

DEVELOPING A GENERAL DESCRIPTION OF THE UNFOLDED STATE OF A PROTEIN

By

Yujie Chen

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Physics

2011

ABSTRACT

DEVELOPING A GENERAL DESCRIPTION OF THE UNFOLDED STATE OF A PROTEIN

By

Yujie Chen

This thesis focuses on both the dynamical and structural aspects of the unfolded states of a protein and develops a general description of these unfolded states over all denaturing conditions. Investigating the unfolded states under physiological conditions can provide insights into the entire folding pathway as they define the most likely conformations before the folding transition. Also, since the initial motion of a protein in solvent early in the folding process is driven by diffusion, the intramolecular diffusion rate may set an upper speed limit for protein folding.

In this thesis, the experimental measurement of the intramolecular diffusion is achieved by a pump-probe spectroscopy using the technique of Trp/Cys contact quenching. Three typical proteins in different folding categories have been studied: the intrinsically disordered apocytochrome C, the aggregation-prone HypF-N and the well-folded ACBP (Acyl-coenzyme A-binding Protein). For ACBP, I also used a microfluidic mixer coupled with the pump-probe spectroscopy to study the intramolecular loop contact formation in its refolding process.

In order to extract the effective diffusion coefficient from the observed contact rates, I employed a polymer theory by Szabo, Schulten and Schulten (SSS) which requires a probability distribution of equilibrium Trp/Cys distances. In an attempt to capture the transient residual structures and intramolecular interactions of unfolded polypeptides under folding conditions, I integrated sequence specificities into a traditional non-overlapping worm-like chain model. The

new model can statistically re-weight the distribution to favor those conformations with more hydrophobic contacts and it yields quite convergent results over many tested sequences. In the thesis, I will also show that this model quantitatively predicted paramagnetic resonance enhancement (PRE) measurements of ACBP and DrkN with no adjustable parameters.

After applying the SSS theory with the Trp/Cys distance distributions produced by the energy re-weighted worm-like chain model to calculating the intramolecular diffusion coefficients of those experimentally studied proteins, I found an intrinsic relationship between the intramolecular diffusion rates and the aggregation propensities and I propose a dynamic range of conformational reorganization within which partially or fully unfolded states are prone to aggregation.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Lisa Lapidus, for her guidance and support through my graduate years. She has been a mentor, a friend, and a great inspiration to my researches. She always does her best to provide me opportunities to explore the field of experimental biophysics. Because of her trust and unselfishness, I was very thankful and honored having been able to work on several promising projects and to work with so many talented colleagues and collaborators. She has also given me enough freedom to be adventurous and creative, which led me to another level as an independent researcher. On a personal level, I'm especially thankful for her kindness and genuine friendship. It has truly been a blessing working with her for the past five years.

I would like to thank Dr. William Wedemeyer for his assistance in my computational projects. His intelligence and broad knowledge in molecular biology were great inspirations to me.

I would also like to thank Terry Ball, who made almost all the protein samples for my projects. Without him, none of these works would be possible.

Thanks to all the group members: Vijay Singh, Steven Waldauer, Stephen Decamp, Michael King, Eric Buscarino, Li Zhu, and Basir Ahmad. It's been a wonderful experience working with you all. Thank you for your friendship and great memories.

Finally, I would like to thank my parents, who have always been there supporting me with every decision I made. Special thanks to my lovely wife, who always give me honest yet constructive opinions. Her love and support was the driving force behind my accomplishments.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
1 Introduction to Protein Folding, Misfolding and Aggregation	1
1.1 Amino Acids, Peptides and Proteins	1
1.2 Protein Folding: Mechanics, Energy and Time Scale	2
1.3 Introduction to Experimental Approaches	9
1.4 The Importance of Studying the Unfolded States	11
1.5 Polymer Models and Computational Methods	12
2 Spectroscopic Study of the Dynamics of Unfolded Proteins.....	16
2.1 Introduction.....	16
2.2 Monitoring Tryptophan Triplet Quenching Using Pump-probe Spectroscopy	18
2.3 Reaction-limited and Diffusion-limited Rates	24
2.4 Data Analysis Using SSS theory and a Wormlike Chain Model	26
2.5 Tryptophan Triplet Quenching Experiments in a Microfluidic Mixer.....	28
3 Experiments on Several Different Proteins Using Trp/Cys Quenching Spectroscopy	33
3.1 Conformational properties of unfolded HypF.....	33
3.2 Intramolecular Loop Formation of Intrinsically Unfolded Proteins.....	45
3.3 Diffusion of Unfolded Acyl-coenzyme A-binding Protein over a Complete Range of Denaturant.....	53
4 Sequence Specific Wormlike Chain Model	58
4.1 Introduction to the Heteropolymer Model	58
4.2 Boltzmann Energy Reweighting.....	61
4.3 Secondary Structure Propensity (SSP).....	66
4.4 Convergence of σ	68
4.5 Intrachain Contacts and Hydrophobic Energy	71
4.6 Comparison between Two Protein L Mutants.....	76
4.7 PRE Prediction.....	79
4.8 Charge Effects.....	82
4.9 Conclusion	86
5 Conclusions.....	89

BIBLIOGRAPHY	98
--------------------	----

LIST OF TABLES

Table 3.1 Observed rates of tryptophan triplet decay of W81F at various times after dilution in 3% TFE	45
Table 3.2 Wormlike chain parameters and effective diffusion coefficients for both mutants studied	45
Table 3.3 Wormlike chain parameters and effective diffusion coefficients for apocytochrome C47	47
Table 4.1 Energy-reweighted wormlike chain parameters and effective diffusion coefficients for drkN	68
Table 4.2 Energy reweighted worm-like chain parameters and effective diffusion coefficients for AavLEA without charge	83
Table 4.3 reweighting with both σ and γ for AavLEA T58C	83
Table 4.4 reweighting with only γ for AavLEA S2C	85
Table 5.1 Several protein sequences studied in our lab	90

LIST OF FIGURES

Figure 1.1 Two amino acids form a peptide bond between the carboxyl group and the amino group. (http://upload.wikimedia.org/wikipedia/commons/6/6d/Peptidformationball.svg)	2
Figure 1.2 Twenty standard amino acid classified by their polarity (hydrophobicity)	3
Figure 1.3 cartoon of the HypF-N protein showing (1) α -helix (red), (2) β -strand (blue), (3) tryptophan 27 and 81 (purple) and (4) cysteine 65 (yellow)	5
Figure 1.4 Time span of protein folding events	9
Figure 1.5 an energy landscape for protein folding created by HP model.....	14
Figure 2.1 The fraction of unfolded molecules (circles) of the T57C mutant of protein L as measured by various spectroscopic probes: fluorescence intensity, wavelength of peak fluorescence, relative amplitude of intramolecular contact formation (fraction of fast rate), and the magnitude of the unquenched tryptophan triplet lifetime (slow rate). The last two probes will be described in detail in the following sections. The lines are fits to two-state models.....	17
Figure 2.2 Schematic electronic energy levels of tryptophan. 289nm UV laser is used to excite the tryptophan to its singlet states. The intersystem cross from S_1 to T_1 is achieved through the reversal of the spin of an excited electron. Triplet states T_1 absorb blue light (442nm) and have a long lifetime	19
Figure 2.3 Cartoon of loop formation and quenching. A tryptophan residue (W) is first excited to the triplet states and then along with a cysteine residue (C) at the other end diffuses towards and away from each other at rates k_{D+} and k_{D-} respectively. q is the quenching rate	20
Figure 2.4 Schematic of the tryptophan triplet quenching experiment.....	23
Figure 2.5 Geometry of the chaotic advection mixing microfluidic-chip. (A) Buffer inlet. (B) Protein solution inlet. (C) Filter section (one of three total per chip). Each filter post is a 10 by 10 μm square and they are spaced 10 μm apart. (D). The serpentine mixing region is composed of five turns and the channel is 30 μm wide. (E) The narrow observation region is 120 μm wide and 1.4 mm long. (F) The wide observation region is 1000 μm wide and 10.4 mm long. (G) Outlet	

port	31
Figure 2.6 Schematic of the tryptophan triplet quenching experiment coupled with a microfluidic mixer	32
Figure 3.1 Kinetics of tryptophan triplet lifetime for the HypF mutant W81F(long loop) at various concentrations of GdnHCl. The fast decays ($\sim 10^5 \text{ s}^{-1}$) are due to intramolecular contact quenching by C65 and the slow decays ($\sim 10^4 \text{ s}^{-1}$) are due to the natural lifetime of the triplet in an aqueous environment	36
Figure 3.2 a) Observed exponential rates of the tryptophan triplet decay versus [GdnHCl] for the two mutants, W81F (white) and W27F (black). The triangles are the slow rate and the circles are the fast rate as illustrated in Figure 3.1. b) The relative amplitude of the fast rate vs. [GdnHCl] for W27F (black circles) and W81F (white triangles). The grey line is a sigmoidal fit to all data for both mutants	37
Figure 3.3 The observed fast rates vs. viscosity for W27F (a,c,e) and W81F (b,d,f) in 6 M (a,b), 2.3 M (c,d) and 1.25 M (e,f) GdnHCl. Each temperature is plotted as a separate color. The lines are fits of all the data in each plot to Eqs. 2.5 and 2.6.....	38
Figure 3.4 Reaction-limited (black circles) and diffusion-limited (white triangles) rates vs. [GdnHCl] for a) W27F (short loop) and b) W81F (long loop) at 20 C and 1 cP. The points at 0 M GdnHCl are those measured in 3% TFE and determined from the slope and intercept of Figure 3.5b.....	40
Figure 3.5 a) Tryptophan triplet decay of W81F in 3% TFE, T=20 C at various times after dilution. b) The observed fast rates vs. viscosity for W81F in 3% TFE for T= 0, 10, 20, 30, and 40 C ~6 minutes after dilution	42
Figure 3.6 Reaction-limited (red triangles) and diffusion-limited (blue triangles) rates vs. [GdnHCl] for apocytochrome C	47
Figure 3.7 Structural cartoon of F_{exch} and U_{exch} of DrkN.....	48
Figure 3.8 Kinetics of tryptophan triplet lifetime for the drkN mutant C60 at various concentrations of GdnHCl	49
Figure 3.9 The observed fast rates vs. viscosity for C2 (c, d) and C60 (a, b) in 6 M (a, c), 2 M (b,	

d) GdnHCl. Each temperature is plotted as a separate color. The lines in (a) and (c) are fits of all the data to Eqs. 2.5 and 3.1.	50
Figure 3.10 Reaction-limited (black) and diffusion-limited (red) rates vs. [GdnHCl] for C2 (triangle) and C60 (circle) at 20 C and 1 cP	51
Figure 3.11 Reaction-limited (filled) and diffusion-limited (open) rates vs. [GdnHCl] for S2C (black circle) and T58C (red triangle) at 20 C and 1 cP. Both loops are equally 28-residue long.	53
Figure 3.12 Trp/Cys contact-quenching studies of ACBP unfolded states. (a) Observed quenching rates vs. time after mixing for I86C. Inset: observed quenching rate vs. viscosity. (b) The effective diffusion coefficients at various GdnHCl concentrations. (c) Reaction-limited (filled) and diffusion-limited (open) rates for I86C (d) Reaction-limited (filled) and diffusion-limited (open) rates for T17C. Black triangles are measured values from molecular dynamics simulations.	55
Figure 3.13 Observed quenching rates vs. time after mixing for T17C. Dashed line represents the natural lifetime of tryptophan triplet state.	57
Figure 4.1 Hydrophobicity scale used in the energy reweighted WLC model. The values are taken from the Miyazawa-Jernigan scale and normalized to span the interval 0 to 1.	61
Figure 4.2 Rank correlations between radius of gyration (R_g) and Trp47/Cys23 (r_{23-47}) (a) and between R_g and Trp47/Cys57 (r_{47-57}) (b)	65
Figure 4.3 Second structure propensity (SSP) of U_{exch} (red) and U_{gdn} (blue) states of drkN.....	66
Figure 4.4 Average total energy $\langle E_{\text{TOT}} \rangle_r$ of wormlike chains with (black) and without (red) secondary structure constraints	67
Figure 4.5 a) Probability distribution of Trp-Cys distances in protein L K23C before (black) and after (red) reweighting for $\sigma = 2.5$. b) Contour plot of probability vs r and E for b) $\sigma = 0.2$ and c) $\sigma = 2.5$	70
Figure 4.6 The values of which best reproduce the measured reaction-limited rates using equation 4.4 for various sequences and [GdnHCl]	71

Figure 4.7 Values of σ that best fit measured reaction-limited rates for various forms of E_{TOT}

$$1/r^2 \text{ potential: } E_{TOT} = - \sum_{|i-j|>1} \frac{e_{i,j}}{(r_i - r_j)^2} \quad \text{Square-well potential: } E_{TOT} = - \sum_{|i-j|>1} e_{i,j}$$

$$\text{Lennard-Jones potential: } E_{TOT} = \sum_{|i-j|>1} e_{i,j} \left[\left(\frac{4}{r_i - r_j} \right)^{12} - \left(\frac{4}{r_i - r_j} \right)^6 \right] \text{ for } d_\alpha \leq |r_i - r_j| \leq 6.5 \text{ \AA}$$

$$1/r, \text{ no cutoff: equation 4.1 in the text for } d_\alpha \leq |r_i - r_j| < \infty \dots\dots\dots 72$$

Figure 4.8 Values of σ that best fit measured reaction-limited rates for various definitions of hydrophobicity 73

Figure 4.9 a) Values of σ that best fit measured reaction-limited rates for hydrophobic-hydrophobic interactions only. b) Comparisons of unweighted (P(r)) and weighted (Z(r)) probability distributions for weighting schemes using hydrophilic interactions (green and red lines) and only hydrophobic interactions (blue line). 74

Figure 4.10 Average number of favorable contacts of each residue in protein L K23C (black) and F22A K23C (grey). The vertical dashed line shows the position of the mutation F22A..... 76

Figure 4.11 Contact maps of protein L K23C F22A (a, c) and K23C (b, c) at $\sigma=0$ (c) and $\sigma=2.5$ (a,b). Contact map (d) is the subtraction of (a) by (b). The lower-right triangle region shows the average distance between residues i and j and the upper-left triangle region gives the standard deviation of that average distance..... 78

Figure 4.12 Measured (bars) and calculated (lines) PRE per residue for drkN SH3 domain (a and b) and ACBP (c and d). In each graph the green lines correspond to a completely random set of wormlike chains and the red lines correspond to an ensemble of wormlike chains with some residual secondary structure given by measured NMR chemical shifts. The blue lines correspond to the ensemble with secondary structure using a modified version of equation 4.1 in which only interactions more than 4 residues apart in sequence are included ($\sigma = 6$). The paramagnetic labels are located at residue 2 (a), the N-terminus (b), residue 36 (c) and residue 65 (d). For (a) $\tau_C=0.5$ ns, a scaling factor in equation 4.7, is adjusted to best match the measured data. Pearson Correlation coefficient between bars and red lines equals to 0.89 (a), 0.41 (b), 0.35 (c) and 0.59(d)..... 81

Figure 4.13 Probability distribution of Trp-Cys distances in AavLEA T58C before (black) and

after reweighting for $\sigma = 2.5$ (red), $\gamma = 31$ (green) and $\sigma = 2.5$, $\gamma = 31$ (blue)..... 85

Figure 4.14 The cumulative probability ($C(a) = N \int_0^a P(x) dx$, N is a normalization constant) of buried surface area for hydrophobic and hydrophilic residues of protein L at $\sigma = 0$ and $\sigma = 2.5$. Buried surface area is defined as the cross-section of overlap between two links on the wormlike chain assuming each link has spherical volume of diameter 6.5 \AA . Since no two links can be closer than $d_{\alpha} = 4 \text{ \AA}$, the maximum buried surface area is 20.6 \AA^2 per residue..... 87

Figure 4.15 The values of D which best reproduce the measured diffusion-limited rates using equation 4.5 with values of σ shown in Figure 4.6..... 88

Figure 5.1 Experimental reaction-limited (black) and diffusion-limited (white) rates measured at various concentrations of denaturant for (a) ACBP T17C (b) ACBP I86C (c) HypF-N W81F, (d) apocytochrome c, (e) Protein L K23C and (f) Protein G E19C. The errors in these rates are typically 10%. The triangular points in (a) and (b) are measured values from molecular dynamics simulations. The triangular points in (e) are the destabilizing mutation K23C F22A in which the unfolded state is observable in as little as 0.25 M GdnHCl. The reaction-limited rate in (e) (black circle) at 0 M GdnHCl was calculated from a Trp-Cys distribution derived from a molecular dynamics simulation³⁷ using equation 2.9. All other points were derived from measured tryptophan triplet decay rates at various temperatures and viscosities using equations 2.5 and 2.6..... 91

Figure 5.2 Intramolecular diffusion coefficients of several different proteins. Long polyQ is an aggregation-prone protein like HypF-N. One single mutation F22A destabilizes well-behaved wildtype protein L tremendously and gives it a tendency to aggregate. 94

Figure 5.3 Formation of aggregated species, Agg , according to the model outlined in the text. For these simulations, $k_{bi} = k_{agg} = 1$, $K = 0.1$ and k_{-1} varies as indicated on the plot 96

Chapter 1

Introduction to Protein Folding, Misfolding and Aggregation

1.1 Amino Acids, Peptides and Proteins

Proteins (polypeptides), like nucleic acids and carbohydrates, are macromolecules. They are constructed in a linear combination of a set of 20 amino acids. Each type of amino acid has a homogenous backbone including a carboxyl group and an amino group (Figure 1.1) and a distinctive side chain (R-group) bonded to the C α atom. Amino acids are usually classified by the chemical properties of their side chains (Figure 1.2). Amino acids in a protein, sometimes also called residues, are linked by peptide bonds between the carboxyl group and the amino group (Figure 1.1). Thus, each peptide contains only one free non-bonded carboxyl group known as the C-terminus and only one free amino group known as the N-terminus. By convention, the amino acid sequence in a protein is numbered from the N terminus to the C-terminus. The linked series of carbon, nitrogen and oxygen atoms excluding the R-group are usually referred to as the main chain or backbone.

To perform physiological functions in cells, proteins need to fold into specific 3-dimensional structures determined by the sequence of amino acids. For a protein with a single peptide chain, its structure can be divided into three different levels: primary structure, secondary structure and tertiary structure (Figure 1.3). Primary structure is just the unique amino acid sequence of this protein. Secondary structure includes α -helix, β -strand and hairpin bend, all of which result from interactions between local amino acids. The arrangement of these secondary structure elements

make up the tertiary structure, which reflects interactions between distant residues.

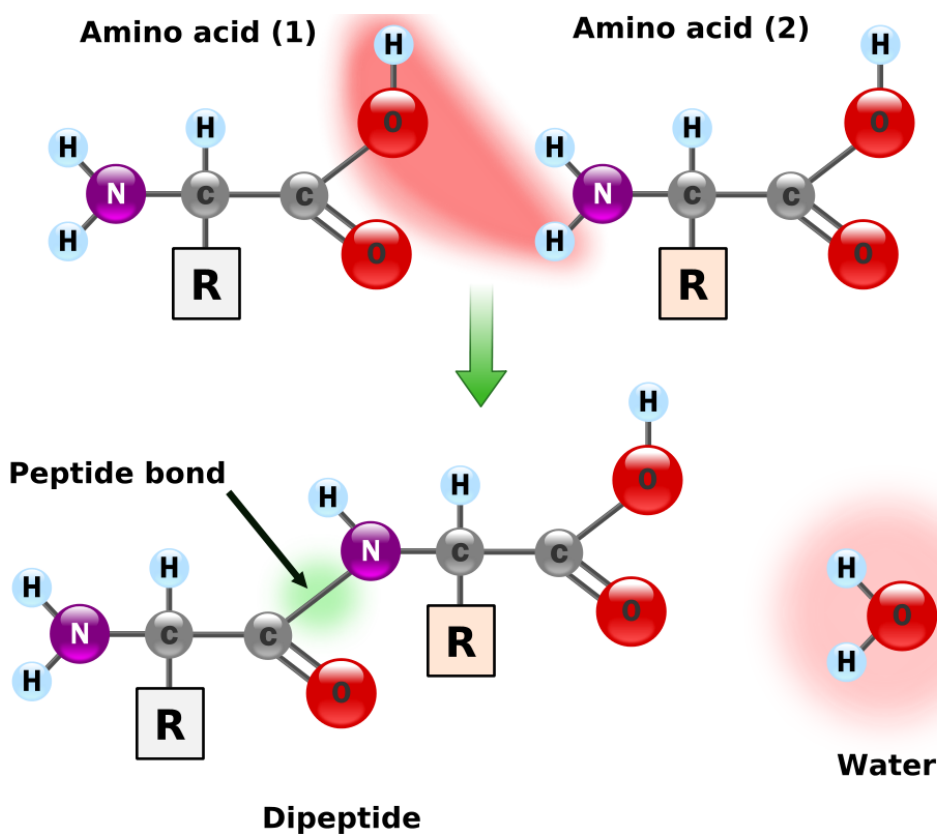


Figure 1.1 Two amino acids form a peptide bond between the carboxyl group and the amino group. (<http://upload.wikimedia.org/wikipedia/commons/6/6d/Peptidformationball.svg>)

For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

1.2 Protein Folding: Mechanics, Energy and Time Scale

Proteins are synthesized in the ribosome through a series of processes including the transcription of DNA to RNA, RNA splicing and then the translation of RNA to proteins. Thus, the 1-D primary structure (sequence) and the final folded structure of a protein can be determined by the genetic codes. Some proteins fold into their native state spontaneously with

little assistance provided by other intracellular mechanisms (chaperones) and only when folded into the correct structure can they perform their intended physiological functions, such as providing cellular structure in the cytoskeleton of cells, acting as the contractile machinery in muscles, and working as enzymes to catalyze reactions¹. Protein misfolding can cause human diseases, such as prion disease, Huntington's chorea and Alzheimer's disease. The famous Parkinson disease is associated with the aggregation of α -synuclein^{2, 3}.

