

BC530 Protein Folding 2012

A. Simple example of folding: Lattice model

B. Folding of larger proteins frequently occurs through populated intermediates

Case study: RNaseH

Methods:

Circular Dichroism

Stopped Flow mixing

Hydrogen Deuterium exchange/NMR

Similar intermediates observed in kinetic and equilibrium experiments

C. Folding of small proteins can occur without detectable intermediates.

Case study: Protein L

Methods:

Fluorescence

Hydrogen Deuterium exchange/ MS

Fitting of thermodynamic and kinetic data to 2-state model

Single molecule measurements

D. Probing protein folding transition states by mutation and kinetic analysis

E. What determines rate of protein folding?

F. Remodeling protein folding pathways by design

Simple example of folding: Lattice model

- 1) useful pedagogically (very concrete)
 - 2) highlight important issues
 - 3) may have some bearing on reality
-

- Model

- two types of residues: O X
 - residues restricted to 2-D square lattice:



- contact energy ϵ between adjacent X residues ($\epsilon < 0$)

Example: Tetramer with sequence X-O-O-X

Chain conformations	Energy	Designation
X--O--O--X	0	"U"
X--O O--X	0	"U"
X--O O X	0	"U"
X--O--O X	0	"U"
X X O--O	ϵ	"N"

$$\Delta G_{fold} = -kT \ln \left(\frac{P_N}{P_D} \right) = \epsilon + kT \ln 4.$$

$$P(N) = \exp (-\epsilon/kT) / Z$$

$$\begin{aligned} Z &= \sum_{\text{states}} \exp (- E_i/kT) = \\ &= 4 \exp (0/kT) + \exp (-\epsilon/kT) \\ &= 4 + \exp (-\epsilon/kT) \end{aligned}$$

Hence

$$P(N) = 1 / (4 \exp (\epsilon/kT) + 1)$$

as $T \rightarrow 0$, $P(N) \rightarrow 1$

as $T \rightarrow \text{infinity}$, $P(N) \rightarrow 1/5$

(all five states equally probable)

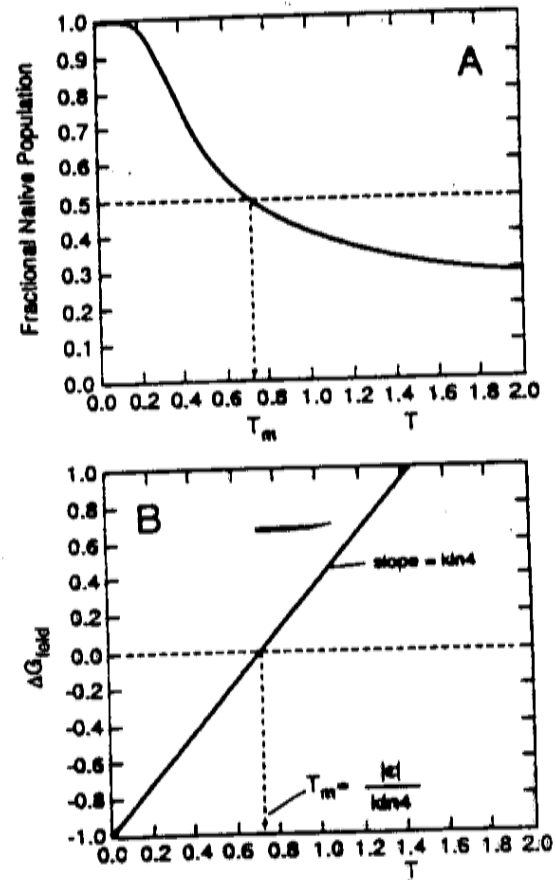
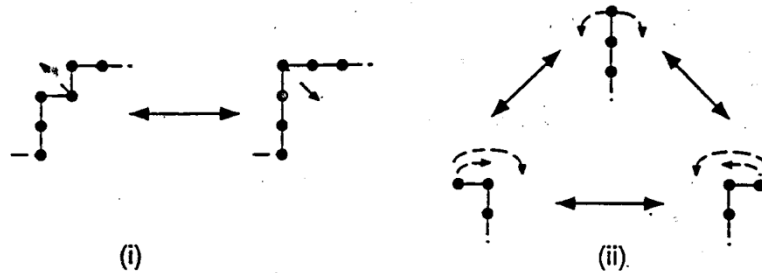


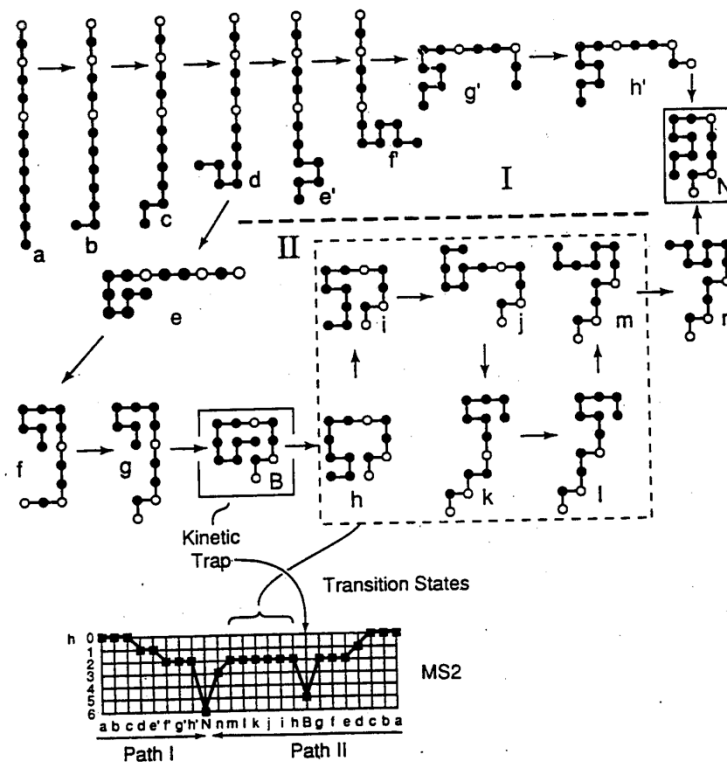
Fig. 19. Denaturation in TTM. Fractional native population (A) and free energy of folding ΔG_{fold} (B). Native state is stable at low temperatures. Protein denatures with increasing temperature. Absolute temperature T is in units of $|\epsilon|/k$, where k is Boltzmann's constant. ΔG_{fold} is in units of $|\epsilon|$. T_m is the mid-point temperature at which half of the chain population is native.

Folding Kinetics in a Lattice Model

First, we need a move set:



Given a move set, we can explore possible folding pathways:



