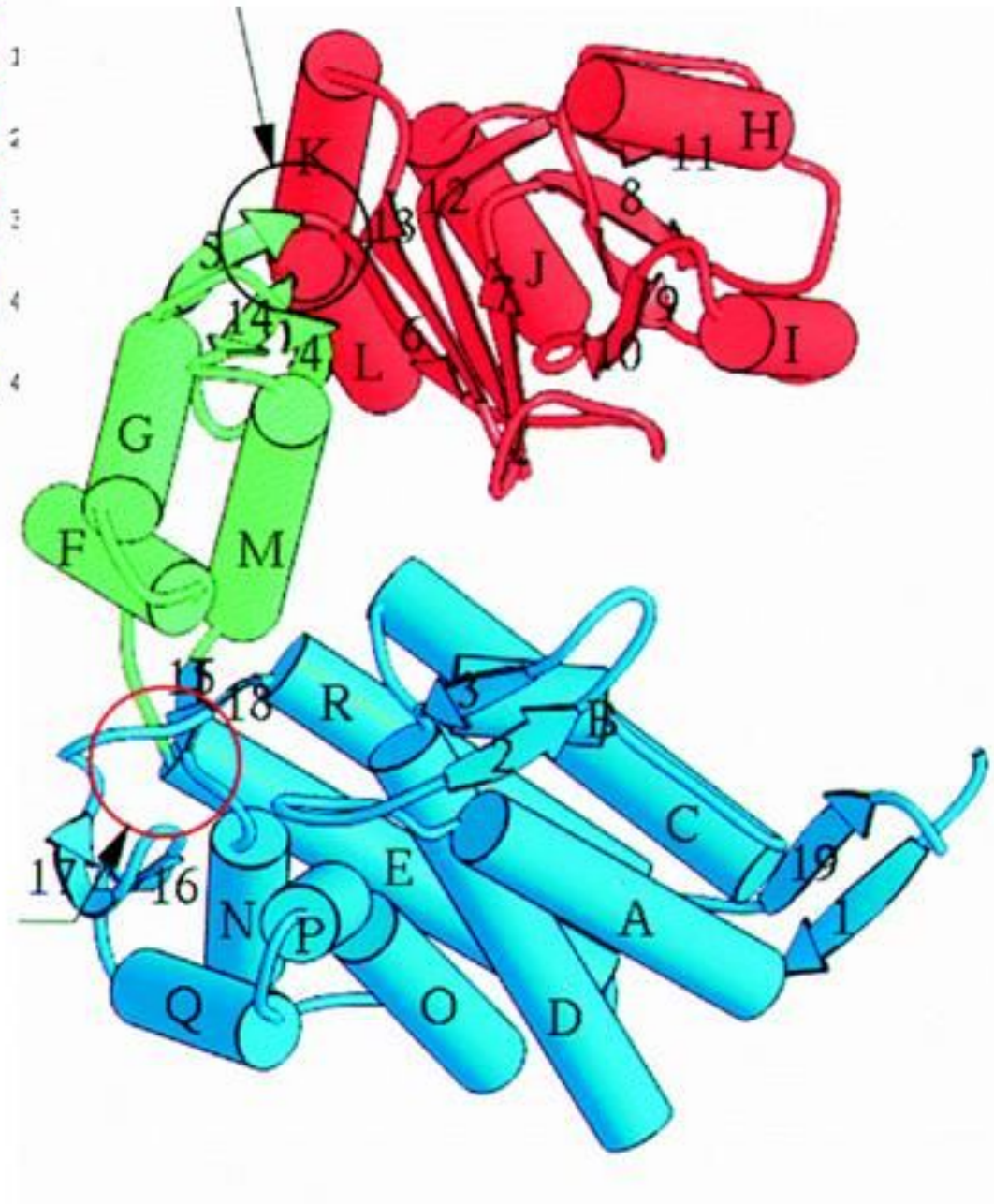
Some concepts from physics and physical chemistry

Multiple folding scenarios, Folding funnels, nucleation-collapse, speed limit, designability and fold
Misfolding and kinetic traps, foldons, roughness, frustration topological and energetic, contact orde
TSE and □-values, pre-factors and Kramers theory, entropy barriers, glass transition, collapse, Flor
Laws, Master Equstion, Finite size effects and consequences

Collapse transition, Folding in terms of changes in distribution functions
SMFS as a tool for folding

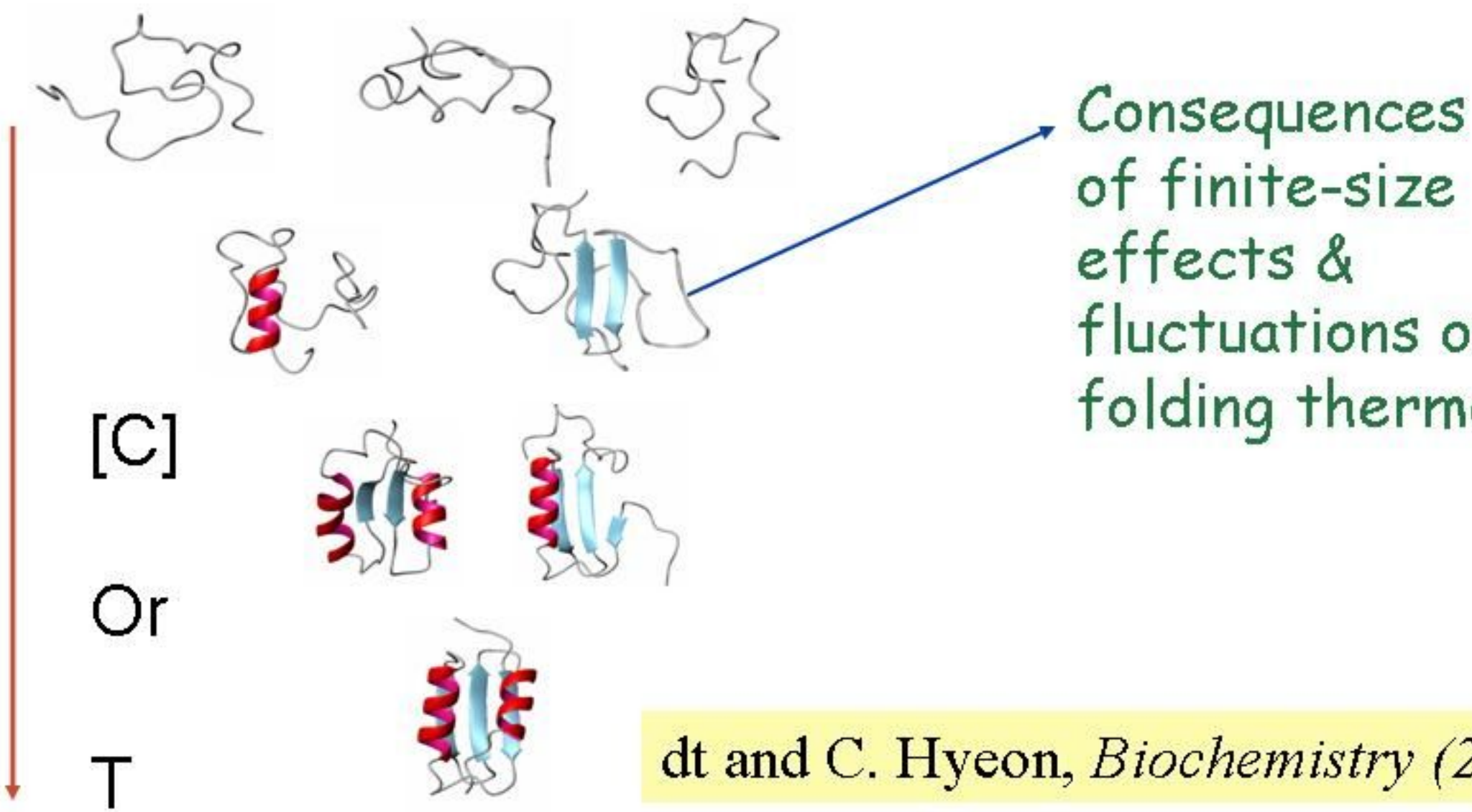
Protein folding problems





- Structure Prediction
- Assembly Mechanism
- Protein Design
- Crowding Effects

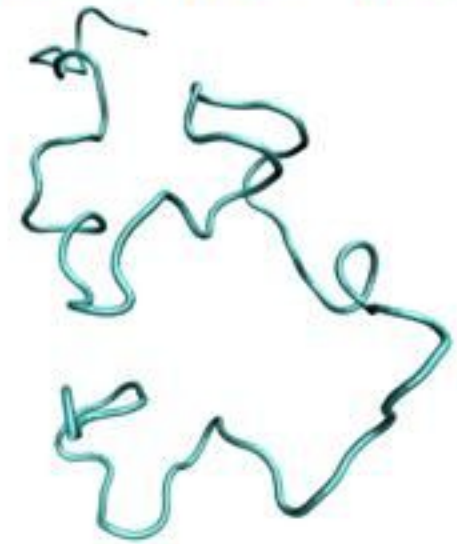
Monomeric (spontaneous) folding Mechanisms



dt and C. Hyeon, *Biochemistry (2005)*

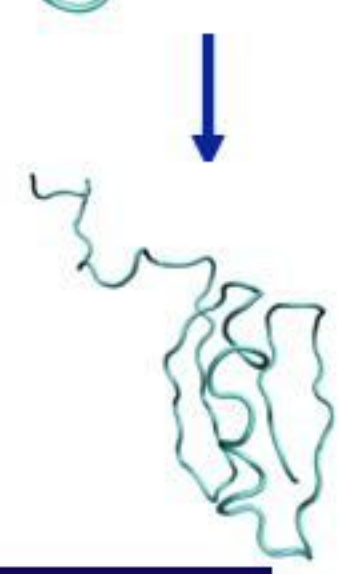
Characteristic Temperatures in Proteins

HIGH T
or [C]



Random
Coil (Flory)

$T \approx T_\theta$
Or $[C_\theta]$



Compact

$T \approx T_F$
$[C_F]$

$\sigma = (T_\theta - T_F)/T_\theta$

**Native
State**

Size (R_g) of Proteins: Flory Laws

Unfolded state

$R_g \approx a_D N^{\nu}$ "good solvent" $\nu \approx 3/(d+2)$

Folded and Collapsed Globule

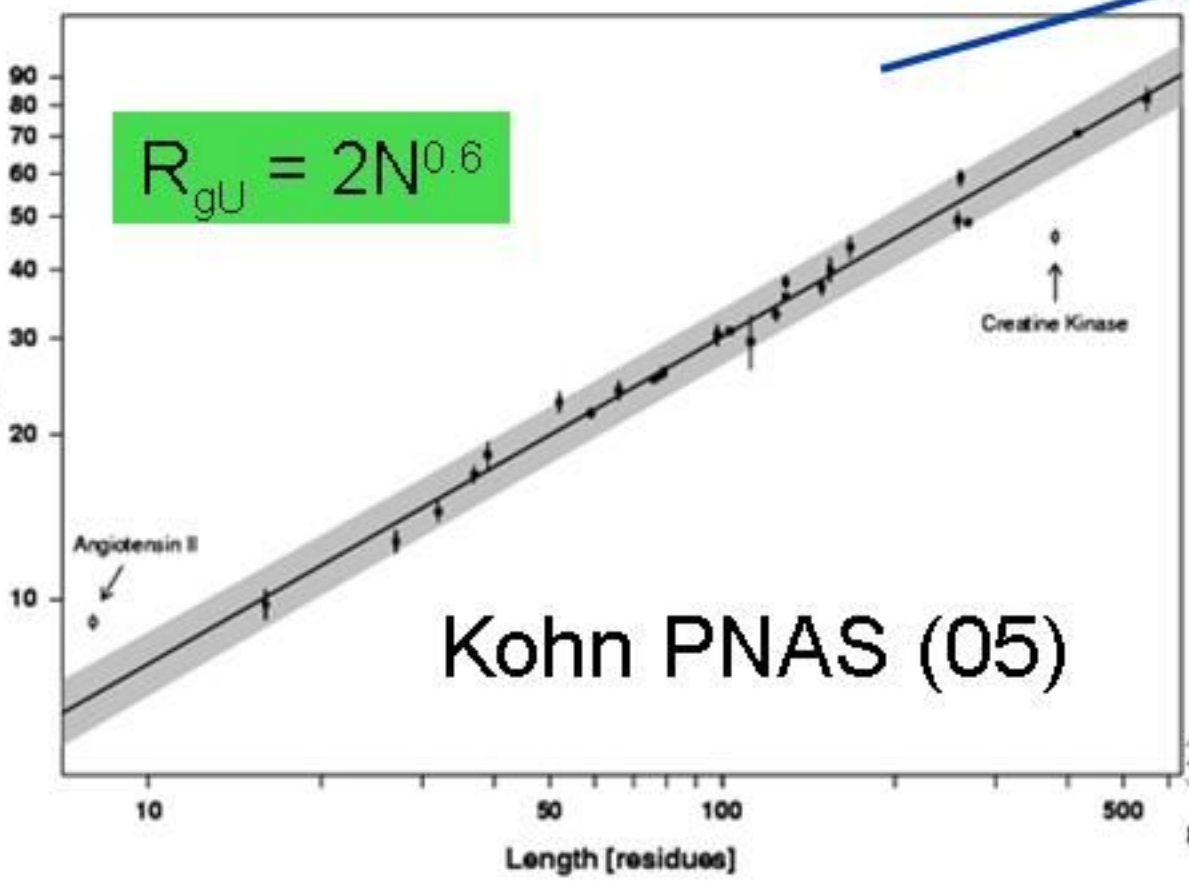
$R_g \approx a_N N^{\nu}$ "poor solvent" $\nu = 1/d$

How good are Flory Laws for proteins?

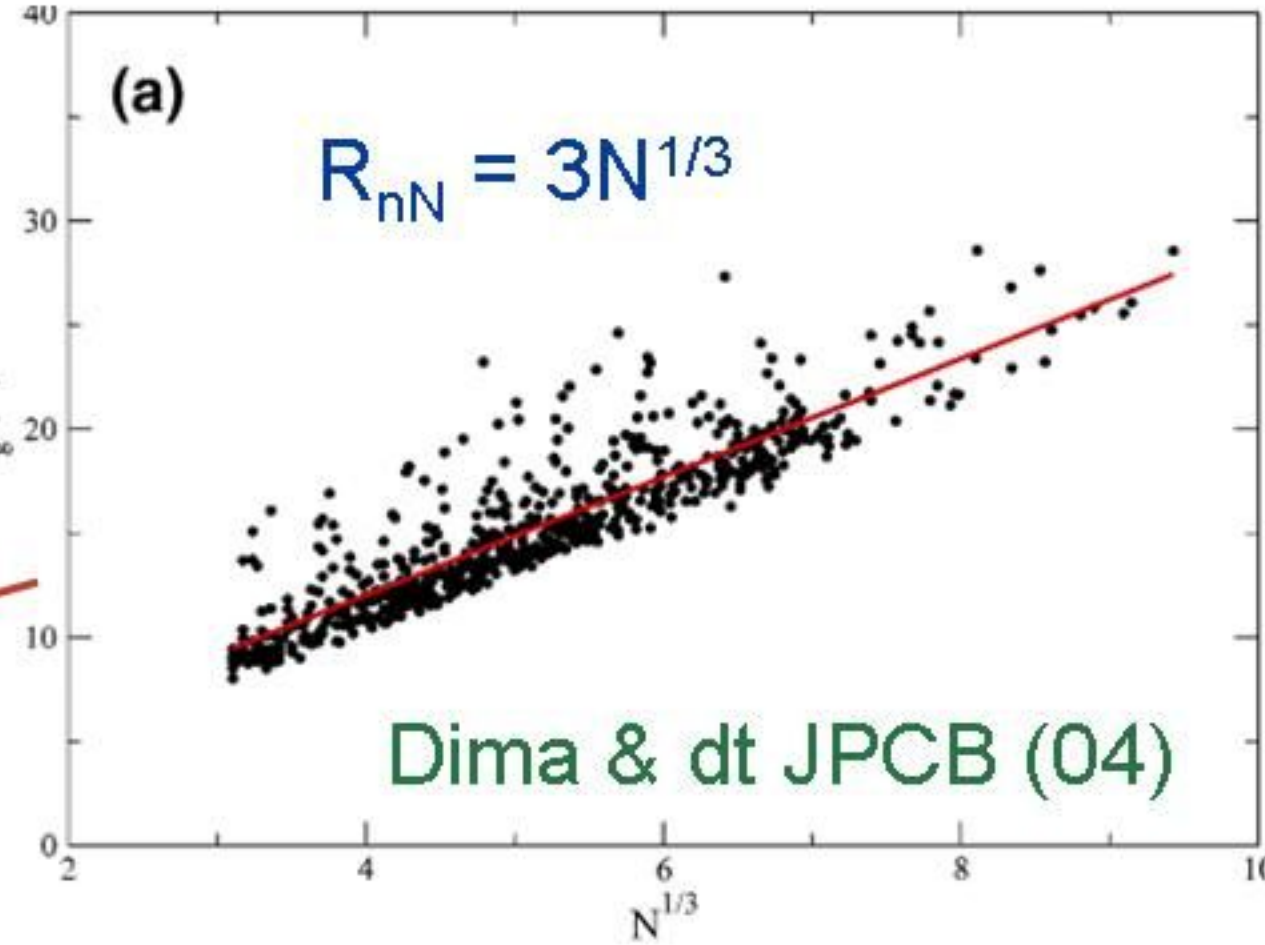
Protein Collapse : R_g follows Flory law

"Unfolded"

$R_{gU} = 2N^{0.6}$

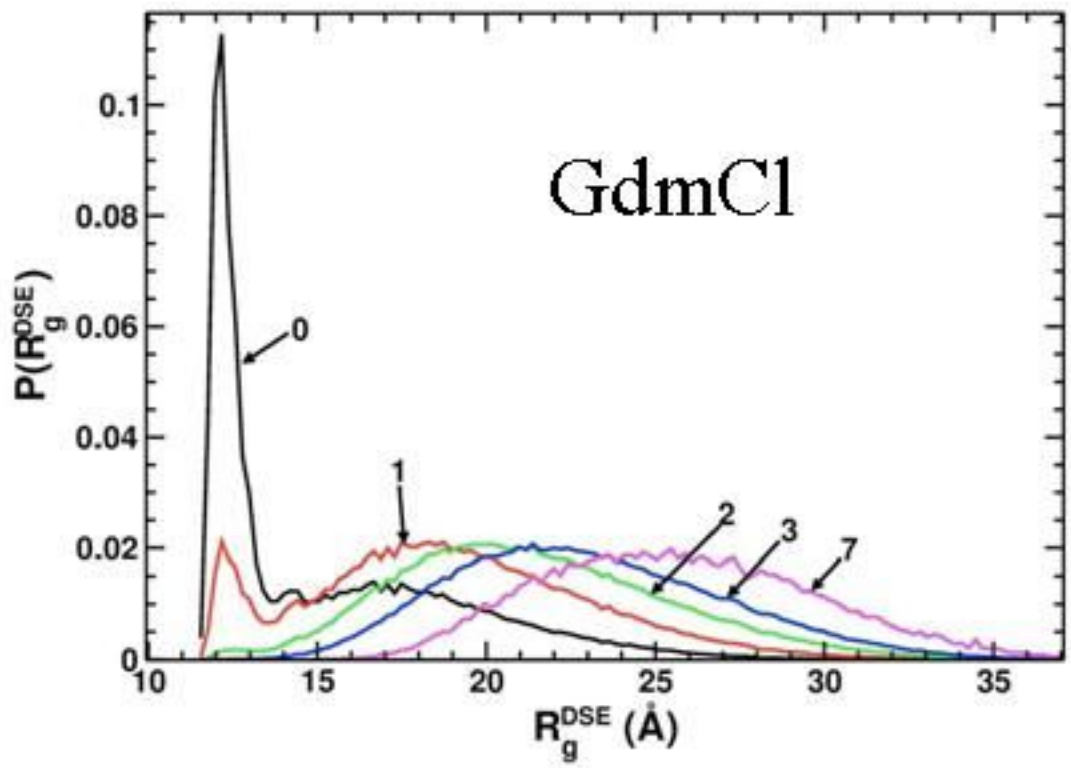


Folded



Folding as Changes in Distribution Function

O'Brien PNAS (2008)



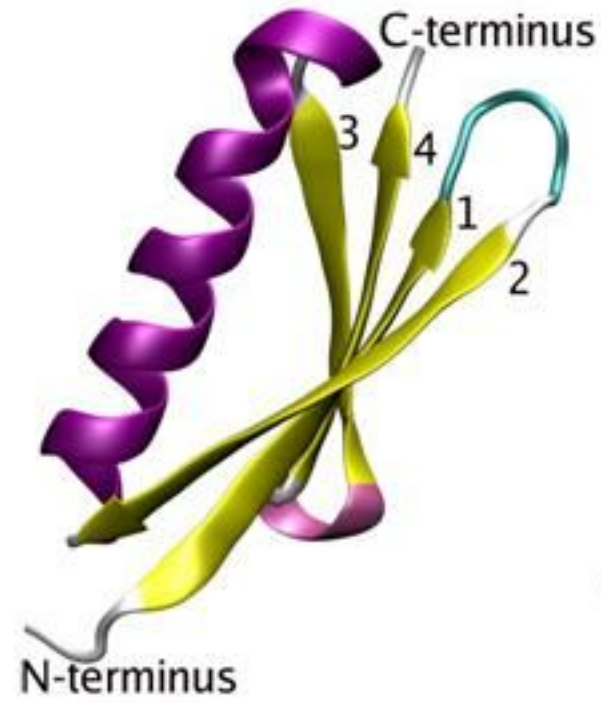


Fig. S2. Distribution of $P(R_g^{DSE})$ for protein L at various GdmCl concentrations at 328 K. The concentration ([C]) of GdmCl in molar units are shown in the curves.

Folding reaction as a phase transition: A rationale N = number of amino acids

Order Parameter Description

$\rho = N/R_g^3$; χ = Overlap with NBA (0 for NBA)

Unfolded (U), Collapsed Globules (CG); Folded (NBA)

U: ρ (small), χ Large ("vapor")

CG: ρ . O(1), χ Large (Dense no order "Liquid")

NBA: ρ . O(1), χ Small (Dense order "Solid")

Developing a "nucleation" picture

Free Energy of Creating a Droplet

$\Delta G(R) \sim -\Delta\mu R^3 + \gamma R^2$

Driving force + Opposing

What are these forces in proteins?