Nonpolar, hydrophobic

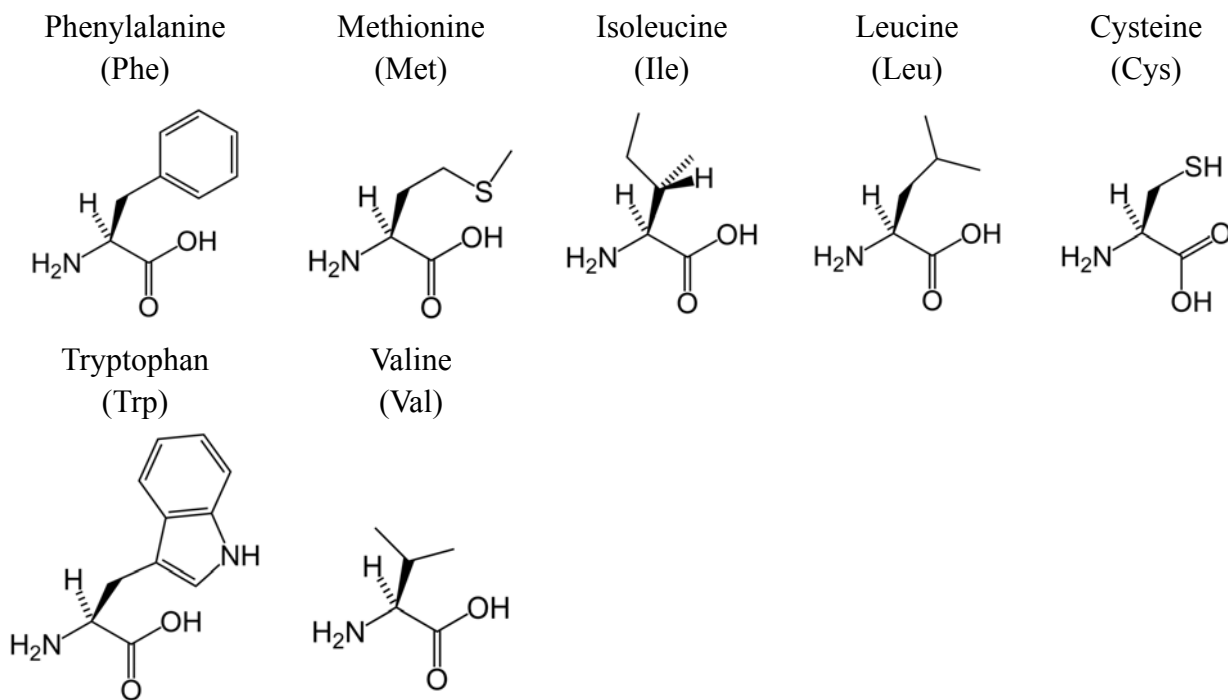
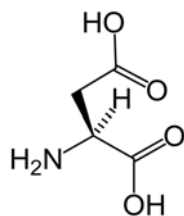


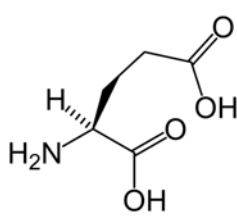
Figure 1.2 Twenty standard amino acid classified by their polarity (hydrophobicity)

Polar, charged

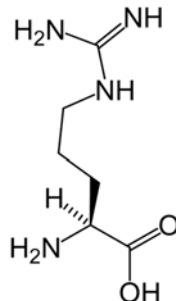
Aspartic acid
(Asp)



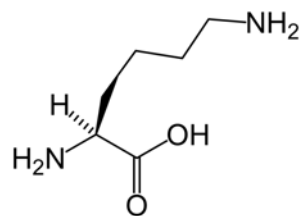
Glutamic acid
(Glu)



Arginine
(Arg)

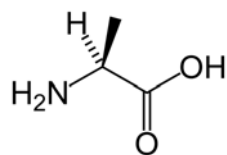


Lysine
(Lys)

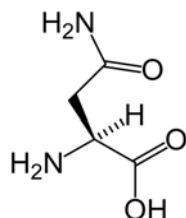


Polar, uncharged

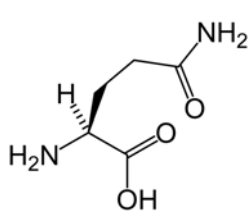
Alanine
(Ala)



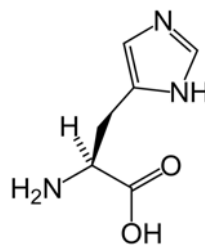
Asparagine
(Asn)



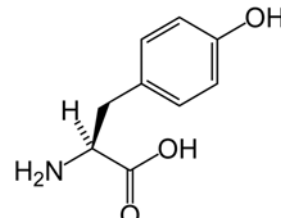
Glutamine
(Gln)



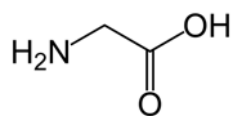
Histidine
(His)



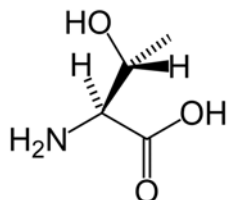
Tyrosine
(Tyr)



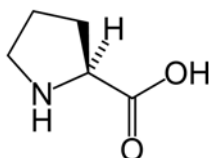
Glycine
(Gly)



Threonine
(Thr)



Proline
(Pro)



Serine
(Ser)

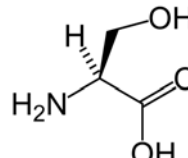


Figure 1.2 (cont'd)

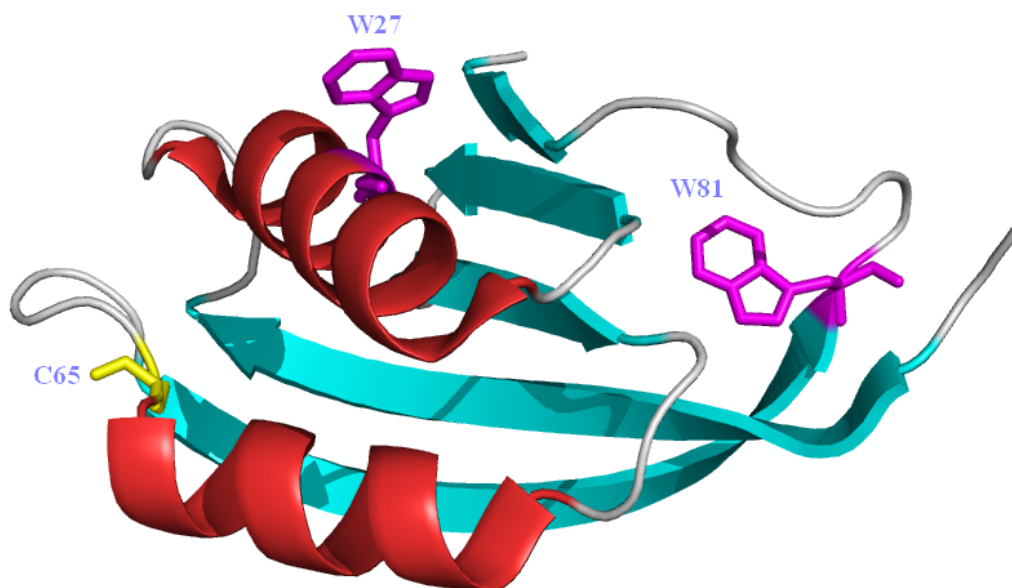


Figure 1.3 cartoon of the HypF-N protein showing (1) α -helix (red), (2) β -strand (blue), (3) tryptophan 27 and 81 (purple) and (4) cysteine 65 (yellow).

Therefore, in order to understand the connection between protein structures and their functions, and between the folding pathways and the misfolding mechanism, it is essential to understand how proteins fold into their native conformations from unfolded states, which is the so called protein folding problem.

If we consider a protein in isolation, the folding process is essentially a competition between enthalpy and entropy. First, there are several interactive forces and solvent effects that can drive proteins towards folding. The hydrophobic effect is a major solvent effect that causes protein collapse, in which non-polar residues cluster into a hydrophobic core surrounded by solvent-exposed polar residues⁴. Other factors include Van der Waals forces, disulfide bonds and electrostatic Coulomb interactions. Hydrogen-bonding is due to Coulomb interactions. It facilitates the formation of α -helices and β -sheets and also stabilizes a protein structure. On the other hand, the entropic effect tends to unfold a protein as self-organization upon formation of secondary and tertiary structures and volume compaction induced by the hydrophobic effect, decreases the configurational entropy. Thus, the Gibbs free energy is widely used in protein folding studies to combine these two opposing effects, which essentially results in the marginal stability of the native state and the complexity of folding pathways.

Does the native state correspond to the free energy minimum? Christian B. Anfinsen, winner of the 1972 Nobel Prize in Chemistry, proposed a positive answer to this question based on his experiments on Ribonuclease A which demonstrated that folding and unfolding is reversible by adding or removing Urea (a chemical denaturant used to unfold proteins)⁵. He stated that the protein would spontaneously fold into the most stable state and hence all the information needed

to determine the protein native conformation was in the sequence itself⁵.

A small protein (<100 residues) usually folds on a millisecond timescale, however it would take longer than the age of the universe for a 100-residue polypeptide to sample all the possible conformations (Levinthal's Paradox)⁶. Therefore, it is much more likely that the folding pathway towards the native state is biased to those with low activation energies in order for the protein to fold in a biologically relevant timescale. Most scientists today take both thermodynamic and kinetic theory into consideration in the analysis of protein folding^{7, 8}.

Proteins fold on a wide range of timescales, from nanoseconds to hundreds of seconds. Figure 1.4 shows a comparison of different protein folding events based on their associated timescales. According to the following Arrhenius equation which was later developed into the transition state theory (TST),

$$k = A \exp(-E_a / RT) \quad (1.1)$$

the rate constant k of chemical reactions is directly connected to the activation energy E_a . However, we generally have little knowledge of the pre-exponential factor A . Therefore, rather than calculating the absolute reaction rates, the transition state theory has been used to derive the free energies, entropies and enthalpies from experimentally measured reaction rates⁹⁻¹⁴. Also the transition state theory was initially developed to describe reactions in the gas phase, so it might be too simplified to model the dynamics of protein folding, in which the interactions between the proteins and solvent play an important role. In this regard, an alternative, the Kramers' theory¹⁵⁻¹⁹, might be more suited because it models the folding reaction as a diffusional crossing of an energy barrier. The barriers described in the Kramers' theory are explored by Brownian

motion and can be crossed repeatedly. In a limit of strong frictional dumping from the solvent, the reaction rate can be simplified as

$$k = (\omega_a \omega_b / 2\pi\gamma) \exp(-E_a / k_B T) \quad (1.2)$$

where, ω_a , ω_b are the curvature of the unfolded well and the barrier top, E_a is the barrier height and γ is the frictional drag coefficient, which is proportional to the solvent viscosity η .

This suggests an inverse linear relationship between the folding rate k and the solvent viscosity η ¹⁸. This diffusion-based theory also inspires researchers to experimentally estimate the upper speed limit of protein folding by measuring the intramolecular loop contact rates of unfolded proteins²⁰⁻²⁴, because a protein cannot fold any faster than the formation of those elementary folding events, such as loops, turns, α -helices and β -structures.

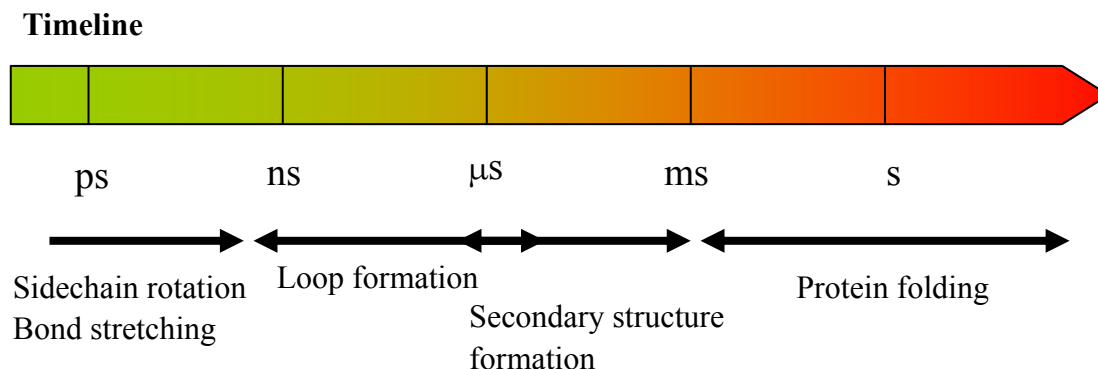


Figure 1.4 Time span of protein folding events

1.3 Introduction to Experimental Approaches

Experiments on protein folding have been performed both *in vivo* and *in vitro*. From the thermodynamic point of view, one should observe similar dynamics in either environment as long as the physiological conditions (pH, salt concentration and so on) are the same. This provides scientists the chance to investigate the folding process of proteins in quartz cuvettes instead of human vessels and on optical tables instead of living organisms. *In vivo* experiments are still irreplaceable in some cases when other intracellular molecules that assist folding are involved. These assistant molecules are called chaperone proteins which to some extent disprove the thermodynamic hypothesis proposed by Anfinsen which states that the native structure is determined only by the protein's amino acid sequence. While molecular chaperones raise very interesting biomolecular questions, this section will focus only on *in vitro* experiments.

In vitro protein folding experiments usually involve a perturbant to unfold or partially unfold

a protein. The perturbant can be a chemical denaturant such as urea and guanidine hydrochloride (GdnHCl), heat or even some sort of mechanical force. Guanidine hydrochloride is commonly used in the denaturation of proteins by decreasing solvent polarity and consequently disrupting non-covalent interactions. An example of heat denaturant is the high energy laser pulse used in T-jump experiment where the protein is unfolded by a sudden change of temperature^{25, 26}. The advantage of using mechanical force as denaturant is that instead of global denaturation, forces can be exerted on designed sites exclusively. Such experiments usually employ techniques such as AFM and optical tweezers and are extremely hard to perform because they require manipulation of proteins on the single molecule level^{27, 28}.

Another key element for protein folding experiments is the probe. While new techniques are always being developed, several major probes widely employed by researchers include NMR, X-ray crystallography, circular dichroism (CD), UV fluorescence and some other optical approaches. The first two are commonly used in determining molecule structures. CD and UV fluorescence, on the other hand, are very sensitive to rapid conformational changes, such as CD to secondary structures and tryptophan fluorescence to its local environment, and hence are suitable for timed folding experiments.

Advanced experimental techniques enable scientists to monitor the course of the entire or partial folding process in real time by removing the perturbant at early times (ns to ms). For mechanical denaturant, this can be simply accomplished by stopping pulling. For chemical denaturant, a mixer device is usually employed to dilute proteins in high denaturant concentrations into aquatic buffer. But no matter how small the ratio you mix the sample into

solution, there will always exist a small amount of denaturant that may shift the protein from its native state. The time resolution of folding kinetics experiments is greatly influenced by the class of probes. However, a continuous-flow microfluidic mixer device can be utilized to mix on the μs timescale^{29, 30}. In such a mixer, solutions flow through the observation channel at a constant flow rate so that observations taken at different positions correspond to different times after mixing. This mixer not only shortens the measurement dead time due to turbulence but also allows for longer integration time for each measurement.

1.4 The Importance of Studying the Unfolded States

It is evident that insights into the protein folding pathways and energetics can be gained from studying the unfolded states of proteins according to the energy landscape theory. If it is impossible for proteins to sample all possible conformations in physiological relevant time as proposed by Levinthal's paradox⁶, there must be some specific folding pathways associated with specific starting unfolded structures that are biased towards the native configuration. And no matter what the folding mechanism is, there will be some unfolded conformations that are more favorable than others. The topology of these favorable unfolded states in folding conditions may favor certain side chain interactions that can trigger or catalyze protein folding. Recent experiments on several proteins using NMR spectroscopy support this model by showing the presence of residual structures in unfolded states³¹⁻³⁵. People used to consider unfolded states to be completely random polymers switching configurations rapidly through diffusion, suggesting a very smooth surface on energy landscape. However, several high resolution experiments and

computer simulations have shown that the diffusion of unfolded states of well behaved proteins is indeed relatively slow, suggesting a rugged energy surface³⁶⁻³⁸. To add these complexities to the current polymer model for describing the distribution of unfolded conformations, an energy reweighted wormlike chain model will be brought into discussion in chapter 4.

Experimentally, monitoring the dynamics of unfolded states is no easy task because of the nanosecond to microsecond relaxation time scale. Traditional equilibrium experiments performed in different denaturing conditions can only be used to calculate thermodynamic parameters, which give no insights into kinetics. The stopped-flow mixer has a mixing time longer than milliseconds so it focuses mostly on the later stages of the folding process. However, a technique with ns time resolution, probing the intramolecular contact frequency between Trp and Cys residues can be utilized to capture the loop formation of unfolded polypeptides, which provides direct insight into the intramolecular diffusional dynamics^{23, 39-41}. Such experiments can be performed in either equilibrium conditions or in a fast mixer³⁸. Chapters 2 and 3 will demonstrate a series of results covering proteins with various aggregation propensities using this method.

1.5 Polymer Models and Computational Methods

Unfolded proteins are one special type of polymer and can be elucidated well using simplified polymer models. If one considers a completely denatured peptide with N residues as a random walk in 3 dimensions with step size l_k (Kuhn length), the end-to-end distance r of this polymer will have a Gaussian distribution.

$$P(r) = \left(\frac{3}{2\pi N l_k^2} \right)^{3/2} \exp \left(-\frac{3r^2}{2N l_k^2} \right) \quad (1.3)$$

This approximation tends to fail for short peptides because unfolded proteins are not real freely-joined chains and they have a certain stiffness due to chemical bonds which can only be bended to limited angles. One may also want to take into account the excluded volume effect of biomolecules to create self-avoiding chains. A wormlike chain model has been used to include these constraints and will be discussed in chapter 2. Another factor worth considering is the heterogeneity of peptides in low denaturant concentrations where intramolecular interactions start to take place.

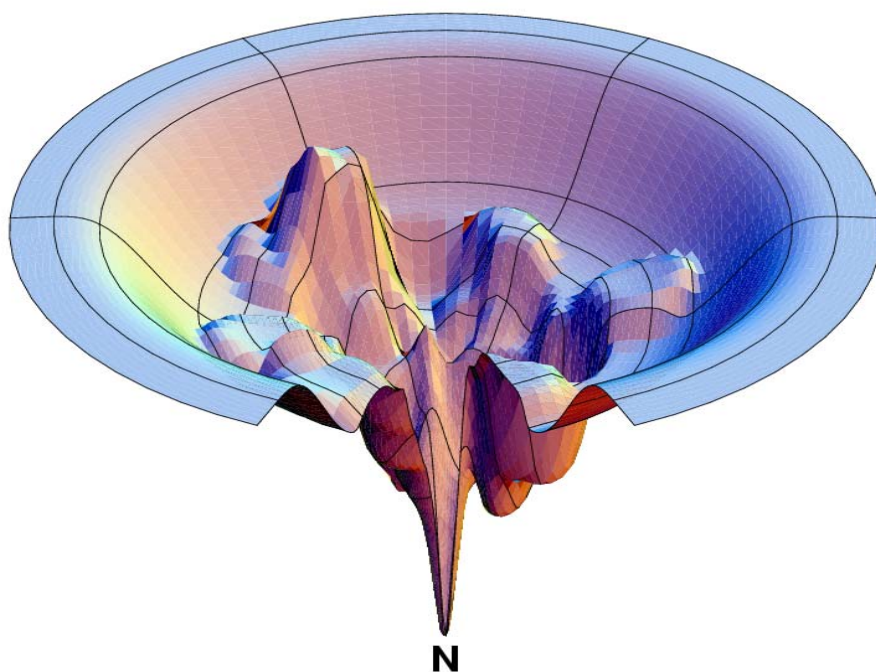


Figure 1.5 an energy landscape for protein folding created by HP model⁴²

More sophisticated models are required to describe the complexity of protein folding dynamics. The energy landscape theory has served as a central theory for many computer-based protein folding simulations and experiments⁴³⁻⁴⁸. It takes a statistical approach to the energetics of protein conformations and demonstrated that proteins can fold into their native states through any arbitrary pathway on a funneled energy landscape (Figure 1.5), rather than through a single 2-state mechanism. Molecular Dynamics (MD) is a powerful computational tool that allows

atoms and molecules to interact in real time based on a physics-based force field and statistical mechanics. MD simulation, however, is limited to the folding of small proteins on the microsecond timescale due to the high computational cost at all-atom level. Enhanced sampling methods applied to help proteins cross energy barriers are commonly used to fold large, slow-folding proteins⁴⁹. Also, hardware improvement, such as the development of GPU computing⁵⁰ and distributed computing^{44, 51}, has made it possible to gather more folding trajectories in less computational time.

Chapter 2

Spectroscopic Study of the Dynamics of Unfolded Proteins

2.1 Introduction

Spectroscopy, microscopy and other optical methods have been widely used in the study of protein folding. Light sources cover the wavelengths from X-ray to IR. In the presence of several aromatic amino acids or with other dyes attached (fluorophore or luminophore), the process of protein folding can be monitored at various time scales with different experimental systems. Through different approaches reflecting different aspects of folding kinetics, a combination of these results should reveal a whole picture of protein folding pathways-the energy landscape.

Conventional *in vitro* equilibrium experiments are not hard to accomplish and they can provide very meaningful results. Fluorescence spectroscopy is a typical probe. For example, Trp47 in B1 domain of protein L under continuous wave excitation at 280 nm or 295 nm yields an emission spectrum with a peak that is monotonically dependent on the local polarity (hydrophobicity). A blue-shift occurs when the tryptophan residues are more buried from water⁵²,⁵³. This provides an excellent tool to study the structural change of proteins under various denaturing conditions. The fraction of unfolded proteins under different denaturing conditions can be determined by plotting the wavelength or intensity of spectrum peaks against denaturant concentrations (Figure 2.1). These fluorescence data fit well with a two-state model.