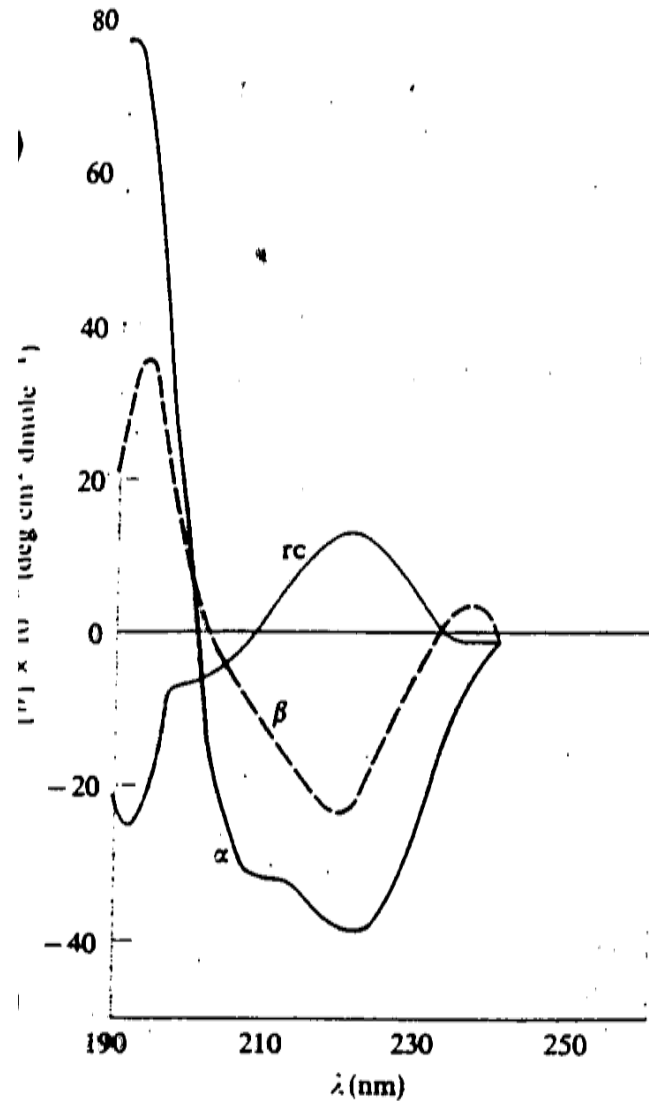
Folding of larger proteins

Case Study:Folding of Rnase H

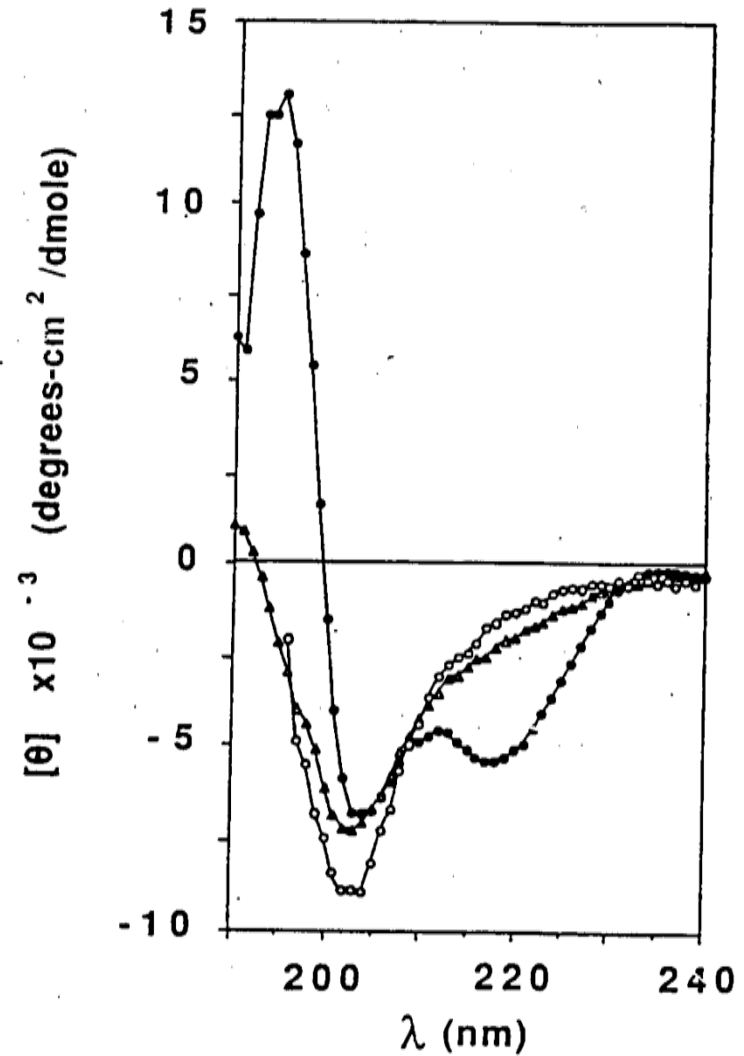
Different experimental methods may be used to monitor folding:

- I. Circular dichroism
 - Differential absorbance of circularly polarized light
 - $CD(L) = \epsilon_L - \epsilon_R$
 - Periodic secondary structure (alpha helices, beta sheets) differentially absorb circularly polarized light and thus have strong CD
- II. Stopped flow mixing
 - Most proteins fold on the ms-second time scale. To initiate folding, rapidly mix denatured protein (in guanidine or urea solution) with buffer lacking denaturant
- III. Hydrogen Deuterium exchange
 - Solvent exposed amide protons can exchange with deuterons
 - Sites of exchange can be identified by NMR or MS.

Circular Dichroism: Probe of Secondary Structure



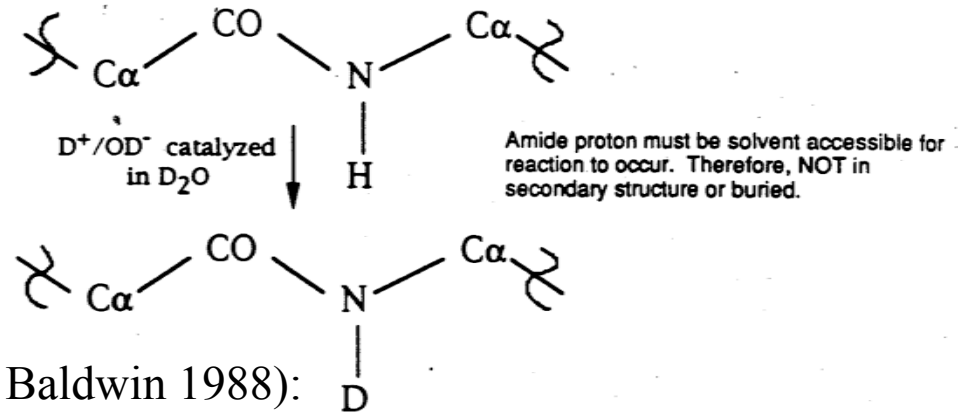
CD spectra : peptide models
of secondary structure elements



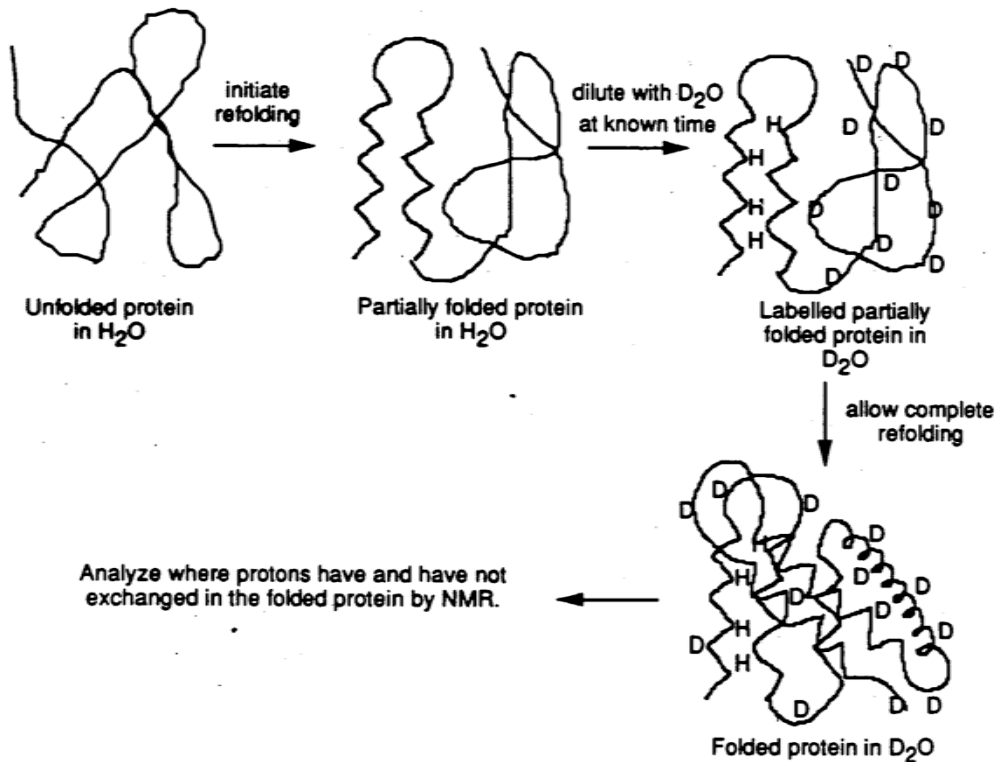
CD spectra for typical
proteins

Hydrogen Deuterium exchange

To identify which portion of the molecule is structured in the intermediate, use HD exchange



The Amide Exchange Experiment (Roder and Baldwin 1988):



Folding of RNase H (142 residues)

- I. At very low pH (< 2), a partially folded "molten globule" like state is populated. CD analysis of this state shows substantial secondary structure (Biochemistry 35, 11951-11958).
- II. At equilibrium at neutral pH, similar partially folded states are populated at very low levels. These states are detectable using native state HD exchange (NSB 3, 782-787)
 - Are these partially formed states populated during refolding?
 - Stopped flow CD shows that an intermediate with substantial secondary structure is formed rapidly after the initiation of refolding:

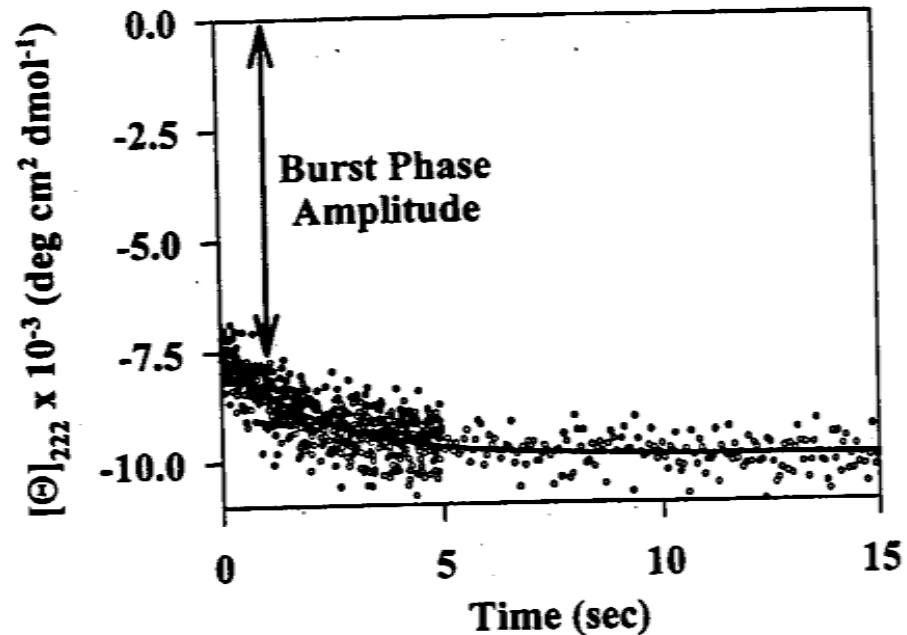


Fig. 1 The refolding of RNase H* shows biphasic kinetics. Refolding was initiated by 11-fold dilution from 7 M urea in the stopped-flow CD and monitored at 222 nm (dead time 12 ms). The line represents the global fit of the data to a three state sequential model (see Methods). The ellipticities of the unfolded, intermediate and native states, as determined from the fit, are 0, -7650 and -10,000 deg cm² dmol⁻¹ respectively.

Fig. 4 Helices A and D and strand 4 are well protected in the kinetic intermediate. *a*, Fractional proton occupancy, determined by pulse labelling hydrogen exchange, for probes representative of each secondary structure element are shown. Colours reflect the protection factors of the amide protons in the kinetic intermediate as described in Fig. 5a. Red circles, strand 1 (F8); red squares, strand 2 (I25); red triangles, strand 3 (Y39); blue circles, helix A (V54); blue diamonds, strand 4 (I66); blue squares, helix D (W104); red inverted triangles, strand 5 (E119); green hexagons, helix E (A140). Dashed line, slow refolding rate from SFC data ($k_{app}=0.45\text{ s}^{-1}$). *b*, The dependence of the fractional proton occupancy on the pulse pH was used to measure protection factors of the well protected amide protons in the kinetic intermediate. The lines represent fits of the data to Eq. 6 (see Methods). Circles and solid line, L56; squares and dashed line, I66; triangles and dotted line, A110.

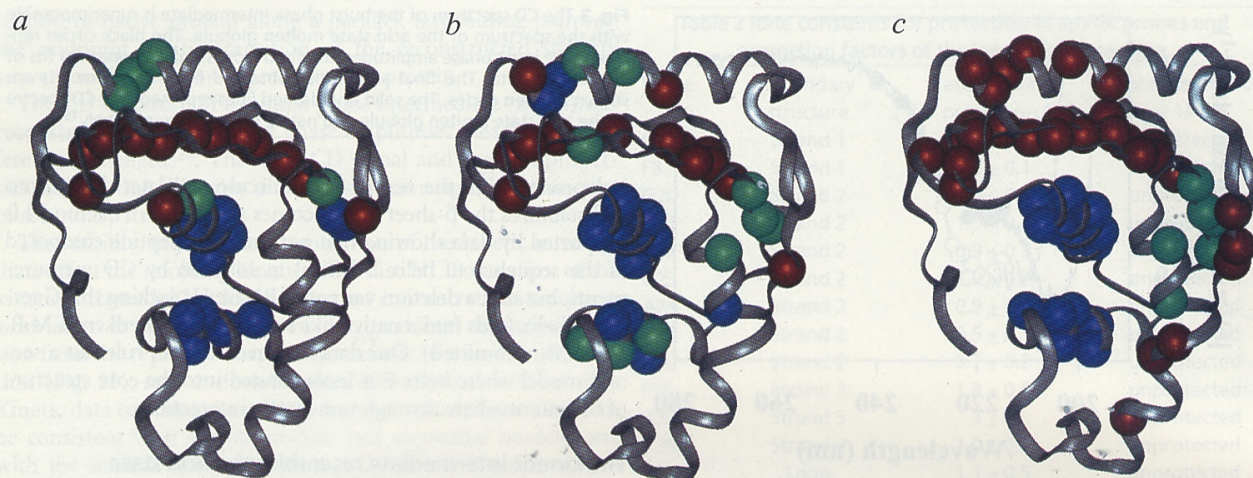
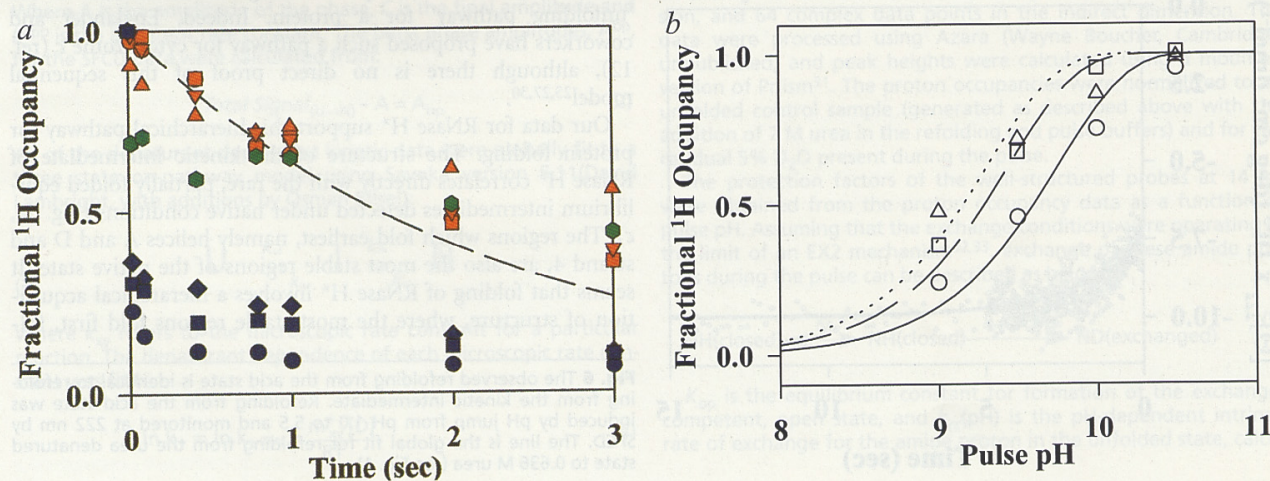


Fig. 5 Comparison of the structures of the kinetic intermediate, the acid state molten globule, and the native state stability of RNase H*. *a*, The protection factors of the kinetic intermediate (14 ms refolding time). Unprotected probes are coloured red. Slightly protected probes are green, and well protected probes for which a protection factor could be measured are blue. *b*, The protection factors of the acid state molten globule 19 . Red, unstructured probes ($P < 5$). Green, probes with intermediate protection ($5 < P < 10$). Blue, well-protected probes ($P > 10$). *c*, Stability of the native state 13 . Red, regions that unfold with an average free energy of 7.4 kcal mol $^{-1}$. Green, regions that unfold with an average free energy of 8.7 kcal mol $^{-1}$. Blue, regions that unfold with an average free energy of 10 kcal mol $^{-1}$. All data are mapped onto the crystal structure of wild type RNase H 14 .

The Hierarchical Formation of Structure ...

- A striking result is that the same part of the protein is ordered
 - 1) in the kinetic folding intermediate
 - 2) in a well populated equilibrium partially folded state at very low pH
 - 3) in very rare partially folded species in equilibrium with the native state at neutral pH
- *Conclusion:* hierarchical formation of structure; regions that are most stable form first.