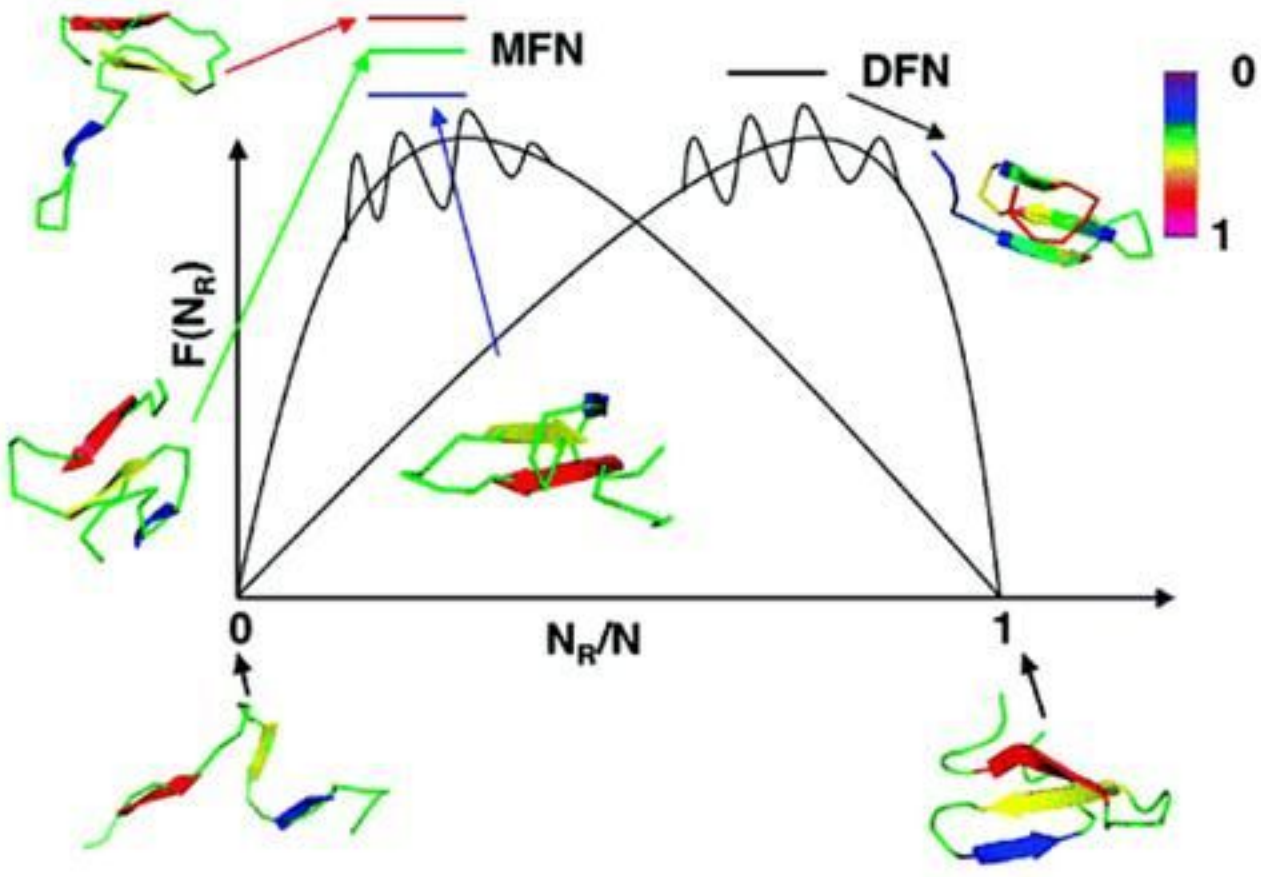
Driving force: Hydrophobic Collapse

Burying H bonds

Opposing: "Droplet with nonconstant γ"

Entropy loss due to looping

Tentative Models + Slight refinement



Cost of creating a region with N_R ordered residues out of N?

Rugged Landscape with Many possibilities

Some phenomenological Models

$$\Delta G_{BW}(N_R) \approx -\Delta f(T) N_R + \gamma a^2 N_R^{2/3}$$

$$N_R^* \approx (8\pi\gamma a^2/3\ \Delta f(T))^3$$

N_R^* too large for typical γ and $\Delta f(T)$ values

$$\Delta G_{GT}(N_R) \approx \varepsilon_h(\varepsilon_h - 1) N_R^2 + \gamma a^2 N_R^{2/3}$$

$$N_R^* \approx (8\pi\gamma a^2/\varepsilon_h)^{3/4}$$

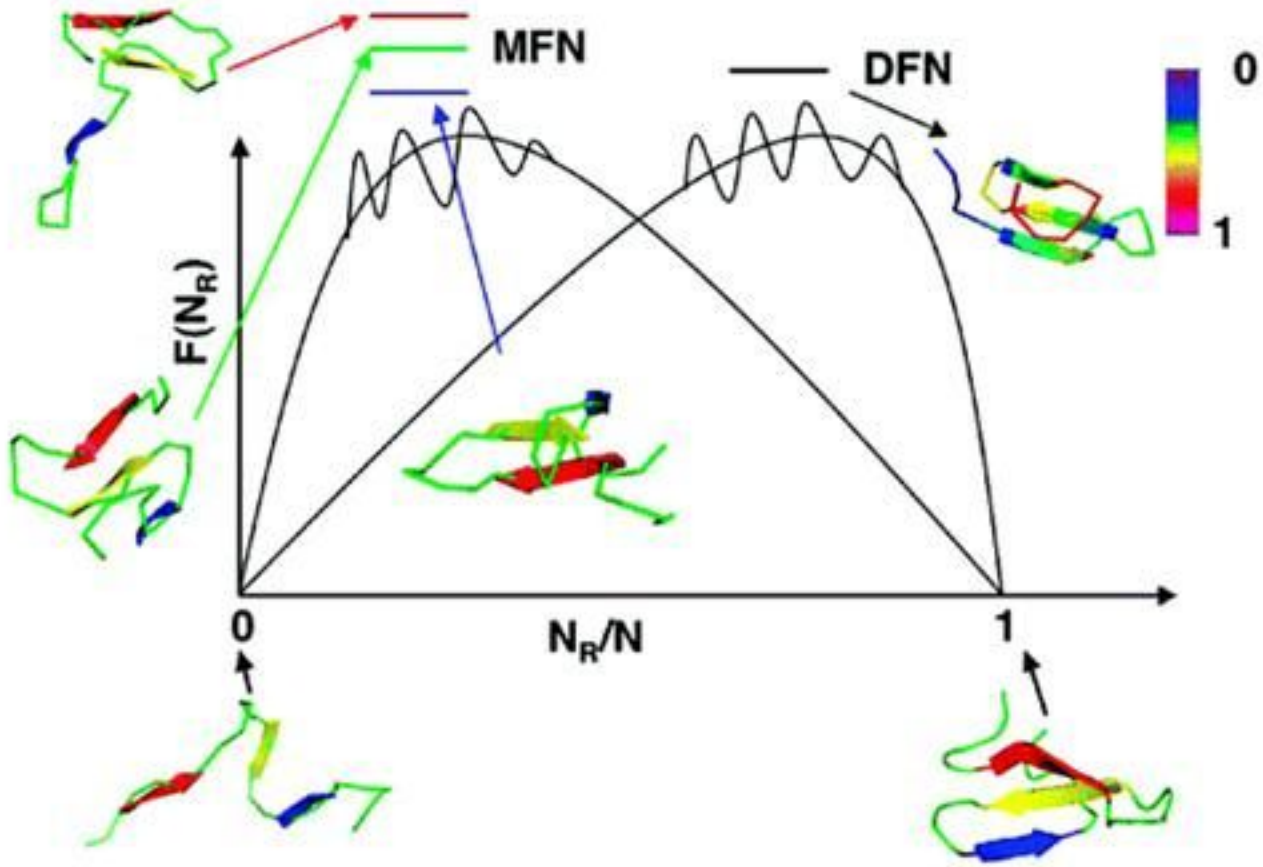
$N_R^* \approx 15$ or so…

Using experimental parameters

$N_R^* \approx 27$ or so..

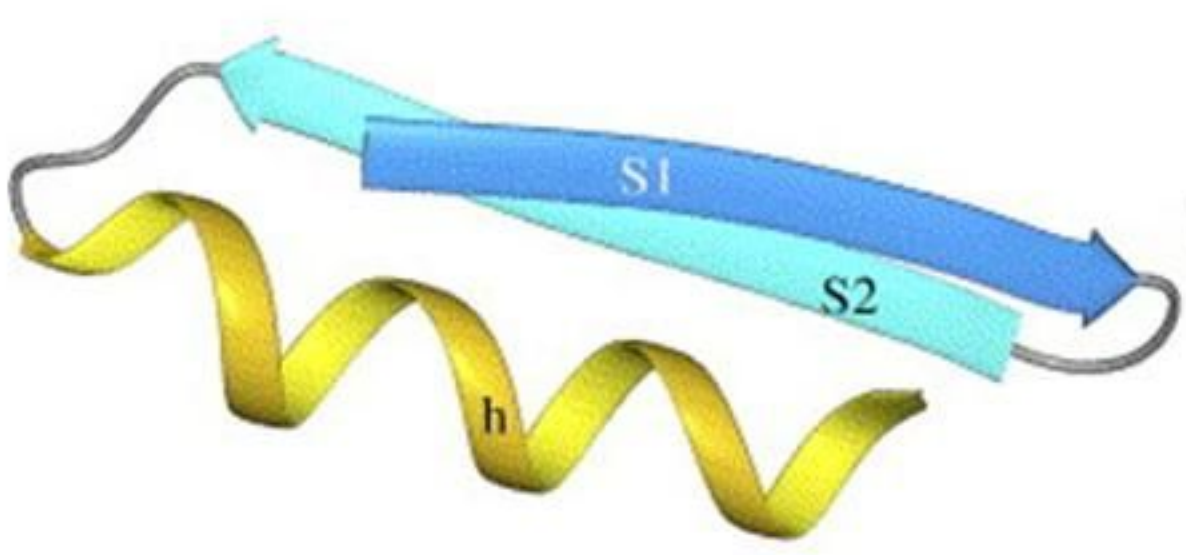
Folding trajectories to MFN to transition state ensemble (TSE)

Structures near Barrier top or TSE Simulations



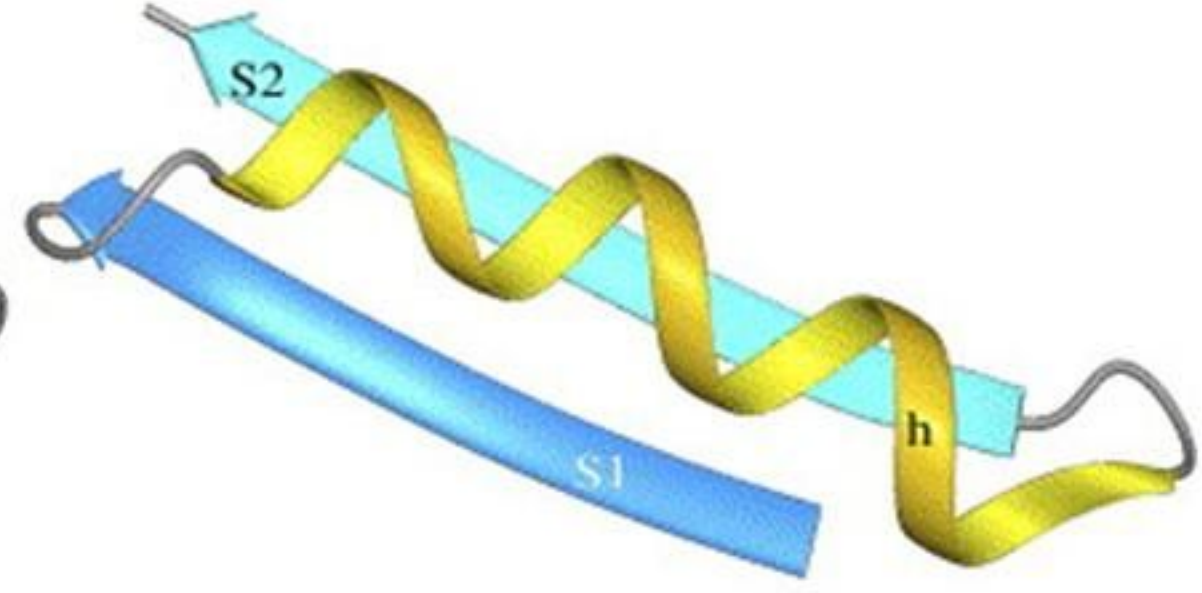
Folding trajectories Simple Topology Brownian Dynamics Simulations

A1



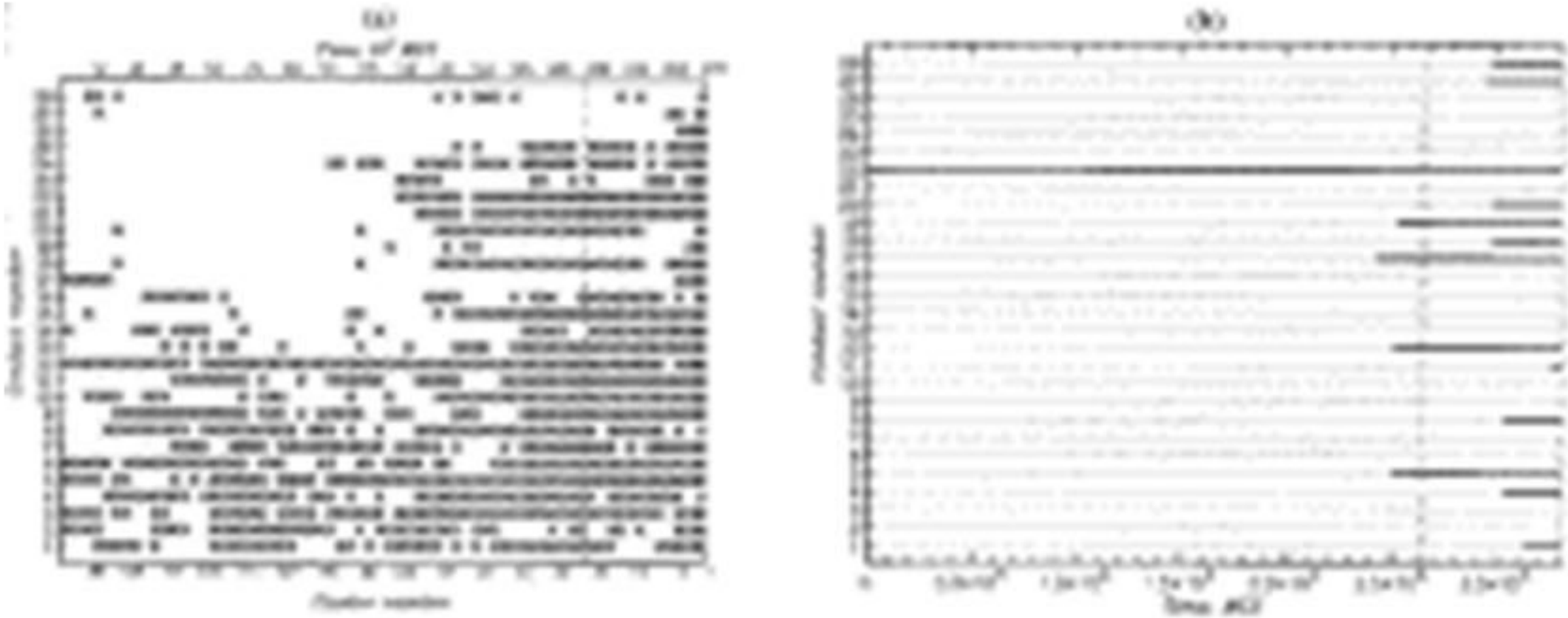
BLBLBLBLBNNNLBLBLBNBLNNNLBBLLBBLLBLLBB

A2

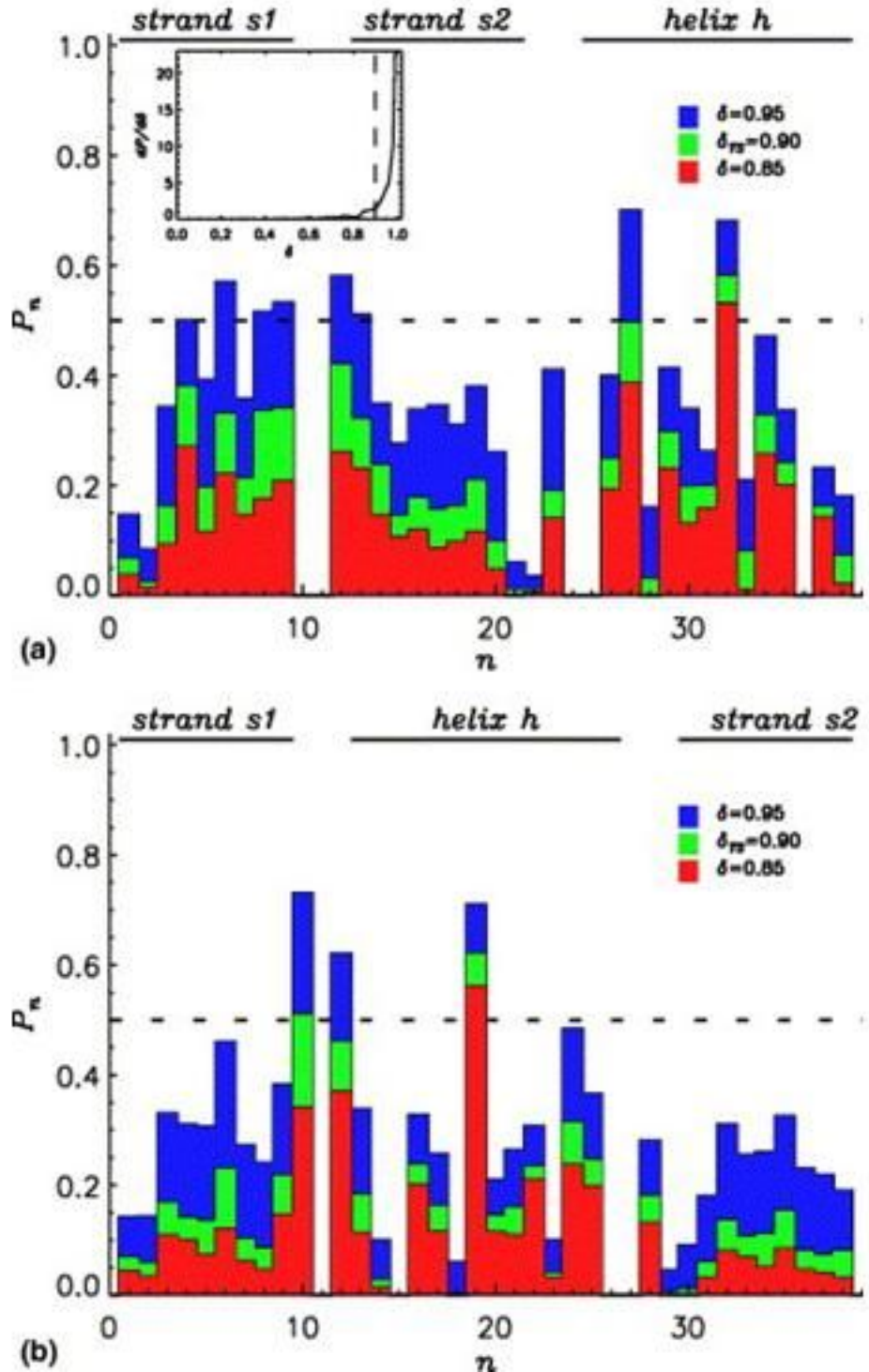


BLBLBLBNBNLNBBLLBLLBBLLBBNLNNLBNBLBLBL

Heterogeneity in Folding Trajectories



Multiple Folding Nuclei (MFN) Model



No unique nucleus

Different molecule Follows distinct Growth pattern

Structural plasticity

Refinement (Hiding Ignorance)

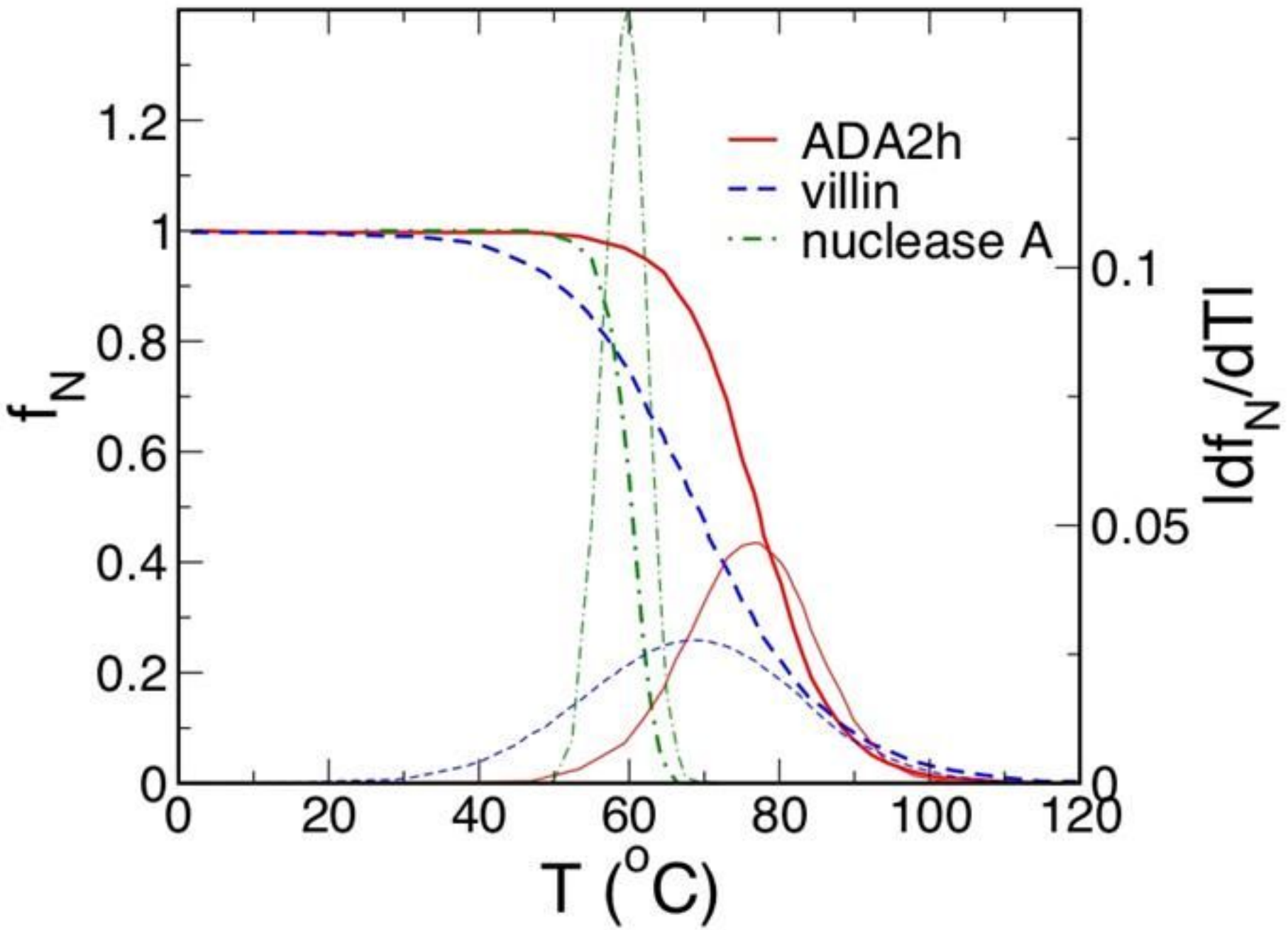
$$\Delta G(N_R) \approx -\alpha_1 N_R^{\delta} + \gamma N_R^{\theta} + S\ (\text{loop})$$