However a recent study of the early folding process (before 100 μ s) of protein L using a rapid continuous-flow mixer revealed the complexity of the folding pathway of this well recognized

two-state folder, interpreted as the ruggedness in folding

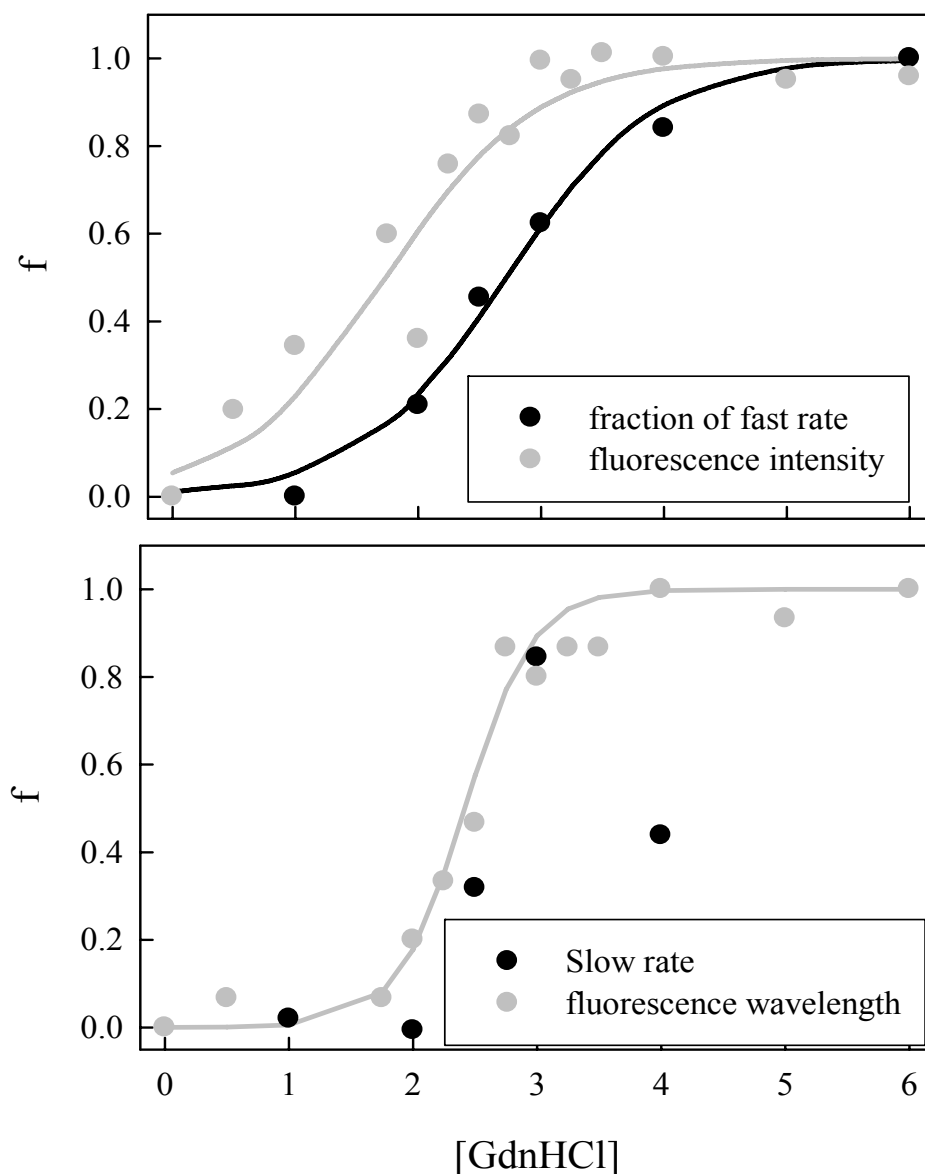


Figure 2.1 The fraction of unfolded molecules (circles) of the T57C mutant of protein L as measured by various spectroscopic probes: fluorescence intensity, wavelength of peak fluorescence, relative amplitude of intramolecular contact formation (fraction of fast rate), and the magnitude of the unquenched tryptophan triplet lifetime (slow rate). The last two probes will be described in detail in the following sections. The lines are fits to two-state models

landscape³⁶. So understanding the early events in protein folding not only helps us find the true folding speed limit but also refines the energy landscape. In this chapter, a pump-probe spectroscopy for measuring intramolecular loop formation will be introduced and followed by the data analysis method. Finally, a microfluidic mixer device is coupled with the pump-probe spectroscopy to discover the intramolecular dynamics of unfolded proteins before they fold.

2.2 Monitoring Tryptophan Triplet Quenching Using Pump-probe Spectroscopy

Nanosecond intramolecular loop contacts can be detected experimentally using a pump-probe laser spectroscopy. The mechanism is the quenching of tryptophan triplet energy state (probe) by cysteine²³ (both tryptophan and cysteine are natural amino acids). Protein samples are engineered and mutated with minimum influence on their original structures and stability to contain only one tryptophan (Trp) and one cysteine (Cys) so that the measured quenching rate reflects only one intrachain contact. Normally, for each protein studied, multiple mutations are made to cover different loops over the full protein length.

Tryptophan can be excited to long-lived triplet energy states upon the absorption of 289 nm UV light (Figure 2.2). It has been shown previously that the tryptophan triplet lifetime is about 40 μ s in water and longer in the hydrophobic core of a folded protein⁵³. The lifetime of the triplet state will be shortened when cysteine is nearby. Cysteine is a very efficient quencher with a quenching rate 400-fold faster than other amino acids at physiological pH. The mechanism of the quenching is electron transfer so the quenching rate has an exponential dependence on the

intramolecular Trp/Cys distance^{40, 54}:

$$q(r) = q_0 \exp[-\beta(r - a_0)] \quad (2.1)$$

where r is the intramolecular distance, a_0 is the distance of closest approach (4 Å), $q_0 = 4.2 \text{ ns}^{-1}$ and $\beta = 40 \text{ nm}^{-1}$.

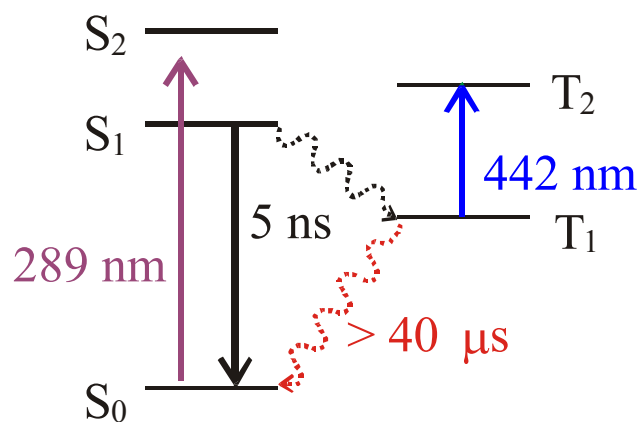


Figure 2.2 Schematic electronic energy levels of tryptophan. 289nm UV laser is used to excite the tryptophan to its singlet states. The intersystem cross from S₁ to T₁ is achieved through the reversal of the spin of an excited electron. Triplet states T₁ absorb blue light (442nm) and have a long lifetime.

The triplet state of tryptophan absorbs visible light at 442nm wavelength, which can be used as a probe to monitor the population of the triplet state. The observed decay rate is a combination of several kinetic rates⁵⁵:

$$k_{obs} = k_0 + \sum_i k_i^{uni} + \sum_i k_i^{bi} [i] \quad (2.2)$$

where k_0 is the decay rate without quencher i , k_i^{uni} and $k_i^{bi} [i]$ are the unimolecular and bimolecular quenching rates respectively. To make the unimolecular rate more significant than others, we first need to keep the sample concentration low ($\sim 30 \mu\text{M}$) so that the contribution of the bimolecular rate ($2 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$) is negligible²³. We also need to make sure that tryptophan and cysteine are engineered close enough in sequence so that the unimolecular contact rate is faster than k_0 ($2 \times 10^4 \text{ s}^{-1}$) and hence distinguishable. With these experimental controls, the kinetics happening on the microsecond time scale can be simplified to a 2-step kinetic model (Figure 2.3).

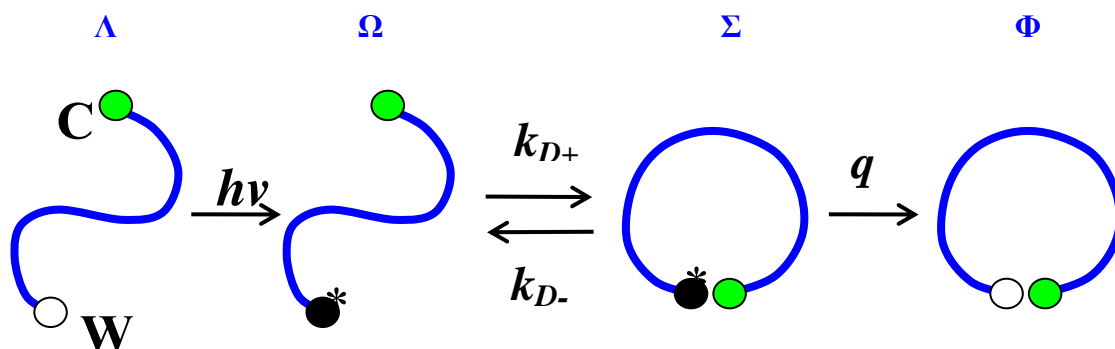


Figure 2.3 Cartoon of loop formation and quenching. A tryptophan residue (W) is first excited to the triplet states and then along with a cysteine residue (C) at the other end diffuses towards and away from each other at rates k_{D+} and k_{D-} respectively. q is the quenching rate.

1. The excited tryptophan and cysteine (quencher) diffuse towards each other at a rate of k_{D+} .
2. Cysteine either quenches the tryptophan at quenching rate q or diffuses away at rate k_{D-} .

By solving the chemical reaction equations with the assumption of no accumulation at the 3rd stage in Figure 2.3,

$$\begin{aligned}\frac{d[\Sigma]}{dt} &= 0 = k_{D+}[\Omega] - (k_{D-} + q)[\Sigma] \\ \frac{d[\Phi]}{dt} &\equiv k_{obs}[\Omega] = q[\Sigma] = q \left(\frac{k_{D+}}{q + k_{D-}} \right) [\Omega]\end{aligned}$$

the observed rate is given as

$$k_{obs} = \frac{k_{D+}q}{q + k_{D-}} \quad (2.3)$$

If $q \gg k_{D-}$, then $k_{obs} = k_{D+}$, but if that condition does not hold then equation 2.3 can be rewritten as

$$\frac{1}{k_{obs}} = \frac{1}{qK} + \frac{1}{k_{D+}} = \frac{1}{k_R(T)} + \frac{1}{k_{D+}(T, \eta)} \quad (2.4)$$

where $K \equiv k_{D+}/k_{D-}$ is the equilibrium constant for forming the Trp/Cys encounter complex and $k_R \equiv qK$, which is the reaction-limited rate. k_{D+} is called the diffusion-limited rate.

Figure 2.4 demonstrates the instrumentation for measuring the transient absorption of the tryptophan triplet states using a transmission pump-probe spectroscopy. Two different laser beams are involved in this collinear geometry setup. The Nd:YAG pulse laser (Continuum Surelite II-10, pulse duration: 8 ns, interval time: 10 ms) is employed to pump the tryptophan residues into their triplet states. The wavelength of the pulse beam is shifted from 266 nm to 289 nm using a Raman cell (filled with methane at 250 psi) to minimize the photodamage⁵⁶. Between each pulse, a continuous wave (CW) 441 nm laser beam (He-Cd) is used to monitor the tryptophan triplet states. It is split into two parts: a reference beam and a sample probe beam. The

transmitted probe beam is aligned collimated to the pulse beam so that it passes through those excited molecules. The intensities of the probe beam and the reference beam are measured using two silicon photodiodes (New-Focus), recorded with digital oscilloscopes (Tektronix TDS), subtracted, amplified and then stored on a computer using GPIB interface. The absorption decay trace reflects the lifetime of the tryptophan triplet states which can range from nanosecond to submillisecond. So two oscilloscopes are used to cover the wide temporal range: one with a 10 μ s window and one with a 10 ms window. Two additional amplifiers (LeCroy DA1886A differential amplifier and SR445A, SRS Inc four channel preamplifier with 5x gain at each stage) have been recently incorporated into the instrument to allow for lower sample concentration and lower pulse beam power.

To eliminate any pulse leakage and high frequency cable noise, a background with the pump but without the probe beam is recorded each time at the beginning of each experiment and thereafter subtracted from all other recorded decay traces. The UV pulse also tends to generate hydrated electrons and neutral radicals that absorb light near 450 nm and decay within 3 μ s^{23, 57}. Furthermore, O₂ is a good quencher of the tryptophan triplet state. Therefore, before each experiment, those free electrons are scavenged and oxygen molecules are removed by degassing the sample solution with N₂O for at least one hour to eliminate from the signal trace the amplitude cause by these photo-effects. Another undesired photo-effect is thermal lensing, in which the heat induced by the pulse laser changes the local density and consequently the refraction of index of the solution, causing the probe beam to slightly reflect and move across the detector. Because the sensitivity is not uniform over the entire detector, thermal lensing may

result in a pseudo-decay trace on the millisecond timescale. Ways to diminish the effect of thermal lensing include decreasing the power of pulse laser, realigning the optics of the pulse and probe beams and adjusting the position of the detector.

The sample cuvette is placed in a Peltier temperature controlled sample holder (Quantum Northwest). Benzophenone has a decay time of about 200 ns at 50 μ M and is regularly used for the initial optical alignment of the instrument. N-acetyl-L-tryptophanamide (NATA) is used to optimize the alignment because it is the uncharged analogue of tryptophan and has a 40 μ s decay time in the presence of deionized water in the absence of any additional quencher⁵³.

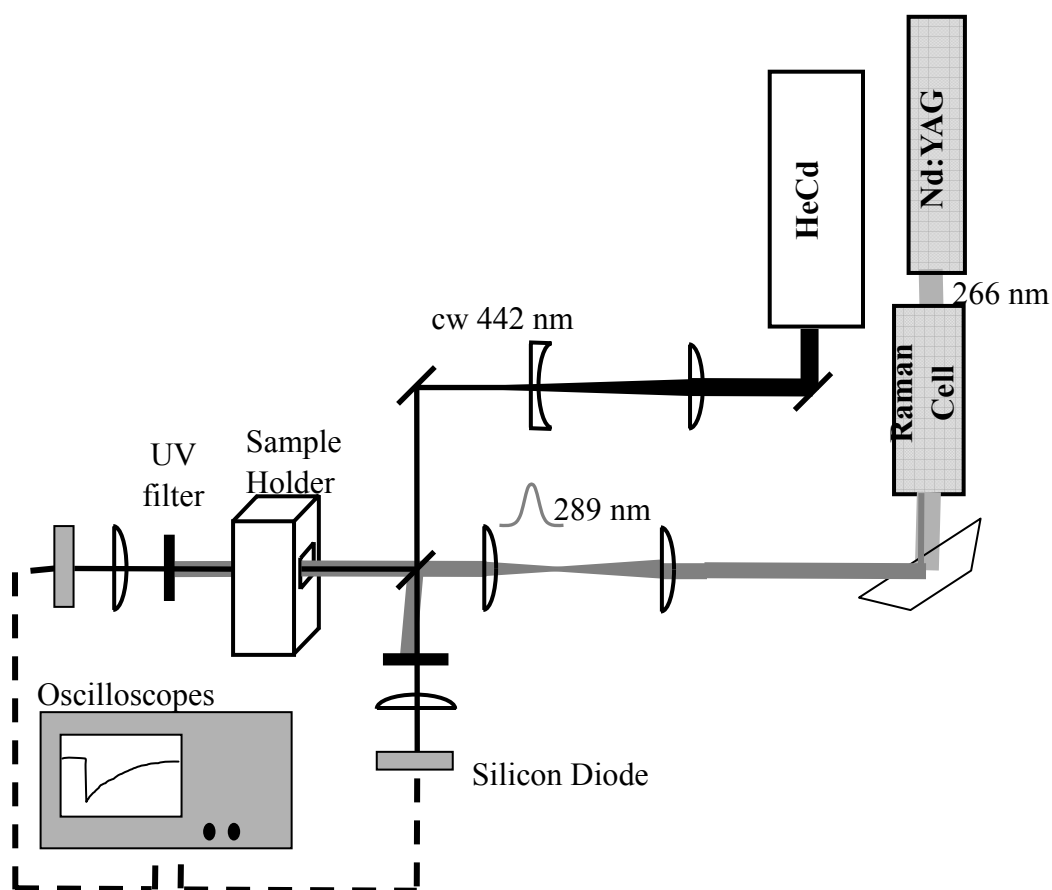


Figure 2.4 Schematic of the tryptophan triplet quenching experiment

2.3 Reaction-limited and Diffusion-limited Rates

In an ideal contact quenching model, quenching is infinitely rapid (diffusion-limited) and cut off at a short distance. Excited probes are immediately quenched upon loop closure and hence the measured quenching rate is diffusion limited. However, for the Trp/Cys system, polymer dynamics is more or less coupled with reaction (quenching) kinetics. For instance, equation 2.1 shows the quenching of the tryptophan triplet by cysteine exhibits an exponential decay with base quenching rate q_0 . Because q_0 is comparable to k_{D-} , there is a fairly good chance for tryptophan to escape the quenching from a close encounter (Figure 2.3). Consequently the decay profile observed in experiment may actually reflect 2 or more processes of loop formation and deformation. In this case, the measured rate is a mixture of reaction-limited and diffusion-limited rates and does not solely reflect diffusional dynamics.

Because of the complexity of the experimentally measured kinetics of loop formation shown above, efforts have been made to extract a meaningful diffusion-limited rate from the measured effective quenching rate. Equation 2.4 enables us to separate reaction-limited and diffusion-limited rates by performing experiments at various temperatures and viscosity. The reaction-limited rate, k_R , depends only on temperature because temperature affects the chemical reaction speed but viscosity does not. We assume the diffusion-limited rate, k_{D+} , depends on both temperature and viscosity. Because we observed a linear relationship between the inverse of the observed rate and the solvent viscosity, we decided to use the following two empirical equations to fit our experimentally measured rates,

$$k_R(T) = k_{R0} \exp\left(\frac{E_0(T - T_0)}{RTT_0}\right) \quad (2.5)$$

$$k_{D+}(T, \eta) = \frac{k_{D+0}T}{\eta T_0} \exp(\gamma(T - T_0)) \quad (2.6)$$

where T is the absolute temperature, $T_0 = 273$ K, $R = 0.00198$ kcal•K⁻¹ mol⁻¹, η is the solution viscosity and k_{R0} , k_{D+0} , E_0 , and γ are fitting parameters. Based on equations 2.4, 2.5 and 2.6, a plot of $1/k_{obs}$ against viscosity η at constant temperature T exhibits a linear relationship. The intercept gives the inverse of reaction-limited rate, $1/k_R$, and the slope is proportional to $1/k_{D+}$. In experiments, observed exponential decay rates were determined using TableCurve 2D v5.01 and fitted with equations 2.5 and 2.6 using Matlab to extract the reaction-limited rate and diffusion-limited rate.

The linear $1/k_{obs}$ vs η relationship may breakdown in some cases when a power-law dependence $1/k_{obs} \propto \eta^\delta$ needs to be applied instead. In systems of 20-30 amino acids long loops, δ has been estimated to be very close to 1⁵⁸. Therefore variations due to these two different fittings are negligible.

A brief analysis of k_R and k_{D+} is as follows. The reaction-limited rate, k_R , is the effective quenching rate that can be observed only when diffusion is sufficiently fast. It should depend on the individual quenching rate, q , and the Trp/Cys distance distribution. The diffusion-limited rate, k_{D+} , is the true rate of bringing two ends of the loop together. It should depend on both the Trp/Cys distance distribution and the diffusional dynamics of this particular chain.

2.4 Data Analysis Using SSS theory and a Wormlike Chain Model

Reaction-limited and diffusion-limited rates can be determined analytically using the polymer theory of Szabo, Schulten and Schulten (SSS), which models intrachain diffusional dynamics as motion on a 1-dimensional potential of mean force which is determined by the probability distribution of intrachain distances⁵⁹. Using a Smoluchowski-like diffusion equation with a distance dependent quenching rate under the potential defined as

$$U(r) = -k_B T \ln P(r) \quad (2.7)$$

where k_B is the Boltzmann constant and $P(r)$ is the equilibrium Trp/Cys distance distribution, the observed rate of bringing two ends of chain in close contact can be expressed as⁵⁴

$$\frac{1}{k_{obs}} = \frac{1}{k_R} + \frac{1}{k_R^2 D} \int_{a_0}^{l_c} \frac{dr}{P(r)} \left\{ \int_r^{l_c} (q(x) - k_R) P(x) dx \right\}^2 \quad (2.8)$$

where a_0 is the closest contact distance, l_c is the contour length of the loop, D is the effective intramolecular diffusion constant and $q(r)$ is the distance dependent quenching rate. The reaction-limited rate is given by,

$$k_R = \int_{a_0}^{\infty} q(r) P(r) dr \quad (2.9)$$

and the second term of equation 2.8 yields the expression for the diffusion-limited rate,

$$\frac{1}{k_{D+}} = \frac{1}{k_R^2 D} \int_{a_0}^{l_c} \frac{dr}{P(r)} \left\{ \int_r^{l_c} (q(x) - k_R) P(x) dx \right\}^2 \quad (2.10)$$

If we take the reaction-limited and diffusion-limited rates from experiment and want to

numerically solve for diffusion constant D , the only unknown information in equations 2.9 and 2.10 is the Trp/Cys distance distribution $P(r)$. One simple way to generate such a distribution is using freely-joined or Gaussian chains. According to the Gaussian probability distribution (equation 1.2) the reaction-limited rate for $r \gg d_\alpha$ is

$$k_R = \frac{4\pi q a_0^3}{\left(\frac{2}{3}\pi \langle r^2 \rangle\right)^{3/2}} \exp\left(-\frac{3a_0^2}{2\langle r^2 \rangle}\right) \quad (2.11)$$

which tells us reaction-limited rate k_R is inversely proportional to the average volume $\langle r^2 \rangle^{3/2}$ of unfolded chains. The diffusion-limited rate of Gaussian chains is given by

$$k_{D+} = \frac{4\pi D d_\alpha}{\left(\frac{2}{3}\pi \langle r^2 \rangle\right)^{3/2}} \quad (2.12)$$

Therefore k_{D+} is directly proportional to the diffusion constant D while it is also inversely proportional to the chain volume. Though a Gaussian chain is a very simplified model, these qualitative relationships between rates and chain properties hold true in more sophisticated studies of intramolecular dynamics.