Folding of small protein

Case Study:Folding of Protein L

Methods:

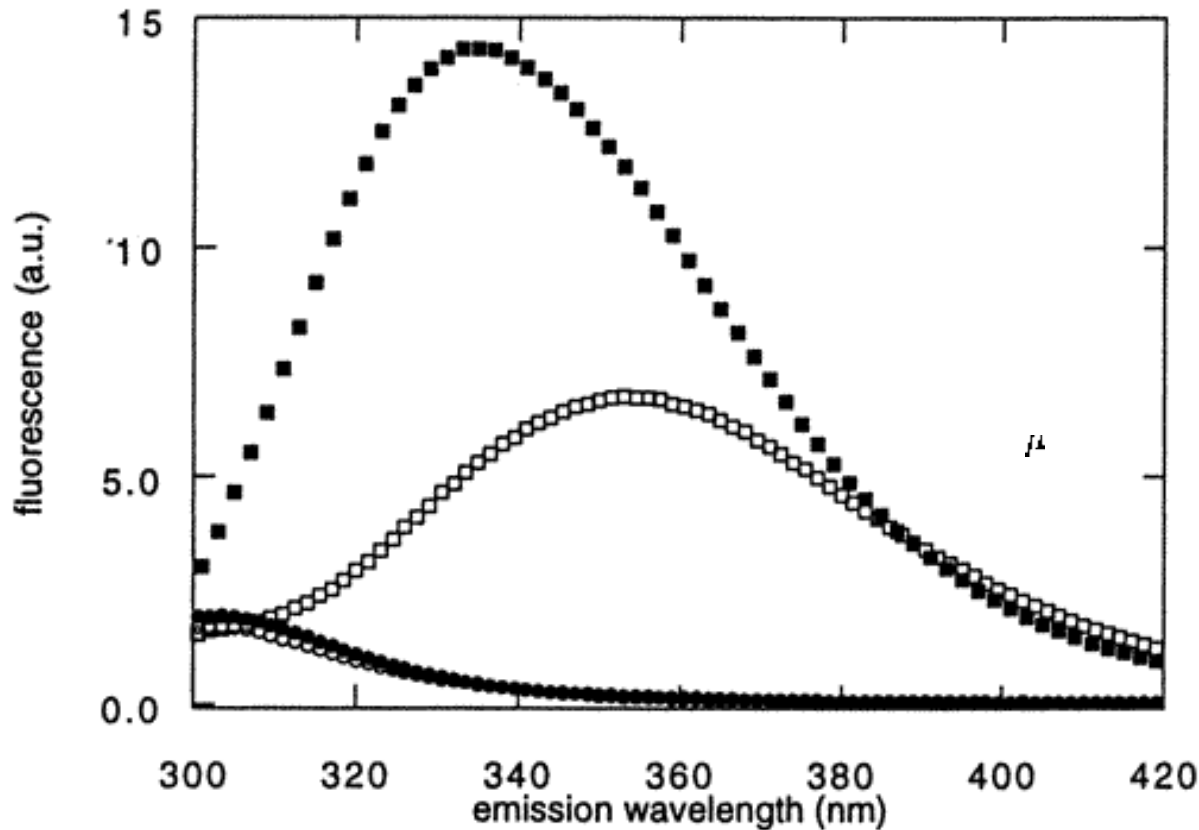
– I. Fluorescence

- Light at one wavelength absorbed, and light at a longer wavelength emitted
- The quantum yield (intensity) and the wavelength of the emitted light are sensitive to the environment.
- The intrinsic fluorescence of tryptophan is particularly useful

– II. HD exchange/Mass Spectrometry

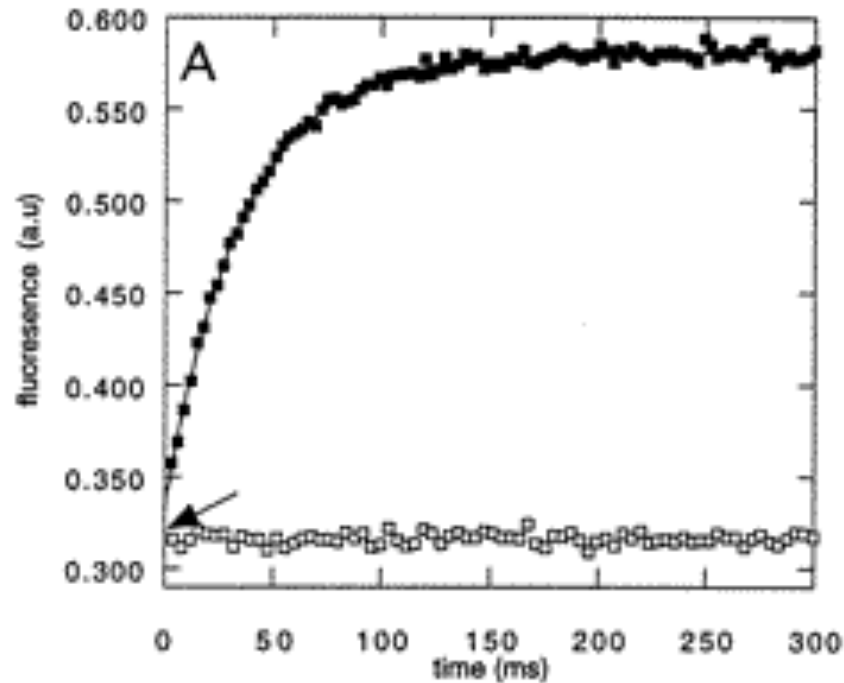
- Partially folded species with a subset of protons protected from exchange will appear as separate peak in Mass Spectrum
- Deuterate all amide protons in denatured protein, rapidly remove denaturant and expose to H₂O after time Δt
- Species present:
 - DDDDDDDDDDD (complete folding within Δt)
 - HHHHHHHHHHH (not folded within Δt)
 - DDHDDHHDHD (partially folded species formed within Δt)

Intrinsic Tryptophan Fluorescence can be used to Monitor Protein Folding (Protein L)



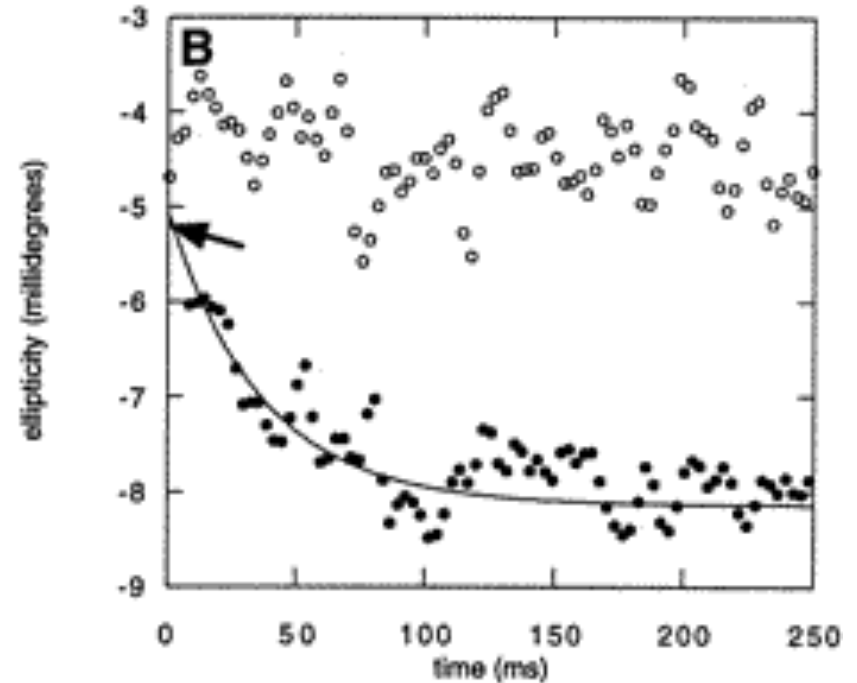
The fluorescence of a tryptophan in protein L is much greater in the folded state (black squares) than in the unfolded state (open squares). Circles are for protein L lacking the tryptophan.