$\delta \approx \theta$ small barrier (downhill folding)

Surface tension cannot be a constant

Multiple Folding Nuclei (Structural Plasticity)

Finite Size Effects on Folding
Order parameters matter



Scaling of Ω_c with N (number of aa)

Two points:

1) T_F = max in $\Delta\chi$ (suceptibility)

$\Delta\chi = T(d\langle\chi\rangle/dh$; h = ordering field (analogy to mag system)
$\Delta\chi$ is dimensionless $\Rightarrow$ h ~ T (in proteins or [C])

2) Efficient folding $T_F \approx T_\theta$ (collapse Temp; Camacho & dt PNAS (1993)) $\Rightarrow \Omega_C$ controlled by protein DSE at $T \approx T_F \approx T_\theta$

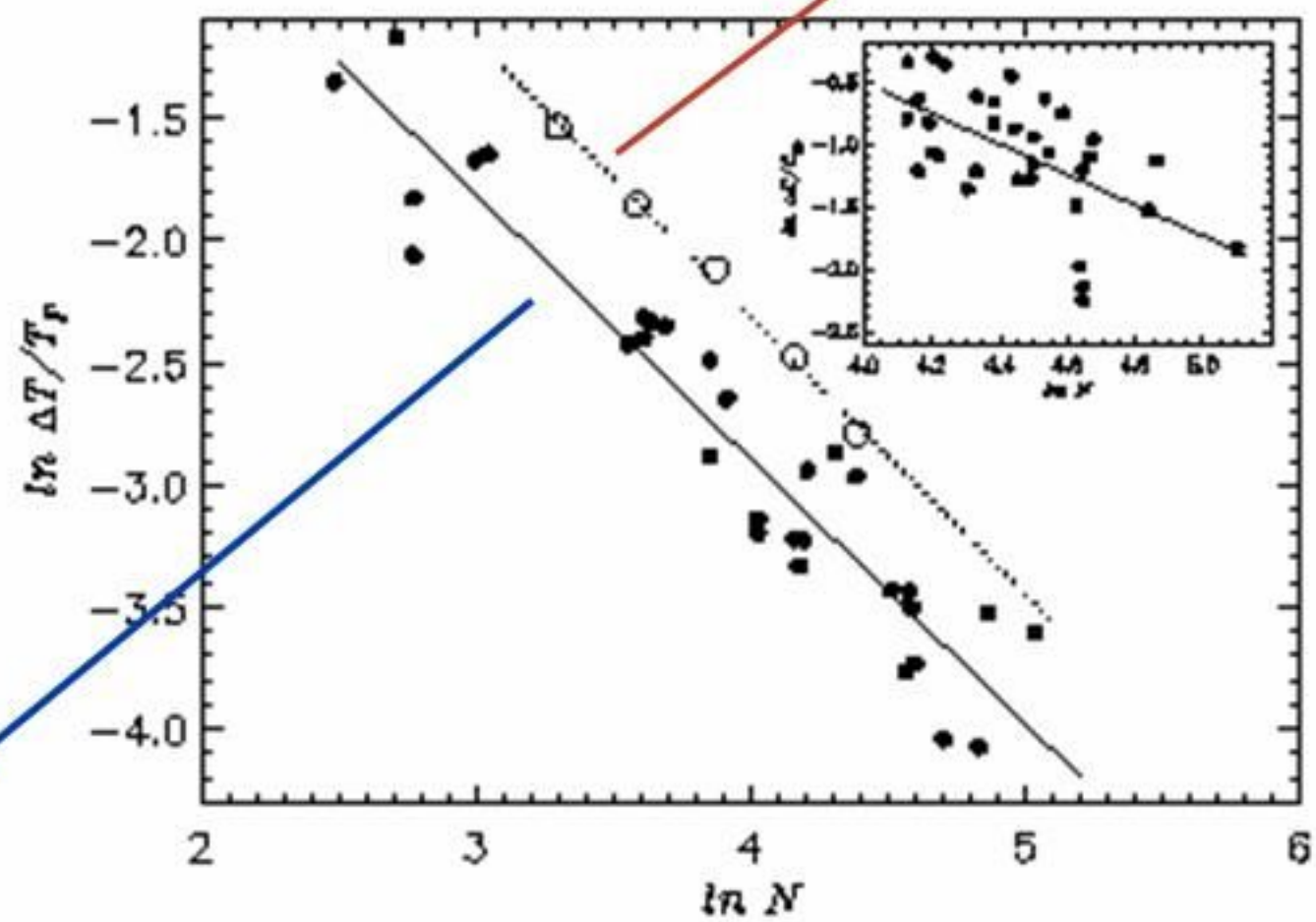
$R_g \sim (\Delta T/T_F)^{-\nu} \sim N^{\nu}$ (DSE a SAW & manget analogy)

$\Delta T/T_F \sim 1/N$ (Result I)

Finite-size effects on T_F

Lattice models
Side Chains

$\Delta T/T_F \sim 1/N$



Experiments

Li, Klimov & DT Phys. Rev. Lett. (04)

Scaling of Ω_c with N
Magnet-Polymer analogy

$$\Omega_c = (T_F/\Delta T)\,[T_F(d\langle\chi\rangle/dT)]$$

"disp in T_F" X "suspectibility"

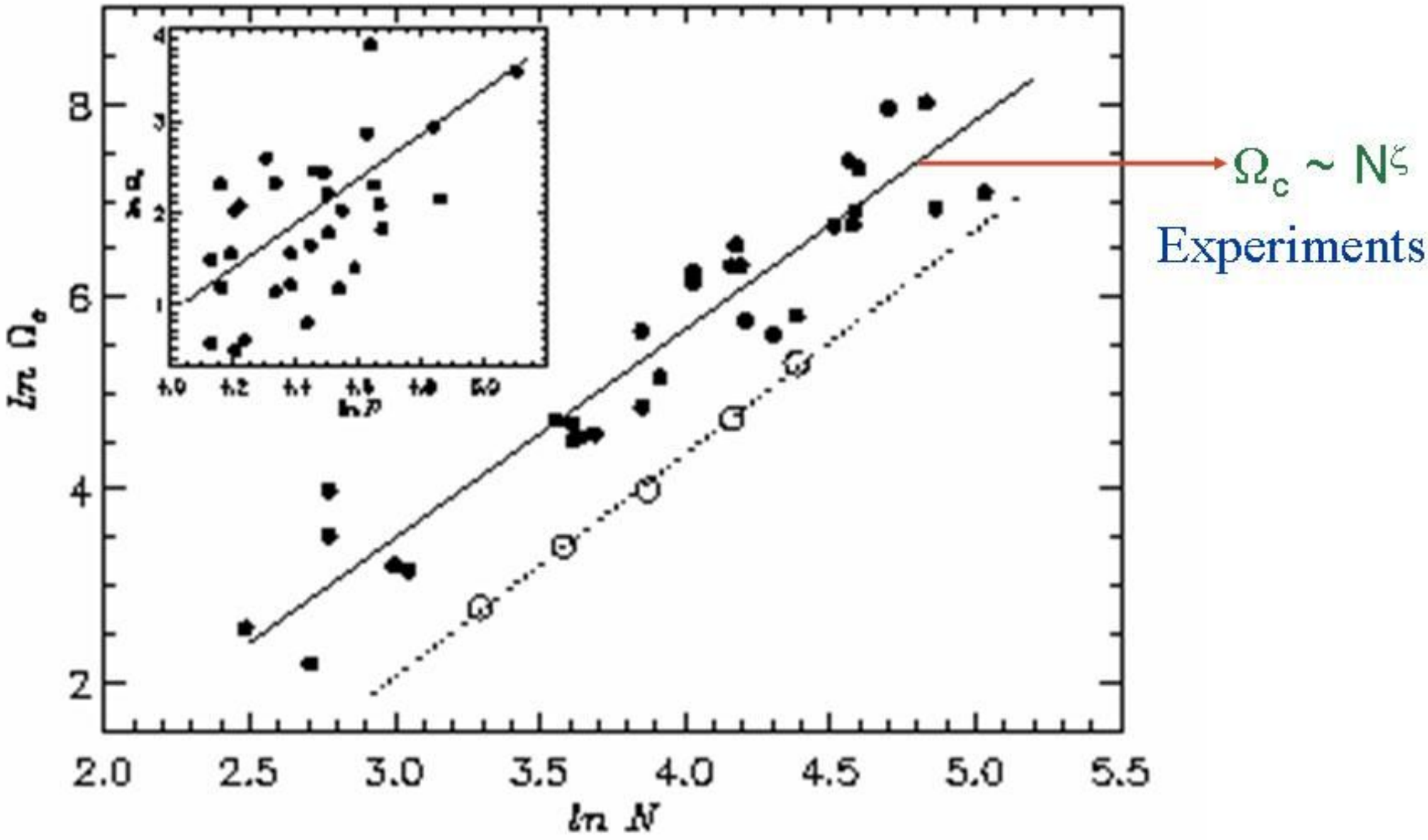
$T \approx T_F \approx T_\theta$

$\Delta\chi \approx N^\gamma$

$\Omega_c \approx N^\zeta$; $\zeta = 1 + \gamma$ (Universal); $\gamma \approx 1.2$

Result II

Universality in Cooperativity

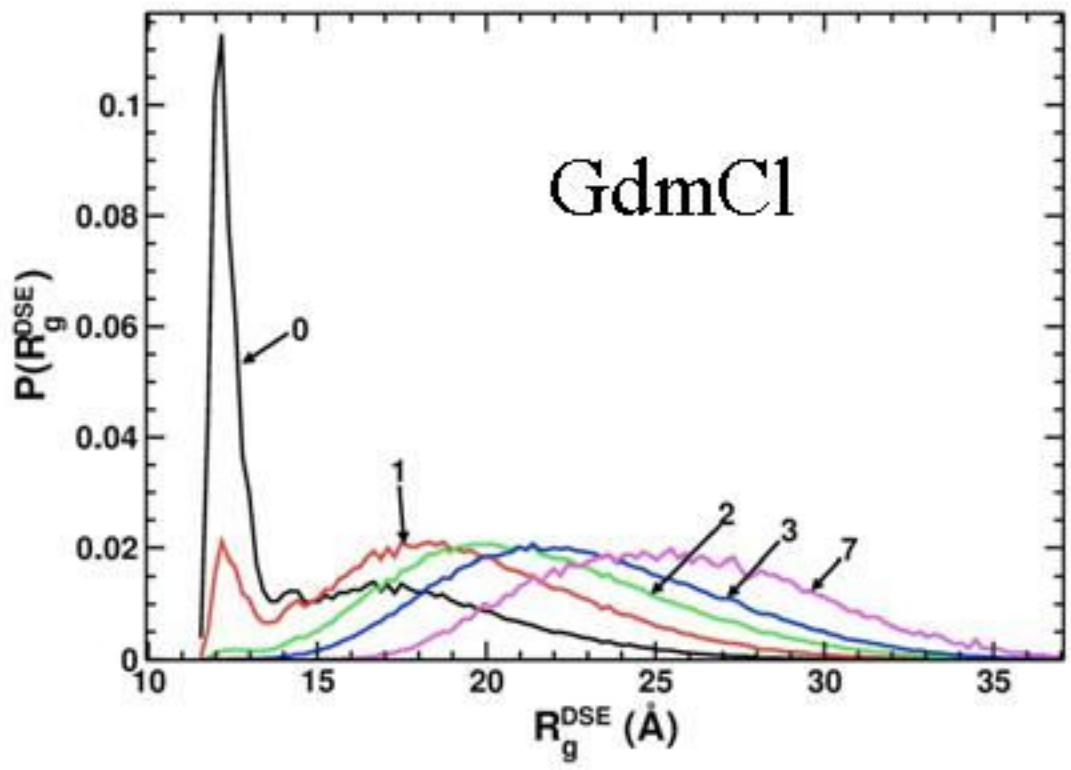


Folding as Changes in Distribution Function (of what?)

O'Brien PNAS (2008)

Finite-size Fluctuations give rise to dispersions in T_m

Different probes (order parameter) can give different information…



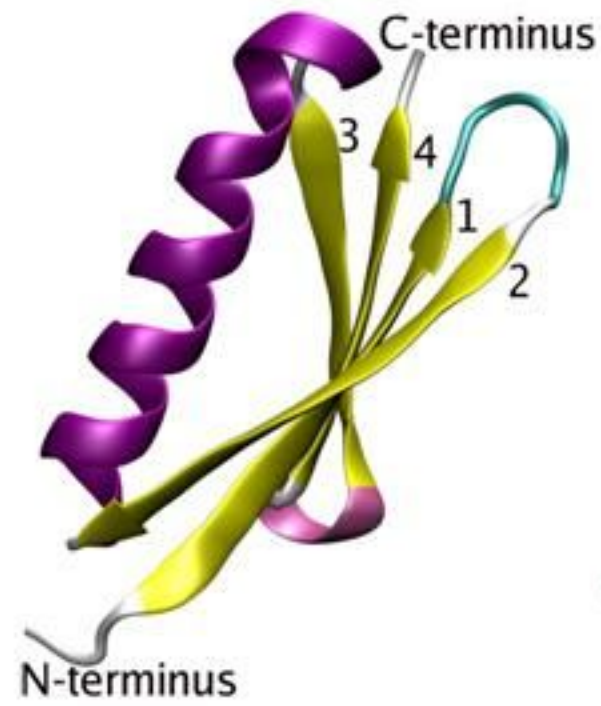


Fig. S2. Distribution of $P(R_g^{DSE})$ for protein L at various GdmCl concentrations at 328 K. The concentration ([C]) of GdmCl in molar units are shown in the curves.

Order parameter description

Macroscopic System

Ferromagnetism M
Nematic Phases $S = P_2(\cos\theta)$
Smectic Phases S,tilt angle

Spin Glasses: M; q_{EA}
Paramagnet $M = 0$. $q_{EA} = 0$
Spin Glass $M = 0$; $q_{EA} \neq 0$
Ferromanet $M \neq 0$; $q_{EA} \neq 0$

Physics dictates OP

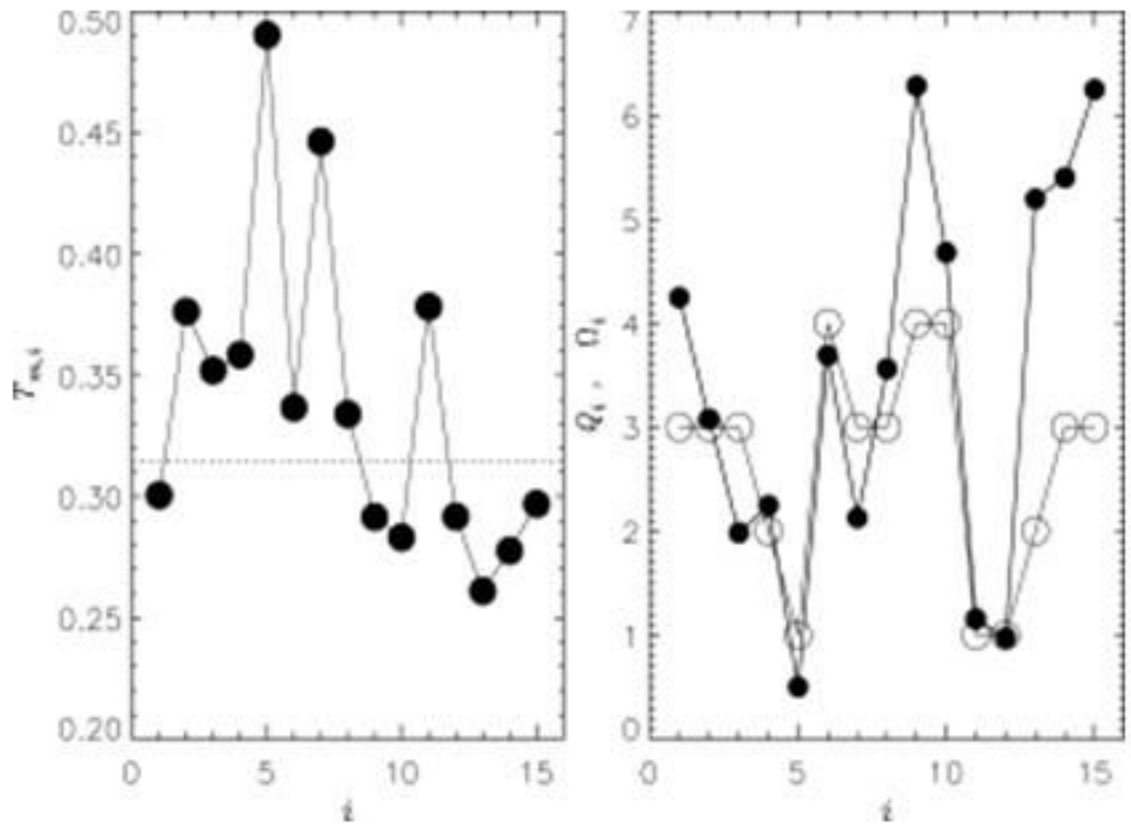
Proteins a lot of choices
OP is in the eye of the beholder
$\rho = N/R_g^3$; χ (overlap)
"unfolded" (Small,big)

Compact non-native
(O(1), big)

Native
(O(1), small)

Other Choices
Helix/sheet content;
Distribution of contacts
.....

Residue-dependent melting T_m-Holtzer Effect

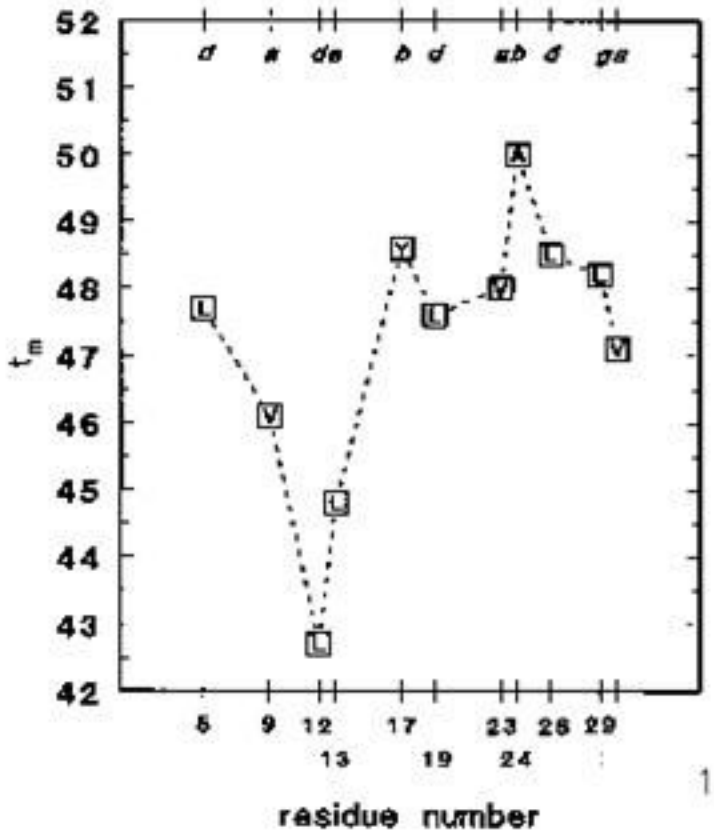


$$f_m(T_{mi}) = 0.5$$

Lattice Models
Side Chains

Klimov & dt J. Comp. Chemistry (2002)

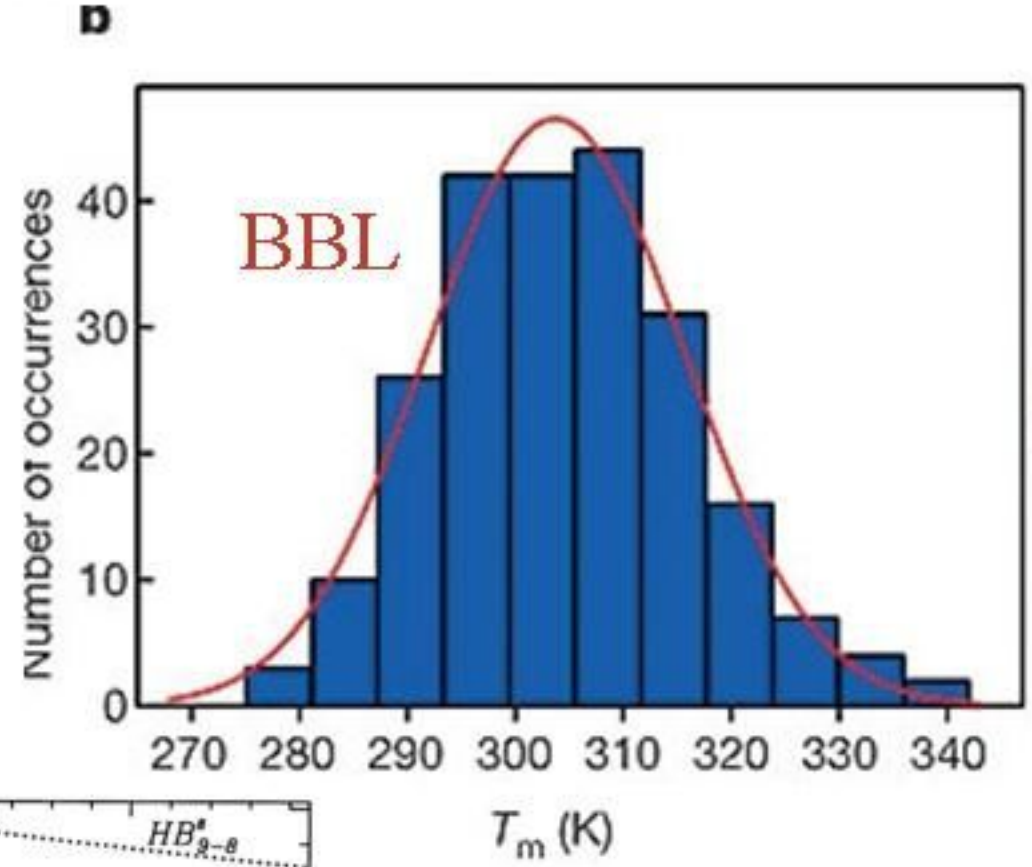
Is the melting temperature Unique?
Finite-size effects!