A wormlike chain model has been widely used in our lab for generating chains^{41, 60}. This is a backbone bead model that excludes all the side-chain effects, but has been proven to be a good approximation because of two important constraints added to an ideal Gaussian chain model. First, we apply a persistence length l_p to characterize the real chain stiffness due to chemical bonds. Second, we account for the excluded volume effect by introducing a hard-sphere shell with diameter d_α to the end of each virtual peptide bond. The distance between two non-neighbor

spheres must exceed d_α . Both persistence length and excluded radius prevent intrachain clashes. Adjusting either of them can change the equilibrium distribution $P(r)$.

Wormlike chains are computationally simulated using the method of Hagerman and Zimm⁶¹ as described in refs.^{41, 54}. We typically generate 2 million chains in order to make a normalized histogram of Trp/Cys distances smooth enough for numerical calculation. Chains are grown from the N-terminus by vectorially adding links (1/10 of the length of a peptide bond) to the chain with a random azimuthal angle (ϕ) between 0 and 2π radians and a axial angle (θ) which is a Gaussian distributed random number with a variance $4 l_c / l_p$ around zero. A Monte Carlo algorithm is applied in the construction of each chain with one modification to speed up the computational time⁴¹: when a clash (less than d_α between any two non-neighbor residues) is detected after the addition of every 10 links, the chain is truncated 3 persistence lengths before the clash and the chain is regenerated, otherwise additional links are added until the full length of the protein is reached. This algorithm has been proven to be statistically equivalent to simply discarding chains with clashes.

2.5 Tryptophan Triplet Quenching Experiments in a Microfluidic Mixer

In this section, I will briefly introduce an instrument first built by Steven Waldauer in our lab to measure the intramolecular diffusion kinetically before the protein folds using a microfluidic mixer coupled to the pump-probe spectroscopy. Figure 2.5 shows the geometry of the serpentine mixing chip used in our experiments. It utilizes two different aspects of chaotic

advection⁶², Dean and corner vortices and features a good mixing efficiency at a Reynolds number higher than 60 and a mixing time of $\sim 250 \mu\text{s}$ at a typical flow rate of $\sim 1 \text{ m/s}$. Details on the serpentine chip's design and fabrication can be found in ref.⁶³.

The instrument for the kinetics experiment is built next to the equilibrium Trp/Cys contact quenching instrument (Figure 2.6), sharing the same pulse and CW lasers. Simply by adjusting two flip mirrors, the pump and probe beams can be redirected into the new instrument, whose optical design is very similar to the equilibrium set up. The microfluidic serpentine chip manifold is mounted on an automated 3 axis translation stage (9066-COM-E, New Focus) which can be controlled either by computer or by hand for motion in the plane normal to the beam but has to be adjusted manually to move in the longitudinal direction. To accurately position the narrow observation region in the beam path, we use a piezo-actuator mechanical screw drive (Picomotor model 8301, New Focus, San Jose CA) capable of 10 nm step increments in both the horizontal and vertical directions.

There is a possible synchronization problem that can be brought into the microfluidic mixing experiment by a two-beam system. That is, the small sample volume excited by the pump beam has to be illuminated by the probe beam long enough to trace the tryptophan triplet population. While this requirement is self-fulfilled after the alignment in an equilibrium experiment, additional optics are necessary for the microfluidic mixer because the volume excited by the pump beam travels out the probe beam volume at a constant flow rate, typically 1 m/s. A UV objective (Objectives for Research LMU-3X-266, Cantwell NJ) is therefore added to focus the pump beam into an even smaller spot on the observation channel. For example, considering a

probe beam with a diameter of 100 μm , if the pump beam has a diameter of 1 μm on the focus plane and has been aligned directly in the center of the probe beam, it will take the excited molecules 50 μs to escape the volume illuminated by the probe beam. The small pump beam diameter also helps increase the time resolution of the instrument. Following the previous example, the time resolution should be considered as 1 μs ($1 \mu\text{m} / 1 \text{ m/s} = 1 \mu\text{s}$) even though the probe beam time window is 100 μs long. In reality, the pump beam has to be focused onto a plane in front of or behind the observation channel plane in order not to burn the microfluidic chip and is about 10 μm in diameter.

Both amplifiers, the LeCroy DA1886A differential amplifier and the SR445A, SRS Inc four channel preamplifier, must be used in order to compensate for the small signal caused by the much shorter path length in a serpentine chip (about 30~40 μm) compared to a 1 cm cuvette. The differential amplifier helps the baseline subtraction and also outputs a 10x amplitude of the original signal. Typically, one or two stages on the preamplifier are used, providing a total gain of 50-250x. Most elements in the instrument, including the pulse laser, the oscilloscope and the syringe pumps which inject the protein solution and mixing buffer into the serpentine chip are all controlled via LabVIEW VIs on a data collecting computer.

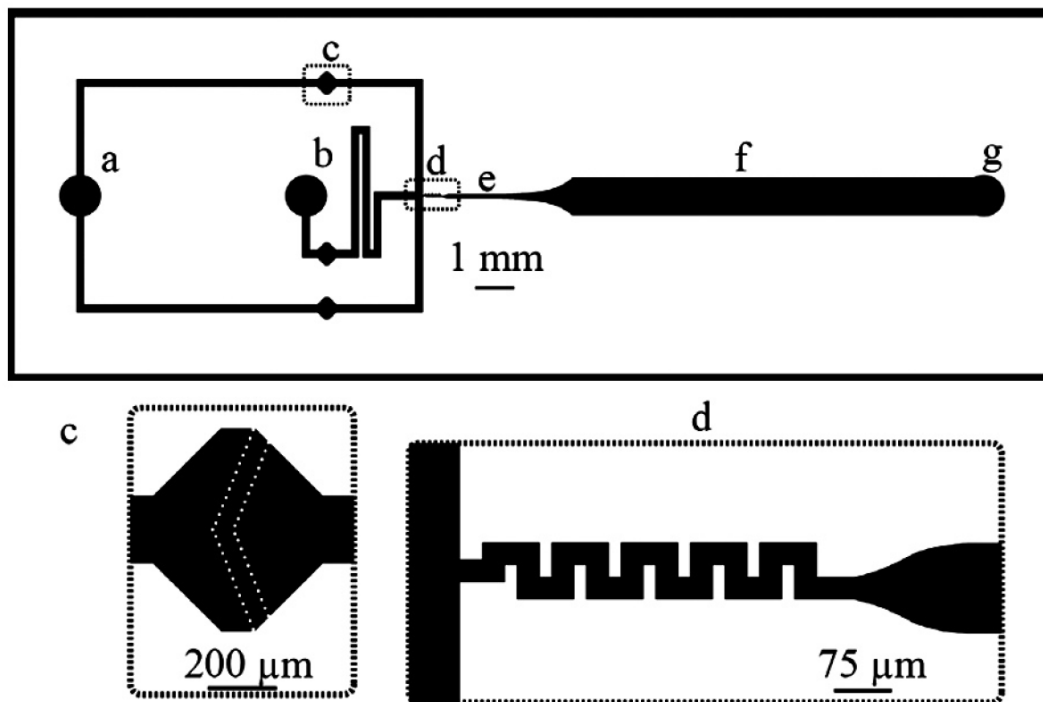


Figure 2.5 Geometry of the chaotic advection mixing microfluidic-chip. (A) Buffer inlet. (B) Protein solution inlet. (C) Filter section (one of three total per chip). Each filter post is a 10 by 10 μm square and they are spaced 10 μm apart. (D). The serpentine mixing region is composed of five turns and the channel is 30 μm wide. (E) The narrow observation region is 120 μm wide and 1.4 mm long. (F) The wide observation region is 1000 μm wide and 10.4 mm long. (G) Outlet port.⁶³

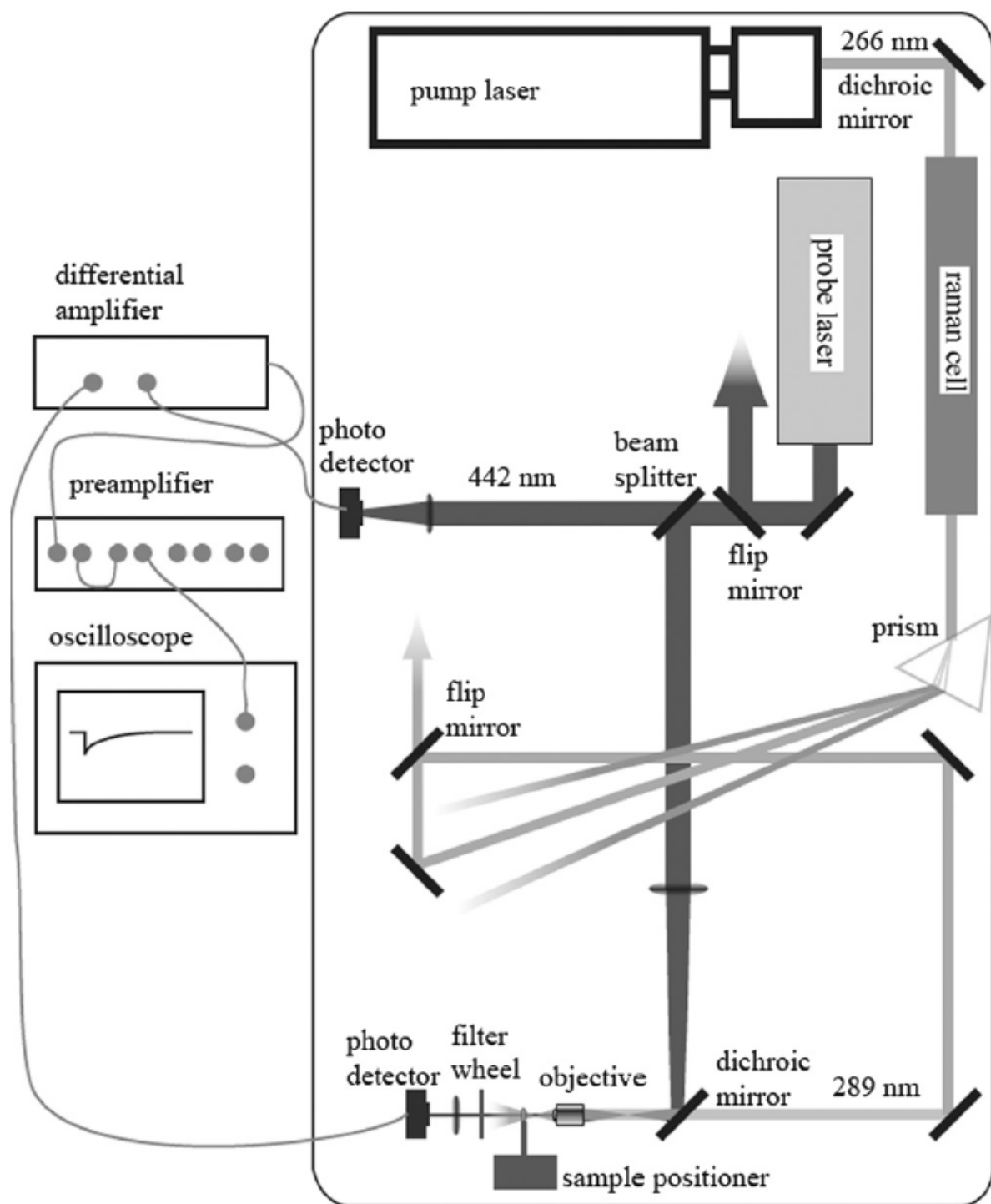


Figure 2.6 Schematic of the tryptophan triplet quenching experiment coupled with a microfluidic mixer⁶³

Chapter 3

Experiments on Several Different Proteins Using Trp/Cys Quenching Spectroscopy

3.1 Conformational properties of unfolded HypF

Amyloidgenic Protein HypF

Amyloid is one type of protein aggregation characterized by a cross-beta sheet quaternary structure. Abnormal accumulation of these insoluble fibrous amyloids in organs can cause serious amyloidosis and some other neurodegenerative diseases. HypF-N⁶⁴, the fragment of maturation factor of prokaryotic hydrogenases⁶⁵, is one of the most amyloidgenic proteins and has been shown to have one of the fastest aggregation rates *in vitro*⁶⁶. It is thought that the disordered oligomers produce the toxic species^{67, 68}, so it is quite relevant to study the conformational properties of the unfolded HypF.

It is of medical importance to answer questions like why some proteins tend to aggregate while some don't and what determines the aggregation rates. Recent work has shown a strong correlation between aggregation rates and hydrophobicity, hydrophobic patterning and charge of a sequence^{66, 69-73}; additionally helix and sheet propensity correlates significantly with mutational changes in aggregation rates within a sequence^{70, 74-77}. From these statistical analyses a general model of the relevant physicochemical parameters has emerged^{66, 70}, but it

does not fully define the physical basis for how aggregation happens. One possible mechanism is that the aggregation precursor is fairly disordered and transiently exposes hydrophobic residues to the solvent for sufficient time to form bimolecular interactions. Therefore there may be some correlation between intramolecular diffusion rates of aggregation precursors and aggregation propensity.

Campioni *et al.* have shown that the conformation populated at low pH can form prefibrillar aggregates and can be described as a pre-molten globule, defined by Uversky as an intermediate between a fully disordered chain and a molten globule⁷⁸. Furthermore, ANS fluorescence measurements during folding have shown that a collapsed state formed within the first 1 ms of folding has a significant number of hydrophobic residues exposed to solvent⁷⁹. These results support the idea that a fairly disordered state, which may even be the unfolded state under folding conditions, is the aggregation precursor for HypF-N, and therefore the rate of intramolecular diffusion should determine the rate of bimolecular association.

Experimental Results for HypF

The mutants of HypF used in this study are a kind gift from Niccolò Taddei and Claudia Parrini. Two mutants were made⁷⁸, W27F/C7S/C40S (referred to as W27F) and W81F/C7S/C40S (W81F), to remove two cysteines and one of these tryptophans so that intramolecular contact is observed between positions 65 and 81 (short loop) and positions 27 and 65 (long loop) independently. For both mutants at every concentration of denaturant, we observe one or two exponential decays of the tryptophan triplet state (Figure 3.1). The fast decay rate

represents the Trp-Cys quenching kinetics of the population of unfolded proteins in the sample and a slow rate corresponds the natural lifetime of tryptophan for folded proteins since the positions of tryptophan and cysteine are designed to be distant in folded structure. Figure 3.1 shows the relative amplitude of the slow rates (fraction of folded states) increases with decreasing concentration of denaturant. The observed contact rates for the short loop are generally faster but similar at 6M and 1.25M GdnHCl compared to the long loop (Figure 3.2(a)). The slow rates for both mutants are similar to that observed for free Trp in water ($\sim 10^4 \text{ s}^{-1}$)²³, suggesting that the tryptophans in both mutants are still solvated in their folded or intermediate states, not buried in the hydrophobic core of the folded protein. Figure 3.2(b) shows the fraction of the unfolded state for both W81F and W27F. Based on this measure of protein stability, the midpoint of folding for HypF is at about 1.25M GdnHCl.

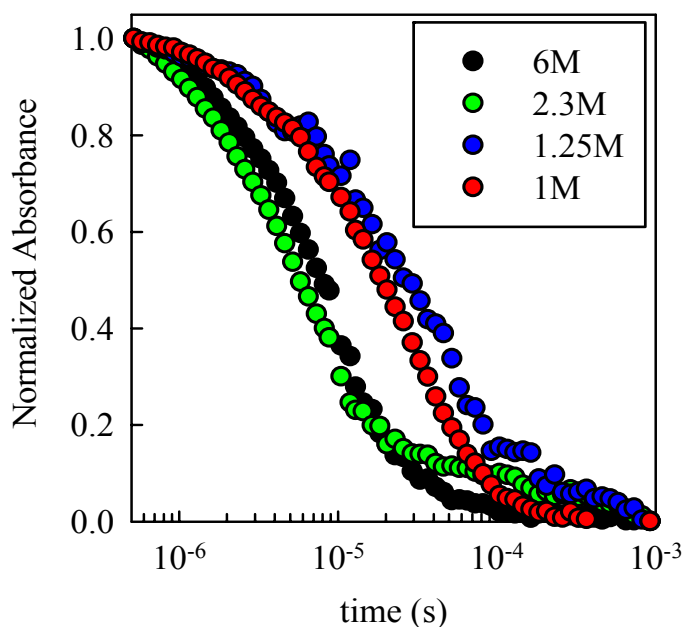


Figure 3.1 Kinetics of tryptophan triplet lifetime for the HypF mutant W81F (long loop) at various concentrations of GdnHCl. The fast decays ($\sim 10^5 \text{ s}^{-1}$) are due to intramolecular contact quenching by C65 and the slow decays ($\sim 10^4 \text{ s}^{-1}$) are due to the natural lifetime of the triplet in an aqueous environment.

The global fitting of measured decay time ($1/k_{obs}$) against viscosity at multiple temperatures guarantees the minimization of uncertainty raised by either a faulty single experiment or possible change of experimental conditions, such as optical alignment, for each measurement. Fits at different concentrations of GdnHCl are shown in Figure 3.3. Each temperature is plotted as a different color and the viscosity is increased by the addition of sucrose. Sucrose has been shown to stabilize folded proteins, but we assume that it won't change the dynamics of the unfolded states at the protein concentrations we use, especially in the presence of GdnHCl. For both mutants at 6 M

GdnHCl, each temperature lies on a different line, indicating that k_R depends strongly on temperature. However, at 1.25 M GdnHCl, W81F still shows separate lines for each temperature while W27F has converged on a single line for all temperatures, suggesting both k_R and k_{D+} have lost any dependence on temperature. The loss of temperature dependence on k_R might be due to the fact that the equilibrium rate, K , had increased far above 1 at low denaturant⁴¹. The loss of temperature dependence on k_{D+} indicates that γ in equation 2.6 is infinitely small and the effect of viscosity change on decay time has become more significant than temperature change (a steeper slope of $1/k_{obs}$). Thus the 65-81 (short) loop appears to be highly collapsed ($1/k_R \sim 0$) at low denaturant and the intramolecular diffusion rate ($k_{obs} \sim k_{D+}$) is slow enough to be independent of temperature.

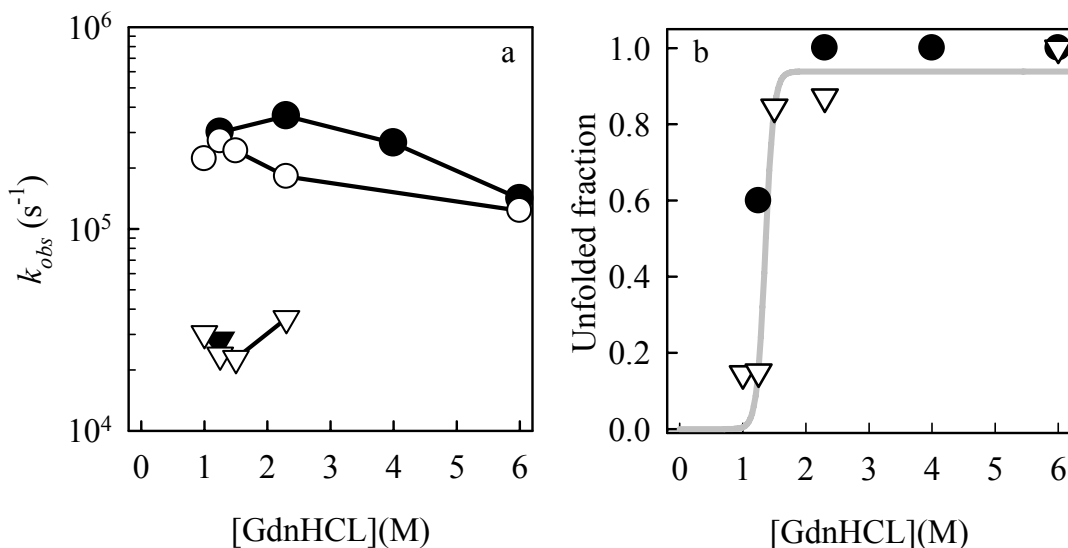


Figure 3.2 a) Observed exponential rates of the tryptophan triplet decay versus [GdnHCl] for the two mutants, W81F (white) and W27F (black). The triangles are the slow rate and the circles are the fast rate as illustrated in Figure 3.1. b) The relative amplitude of the fast rate vs. [GdnHCl] for W27F (black circles) and W81F (white triangles). The grey line is a sigmoidal fit to all data for both mutants.

Figure 3.3 The observed fast rates vs. viscosity for W27F (a,c,e) and W81F (b,d,f) in 6 M (a,b), 2.3 M (c,d) and 1.25 M (e,f) GdnHCl. Each temperature is plotted as a separate color. The lines are fits of all the data in each plot to Eqs. 2.5 and 2.6.