Protein Folding Rates can be Measured by Stopped-Flow Fluorescence and CD



Fluorescence:

- refolded conditions (closed symbols)
- unfolded conditions (open symbols)



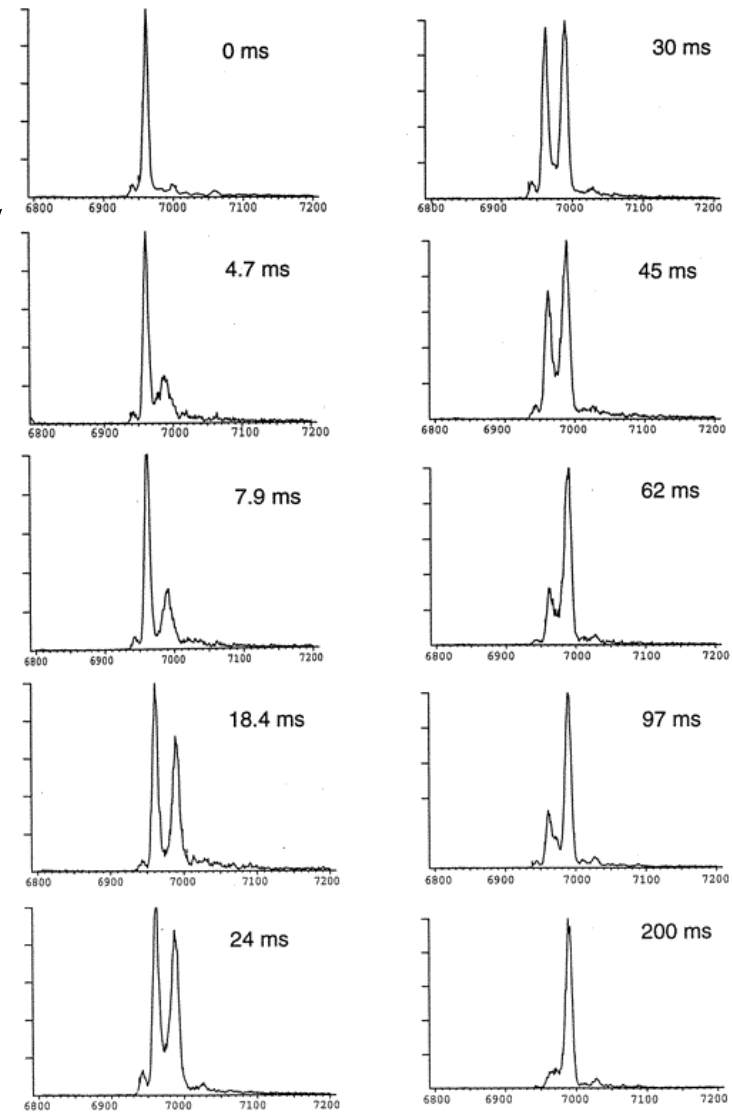
CD:

- refolded conditions (closed symbols)
- unfolded conditions (open symbols)

HD Exchange Experiment Coupled to Mass-Spec (Protein L)

Refolding monitored by quenched flow HD exchange coupled to mass spectrometry

Unfolded conditions



After complete refolding

Two State Model

- Only two states are significantly populated during folding (U and N)
- Changes in temperature or [denaturant] only affect relative abundance of U and N.

[denaturant]	Energy		Population
	U__	N__	
0M			99% N
2M	U__	N__	50% N, 50% U
4M		N__	99%U
	U__		

- U is unstable in 0M denaturant because of exposed hydrophobic groups; increasing [denaturant] reduces strength of hydrophobic effect.

Studying Protein Folding Using Single Molecule Measurement

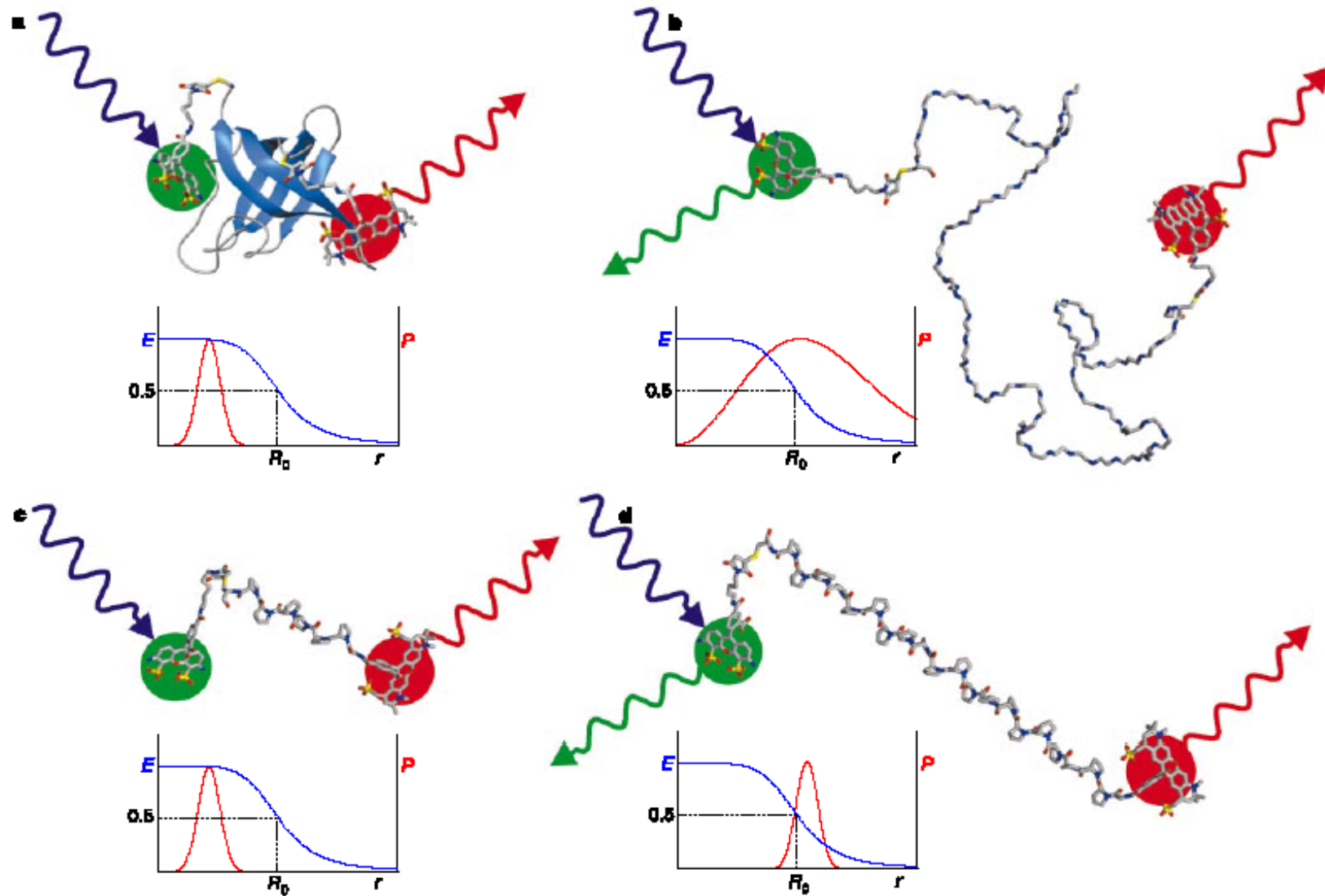


Figure 1 Schematic structures of protein and polyproline helices labelled with donor (Alexa 488) and acceptor (Alexa 594) dyes (using the program MOL/MOL). **a**, Folded Csp 7m, a 66-residue, five-stranded β -barrel protein (Protein Data Bank accession code 1G6P)³¹; **b**, unfolded Csp 7m; **c**, (Pro)₆; and **d**, (Pro)₂₀. A blue laser excites the green-emitting donor dye, which can transfer excitation energy to the red-emitting acceptor dye

at a rate that depends on the inverse sixth power of the inter-dye distance^{1,8}. In each case, the functional form of the FRET efficiency E versus distance (blue curve) is shown, as well as a representation of the probability distribution of distances between donor and acceptor dyes, P (red curve).