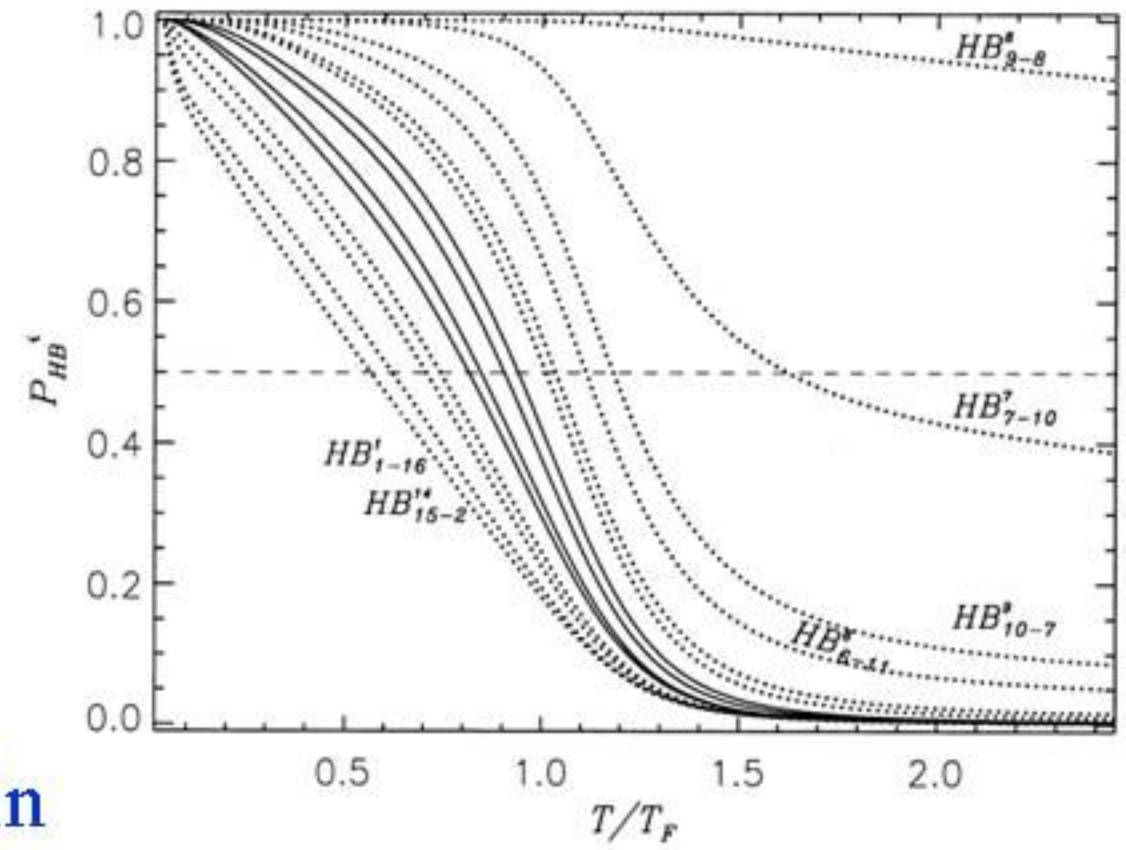
Holtzer
Leucine
Zipper
Biophys J
1997

Udgaonkar
Barstar Monnelin

ΔT large



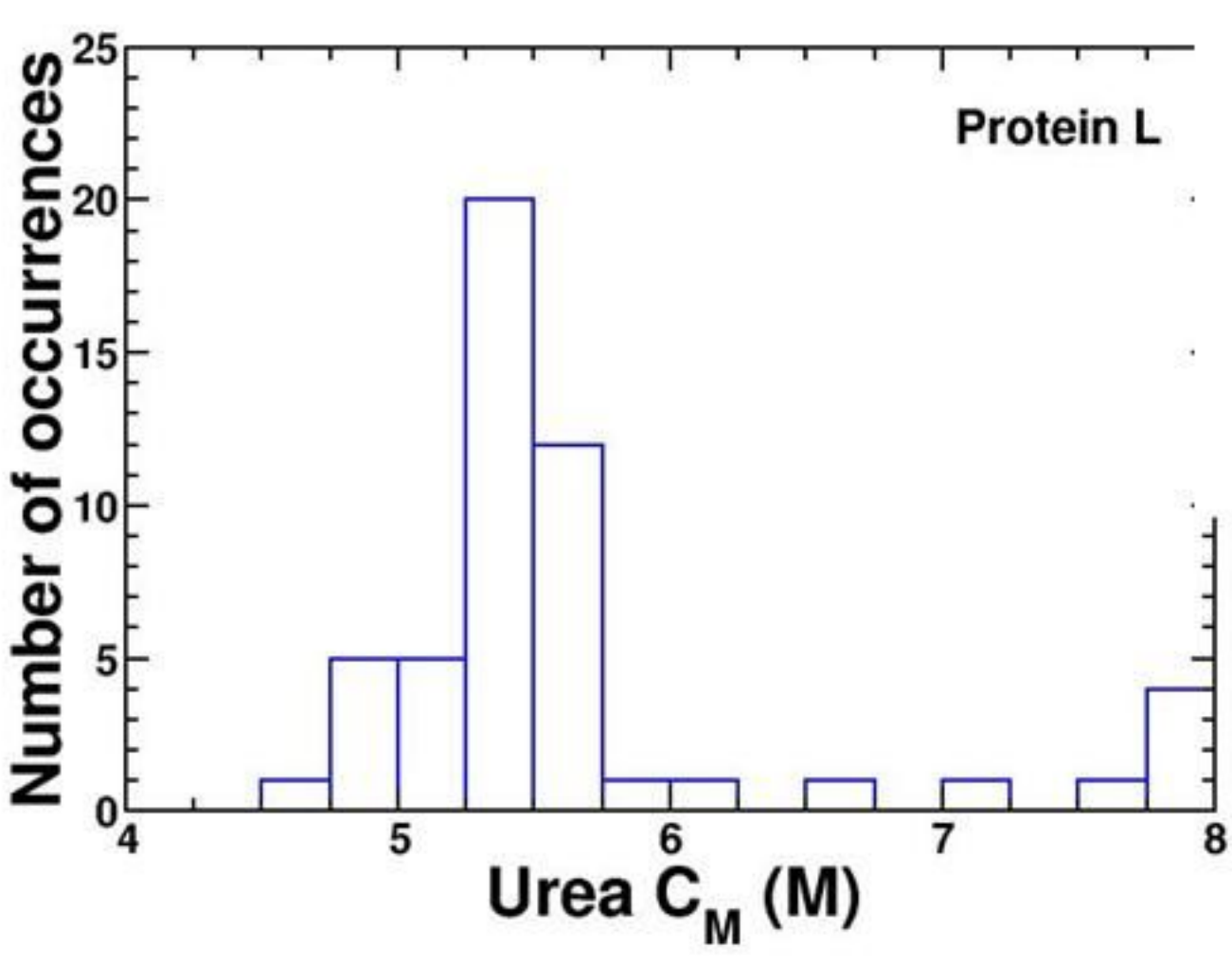
Munoz
Nature
2006

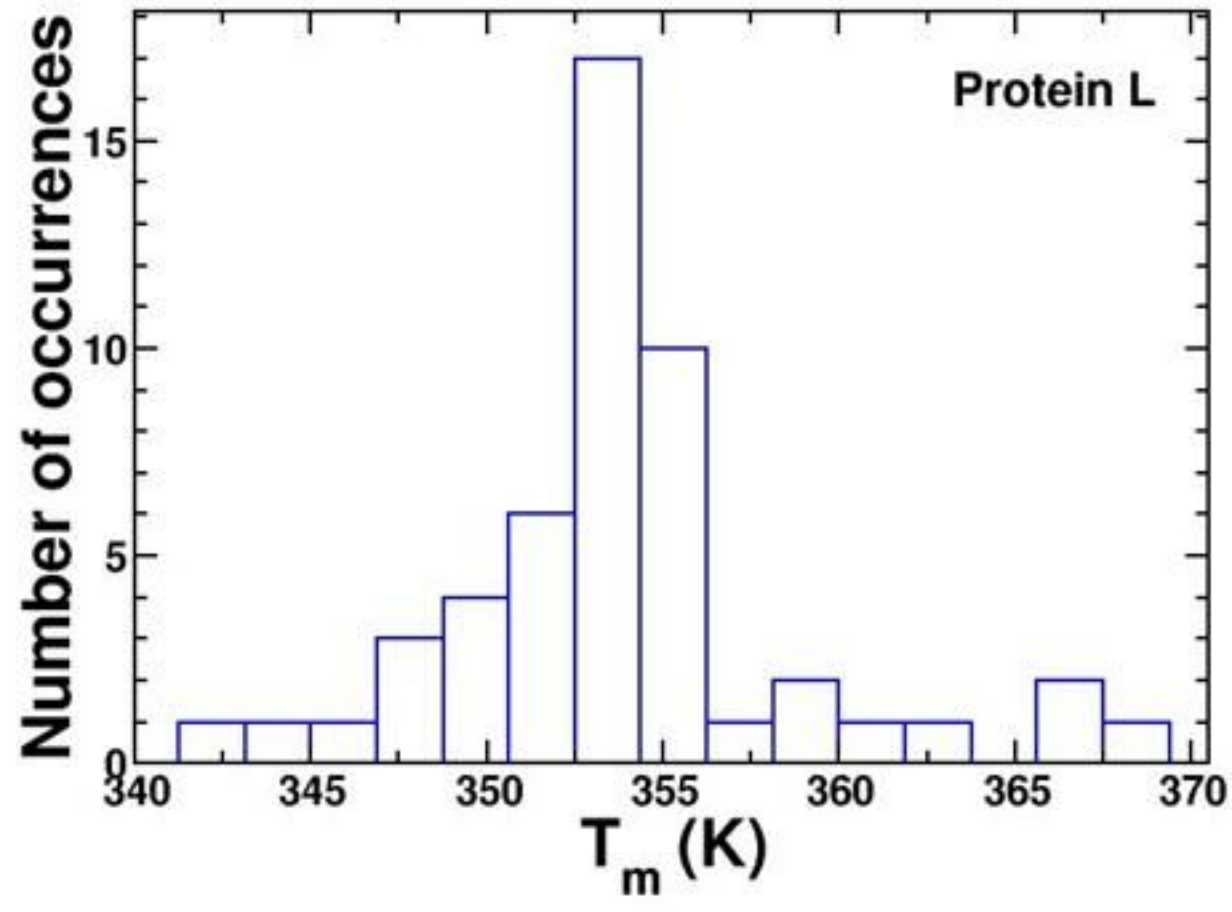


β-hairpin
PNAS 2000
Klimov & dt

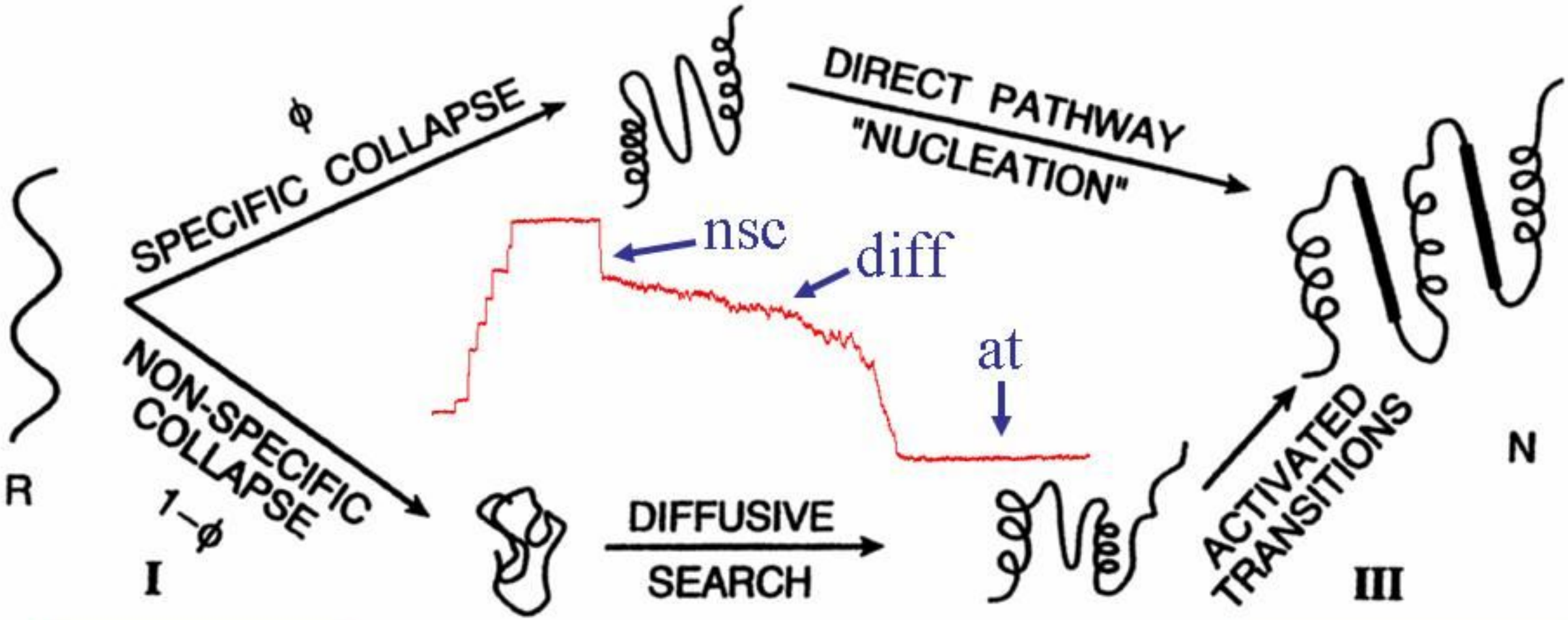
Residue dependent ordering Protein L
O'Brien, Brooks & dt Biochemistry (2009)

Spread decreases as N decreases….finite-size effects





Folding Pathways and times



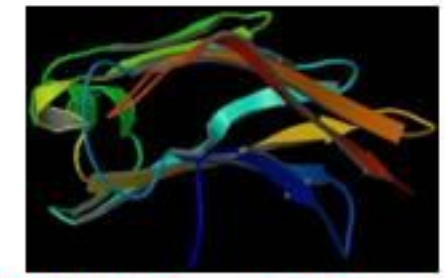
$$\tau_{nsc} \cong \tau_{nsc} n^2$$

$$\tau_{diff} \cong \tau_{diff} n^4$$

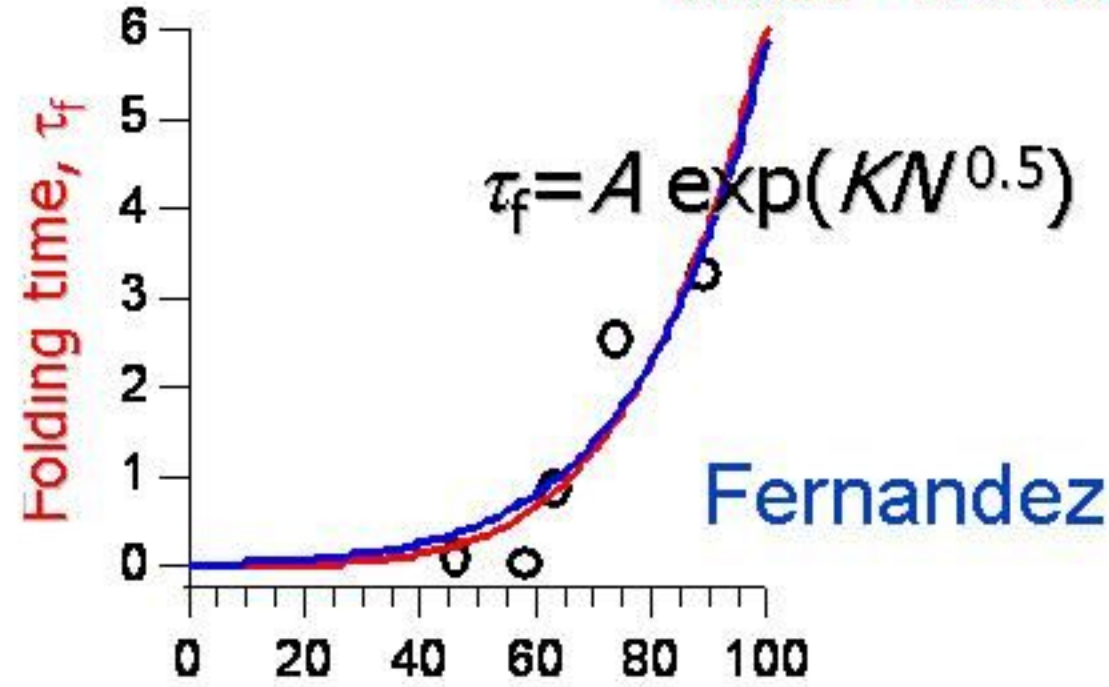
$$\tau_{at} \cong \tau_{at} e^{0.6\sqrt{n}}$$

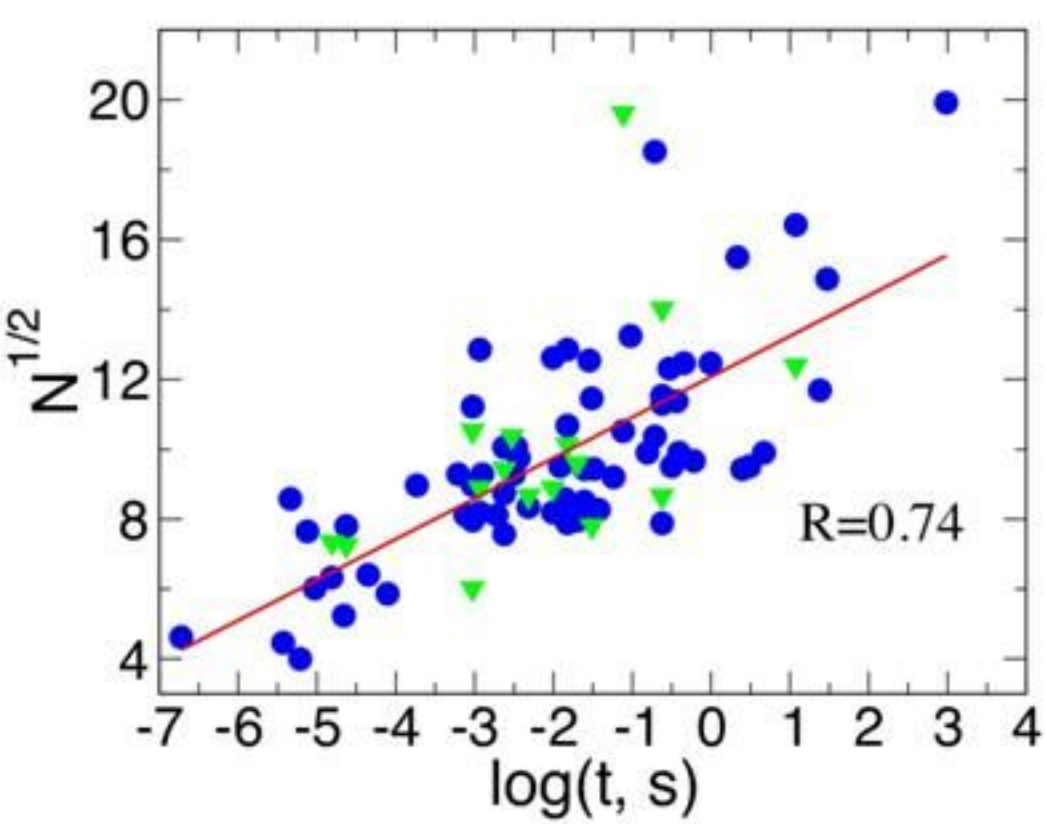
dt: *J. Phys. (France)* 5: 1457-1467

Folding Time vs N



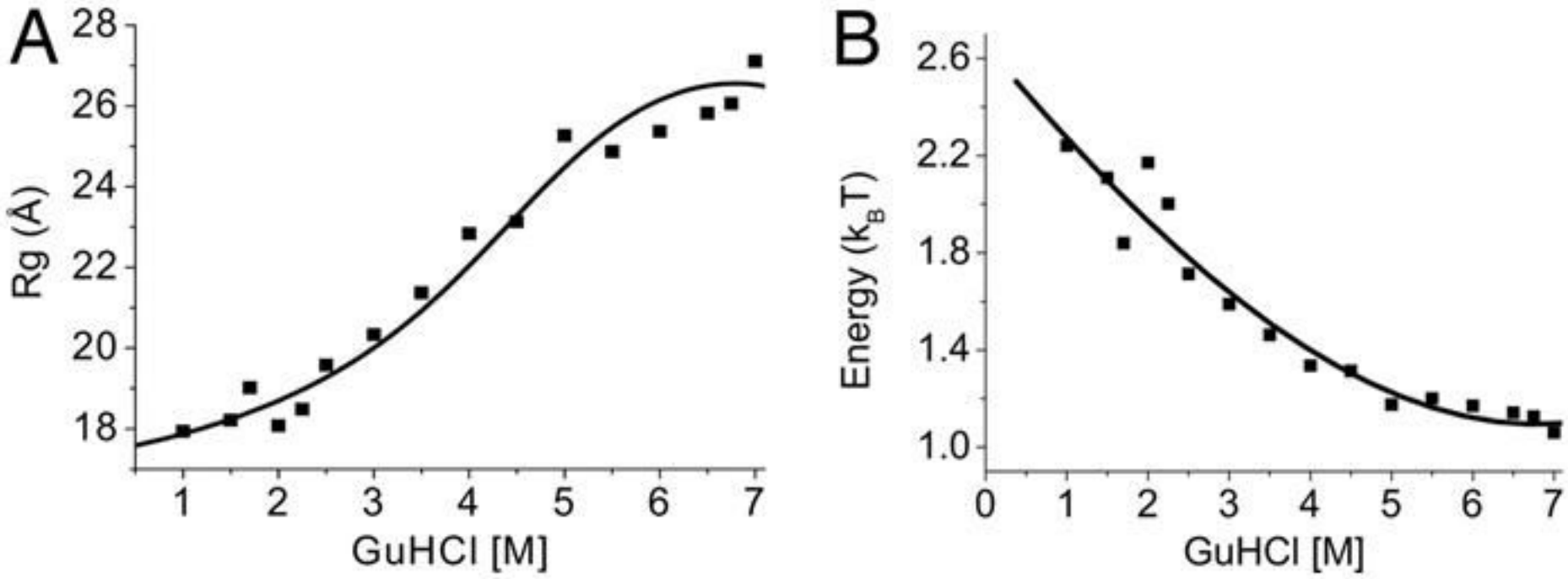
CNA19 –150 a.a.