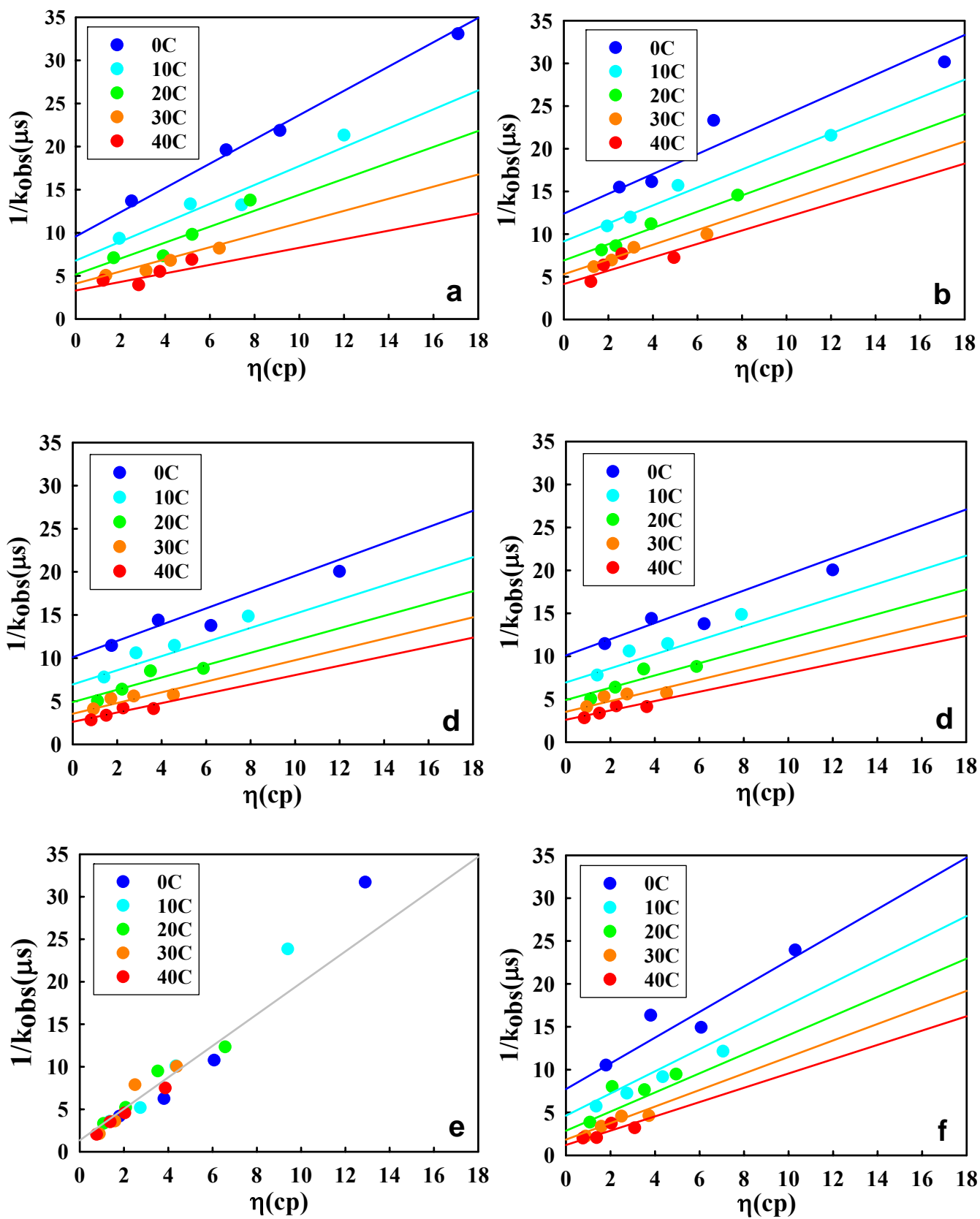


Figure 3.3

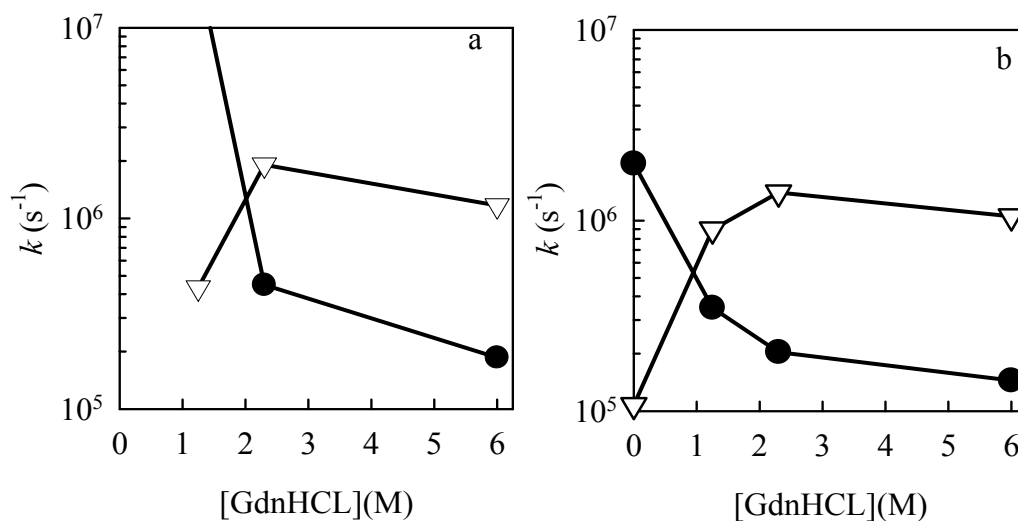


Figure 3.4 Reaction-limited (black circles) and diffusion-limited (white triangles) rates vs. [GdnHCl] for a) W27F (short loop) and b) W81F (long loop) at 20 C and 1 cP. The points at 0 M GdnHCl are those measured in 3% TFE and determined from the slope and intercept of Figure 3.5b.

Figure 3.4 shows the reaction-limited (red) and diffusion-limited (blue) rates after the decomposition using equations 2.5 and 2.6. Between 6 M and 2.3 M GdnHCl, k_R and k_{D+} for both mutants increase slightly, but below 2.3 M GdnHCl, the diffusion-limited rates start to decrease, more significantly for W27F than for W81F. For the W81F (long loop) mutant, the reaction-limited rate continues to increase, while the W27F (short loop) mutant has no measurable reaction-limited rate and is presumed to be greater than $1 \times 10^7 \text{ s}^{-1}$ (see the y-axis intercept in Figure 3.39(e)). Measurements of k_R and k_{D+} were not possible below 1.25 M GdnHCl because the addition of sucrose stabilized the folded state of the protein, preventing the observation of quenching in the unfolded state.

Loop Formation Kinetics in TFE

2,2,2-Trifluoroethanol (TFE) at low concentrations acts as a protein denaturant. However, unlike GdnHCl, TFE unfolds proteins without providing solvent protection of hydrophobes so it promotes bimolecular association and consequent aggregation^{74, 78, 80}. Equilibrium measurement of Trp/Cys quenching in TFE enables us to estimate the polymer dynamics of the unfolded state, which may also be the aggregation precursor, in aqueous solution. In these experiments, the protein was dissolved and measurements began within a few minutes. Distinguishable fast decay traces have only been observed for W81F, probably because the 65-81 loop remains same compact and viscous as in 1.25M GdnHCl. Two TFE concentrations (v/v) 6% and 3% have been used for W81F and there was visible aggregation in the sample within 30 minutes and 2 hours, respectively. The observed kinetics is similar between 3% and 6% TFE concentrations but the lower concentration allows sufficient time for temperature dependent measurements. Figure 3.5 shows the tryptophan triplet decay at various times after dissolution in 3% TFE solution. The first trace shows that the majority of the population decays before 10 μ s while later measurements still have significant population at that time. Each of these traces can be fit to 2 exponentials between 400 ns and 1 ms. The relative amplitude of unfolded state and decay times of these fits are shown in Table 3.1.

We attribute the slow decay to solvent quenching of the tryptophan in either the folded or aggregated state. Since W27F in W81F mutant is solvent exposed in the native state⁷⁸ with a lifetime of 40 μ s, I assume the increase in the slow lifetime over the measurement period is due to a contribution from an aggregate in which the Trp is somewhat buried from solvent.

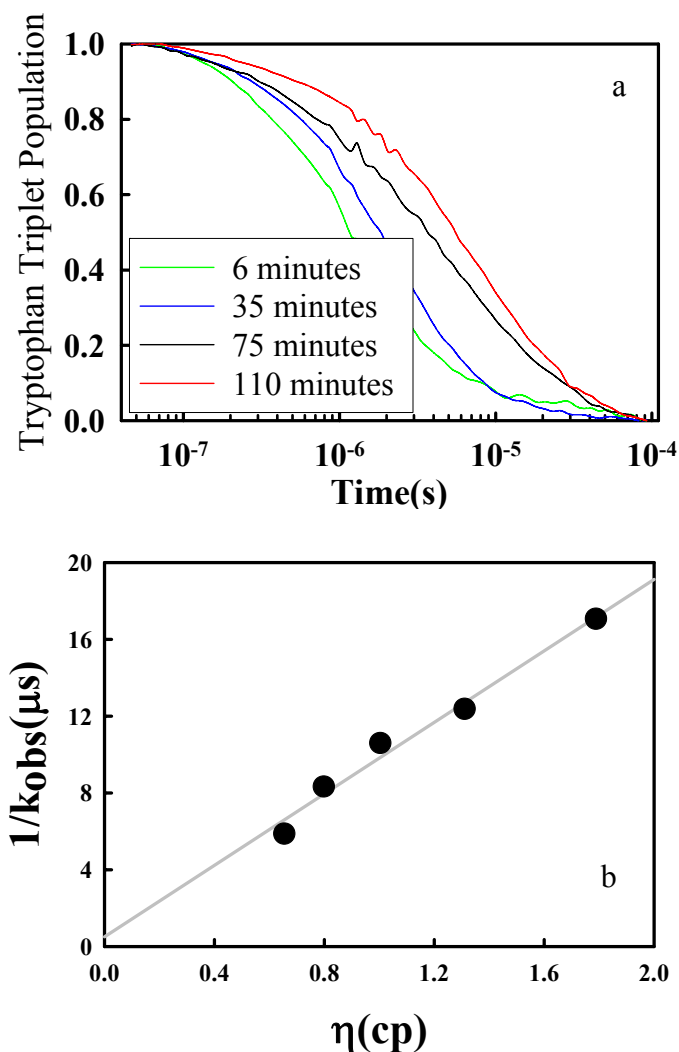


Figure 3.5 a) Tryptophan triplet decay of W81F in 3% TFE, T=20 C at various times after dilution. b) The observed fast rates vs. viscosity for W81F in 3% TFE for T= 0, 10, 20, 30, and 40 C ~6 minutes after dilution.

I assume that the fast decay in 3% TFE is a transiently populated unfolded state that is the precursor to aggregation. Directly determining the reaction-limited and diffusion-limited rates of the unfolded state was not possible because the addition of sucrose stabilized the folded state of the protein. Therefore I measured the triplet lifetime at various temperatures and assumed that there was a minimal temperature dependence for k_R or k_{D+} and fit the rates at all temperatures to a

single line (Figure 3.5b). Assuming 3% TFE approximately represents the solvent conditions in the absence of GdnHCl, I plot these rates at $[\text{GdnHCl}] = 0 \text{ M}$ on Figure 3.4b. If there is a temperature dependence for k_R and k_{D+} such that k vs. η can be fit to separate lines at different temperatures, then our fit in Figure 3.5b of the intercept is underestimated and the slope is overestimated. Therefore the points at 0 M GdnHCl in Figure 4 represent an upper limit on k_R and a lower limit on k_{D+} . I conclude that the intramolecular diffusion rate for W81F decreases by no more than a factor of 10 between 6 M and 0 M GdnHCl. This means the long loop remains fairly diffusive in native conditions.

Diffusional Dynamics of Unfolded HypF

The underlying polymeric dynamics that produces the observed contact rates was determined using SSS theory and WLC model (Chapter 2.4). Chains with full peptide length of 91 residues have been constructed. Since the reaction-limited rate is relatively insensitive to the persistence length⁴¹, it is again kept fixed at $l_p = 4 \text{ \AA}$ and the excluded diameter (d_α) is varied with denaturant concentration. In detail, once two million chains were generated using the method described in chapter 2.4 at a chosen d_α , the corresponding distribution of Trp/Cys distance was plugged into equation 2.9 for numerical integration to find k_R . The next procedure is to compare this theoretically calculated k_R to the experimental value and if they match each other, we believe that an accurate distribution has been found for that particular GdnHCl concentration. Otherwise, the value of d_α must be adjusted accordingly until agreement is reached. One needs to decrease d_α to

match a larger k_R since the k_R is inversely proportional to chain volume. The parameters of the appropriate distribution $P(r)$ for each denaturant concentration and the correspondingly determined effective diffusion coefficient, D , are listed in Table 3.2. One might notice that excluded diameter d_α needs to be adjusted to 0 to match the fastest k_R . Since d_α cannot be negative, this means the limitation of the WLC model on compaction of the chain conformations has been reached. In order to overcome it, additional attractive potentials that can further collapse WLC conformations need to be introduced (see chapter 4).

From $\langle r \rangle$ and D in Table 3.2, we see that the degree of compaction and loss of diffusivity is different in separate loops in the chain below 2.3 M GdnHCl. The C65-W81 loop compacts and loses diffusivity significantly more than the W27-C65 loop. The W27-C65 loop encompasses one of the putative amyloidogenic regions of the protein, based on mutational studies of the homologous protein acylphosphatase, while the C65-W81 loop does not^{81, 82}, so these results suggest that the amyloidogenic region of the chain is more diffusive than surrounding regions.

The physical basis for the increased diffusivity of the long loop could be due to the fact that there is significant net charge in the long loop while the short loop, like the whole chain, is almost neutral. This would force the long loop to adopt more extended conformations. However, the long loop still contains many hydrophobic residues. Using the Kyte-Doolittle scale⁸³, the mean hydrophobicity (-0.41) and the standard deviation (3.36) of the long loop is almost the same as the short loop or the entire chain. Also, the hydrophobic pattern of the entire chain oscillates rapidly, with no large clusters of hydrophobic residues to direct collapse. Therefore the W27-C65 region of the chain, forced into extended conformations by electrostatic repulsion that are not

counterbalanced by large hydrophobic clusters pulling the chain together, will have hydrophobes that are more often exposed to solvent and to other, nearby chains, making bimolecular hydrophobic interactions more likely.

Table 3.1 Observed rates of tryptophan triplet decay of W81F at various times after dilution in 3% TFE

time(min)	fast rate (s^{-1})	slow rate (s^{-1})	fraction fast rate
6	94364	24102	0.70
35	123948	13157	0.27
72	113546	11647	0.27
110	113459	10630	0.24

Table 3.2 Wormlike chain parameters and effective diffusion coefficients for both mutants studied

[GdnHCl]	k_R (s^{-1})	k_{D+} (s^{-1})	l_p (Å)	d_α (Å)	$\langle r \rangle$ (Å)	D ($\times 10^6$ cm ² s ⁻¹)
W81F						
0 ^a	1990089	107408	4	0	30.6	0.02
1.25	347950	896190	4	2.75	35.5	0.51
2.3	203920	1401100	4	3.20	37.0	1.15
6	144830	1049700	4	3.50	38.1	1.07
W27F						
1.25	4.8E+14	433088	4	0	19.9	0.02
2.3	448200	1913100	4	3.20	24.2	0.58
6	185960	1167500	4	4.10	25.1	0.47

^a The solution contained 3% TFE and was measured 6 minutes after dilution.

3.2 Intramolecular Loop Formation of Intrinsically Unfolded Proteins

Apocytochrome C

Intrinsically unstructured (natively unfolded) proteins provide the possibility to study the dynamics of unfolded polypeptides under both denaturing and physiological conditions. After the heme is removed from horse heart cytochrome C, this protein becomes unstructured⁸⁴. The wildtype sequence of apocytochrome C has two cysteine residues at position 14 and 17, either of which can make contact with the tryptophan residue at position 59. Since these two cysteine residues are close, the formation of either loop will demonstrate very similar dynamics. Apocytochrome C is unfolded at all concentrations of denaturant so measurement of tryptophan triplet lifetime yields one exponential decay rate due to contact quenching.

As usual, experiments were carried out at four different GdnHCl concentrations (0, 1.5 M, 2.3 M and 6 M) and the decay trace of the tryptophan triplet absorbance at 441nm was recorded at various temperatures (0 C, 10 C, 20 C, 30 C and 40 C) and viscosities (v/v : 0%, 10%, 20%, 30% and 40% of sucrose) for each GdnHCl concentration. From Figure 3.6, we can see k_R increases as GdnHCl concentration decreases, indicating that unfolded apocytochrome C molecules become more compact in the absence of denaturant. On the other hand, the diffusion-limited rate decreases as GdnHCl concentration decreases, indicating a less diffusive natively unfolded state.

When generating wormlike chains for apocytochrome C, I set the contour length of the Typ-Cys loop to be 42 residues and each terminal tail to be 5 residues since it was proven that the reaction-limited rate is relatively insensitive to the length of the tail⁸⁵. As a result, each chain was only 52 residues long after the truncation and computational time decreased about a half.

WLC parameters and corresponding D for different GdnHCl concentrations are listed in Table 3.3. We can see that D decreases by a factor of 5 from 6 M to 0 M GdnHCl. However compared to that of protein L in the same nondenaturing condition⁴¹, this effective diffusion coefficient is still 2 orders greater, indicating that intrinsically unstructured protein apocytochrome C is much more diffusive than well-folded protein L in aqueous solution.

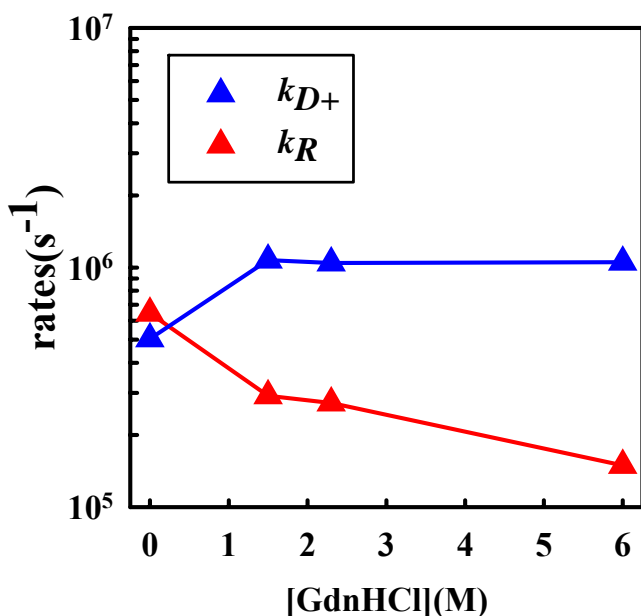


Figure 3.6 Reaction-limited (red triangles) and diffusion-limited (blue triangles) rates vs. [GdnHCl] for apocytochrome C

Table 3.3 Wormlike chain parameters and effective diffusion coefficients for apocytochrome C

[GdnHCl]	k_R (s ⁻¹)	k_{D+} (s ⁻¹)	l_p (Å)	d_α (Å)	$\langle r \rangle$ (Å)	D (x10 ⁶ cm ² s ⁻¹)
0	626989	504390	4	1.98	35.2545	0.2100
1.5	287532	1074200	4	2.86	37.7664	0.7700
2.3	267253	1043900	4	2.90	37.9181	0.7600
6	157440	1051700	4	3.32	39.4323	1.0900

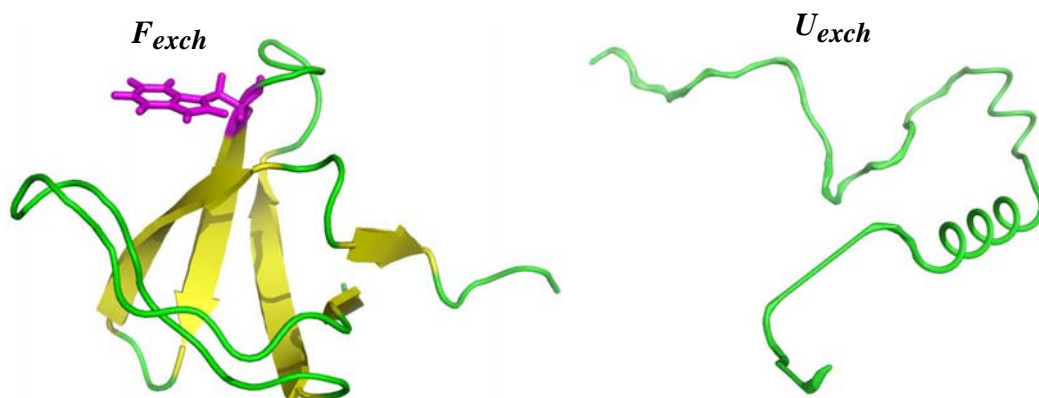


Figure 3.7 Structural cartoon of F_{exch} and U_{exch} of DrkN

DrkN

Drosophila drk N-terminal (drkN) SH3 domain is a metastable protein, which exists in approximately 1:1 equilibrium between folded (F_{exch}) and unfolded (U_{exch}) states in water^{86, 87} (Figure 3.7). Thus, I expected to see some fast kinetics due to contact quenching under nondenaturing conditions. However, I didn't. Both mutants of drkN (C2 and C60) diffuse so slowly that I can still see a slow decay ($\sim 10^4 \text{ s}^{-1}$) even in 6 M GdnHCl (Figure 3.8). As a result, I was only able to conduct temperature and viscosity dependent measurements of these two mutants in 2 M and 6 M GdnHCl. In both conditions, the fraction of the amplitude of fast rate to slow rate is about 1:1 (Figure 3.8). Because it is almost impossible to determine whether the slow rate is due to contact quenching by slowly approaching cysteine residues or just a reflection of the tryptophan natural lifetime, I decided to extrapolate the diffusion-limited and reaction-limited

rates only from the fast observed rates. Another interesting finding is the nonlinear dependence of k_{obs}

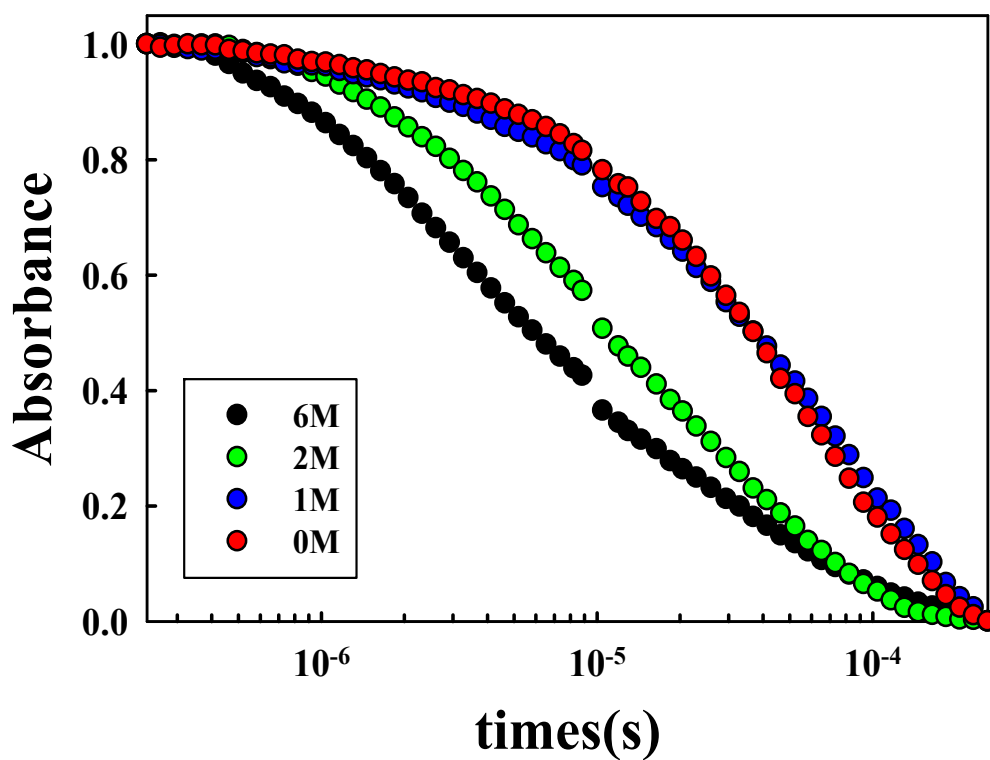


Figure 3.8 Kinetics of tryptophan triplet lifetime for the drkN mutant C60 at various concentrations of GdnHCl

on $1/\eta$ for both mutants in 6 M GdnHCl as shown in Figure 3.9, consequently equation 2.6 was modified to account for the observed curvature⁸⁵.

$$k_{D+}(T) = \frac{k_{D+0}T}{(\eta - A\eta^2)T_0} \exp(\gamma(T - T_0)) \quad (3.1)$$

In 2 M GdnHCl, both proteins seem to converge on a single straight line for all temperatures, similar to what we saw for the W27F mutant of HypF. This suggests a complete loss of temperature dependence of k_R and k_{D+} .