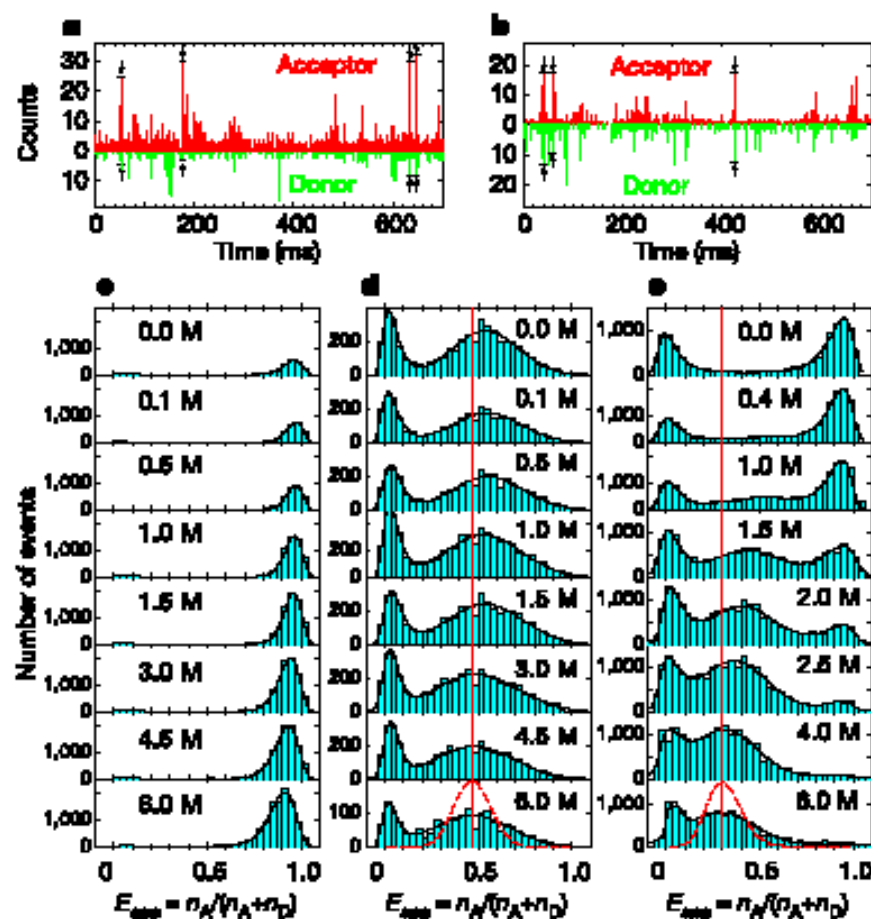


Figure 2 FRET trajectories and histograms. **a, b**, Donor (green) and acceptor (red) channel time traces using 1-ms bins for labelled (Pro)₆ (**a**) and (Pro)₂₀ (**b**). Arrows indicate photon bursts for which the sum of the counts in the two channels is greater than 25. **c–e**, Histograms of measured FRET efficiencies (E_{app}) at various GdmCl concentrations for labelled (Pro)₆ (**c**), (Pro)₂₀ (**d**) and CspTm (**e**). The black curves are the best fits to the data using lognormal and/or gaussian functions. The red dashed curves were calculated from the β -distribution¹⁵ $P(E_{app}) = E_{app}^{\langle n_A \rangle} (1 - E_{app})^{\langle n_D \rangle}$, where $\langle n_A \rangle$ and $\langle n_D \rangle$ are the average number of detected acceptor and donor photons in the significant bursts.

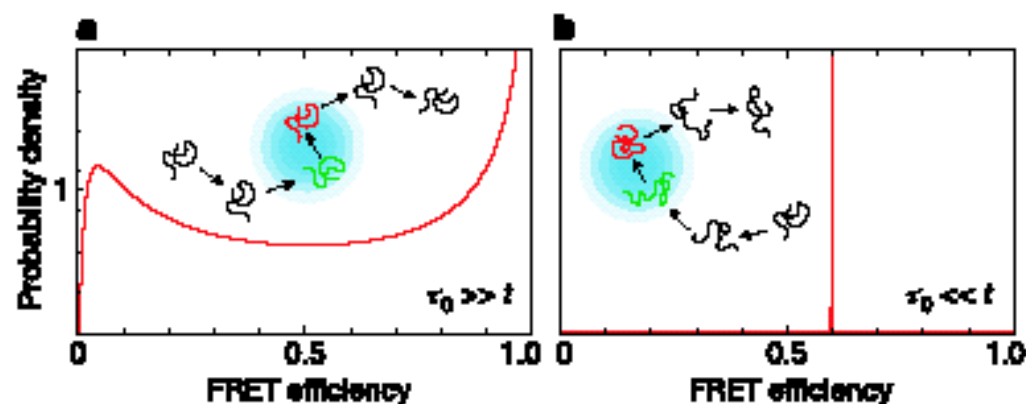


Figure 4 Two limiting cases for polypeptide dynamics in experiments on freely diffusing molecules. **a**, If the end-to-end distance of the protein does not change during the observation period in the illuminated volume (blue), then the distribution of transfer efficiencies reflects the equilibrium end-to-end distance distribution of the molecules and consequently results in a very broad probability distribution of FRET efficiencies, shown here for a gaussian chain (red curve, see Supplementary Information). **b**, If the molecule reconfigures fast relative to the time it takes to diffuse through the illuminated volume, then the FRET efficiency averages completely and (in the absence of shot noise and other broadening effects) is the same for every molecule.

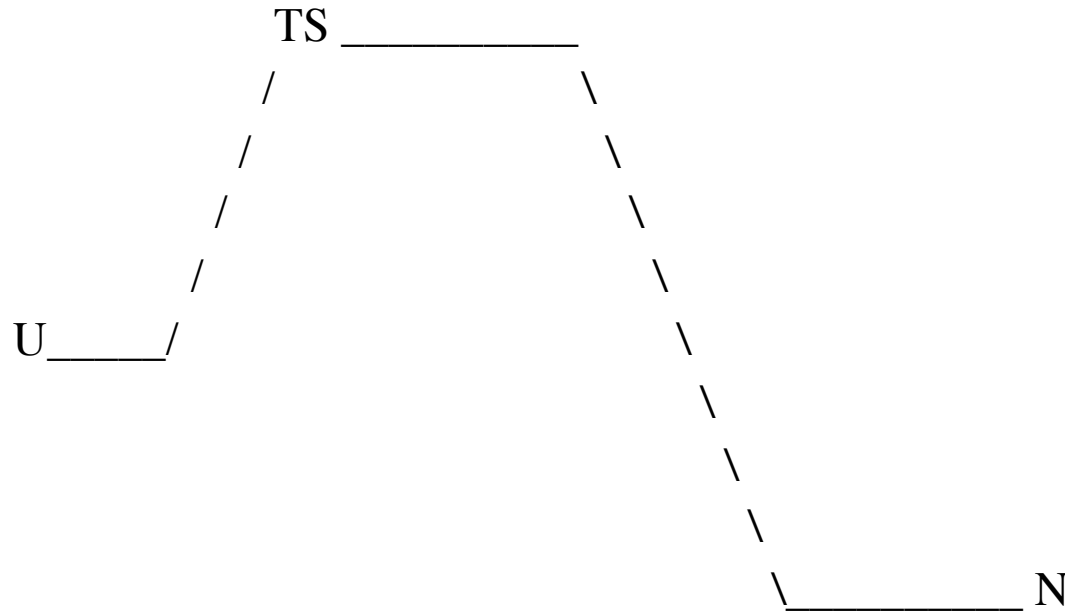
III. Is there a hierarchy to structure formation in the folding of small ($< 100\text{aa}$) proteins?

Method:

- Kinetic analysis of side chain truncation mutants using simple transition state theory (pioneered by Alan Fersht's group)
- Determine the fraction of the interactions made by each residue in the native state that are also made in the folding transition state
- Experimentally, determine the ratio $\Delta\Delta G_{\text{U-TS}} / \Delta\Delta G_{\text{U-N}}$. This is often called the Φ value in the literature.

TS theory for protein folding

- Simple transition state theory is often used to treat protein folding



- rate of folding/unfolding proportional to fraction of population at the transition state