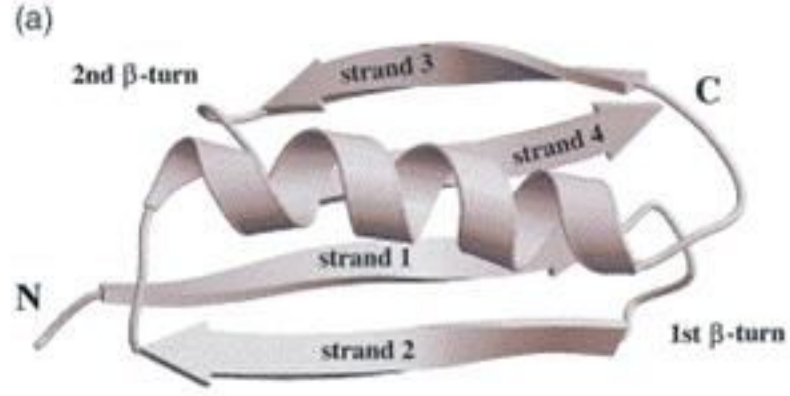




Munoz

The CG transition in denatured protein L smFRET measurements: Collapse

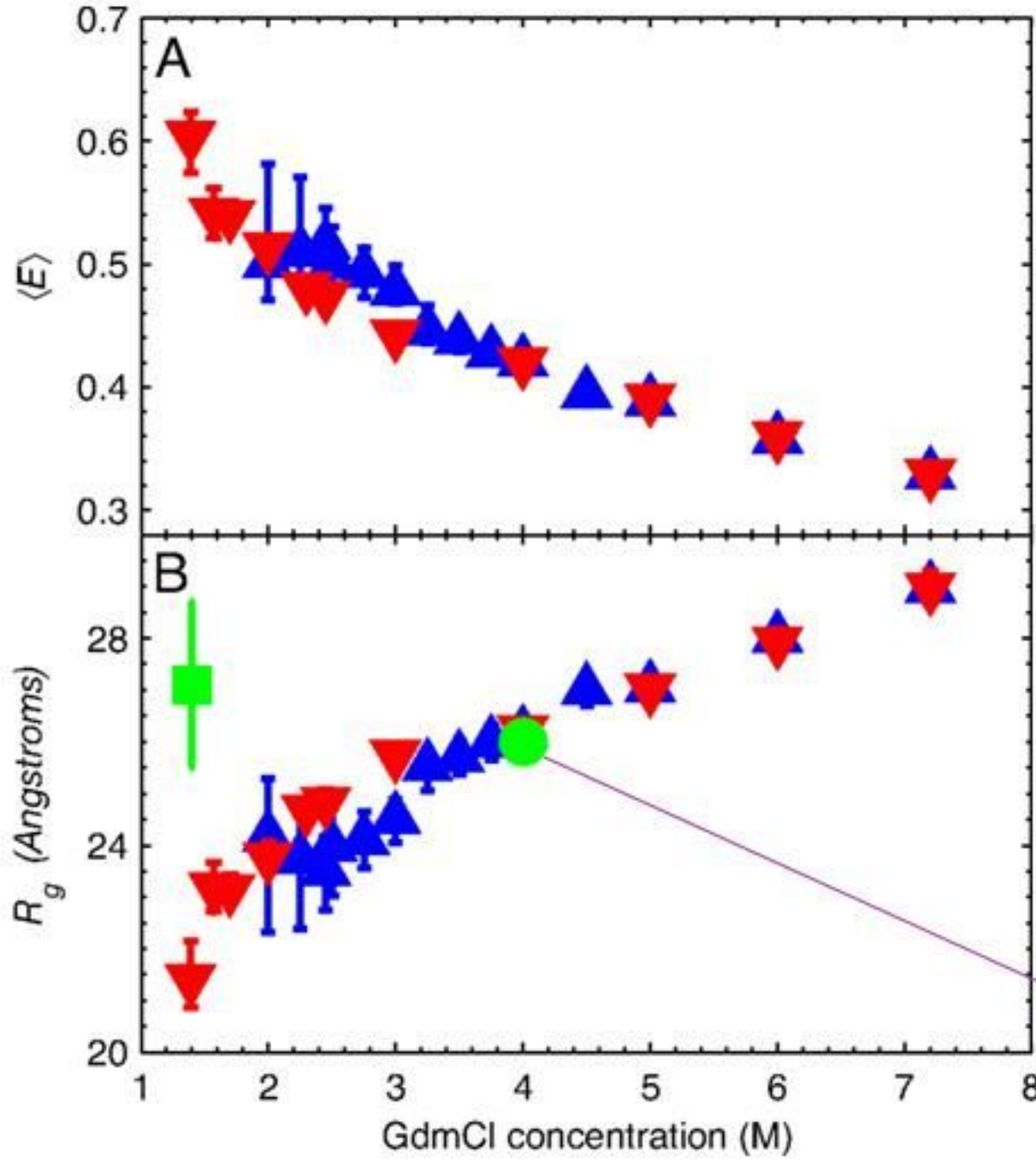




Sherman, Eilon and Haran, Gilad (2006) Proc. Natl. Acad. Sci. USA 103, 11539-11543

Dependence of FRET efficiency and Rg on denaturant concentration

Similar N (65 and 66)



Blue: Prot L

RED: CspTm

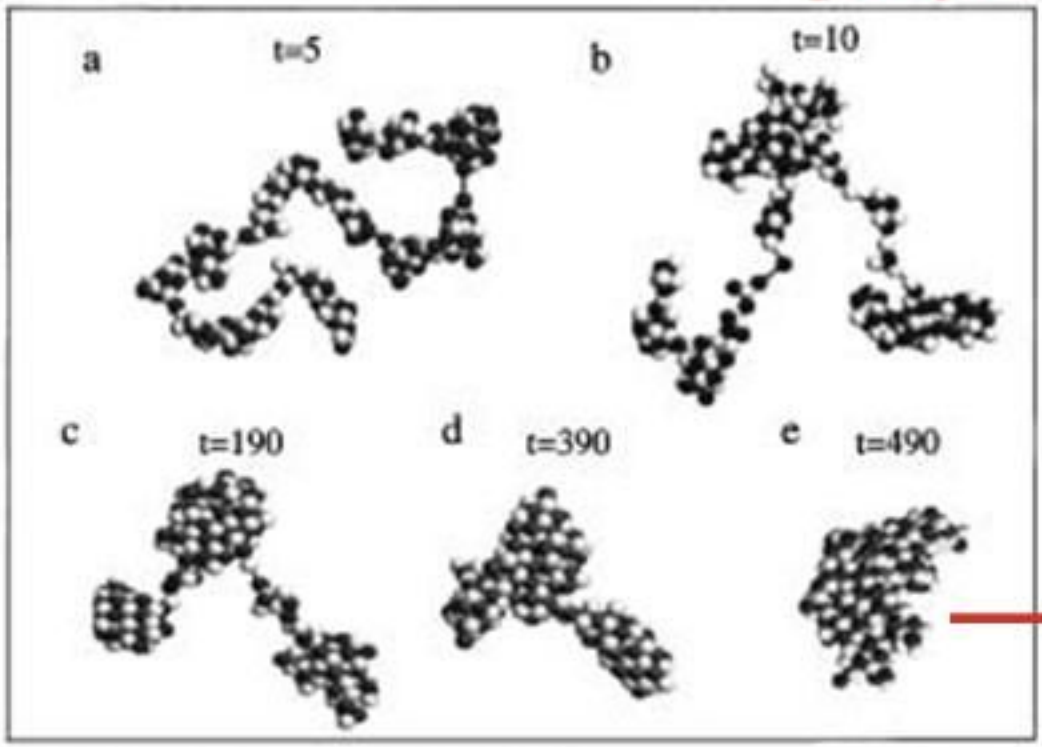
SAXS (1999)

Merchant, Kusai A. et al. (2007) Proc. Natl. Acad. Sci. USA 104, 1528-1533

Collapse Kinetics of Polyampholytes and polyelectrolytes

Lee and DT
JCP (2000)

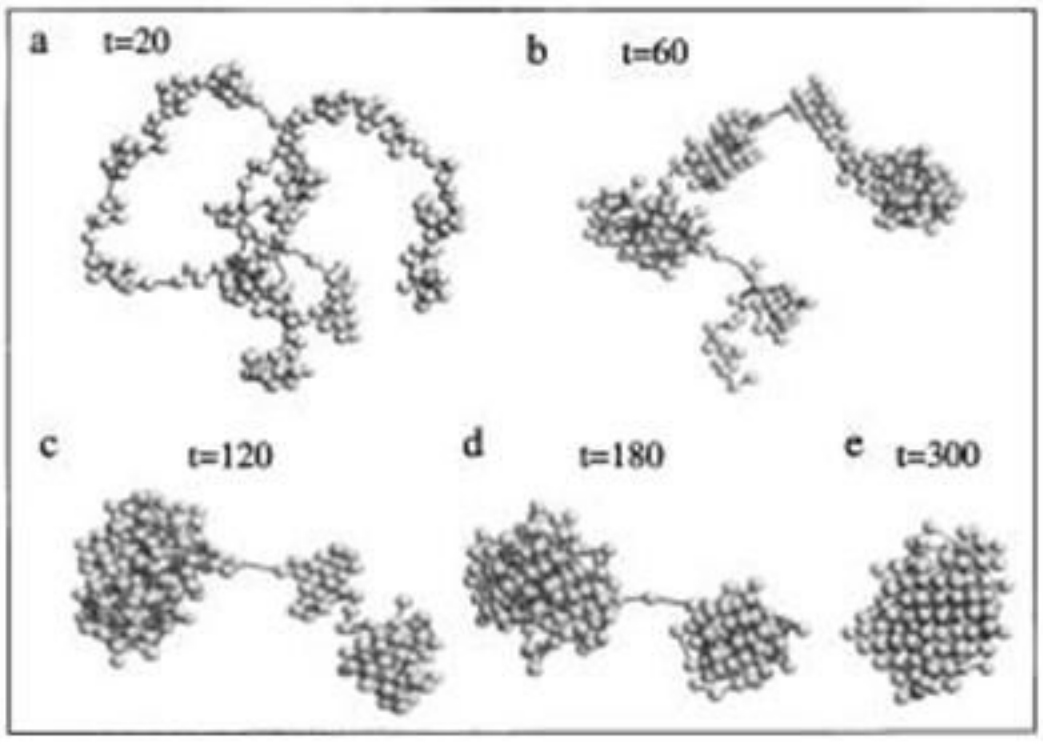
PA



$Q < Q_c = \pm\sqrt{N}\, e$ PA collapses (Kardar, Rubenstein, Obukhov..)

Wigner crystal

PE

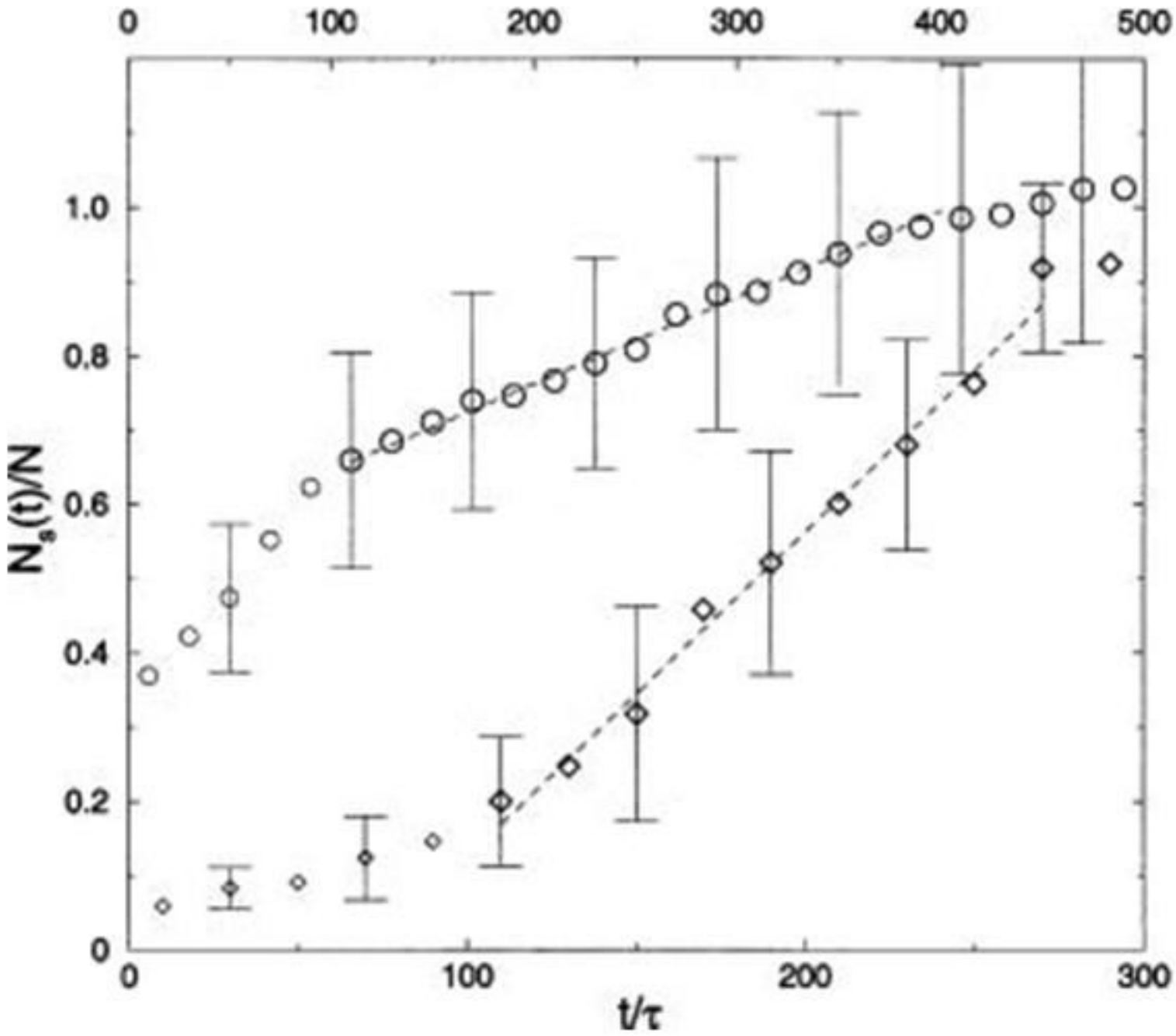


Two-stage transition:

1) Pearl-necklace structures form.

2) Pearls merge to form Globules.

Second Stage: Globules grow by Lifshitz-Slazov mechanism



Lee and dt (2000)

$$\tau_P \approx \mathcal{N}^{4/5}$$

(STAGE 1)

$$\tau_C \approx \mathcal{N}^{\alpha}$$

(stage 2)

de Gennes, Orland, NGS, others for homopolymers

"Two-State" Folders: Stages in Collapse

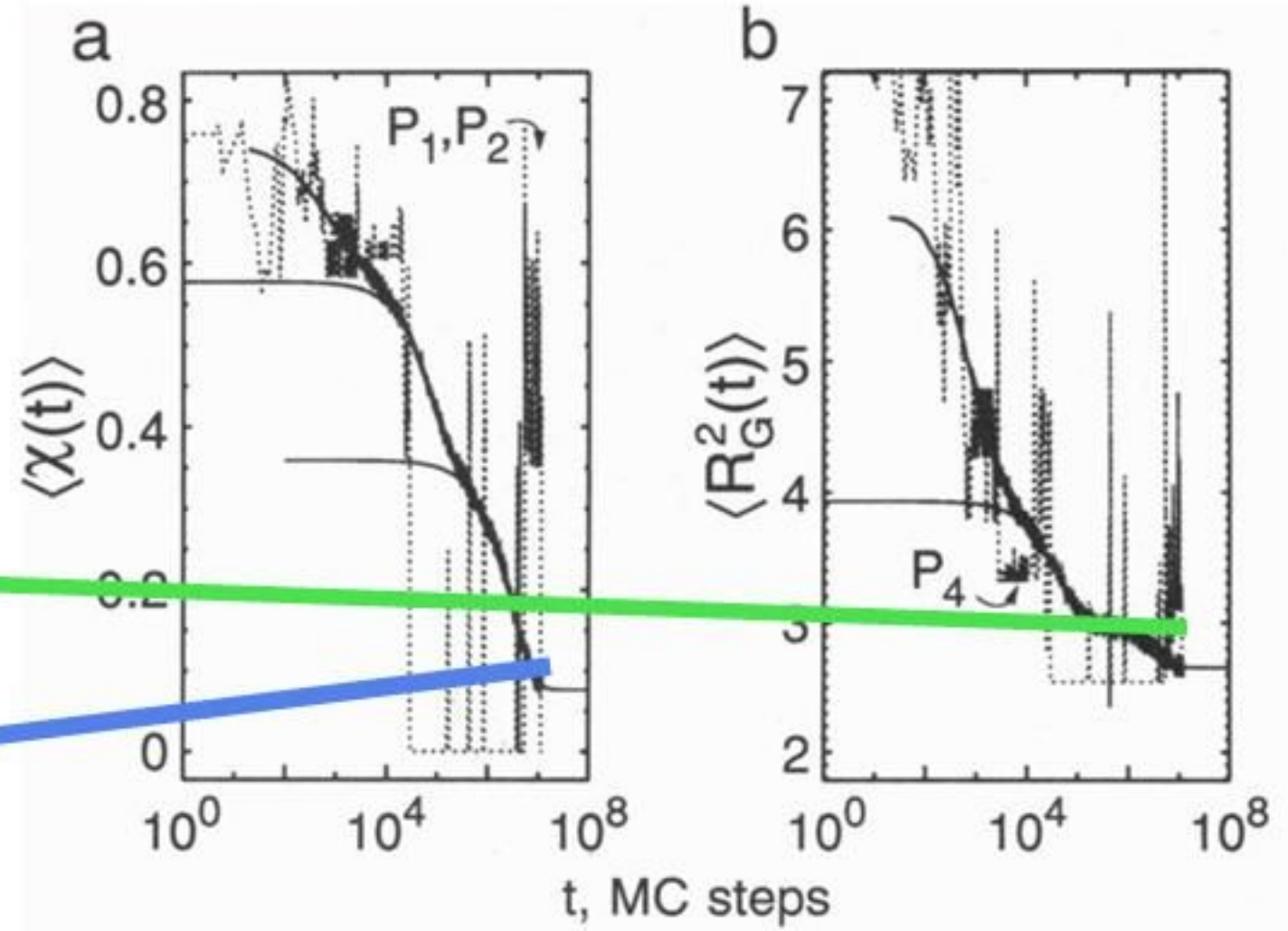
Camacho and dt, PNAS (1993)

Random Coil $\xrightarrow{\tau_C}$ "Specific Collapse" $\xrightarrow{\tau_F}$ Native State

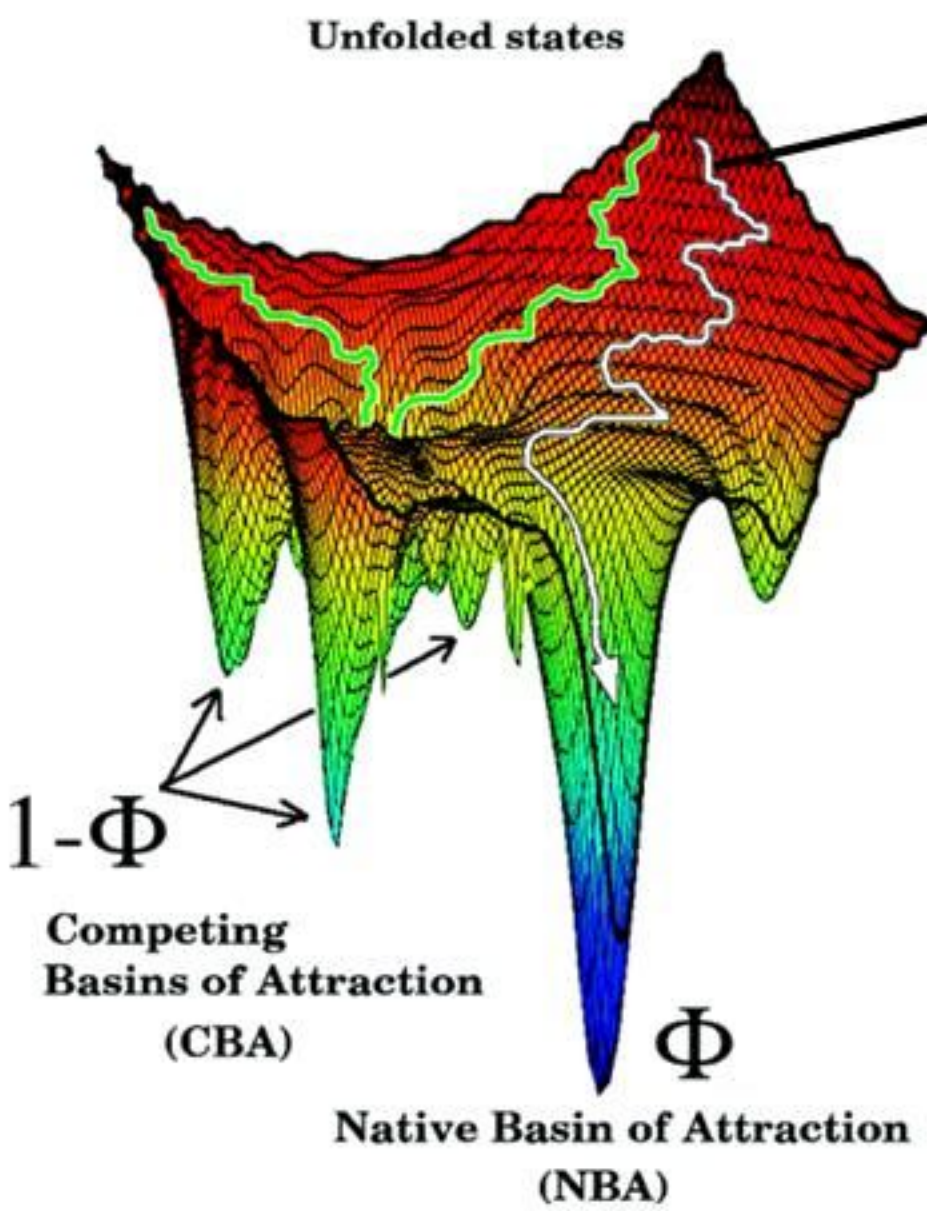
$\tau_F / \tau_C \approx (3 - 5)$

τ_C

τ_F

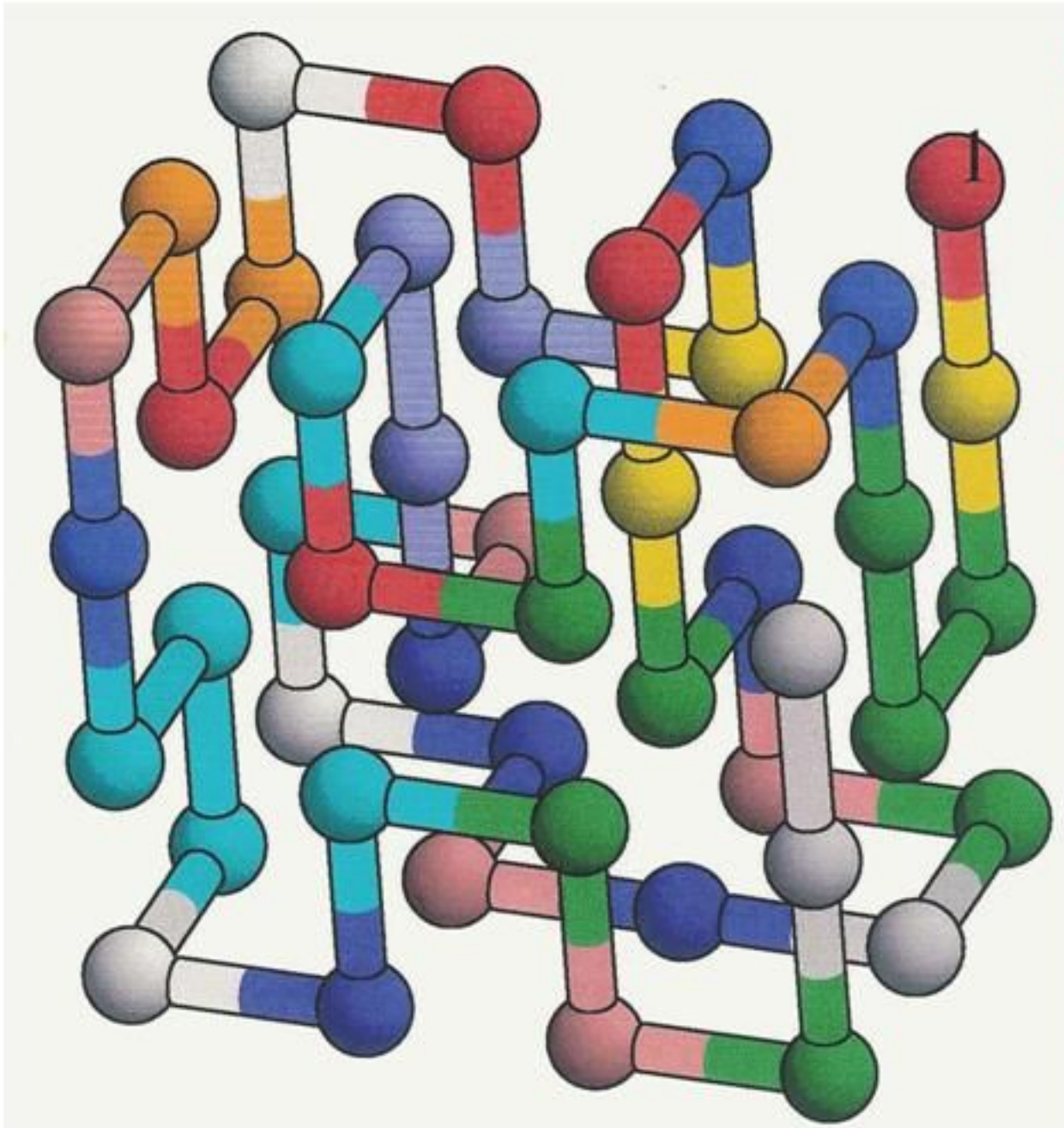


Folding in rugged Landscape: KPM different collapse scenarios



	Φ	τ_{fast}	Ref.
Lysozyme	0.25	50ms	PNAS (95) Kiefhaber
Rubisco	0.05	-	PNAS (96) MT/GL/DT
Pre r-RNA	0.10	1s	Pan/DT/ Woodson JMB (97)