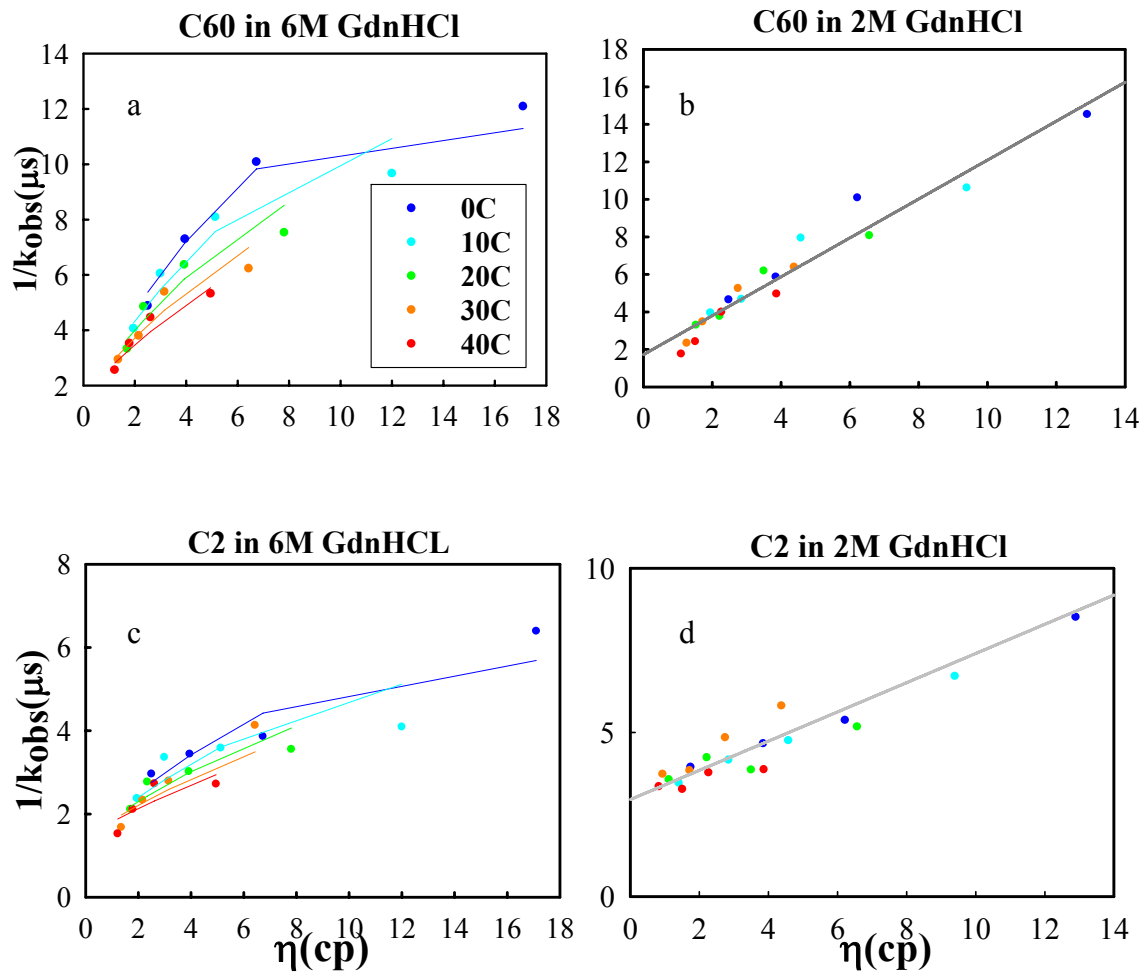


Figure 3.9 The observed fast rates vs. viscosity for C2 (c, d) and C60 (a, b) in 6 M (a, c), 2 M (b, d) GdnHCl. Each temperature is plotted as a separate color. The lines in (a) and (c) are fits of all the data to Eqs. 2.5 and 3.1.

The extracted reaction-limited and diffusion-limited rates are shown in Figure 3.10. We see few changes of k_R and k_{D+} for both C2 and C60 between 2 M Gdn and 6 M GdnHCl. This tells us

both portions of the drkN protein have the same compaction and diffusivity as the GdnHCl concentration decreases. Thus, due to the slow and invariable kinetics, this protein seems not as interesting as it is expected to be for the intramolecular loop formation study. But some additional information available for this protein, such as the secondary structure propensity determined by experimentally measured chemical shifts, does make it a good calibration object for an energy reweighted WLC model, which will be described in detail in chapter 4.

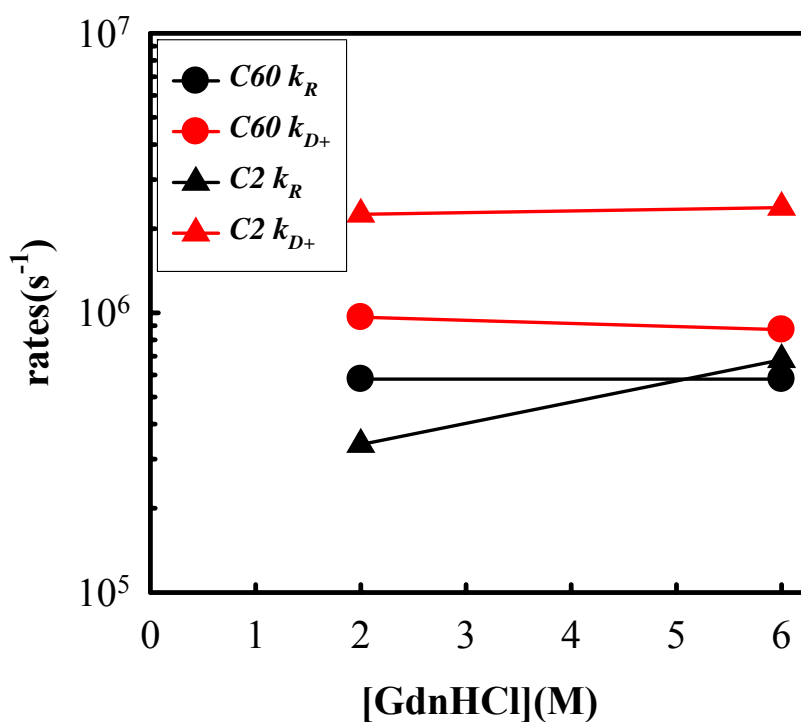


Figure 3.10 Reaction-limited (black) and diffusion-limited (red) rates vs. [GdnHCl] for C2 (triangle) and C60 (circle) at 20 C and 1 cP

AavLEA

Late embryogenesis abundant (LEA) is a special group of intrinsically disordered proteins. They are natively unstructured in water but do begin to aggregate when undergoing dessication

or when they are placed in a non-polar solution⁸⁸. LEA proteins were first found in maturing plant seeds and have since been found in invertebrates and bacteria^{88, 89}. Most organisms are dependent on the presence of water in a cell, however LEA can not only survive desiccation but return to their pre-desiccation states of function after rehydration. This process of desiccation tolerance is known as anhydrobiosis and LEA proteins take on a significant role in anhydrobiosis. More importantly, they can also prevent aggregation in numerous other proteins during desiccation⁸⁸.

The particular LEA protein used in our experiments was AavLEA1 which is found in the nematode *Aphelenchus avenae*⁸⁸. This intrinsically disordered protein has a very high content of charged residues that can prevent intrachain collapse. The measurements were not taken directly by me but my lab colleagues Eric Buscarino and Basir Ahmad. Two mutants S2C and T58C were made. Tryptophan was located at position 30 so the chain lengths between the tryptophan and cysteine were 28 for both mutants. Due to their intrinsically disordered nature, AavLEA1 was not at risk to aggregate quickly but to maintain consistency in our measured rates we used each sample within one - three hours of thawing and dilution. The experiments were performed at pH = 7.5 in order to neutralized histidine residues. Figure 3.11 shows the reaction-limited and diffusion-limited rates of the two mutants of LEA. The kinetics of these two loops is very similar and this protein is still highly diffusive and less compact even in 0 M GdnHCl. We believe the 50% charged residues in the protein sequence account for its extended and unstructured nature. To better model this protein theoretically, I will introduce a method to include the charge effect in the WLC model in the next chapter.

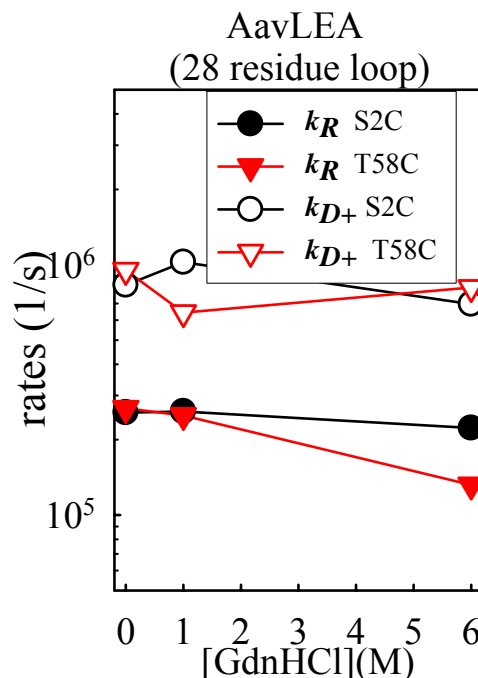


Figure 3.11 Reaction-limited (filled) and diffusion-limited (open) rates vs. [GdnHCl] for S2C (black circle) and T58C (red triangle) at 20 C and 1 cP. Both loops are equally 28-residue long.

3.3 Diffusion of Unfolded Acyl-coenzyme A-binding Protein over A Complete Range of Denaturant

In order to have a complete knowledge of the intramolecular diffusion of proteins with different aggregation propensities, well-behaved acyl-coenzyme A-binding protein (ACBP) has been studied over a wide range of denaturant in both equilibrium and using the microfluidic mixer described in chapter 2. This 86-residue protein contains four helices and folds on a 10ms timescale⁹⁰. Even though ACBP has long been established as a two-state folder^{90, 91}, Teilum et al. revealed the presence of a $\sim 80 \mu s$ phase, suggesting a three-state model with a partially-folded intermediate^{92, 93}. Recent theoretical study using the Markov state models built from atomistic

simulations of ACBP, however, revealed a highly diffusive network of metastable unfolded states instead of a single well-defined intermediate. Therefore, measuring intramolecular loop formation using Trp/Cys contact quenching technique may provide an excellent tool to reveal the underlying dynamics.

Two mutants of ACBP (T17C/W55A and W55A/I86C) were used in this study. The only tryptophan after mutation sits at position 58 and can make contact with cysteine at either position 17 or position 86. Equilibrium Trp/Cys quenching experiments were performed at 6 M, 2.3 M and 1.5 M GdnHCl for both T17C and I86C. Because I86C has a denaturation midpoint slightly lower than the other mutant³⁵, I was also able to collect equilibrium data also at 1.25 M GdnHCl for this mutant. Reaction-limited and diffusion-limited rates were extracted from the observed quenching rate using equations 2.5 and 2.6 and plotted in Figure 3.12 (c and d). As the denaturant concentration decreases from 6 M to 1.25 M, these two mutants yield very similar trends, an increasing reaction-limited rate and a decreasing diffusion-limited rate, which is also observed for most other proteins studied using Trp-Cys quenching.

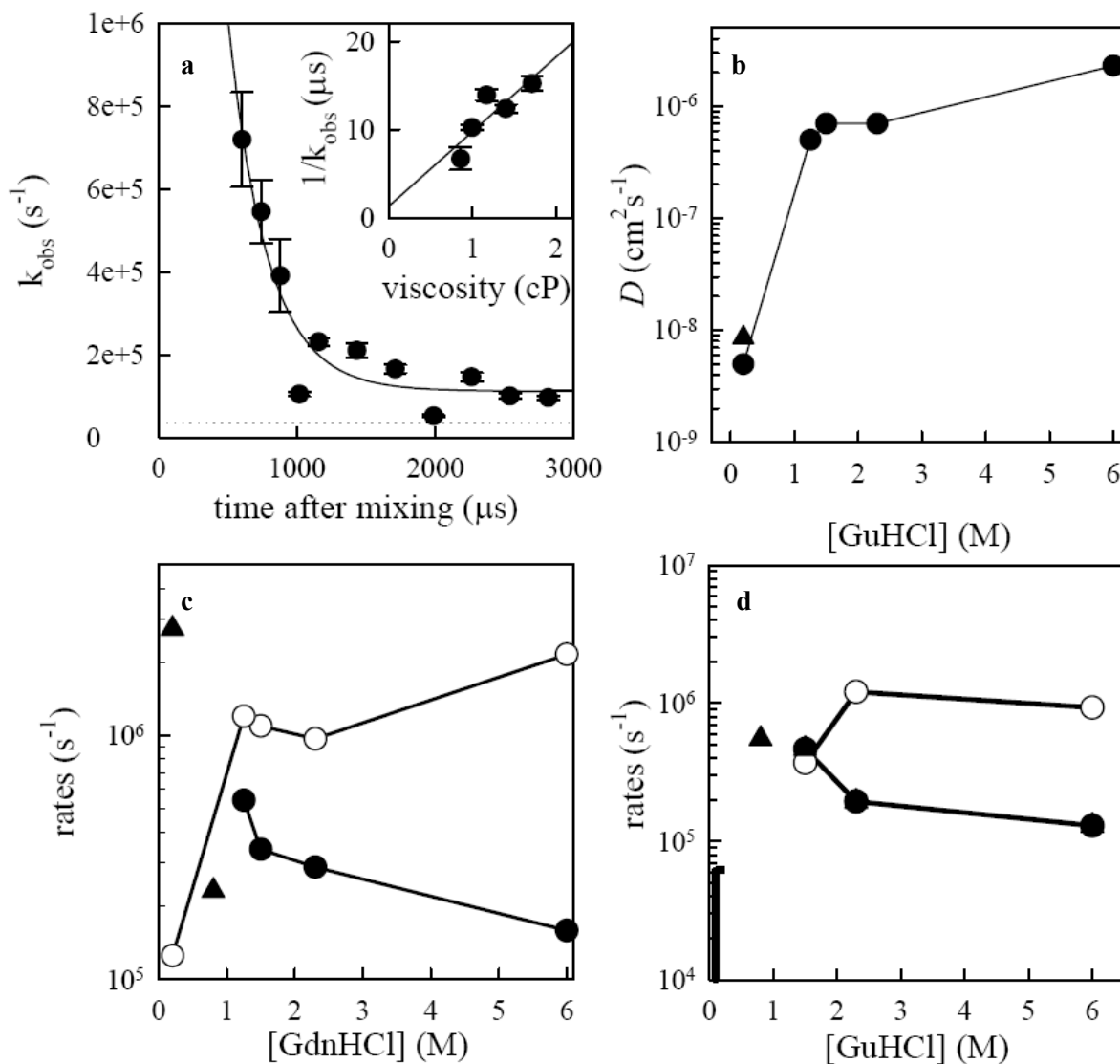


Figure 3.12 Trp/Cys contact-quenching studies of ACBP unfolded states. (a) Observed quenching rates vs. time after mixing for I86C. Inset: observed quenching rate vs. viscosity. (b) The effective diffusion coefficients at various GdnHCl concentrations. (c) Reaction-limited (filled) and diffusion-limited (open) rates for I86C (d) Reaction-limited (filled) and diffusion-limited (open) rates for T17C. Black triangles are measured values from molecular dynamics simulations.

An ultrarapid mixer as previously described in chapter 2.5 has also been employed to capture the dynamics of ACBP before folding. Unfolded 1 mM ACBP protein in 4 M GdnHCl was pushed through the mixer and combined with buffer without denaturant at a ratio of 1:20. After passing

through the serpentine mixing region, the protein enters the wide observation channel with a final GdnHCl concentration of ~0.2 M. The protein itself is also diluted to 50 μM after mixing. The total flow rate is 210 $\mu\text{L}/\text{min}$ ((200+10) $\mu\text{L}/\text{min}$). Most observed rates measured along the exit channel can be well fit with single exponential decay. Figures 3.12 (a) and 3.13 show the observed rate of ACBP I86C and T17C, respectively, as a function of time after mixing based on the position in the channel. The rates of T17C are overall slower and comparable to the natural lifetime of tryptophan (gray dashed line), suggesting that the 17-58 loop might be too long for this kinetics study. But the 58-86 loop (I86C) yields rates $> 10^5 \text{ s}^{-1}$. The decrease of the rate over time after mixing is probably due to the collapse of protein during mixing and was fit into a single exponential decay. The inset in Figure 3.12 (a) shows the observed rate of I86C 1432 μs after mixing as a function of viscosity. The rates in various sucrose concentrations (v/v: 0%, 5%, 10%, 15% and 20%) can be fit to a straight line with a zero intercept within error, indicating that the observed rate is diffusion-limited. So the observed rate at $\eta = 1\text{cp}$ directly gives $k_{D+} = 125000 \text{ s}^{-1}$. The reaction-limited rate at 0.2 M GdnHCl in Figure 3.7 (c) is obtained from molecular dynamics simulations performed at 300K and is used to determine the effective diffusion coefficient using SSS theory, which gives $D = 5 \times 10^{-9} \text{ cm}^2 \text{ s}^{-1}$. The effective diffusion coefficients at other denaturant concentrations are exhibited in Figure 3.12 (b). So the experimental results exhibits similarly slow diffusion as seen in molecular dynamics simulation, which gives $D = 8.579 \pm 1.95 \times 10^{-9} \text{ cm}^2 \text{ s}^{-1}$ obtained by measuring the mean-squared displacement over time³⁷. Given the average radius of gyration of the unfolded states of ACBP (2~3 nm), the reconfiguration time through intramolecular diffusion can be estimated to be $\sim 10 \mu\text{s}$. These results validate the

suggestion of recent simulational study that the diffusion of the unfolded states of ACBP is too slow to form an intermediate in the early microsecond timescale. This protein, however, is still shown to be more diffusive than protein L in physiological conditions³⁸, suggesting a strong sequence dependence of intramolecular diffusion.