$$k_{\text{fold}} \sim e^{-\Delta G_{\text{u-ts}}/RT}$$

$$k_{\text{unfold}} \sim e^{-\Delta G_{\text{n-ts}}/RT}$$

$$K = e^{-\Delta G_{\text{u-n}}/RT}$$

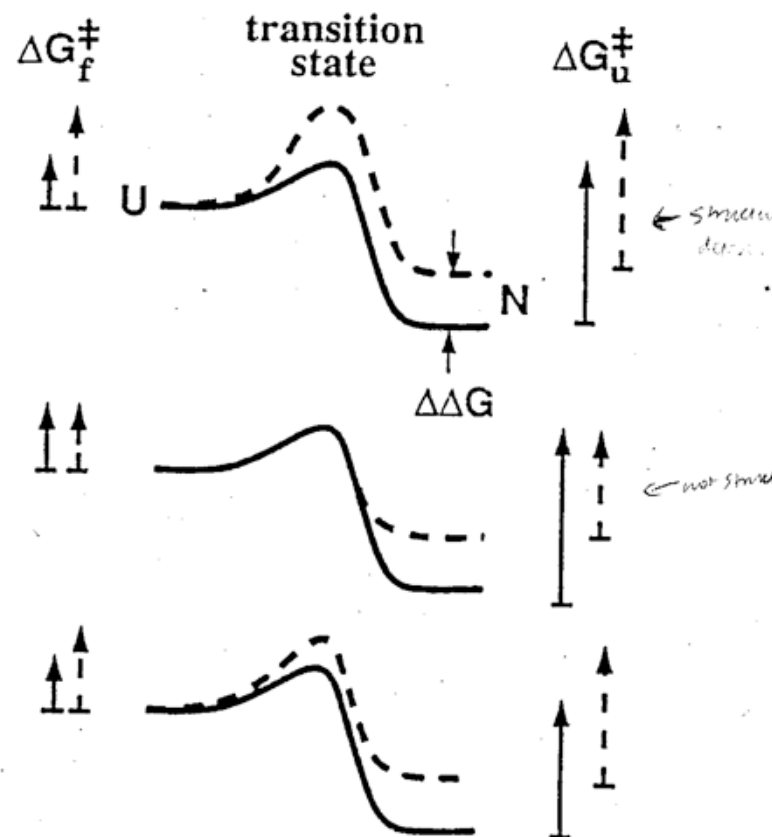
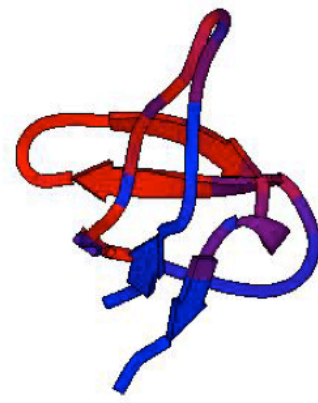
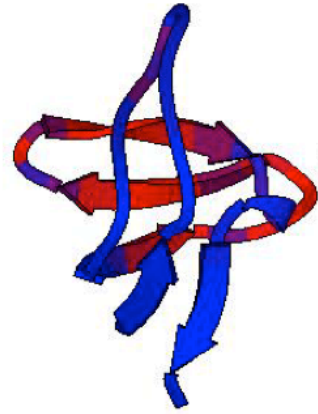
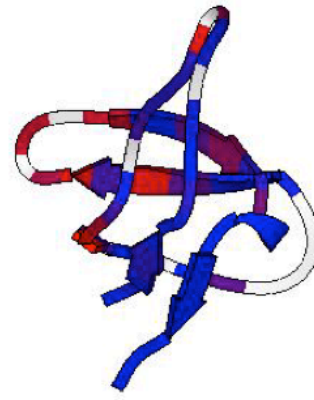


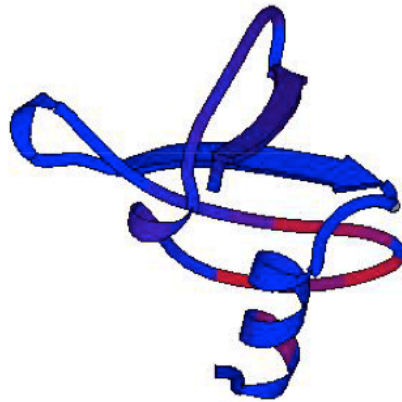
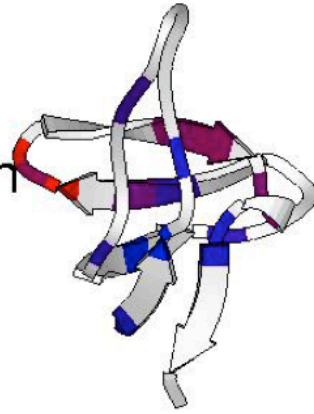
Fig. 1. Free energy diagrams illustrating the possible effects of altered helix stability on folding and unfolding rates. Activation energies for folding and unfolding of the wild-type sequence (solid lines) and variant (dashed lines) are indicated by the height of the arrows. A decrease in equilibrium stability (helical propensity) may be expressed kinetically as a decrease in folding rate (top, $\Phi_t = 1$), an increase in unfolding rate (middle, $\Phi_t = 0$), or a combination of both (bottom, $0 < \Phi_t < 1$) depending on whether helix is present, absent, or partially formed in the transition state at the site of the substitution. The energetic contribution of helical structure to the transition state can be quantified by the Φ_t parameter (Eq. 4) given by $\Delta\Delta G_t^\ddagger$, the difference in height of the folding-side arrows, divided by $\Delta\Delta G^\circ$, the change in equilibrium stability.



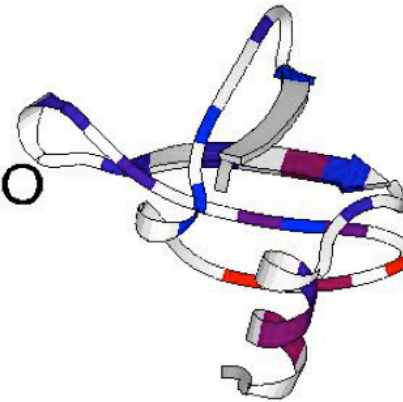
Src



Spectrin



SSO



IV. Investigating protein folding using simple models and wealth of published data on folding of small proteins

Two-State Folding Proteins:

Cytochrome B ₅₆₂	160 000 s ⁻¹	Gray and Winkler, Pers Com
Cytochrome-C	6 350 s ⁻¹	Mines <i>et al.</i> , 1996
λ-repressor	3 640 s ⁻¹	Huang and Oas, 1995
Ubiquitin	1 530 s ⁻¹	Khorasanizadeh <i>et al.</i> , 1995
CspB	1 075 s ⁻¹	Schindler and Schmid, 1996
ADA2h	890 s ⁻¹	Villegas <i>et al.</i> , 1995
Amino Domain of L-9	720 s ⁻¹	Khulna and Raleigh, Pers Com
ACBP	690 s ⁻¹	Kraglund <i>et al.</i> , 1996
Protein G	490 s ⁻¹	Park <i>et al.</i> , 1997
¹⁰ FNIII	155 s ⁻¹	Plaxco <i>et al.</i> , 1997
Fyn SH3	95 s ⁻¹	Plaxco <i>et al.</i> , In Press
Protein L	66 s ⁻¹	Scalley <i>et al.</i> , 1997
CI-2	48 s ⁻¹	Jackson and Fersht, 1991
HPr	15 s ⁻¹	van Nuland <i>et al.</i> , 1998
Acyl Phosphatase	0.2 s ⁻¹	F. Chiti, Pers Com

Simple Model for Folding

Free energy landscape dominated by tradeoff between formation of attractive native interactions and loss of chain configurational entropy (non-native interactions ignored)

All contacts equal in energy

Free energy of ordering segments of same length equal

Contact order: average sequence separation between residues in contact in native structure