Lattice model with small and moderate values of σ

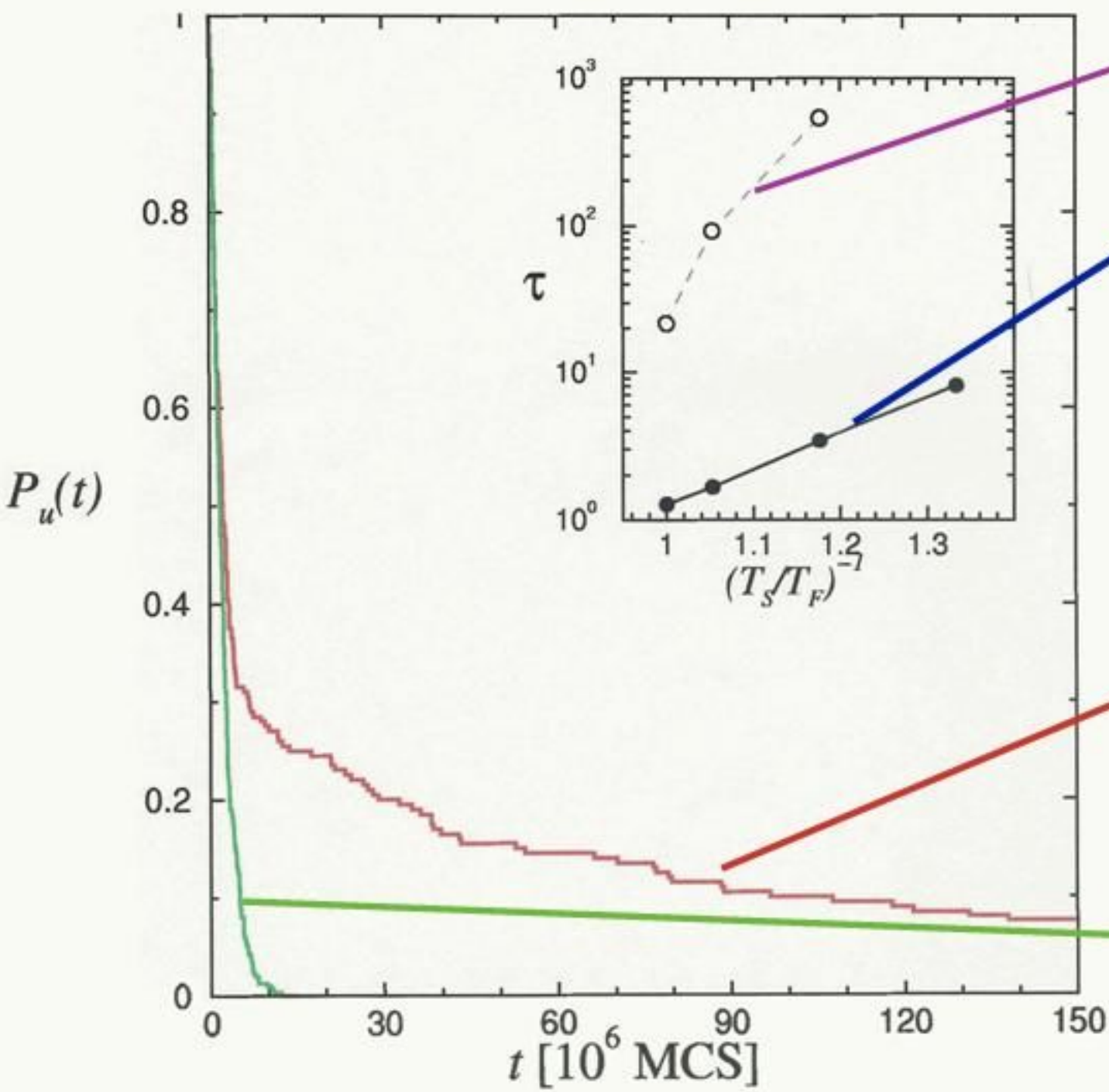


Two sequences
Small σ (specific Collapse)

Moderate σ (KPM)

Fast and slow tracks

Sequences with moderate σ value fold by KPM



τ_{slow}

τ_{fast}

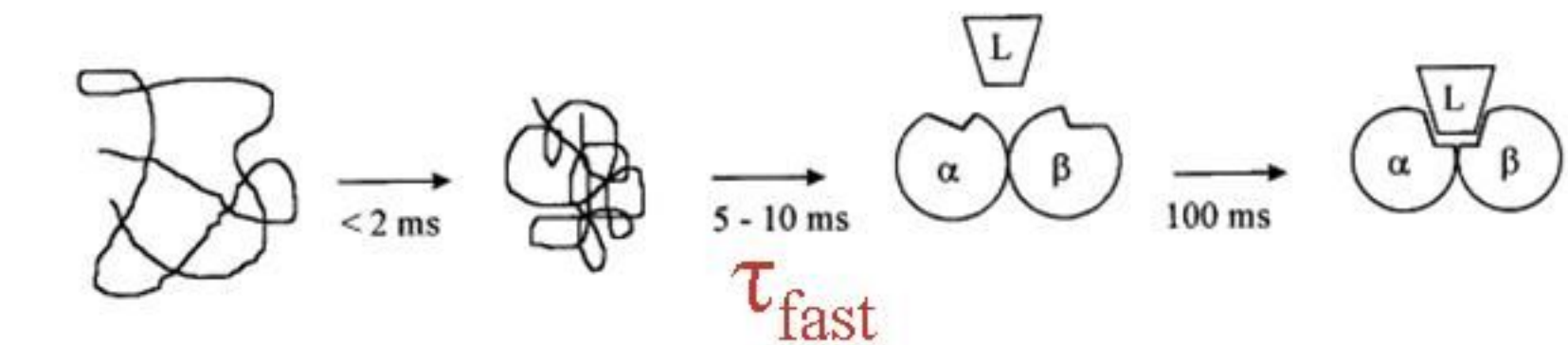
$\tau_{slow} \approx 100\ \tau_{fast}$

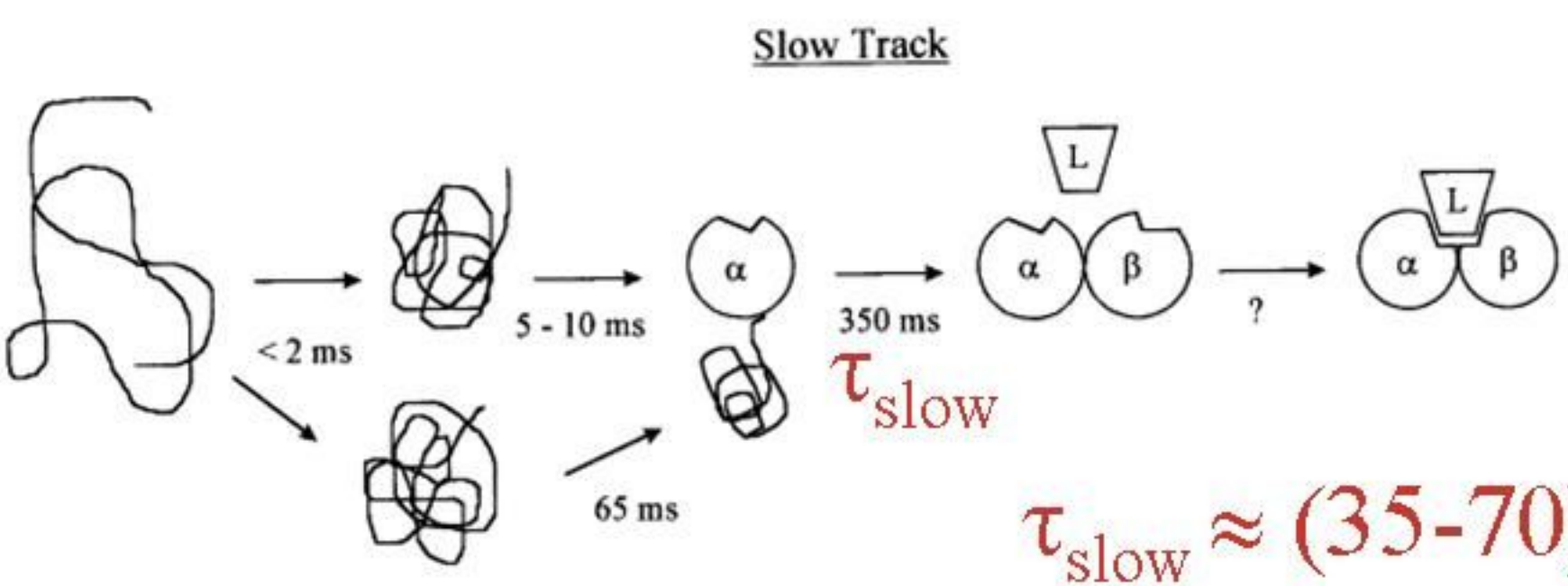
Moderate σ

Small σ

KPM for Lysozyme

Matagne, Radford, & Dobson JMB (1997)

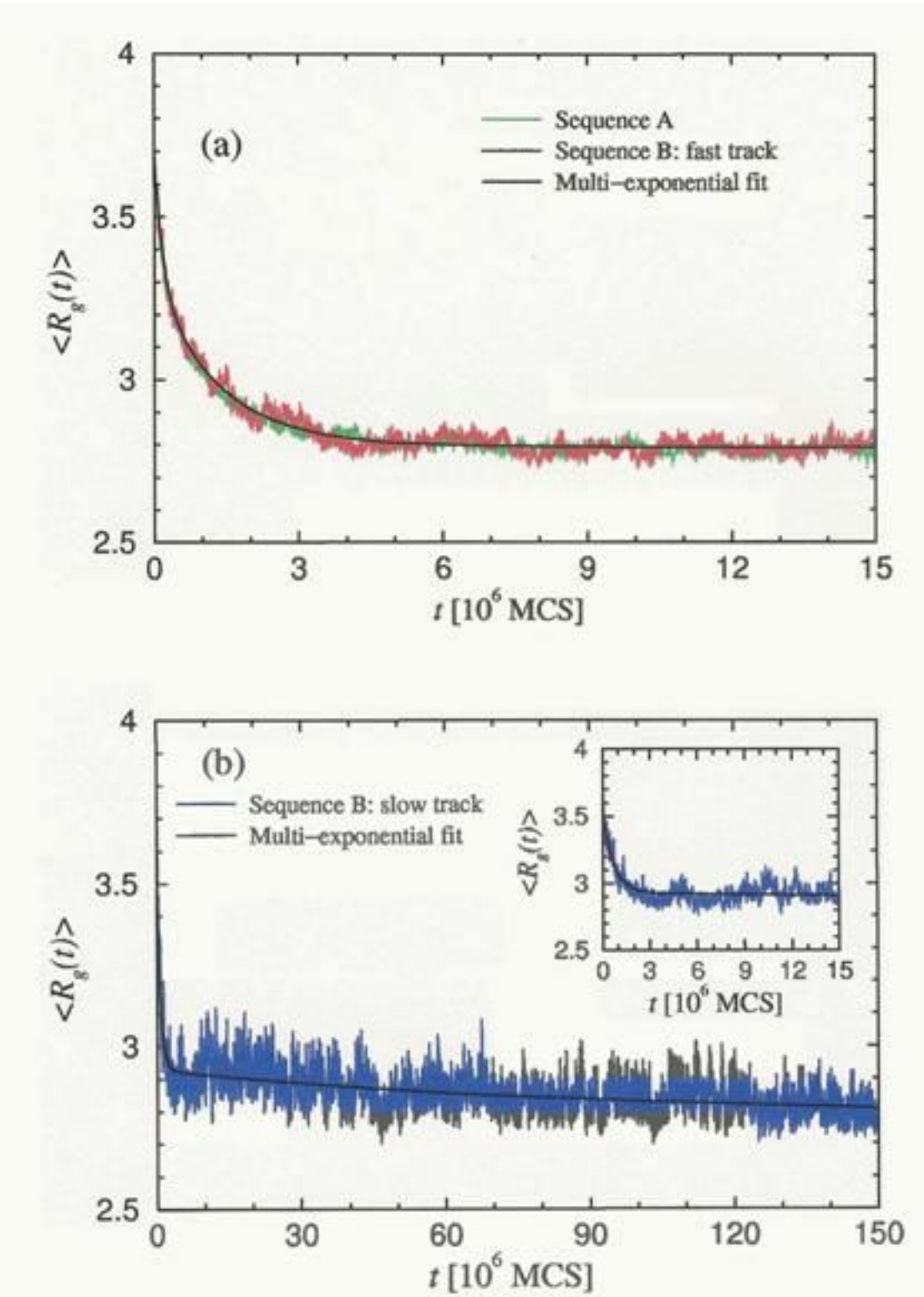




$\tau_{slow} \approx (35\text{-}70)\ \tau_{fast}$

agrees with LM Simulations

Collapse kinetics in fast and slow track molecules are different



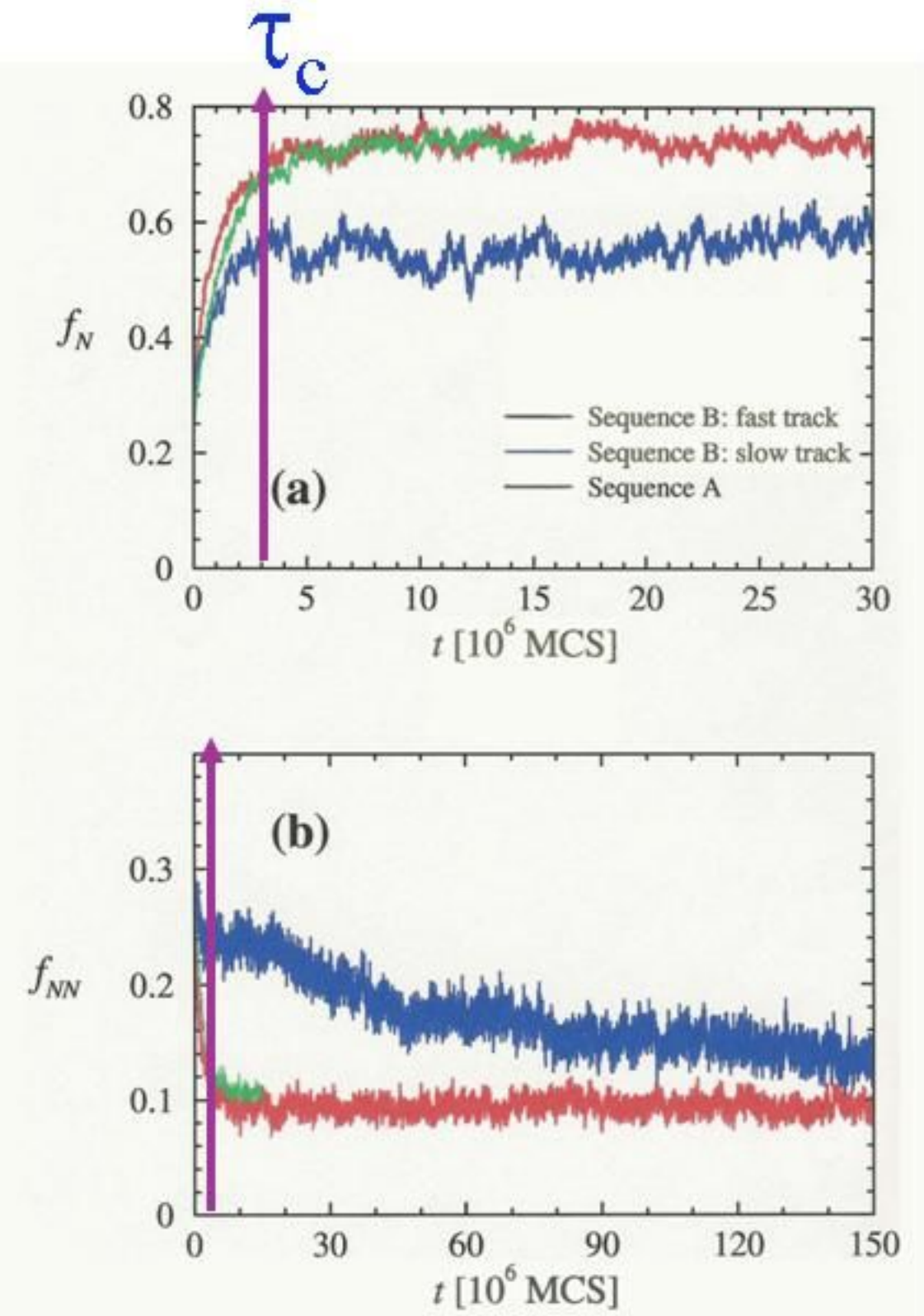
- Fast track collapse similar to small σ sequence
- Both exhibit 3 stage compaction

o Collapse kinetics greatly impeded

o Early events are similar but produce "incorrect" structures

o Partitioning occurs early

Fast track molecules undergo specific collapse



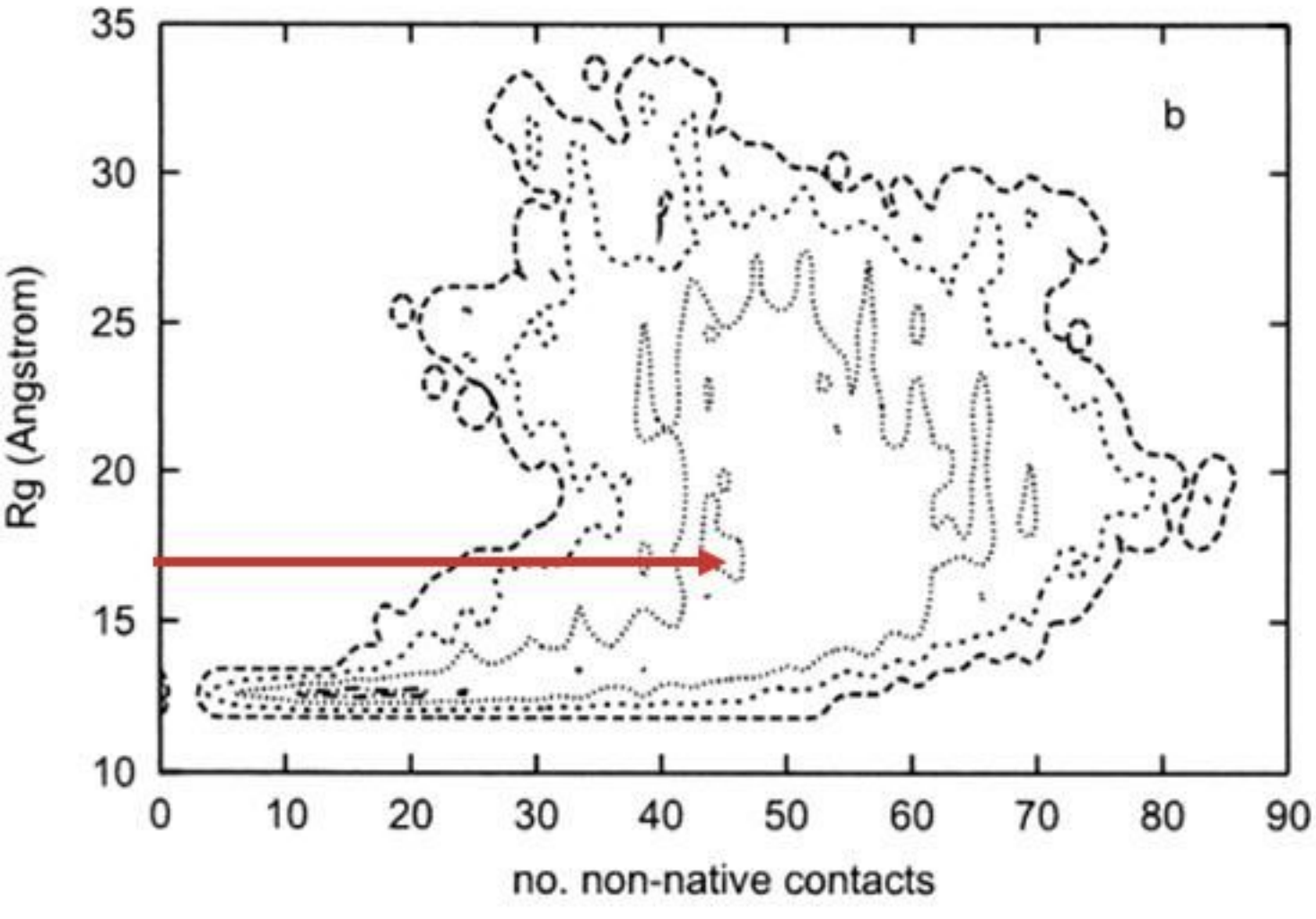
Collapse in small σ sequence and fast track molecules have Substantial fraction native contacts on τ_C.

Slow track molecules: Non-specific collapse with non-native contacts. Annealing takes much longer time.