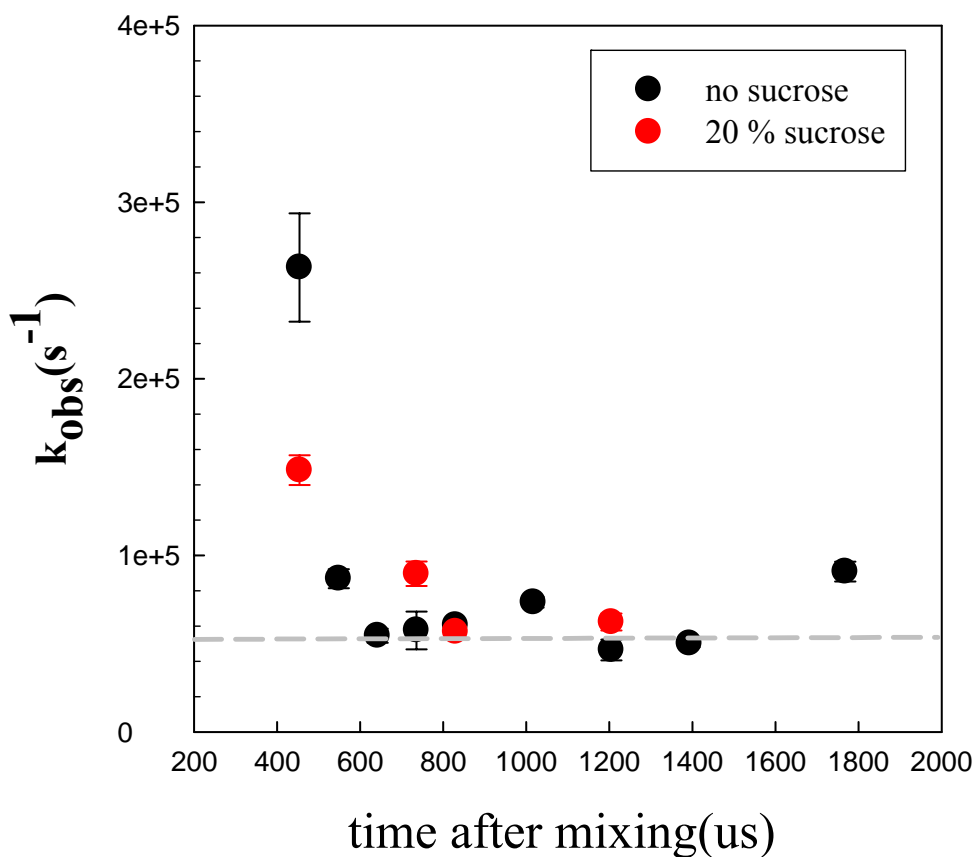


Figure 3.13 Observed quenching rates vs. time after mixing for T17C. Dashed line represents the natural lifetime of tryptophan triplet state.

Chapter 4

Sequence Specific Wormlike Chain Model

4.1 Introduction to the Heteropolymer Model

There is considerable evidence that highly denatured states act globally like random, non-interacting chains restricted by excluded volume of the chain and that this chain reconfigures on the nanosecond-microsecond timescale⁴⁴. Thus, a wormlike chain model with SSS theory is sufficient to describe these states and has been shown to yield a corresponding Trp/Cys distance distribution shift as one gradually decreases the excluded diameter (d_α) with denaturant concentration. It is not clear if such a model can be extended to physiologically relevant conditions. The model has been shown to lead to extremely small values of excluded volume for solvent conditions that favor folding and we have found that for some sequences even $d_\alpha = 0$ does not make a chain compact enough to match experimental k_R . A growing number of unfolded proteins have been shown by NMR methods to have some residual structure under folding conditions³¹⁻³⁵, but these methods do not fully characterize the distribution of conformations and cannot measure the reconfiguration time within the distribution. The experimental work in this thesis shows that unfolded proteins under folding conditions are much more compact than proteins in high denaturant. All these observations in low denaturants suggest more complicated unfolded states with transient intrachain interactions and secondary structure formation. Therefore a homogenous

wormlike chain model is obviously too simple to describe an unfolded protein under folding conditions.

This chapter shows that a sequence-specific wormlike chain model that is statistically reweighted to favor highly collapsed conformations of the chain can reproduce experimental contact quenching rates measured under folding conditions. The motivation for this work is to find a realistic description of the probability distribution of unfolded states that did not require enormous amounts of computation. Similar efforts have been successful in comparing NMR data to randomly generated chains or molecular dynamics simulations^{31, 94}. But because I intended to calibrate my computational model with an experimental method that is only sensitive to the tail of the distribution of random conformations, we required many more than the ~10000 conformations generated by previous methods in order to have statistical accuracy. Furthermore, I did not want to bias our distribution with native-like structure that may not be representative of the full range of accessible conformations. Therefore, I chose a simple wormlike chain model with excluded volume and a bias for short-range contacts between residues of similar hydrophobicity. This model reproduces experimental rates from different proteins under a variety of denaturant concentrations that vary over more than an order of magnitude. After determining the magnitude of reweighting from contact quenching, I then show that this model successfully predicts measured values of paramagnetic resonance enhancement for two different proteins.

I introduce this model by first placing it in context of previous models and methods. Almost 20 years ago, Dill and coworkers sought to describe protein folding in terms of the balance

between preferential de-solvation of different amino acids and chain conformational entropy⁹⁵. In most subsequent work, the effect of denaturant is then modeled with a change in the solvent accessible surface area (ASA) of various amino acids sidechains and/or backbone⁹⁶⁻⁹⁹. A series of papers by Thirumalai and coworkers elucidates these ideas very well¹⁰⁰⁻¹⁰². They first simulate the dynamics of a coarse-grained model of a protein without denaturant and then use the Tanford transfer model to produce statistical distributions of conformations at various denaturant concentrations. This model has been remarkably successful in reproducing average intramolecular distances of denatured proteins measured by single-molecule FRET¹⁰².

For this work I have taken a complimentary approach. Because coarse-grained atomistic models cannot produce sufficiently random conformations in high denaturant, I have chosen to start with a wormlike chain model that has proven to be a realistic model of many highly denatured proteins over a range of sampled conformations e.g. the average radius of gyration from SAXS, the longest extensions from single-molecule stretching and the shortest distances from intramolecular quenching^{44, 85, 103}. Furthermore, because a wormlike chain model is inherently random, I needed a method to choose those conformations which would be favored when denaturant is decreased. Simply decreasing ASA does not explicitly do that because it does not contain specific intramolecular interactions. Therefore I assign an energy to each conformation based on favorable interactions between residues of similar hydrophobicity.

4.2 Boltzmann Energy Reweighting

Wormlike chains are computationally generated in the same way as described in Chapter 2.

For each wormlike chain conformation of a particular sequence, I define a total energy, E_{TOT} , that is the sum of pairwise interactions between the C^α atoms,

$$E_{TOT} = - \sum_{|i-j|>1} \frac{e_{i,j}}{|r_i - r_j|} \quad (4.1)$$

$$e_{i,j} = \begin{cases} 0, & |h_i - h_j| > 0.3 \\ \sigma, & |h_i - h_j| \leq 0.3 \end{cases}$$

where, i and j are the indices of two non-adjacent residues, $|r_i - r_j|$ is the distance between these two residues which must be larger than d_α and less than 6.5\AA (a reasonable cutoff distance for hydrophobic associations¹⁰⁴), and h_i and h_j are the normalized hydrophobicities of the two residues as determined by the Miyazawa

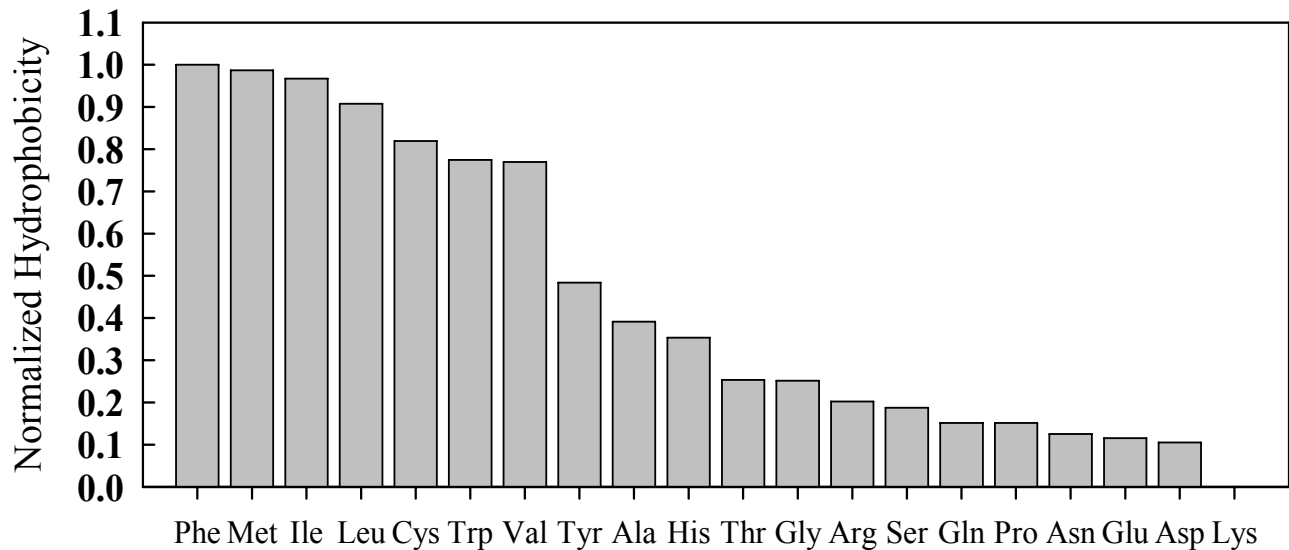


Figure 4.1 Hydrophobicity scale used in the energy reweighted WLC model. The values are taken from the Miyazawa-Jernigan scale and normalized to span the interval 0 to 1.

Jernigan scale (see Figure 4.1). e_{ij} is the strength of the hydrophobic potential between two residues which equals zero if one is hydrophobic and one is hydrophilic and equals an adjustable parameter, σ , if both are hydrophobic or hydrophilic. I used 0.3 as the cutoff value because it corresponds to the spread of the normalized hydrophobicity distribution of all amino acids. σ (in units of $kT \cdot \text{\AA}$) exhibits the strength of the hydrophobic interaction and is tunable according to various solvent conditions.

Using E_{TOT} and r as independent reaction coordinates, I can define a 2-D probability distribution $P(r, E_{TOT})$ so that

$$1 = \int dr \int dE_{TOT} P(r, E_{TOT}) \quad (4.2)$$

and the 1-d Trp-Cys distance distribution is defined as $P(r) = \int P(r, E_{TOT}) dE_{TOT}$. Let $Z(r)$ represent the energy reweighted distribution of distance r between the Trp and Cys C^α position. $Z(r)$ can be written as

$$\begin{aligned} Z(r) &= N \cdot \int P(r, E_{TOT}) \exp(-E_{TOT} / kT) dE_{TOT} \\ &= N \cdot \int P(r, E_{TOT}) dE_{TOT} \cdot \frac{\int P(r, E_{TOT}) \exp(-E_{TOT} / kT) dE_{TOT}}{\int P(r, E_{TOT}) dE_{TOT}} \\ &= N \cdot P(r) \cdot \langle \exp(-E_{TOT} / kT) \rangle_r \\ \frac{1}{N} &= \int_{d_\alpha}^{l_c} P(r) \cdot \langle \exp(-E_{TOT} / kT) \rangle_r dr \end{aligned} \quad (4.3)$$

N is a normalization constant. For any given set of conformations with the same r (within 0.1 \AA), the distribution of E_{TOT} is typically Gaussian (central limit theorem) but there are often outlying conformations that significantly skew $Z(r)$ due to the exponential averaging. Therefore, the limits

of the integral in equation 4.3 are set at two standard deviations below and above the mean, eliminating ~2% of conformations.

To compare $Z(r)$ to experimental measures of intrachain distance I again need to use Szabo, Schulten and Schulten (SSS) theory (Chapter 2.4). After substitution of $P(r)$ by $Z(r)$, equations 2.9 and 2.10 become

$$k_R = \int_{d_\alpha}^{\infty} q(r)Z(r)dr \quad (4.4)$$

$$\frac{1}{k_{D+}} = \frac{1}{k_R^2 D} \int_{d_\alpha}^{l_c} \frac{dr}{Z(r)} \left\{ \int_r^{l_c} (q(x) - k_R)Z(x)dx \right\}^2 \quad (4.5)$$

where r is the distance between the tryptophan and cysteine and $Z(r)$ is given by equation 4.3. D is the effective intramolecular diffusion constant and $q(r)$ is the distance dependent quenching rate.

An important question one might ask is “Is the Trp/Cys distance r a good reaction coordinate?” The answer will be certainly yes if the distribution is Gaussian, because a Gaussian chain should yield the same scaling for the radius of gyration, the end-to-end distance, any intrachain distance or other length measurements. For sufficiently long wormlike chains (> 10 residues) with persistence length $l_p = 4 \text{ \AA}$, it has been shown that the distributions of the end-to-end distance are well-described by a Gaussian distribution⁵⁴. Since the radius of gyration has been established as a good measure of the volume of polymers, especially proteins, how well the Trp/Cys distance is correlated to radius of gyration should help us determine if this measure is also a good reaction coordinate. Figure 4.2 shows the rank correlation between

radius of gyration (R_g) and Trp47/Cys23 (r_{23-47}) (a) and between R_g and Trp47/Cys57 (r_{47-57}) (b) of protein L. These figures are created in the following way: the Trp/Cys distance and R_g of chain i are each converted to a rank over 2 million chains, and then I use these two ranks to generate a 2-D histogram. While r_{23-47} yields a very clear positive correlation with R_g , the correlation between r_{47-57} and R_g is not so obvious. This is probably because the Trp47/Cys57 loop is so short that it might be heavily constrained by the stiffness of the chain. But as long as these two ranks do not correlate in an opposite way, we are safe to use the Trp/Cys distance as the reaction coordinate.

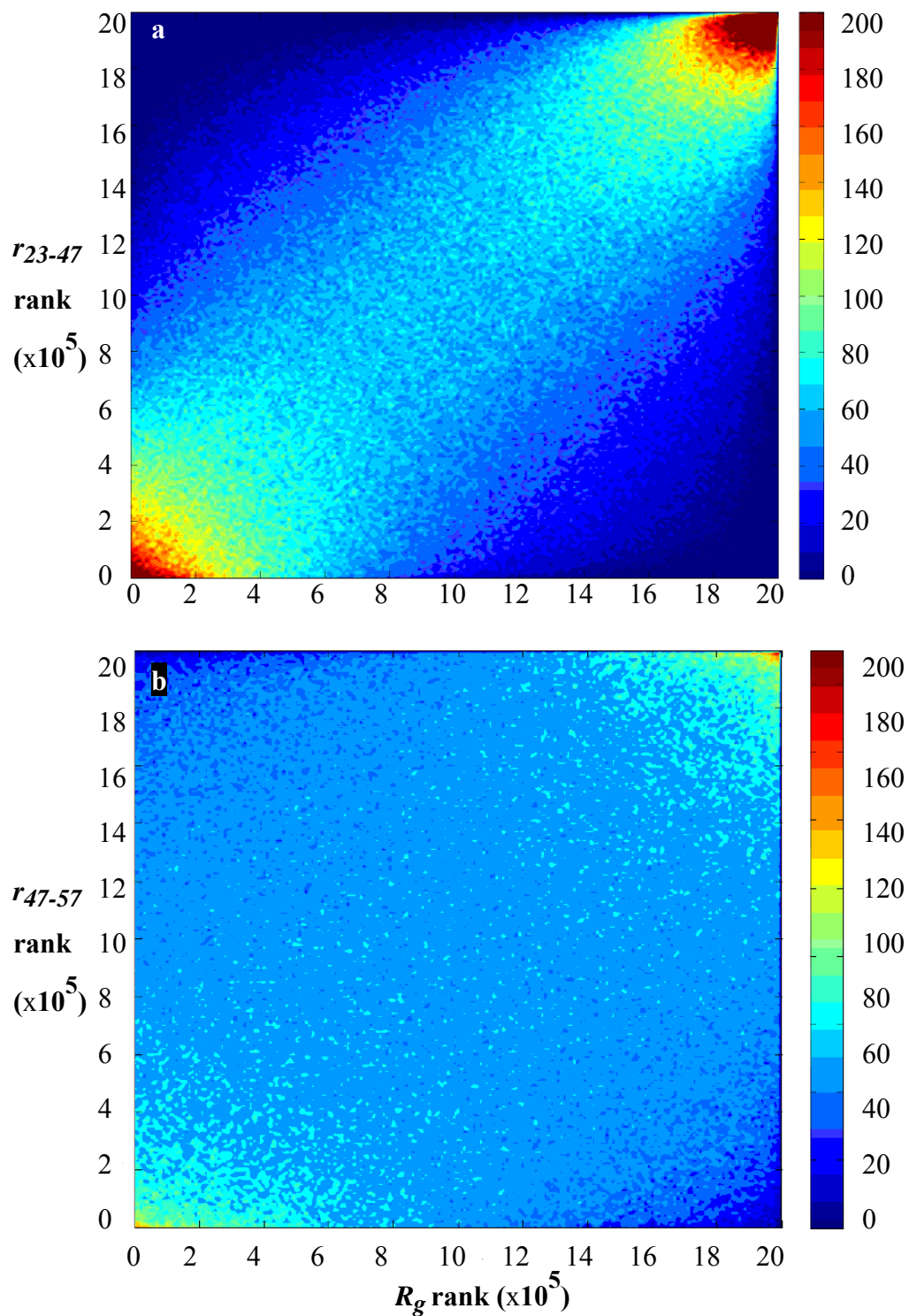


Figure 4.2 Rank correlations between radius of gyration (R_g) and Trp47/Cys23 (r_{23-47}) (a) and between R_g and Trp47/Cys57 (r_{47-57}) (b)

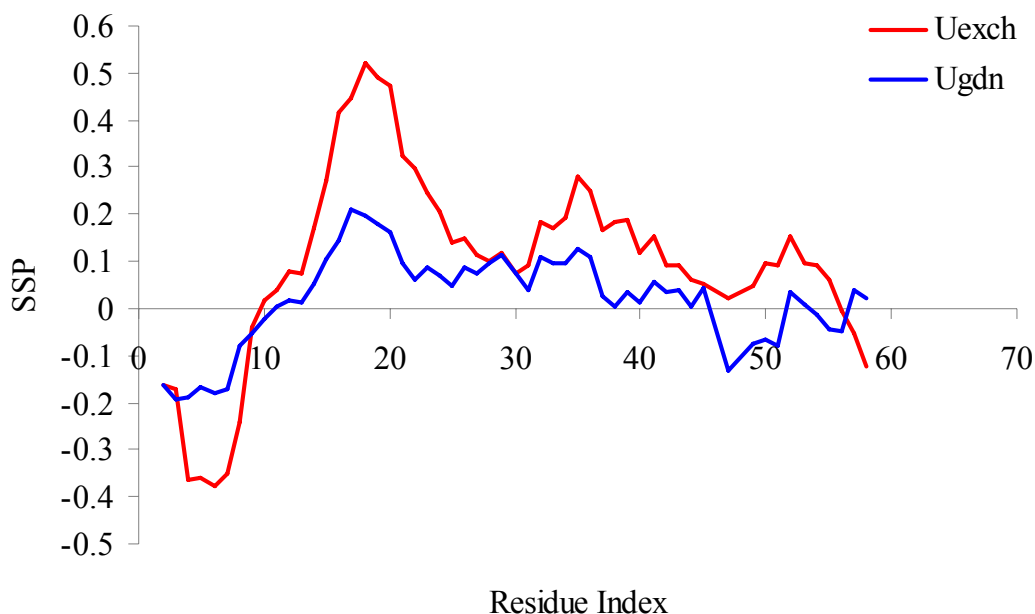


Figure 4.3 Second structure propensity (SSP) of U_{exch} (red) and U_{gdn} (blue) states of drkN

4.3 Secondary Structure Propensity (SSP)

It is relatively straightforward to force certain links of the standard wormlike chain to have axial and azimuthal angles of α -helix or β -strand secondary structure. Since secondary structure is not stable in the unfolded state, each link in the chain can be assigned a secondary structure propensity such as determined by relative NMR chemical shifts of an unfolded protein^{31, 105}. If a random number falls below this propensity, then all 10 links associated with one amino acid adopt the canonical angles for helical ($\theta = 9.2^\circ$, $\phi = 176^\circ$) or strand structure ($\theta = 0^\circ$ and $\phi = 0^\circ$).

Secondary structure was added to the WLC model of drkN. The SSPs of drkN were calculated from the combined C^α and C^β secondary chemical shifts³¹ (deviations of

experimental chemical shifts from their expected random-coil values). (see Figure 4.3) Figure 4.4 shows the comparison of the average hydrophobic energy $\langle E_{TOT} \rangle_r$ as a function of Trp/Cys distance between chains with and without secondary structure constraints and we can clearly see that adding secondary structure constraint lowers the total energy to some extent. However, if we just compare the results in Table 4.1(a) (without SSP) and 4.1(b) (with SSP), such as the average Trp/Cys distance, $\langle r \rangle$, and the efficient diffusion constant, D , we can barely observe differences. Therefore, for typical residual structure observed in unfolded proteins, this added feature has very little effect on the average radius of gyration or other coarse measures of the unfolded ensemble, but as will be shown below, it does improve prediction of more specific measures, such as paramagnetic resonance enhancement.

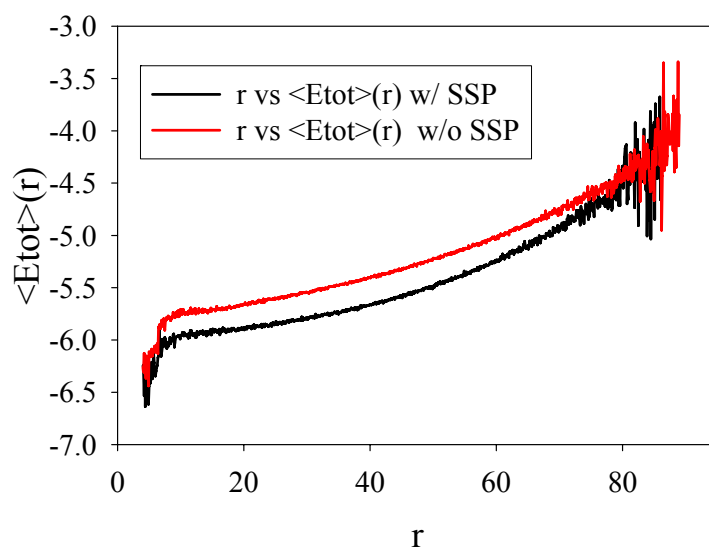


Figure 4.4 Average total energy $\langle E_{TOT} \rangle_r$ of wormlike chains with (black) and without (red) secondary structure constraints.