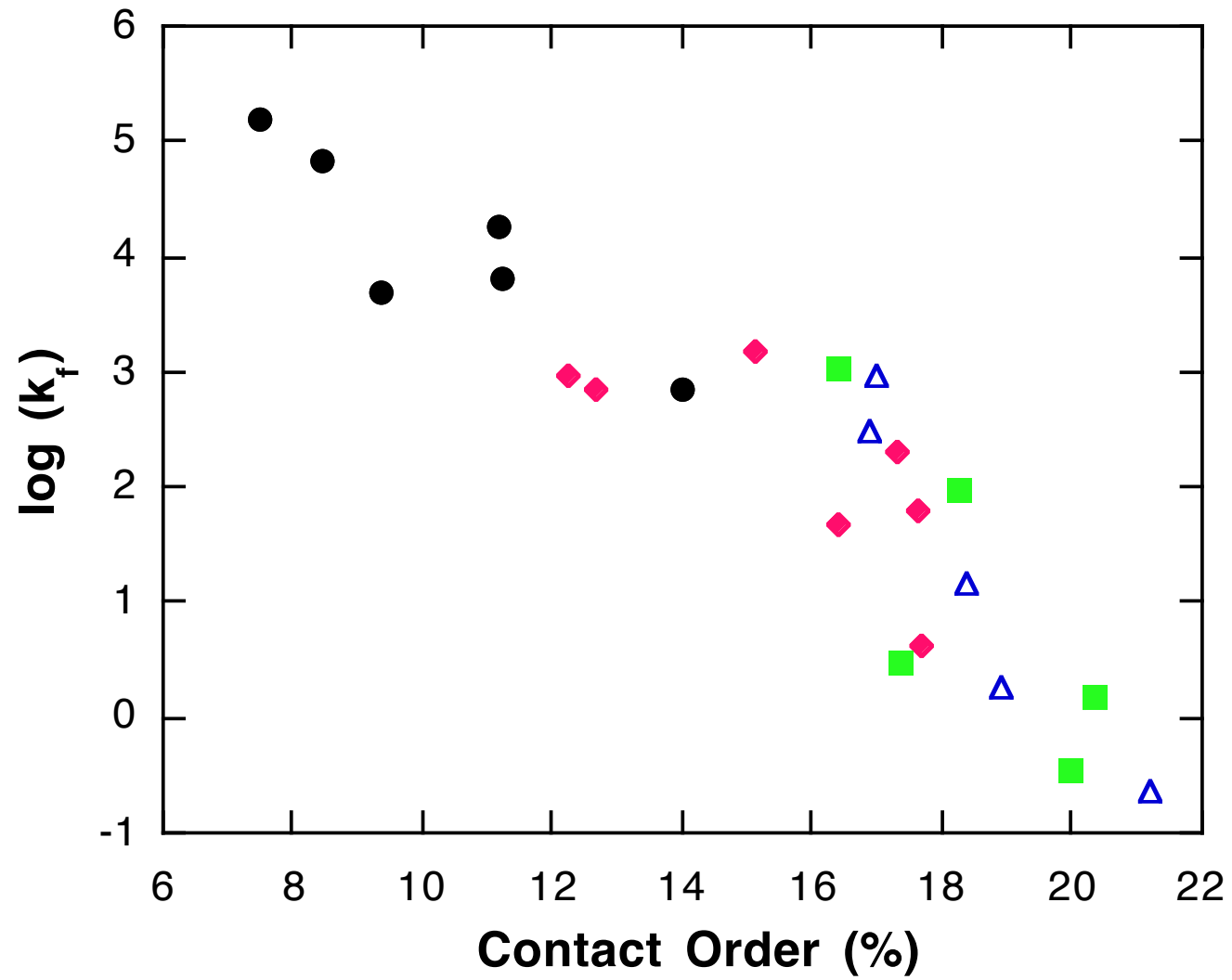
LOW CONTACT ORDER



HIGH CONTACT ORDER

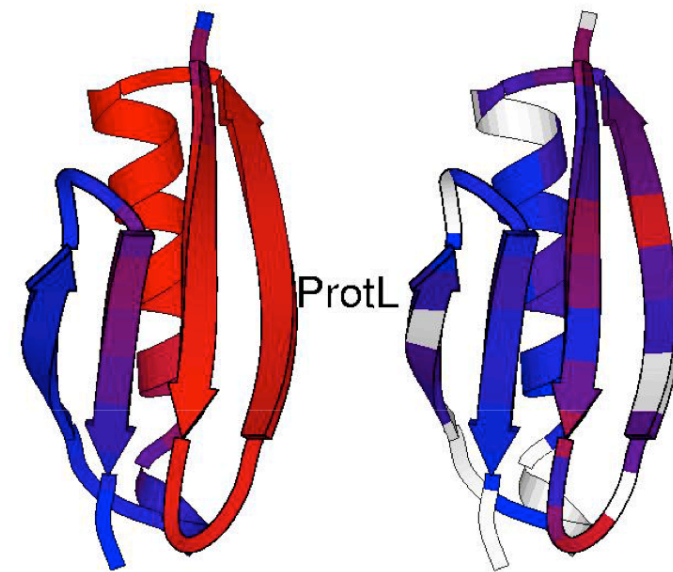
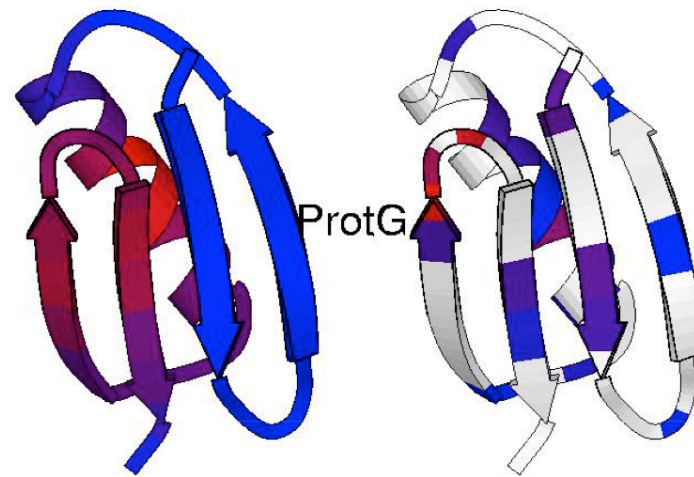


Proteins with low contact order fold faster

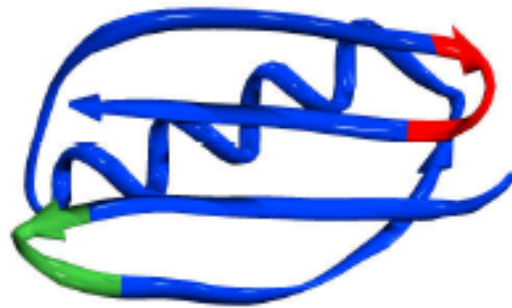


V: Computational remodeling of protein folding pathways

- Topology appears to be major determinant of protein folding mechanisms
- However, in proteins with symmetry, factors such as the relative stabilities of local structural elements can have large roles
- Can folding pathways be altered by modulating stabilities of local structural elements?

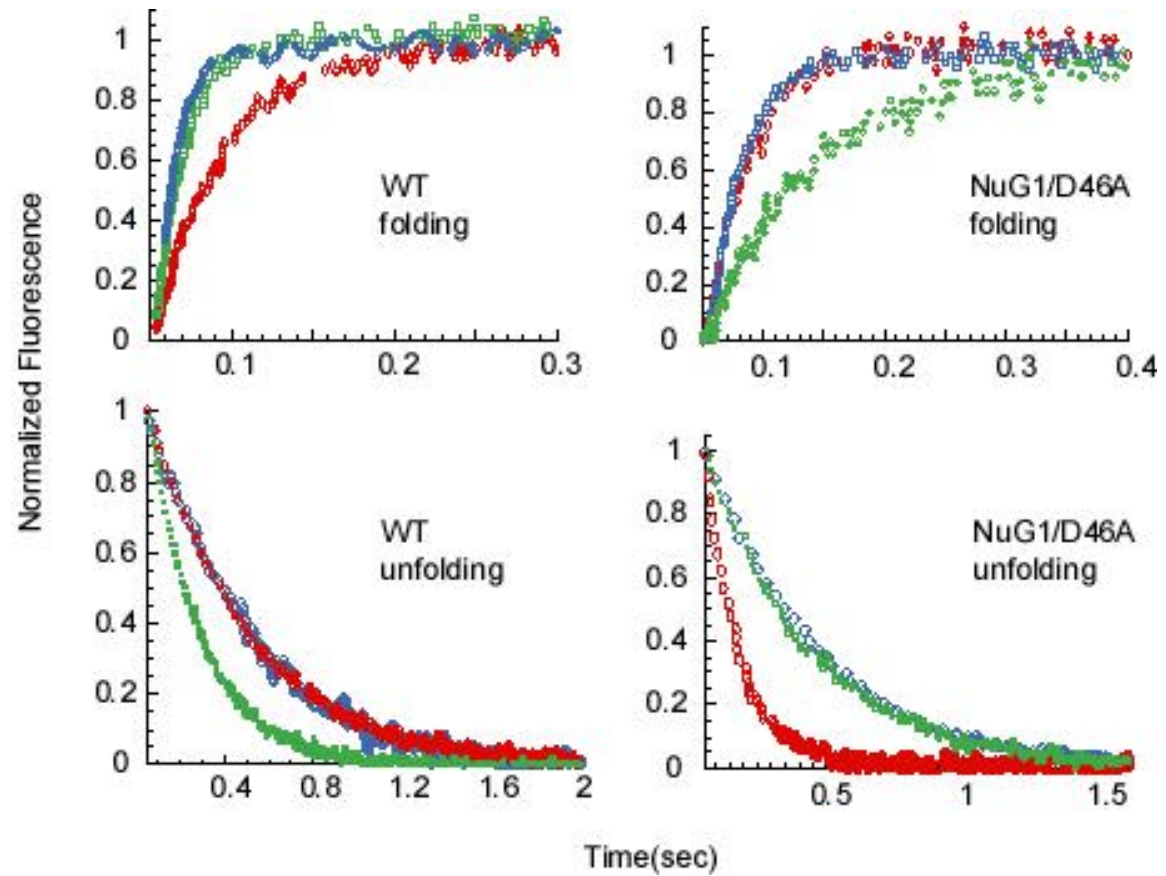


Switched folding mechanism of protein G



2nd hairpin folds first

1st hairpin folds first



References

There is a huge body of literature on folding....

In the books recommended for the class:

Creighton, Ch 7, paragraph 7.5 covers folding kinetics

Fersht, Ch 18 and 19 have a more thorough description