Non-Native contacts in early stages of Cyt C folding

Serdal, Cardenas, dt, Elber

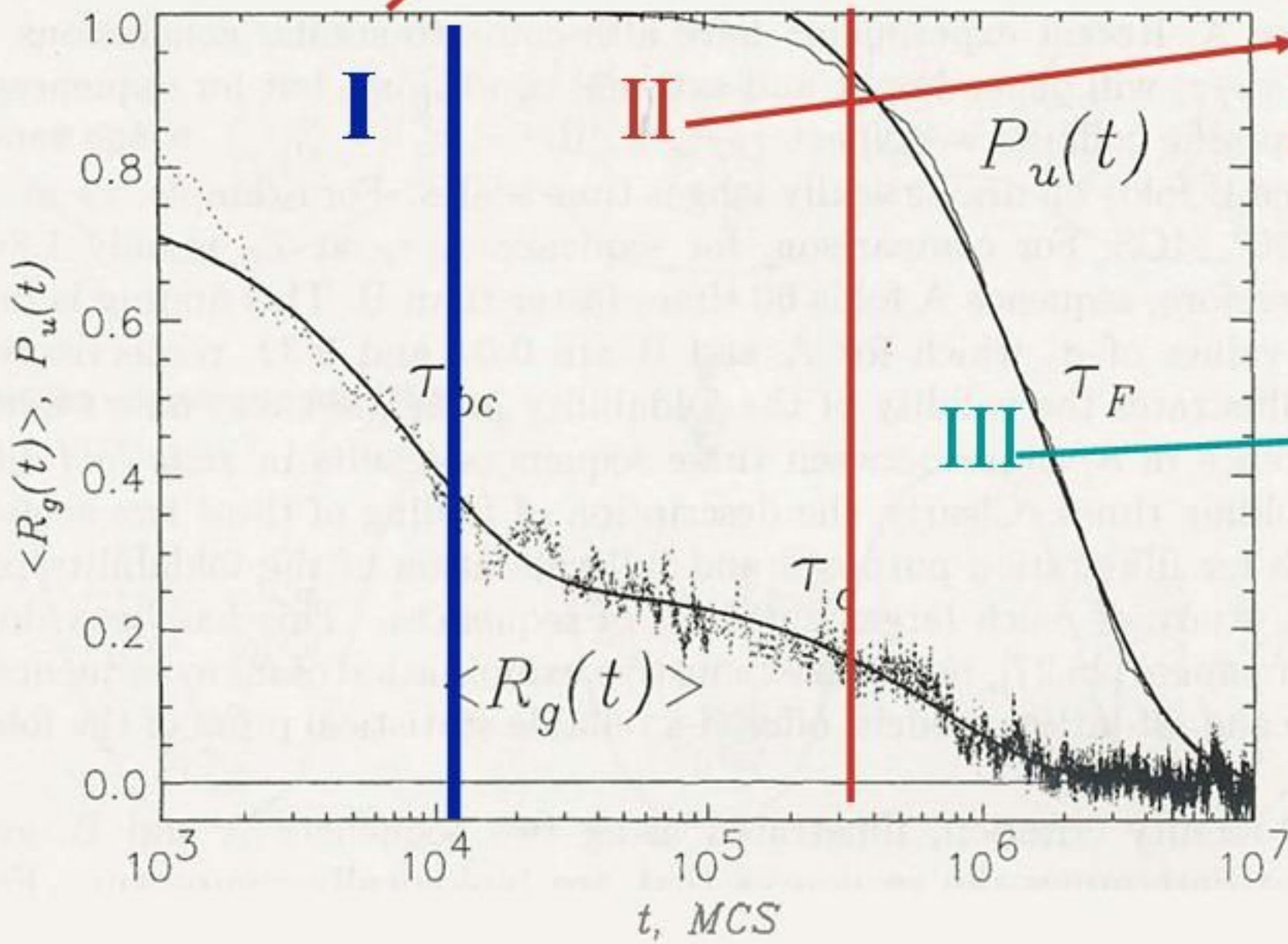
All atom simulations



"Molten Globule"
(40% helix content;
< 10 native contacts)

Relating Collapse and folding times
(Lattice Models with side chains)

Non-native contacts form

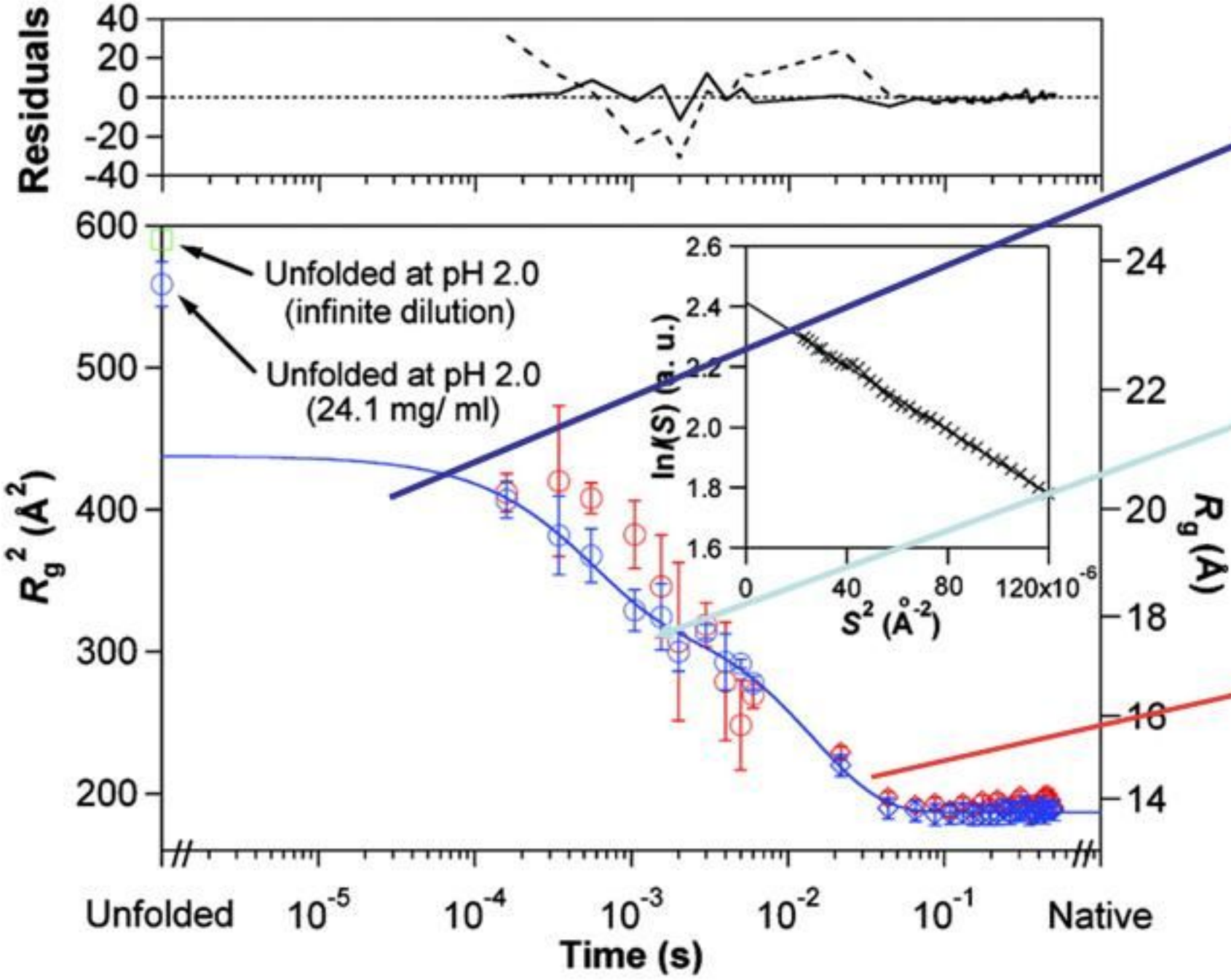


NN contacts anneal

Specific Collapse few NN contacts

$\tau_F = 3\ \tau_C$

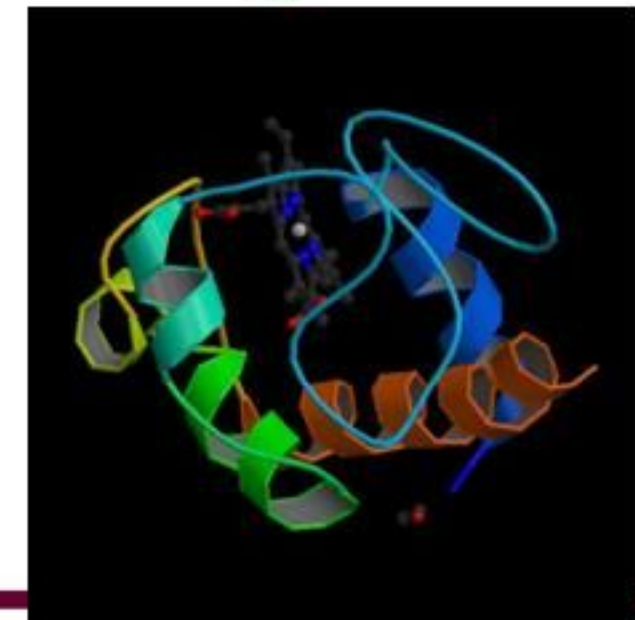
Collapse kinetics of Cyt c (three stages)



τ_B (non native)
(not resolved)

τ_C

τ_F

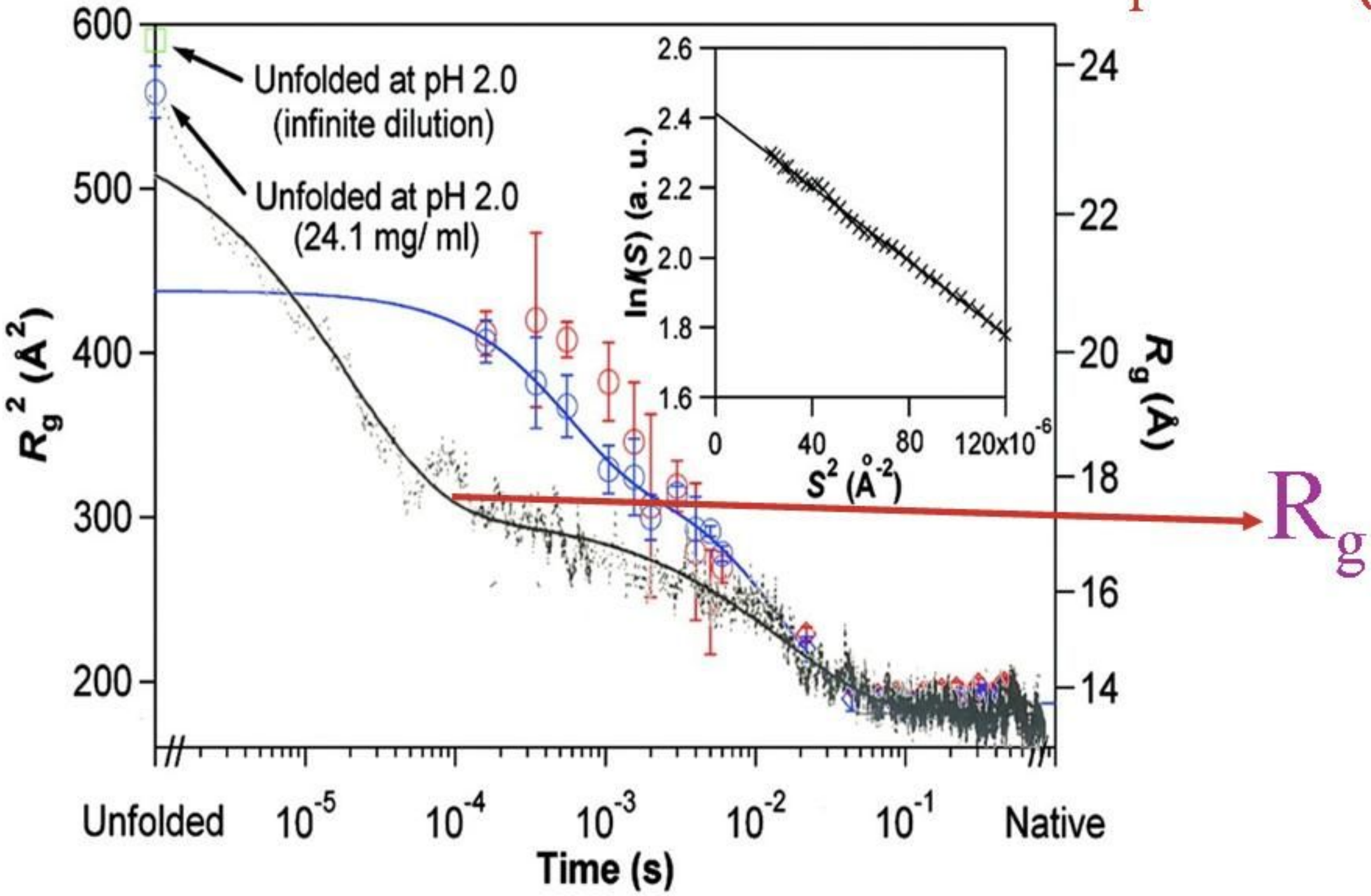


Akiyama, Shuji et al. (2002) Proc. Natl. Acad. Sci. USA 99, 1329-1334

Comparing Cyt C data and Lattice Model with side chains

Akivama PNAS (2002)

$\tau_F \approx 10\ \tau_C$



Some Specific Unresolved Questions

- Is 8M urea a good solvent in Flory sense?

Vitalis, dt & Pappu: No…Means partially compact

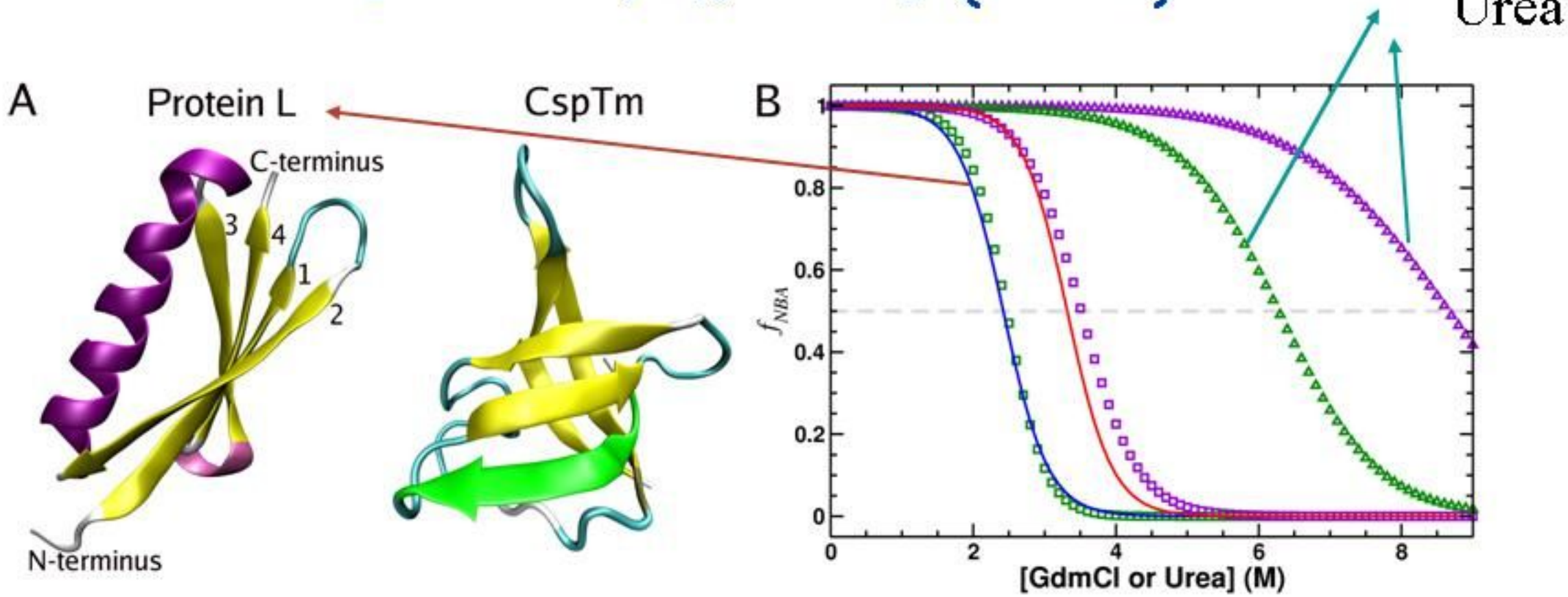
- T vs [C] collapse…Schuler type experiments

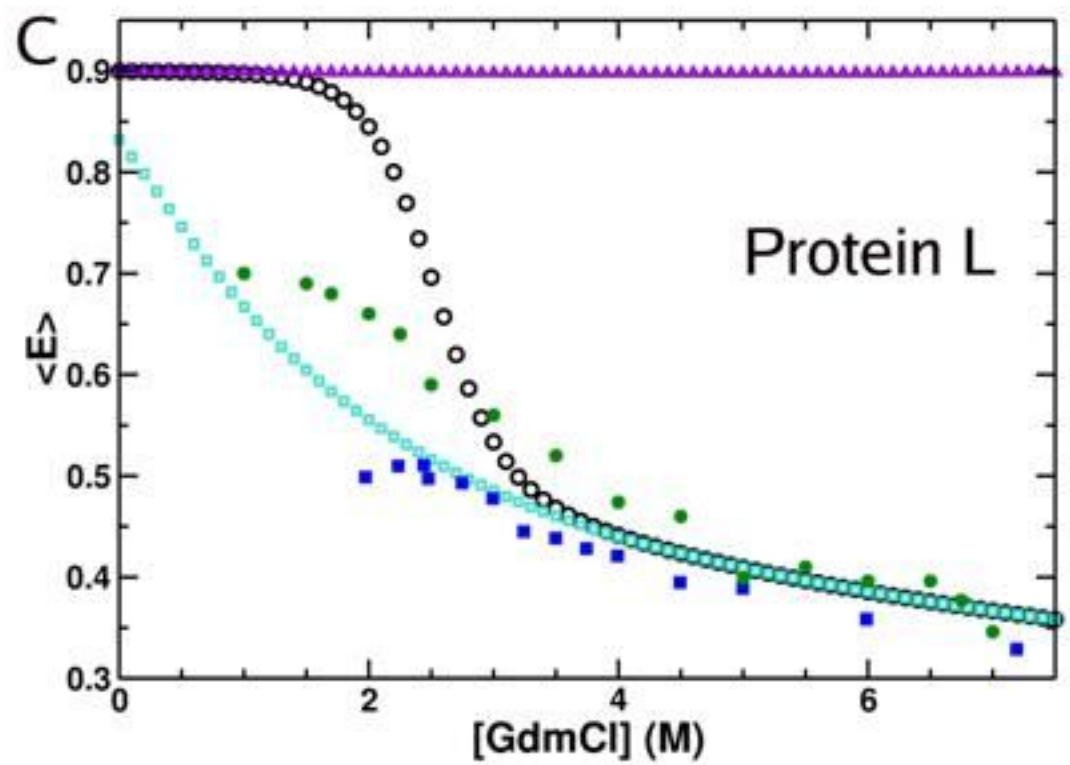
- DSE collapse SM FRET or using f. What about SAXS?

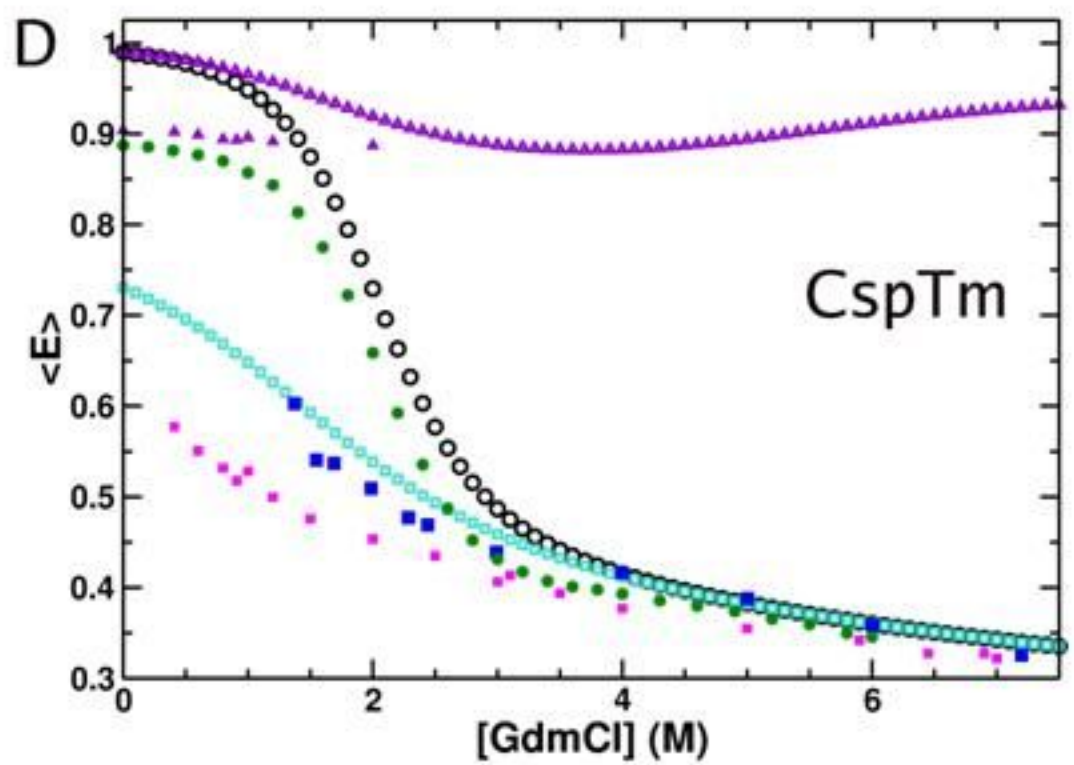
- For simulations: Models for direct comparison

With experiments. MTM (O'Brien) is a great start

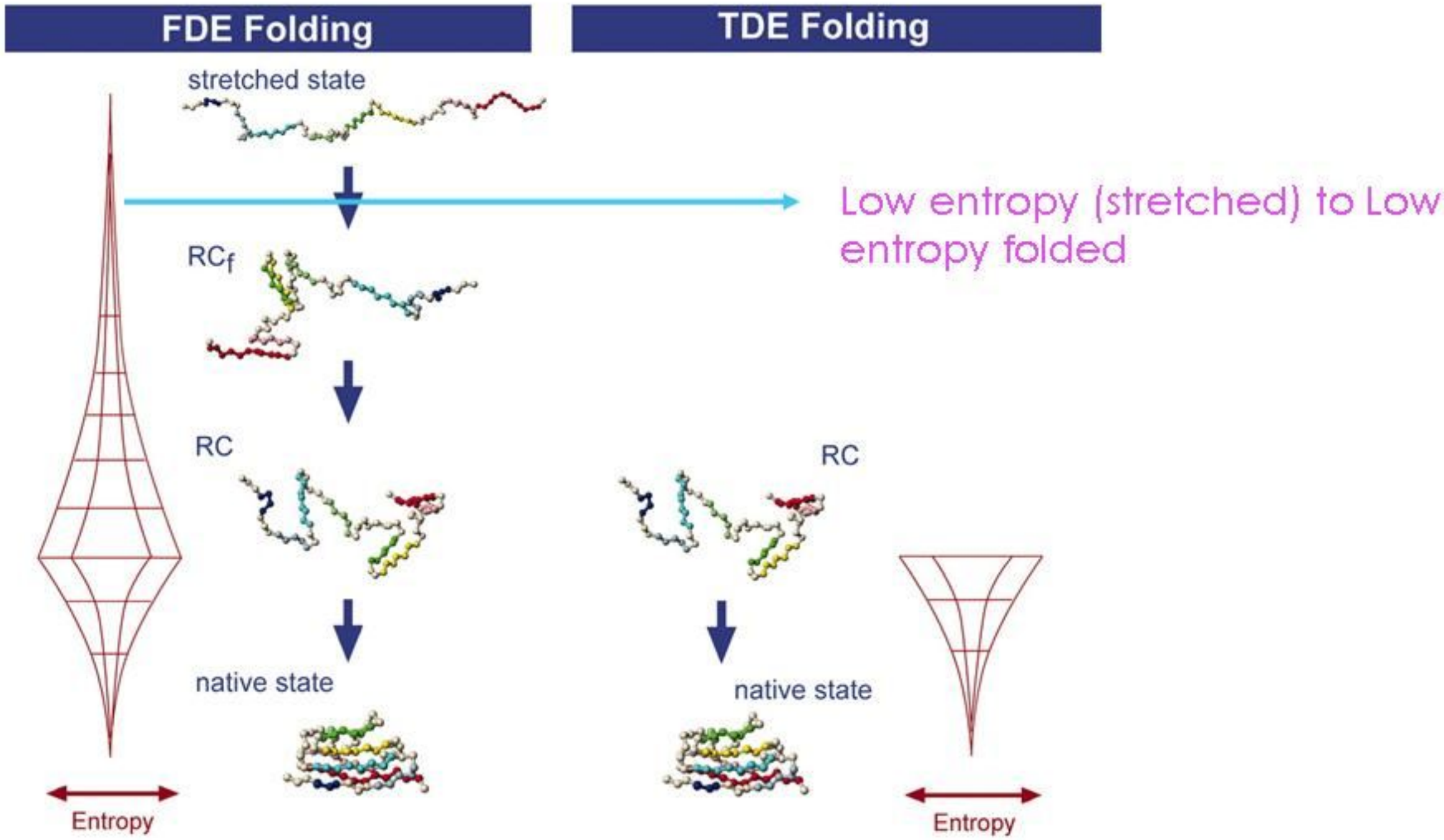
MTM simulations & single molecule exps
O'Brien et. al. PNAS (2008)







Differences in FDE and TDE refolding

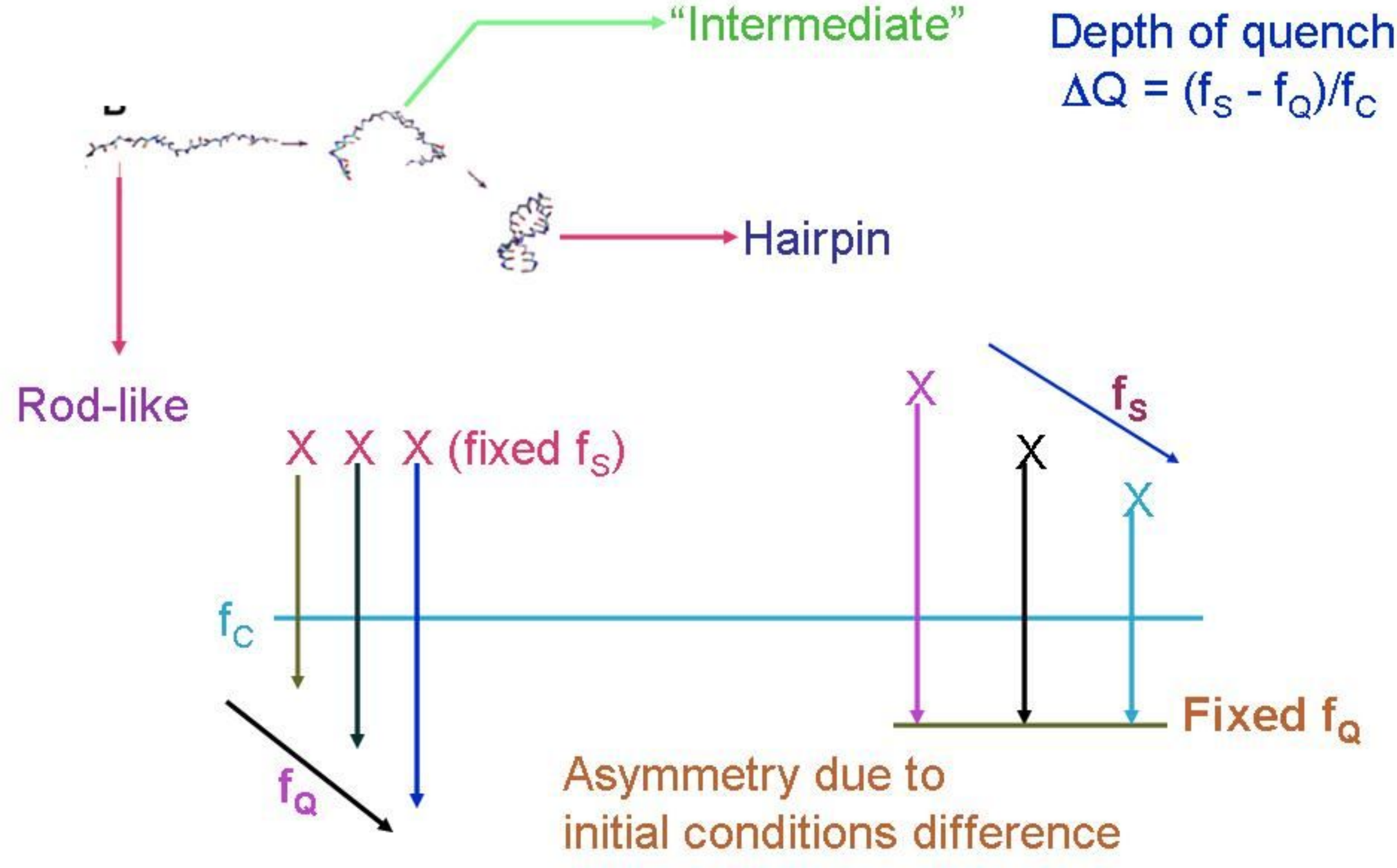


Fernandez J TIBS (2000)

Li, Mai Suan et al. (2006) Proc. Natl. Acad. Sci. USA 103, 93-98 PNAS

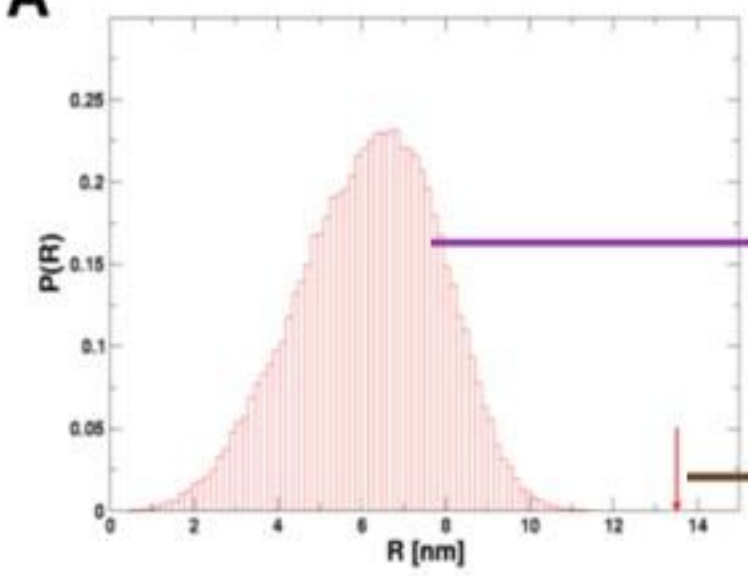
Refolding upon force quench

Hyeon, Morrison, Pincus &dt PNAS (2009)



Refolding upon force quench

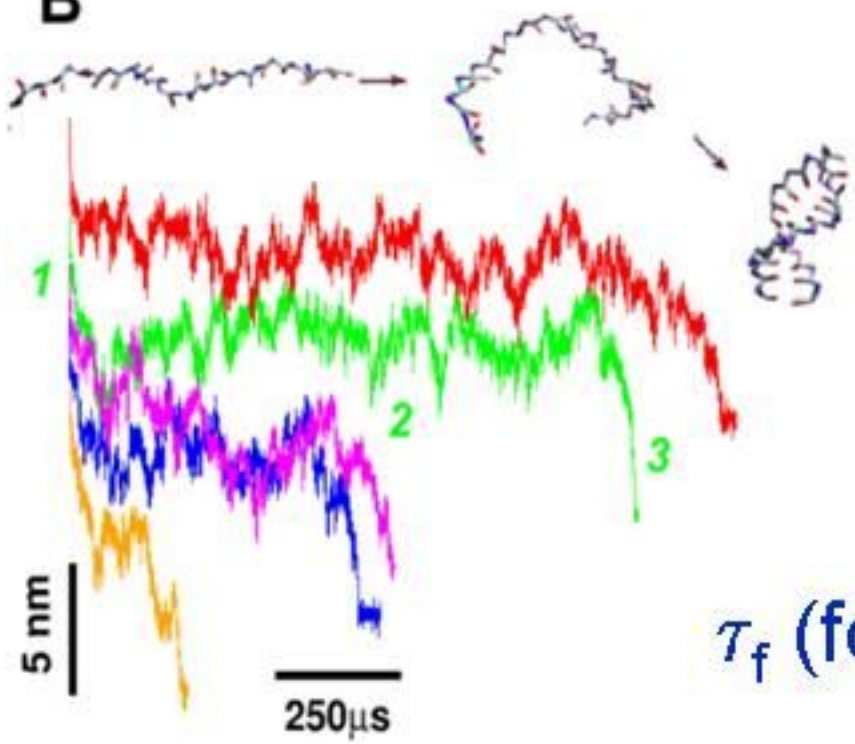
A



End-to-end distribution function

Fully stretched state

B

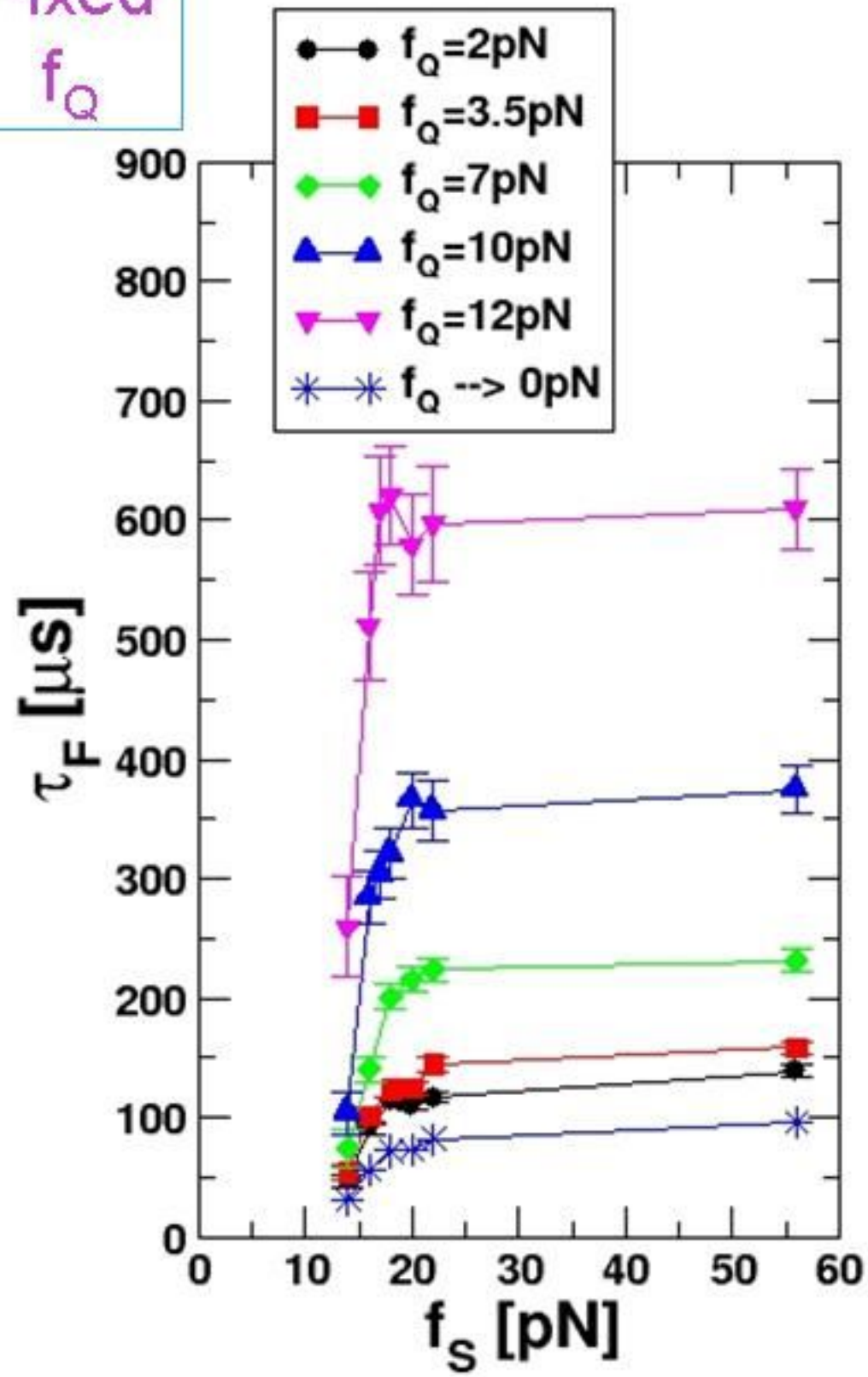


$f_S \rightarrow f_Q = 0$ pN

τ_f (force quench) >> τ_T (thermal quench)

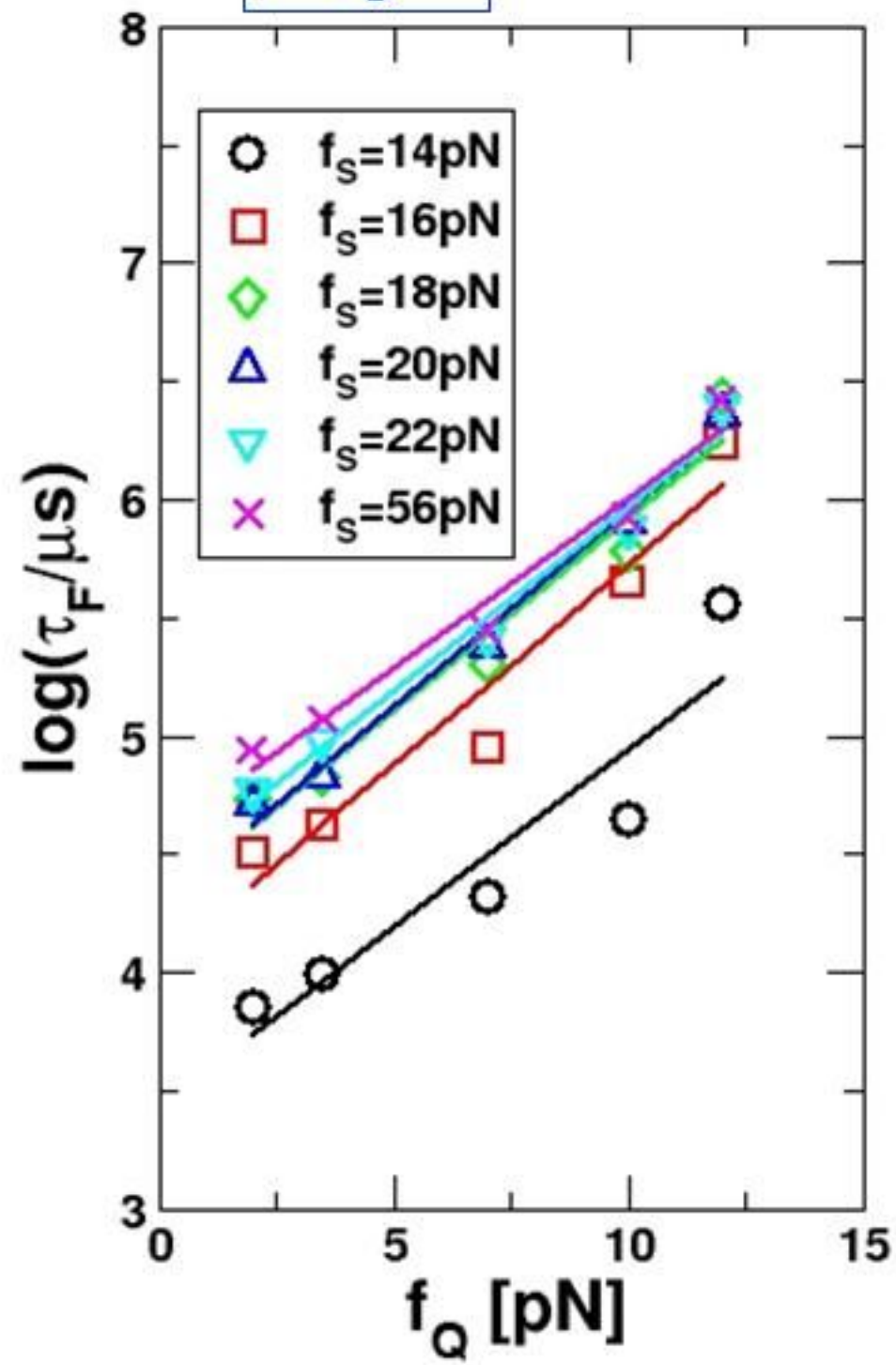
Force quench kinetics of P5GA

Fixed f_Q



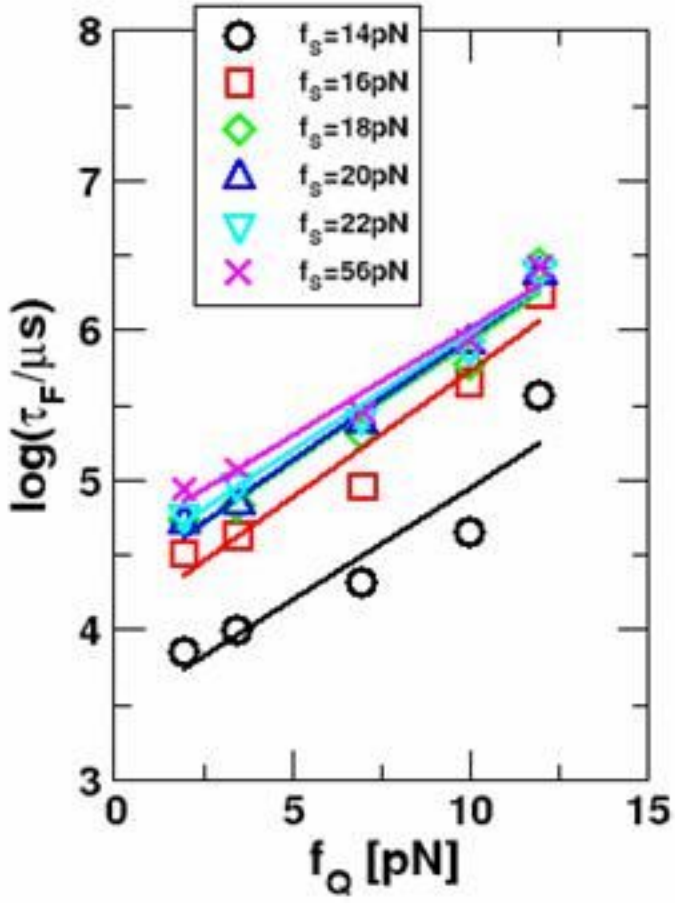
Fixed f_S

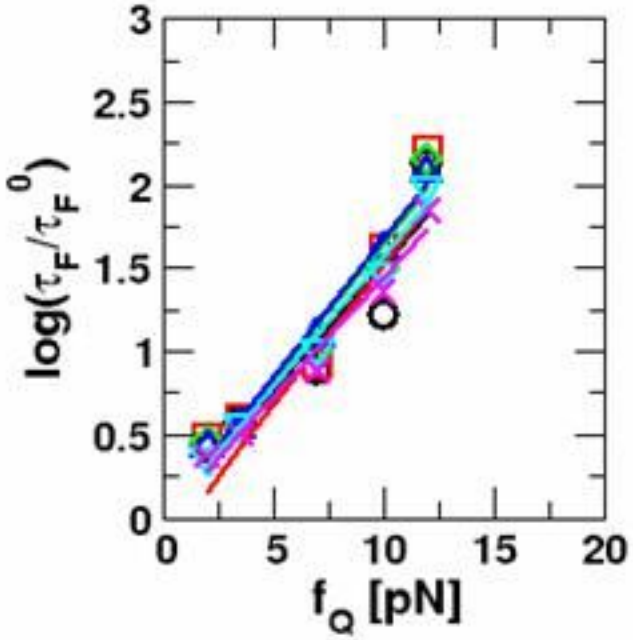
f_C = 14pN



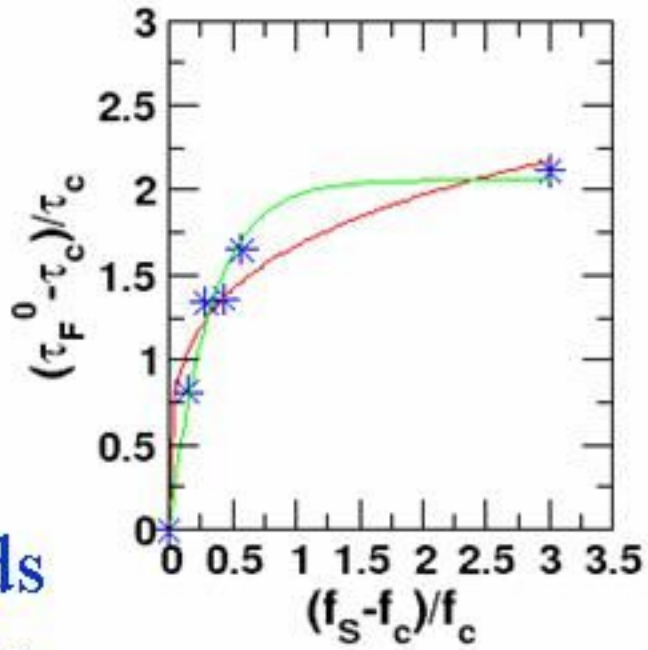
Force quench refolding: Role of f_C

$$\tau_F = \tau_F^0(f_S,f_C)\ \exp[(f_Q\Delta x)/k_BT]$$





Complete Theory
PNAS (2009)



τ_C hopping time at f_C

FQ Protocol
Multiple quench methods
Force Corr Spectroscopy
Barsegov & dt PRL 2006