Table 4.1 Energy-reweighted wormlike chain parameters and effective diffusion coefficients for drkN

(a) without SSP constraints

[GdnHCl]	k_R (s ⁻¹)	k_{D+} (s ⁻¹)	l_p (Å)	d_α (Å)	$\langle r \rangle$ (Å)	σ (1/T)	D (x10 ⁻⁶ cm ² s ⁻¹)
C2							
2	338000	2250000	4	4	34.7435	1	1.4
6	680000	2380000	4	4	32.7247	1.6	1
C60							
2	580000	964000	4	4	29.6232	0.3	0.37
6	580000	870000	4	4	29.6232	0.3	0.34

(b) with SSP constraints

[GdnHCl]	k_R (s ⁻¹)	k_{D+} (s ⁻¹)	l_p (Å)	d_α (Å)	$\langle r \rangle$ (Å)	σ (1/T)	D (x10 ⁻⁶ cm ² s ⁻¹)
C2							
2	338000	2250000	4	4	34.2501	1.3	1.4
6	680000	2380000	4	4	32.1777	2	0.9
C60							
2	580000	964000	4	4	29.3944	0.4	0.38
6	580000	870000	4	4	29.3944	0.4	0.34

4.4 Convergence of σ

This model has only one adjustable parameter, σ , which determines the strength of reweighting due to finding two similar amino acids in close proximity within a wormlike chain. Figures 4.5(b) and (c) shows the contour plots of the probability vs. E_{TOT} and r for $\sigma = 0.2$ and 2.5 for protein L K23C. Increasing σ lowers the energy of the smallest r relative to the largest r which then differentially biases the probability at low r after reweighting, as shown in Figure 4.5(a). After

reweighting, the resulting probability function, $Z(r)$, is used in equation 4.4 to determine the reaction-limited rate of Trp-Cys contact quenching which can be compared to experimental values under various solvent conditions. Figure 4.6(a) plots the values of σ that best fit experimental values of k_R for various proteins and mutants. The standard deviation in σ is about 30% at 2.3 M GdnHCl and 20% at 0 M GdnHCl. At the highest values, k_R is very sensitive to σ ; for a 10% change in σ , k_R changes by about 25%, much larger than our uncertainty. There is good convergence between sequences which suggests that this model is effective at capturing the apparent collapse to varying degrees of all sequences in decreasing denaturant.

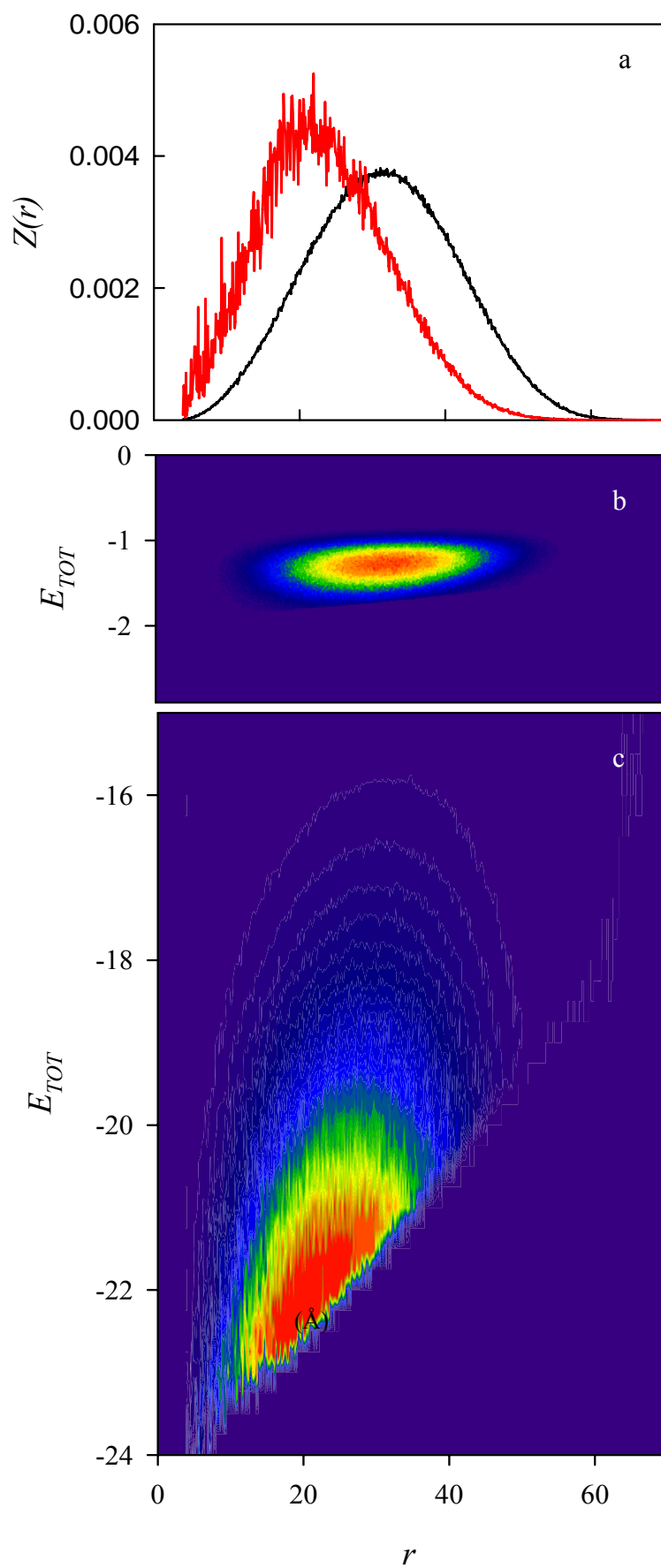


Figure 4.5 a) Probability distribution of Trp/Cys distances in protein L K23C before (black) and after (red) reweighting for $\sigma = 2.5$. b) Contour plot of probability vs r and E for $\sigma = 0.2$ and c) $\sigma = 2.5$.

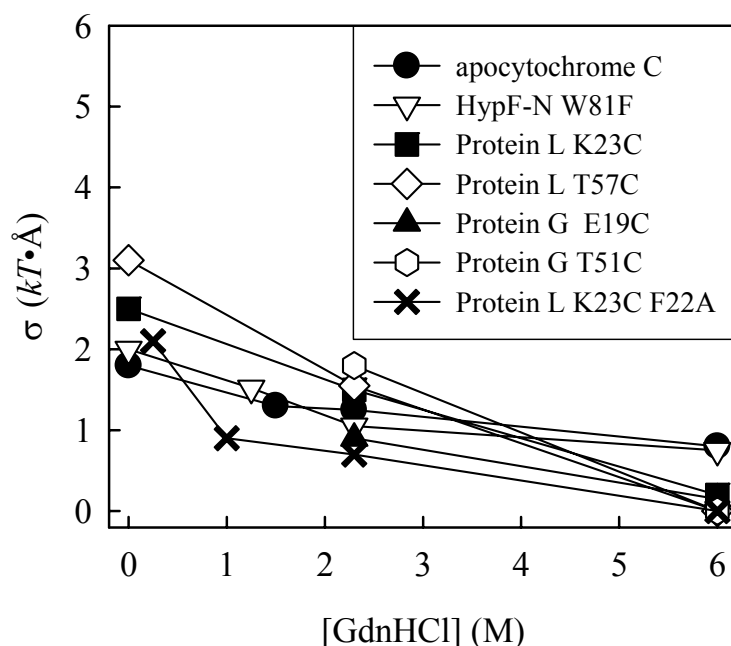


Figure 4.6 The values of σ which best reproduce the measured reaction-limited rates using equation 4.4 for various sequences and [GdnHCl].

4.5 Intrachain Contacts and Hydrophobic Energy

As shown in equations 4.1 and 4.3, I reweight the probability distribution based on close ($1/r$ with a cut off at 6.5 \AA) interactions of residues of similar hydrophobicity. I find that this model gives similar trends and good convergence for many different variations of the distance dependence of E_{TOT} . Figure 4.7 shows σ versus [GdnHCl] for $1/r^2$ distance dependence, a Lennard-Jones distance dependence and no distance dependence (square well). While the magnitude of σ varies from one model to another,

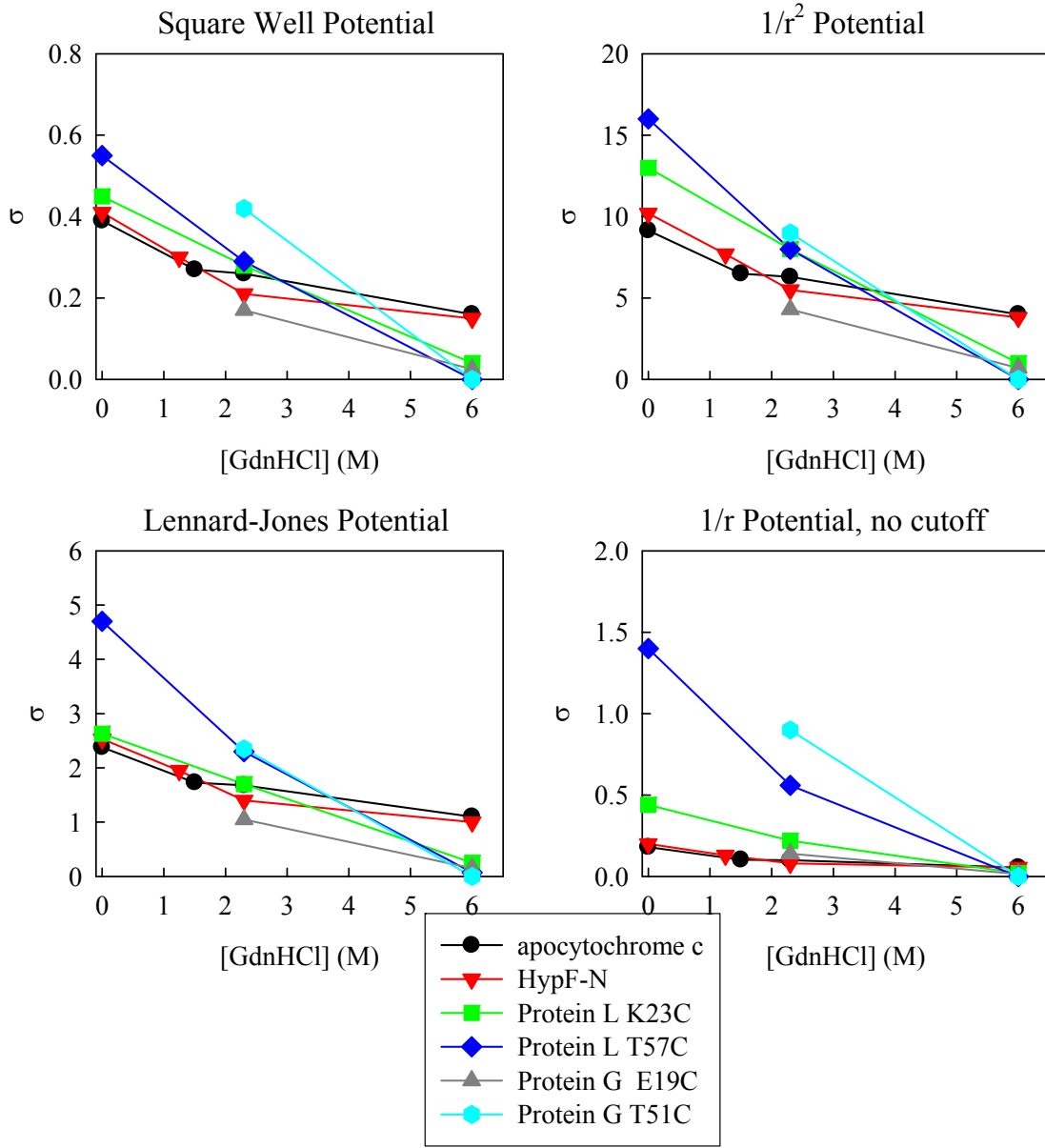


Figure 4.7 Values of σ that best fit measured reaction-limited rates for various forms of E_{TOT} .

$$1/r^2 \text{ potential: } E_{TOT} = - \sum_{|i-j|>1} \frac{e_{i,j}}{(r_i - r_j)^2}$$

$$\text{Square-well potential: } E_{TOT} = - \sum_{|i-j|>1} e_{i,j}$$

$$\text{Lennard-Jones potential: } E_{TOT} = \sum_{|i-j|>1} e_{i,j} \left[\left(\frac{4}{r_i - r_j} \right)^{12} - \left(\frac{4}{r_i - r_j} \right)^6 \right] \text{ for } d_\alpha \leq |r_i - r_j| \leq 6.5 \text{ \AA}.$$

$$1/r, \text{ no cutoff: equation 4.1 in the text for } d_\alpha \leq |r_i - r_j| < \infty$$

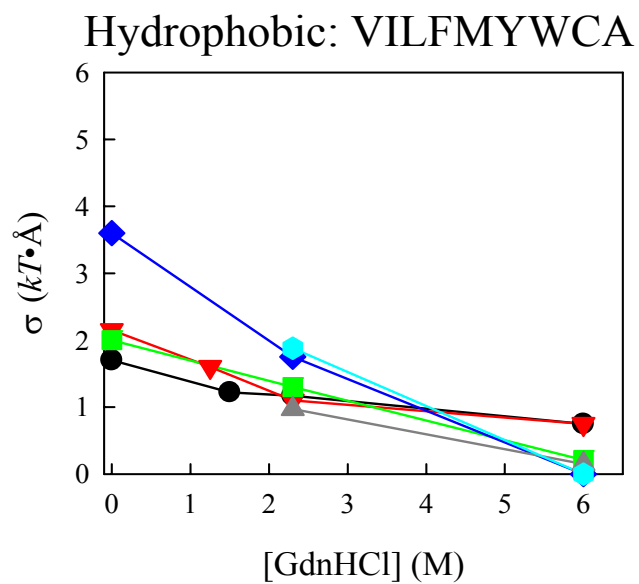
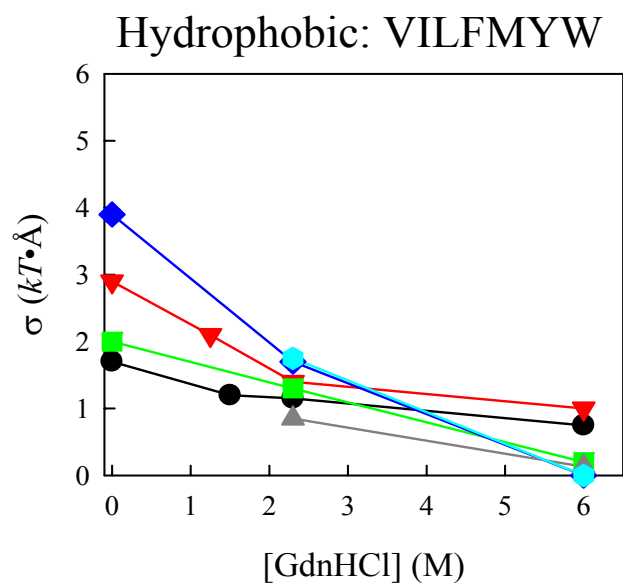
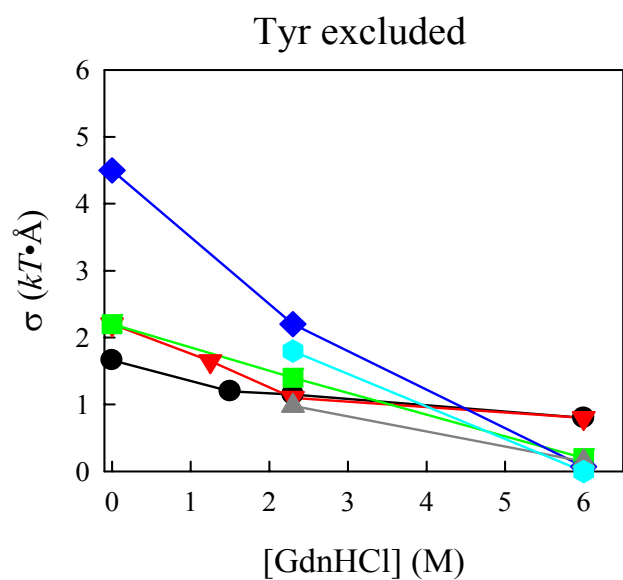


Figure 4.8 Values of that best fit measured reaction-limited rates for various definitions of hydrophobicity.

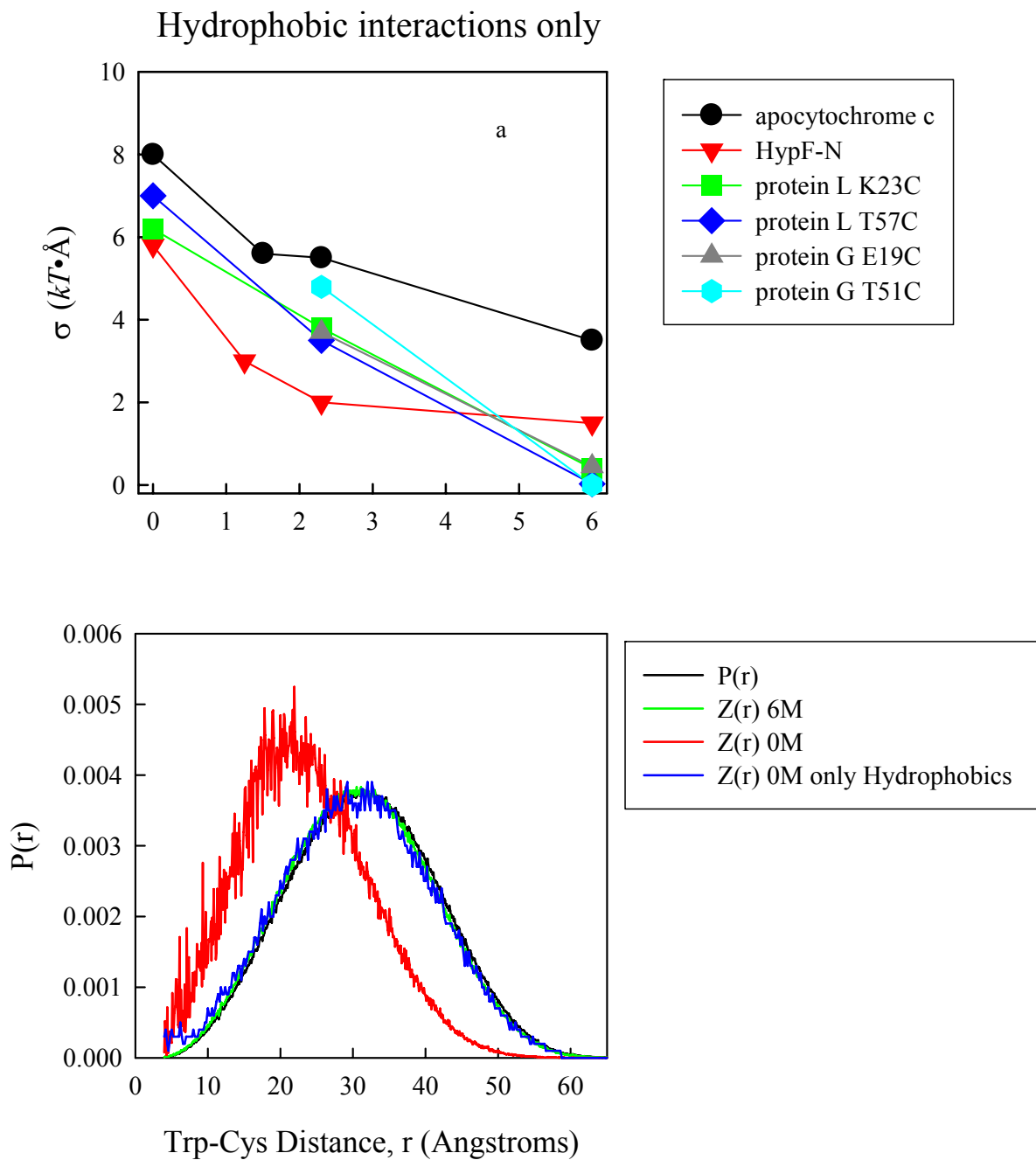


Figure 4.9 a) Values of that best fit measured reaction-limited rates for hydrophobic-hydrophobic interactions only. b) Comparisons of unweighted ($P(r)$) and weighted ($Z(r)$) probability distributions for weighting schemes using hydrophilic interactions (green and red lines) and only hydrophobic interactions (blue line).

the trend of each sequence with denaturant concentration is similar. However, if the energy function does not cut off at 6.5 Å, then the convergence in σ between sequences degrades. This suggests that the inclusion of longer range interactions introduces sequence-dependent complexity to the reweighting function.

In my model I use a normalized Miyazawa-Jernigan scale to assign a numerical hydrophobicity value to each residue in the sequence. I assume that residues of similar hydrophobicity will prefer to be near each other, regardless of whether the pair is hydrophobic or hydrophilic. Therefore, only pairs with hydrophobicity values within 30% of each other are included in the energy function in equation 4.1. Figure 4.8 shows that this model is somewhat sensitive to the method of defining hydrophobicity. In the Miyazawa-Jernigan scale, amino acids fall into two well separated regimes except for tyrosine. If interactions with tyrosine are given no weight ($e_{ij} = 0$), there is almost no change in σ except for the T57C mutant of protein L, which covers a very short loop (10 residues) that includes one tyrosine. Also, if the most hydrophobic residues are given a hydrophobicity value of 1 and all other residues a value of 0 (based on categories used in hydrophobic cluster analysis), there is a greater dispersion in values of σ for different sequences at the same [GdnHCl]. I have also considered the possibility that only interactions between hydrophobic residues drive the collapse, as shown in Figure 4.9(a). The convergence of σ is not as good as including all interactions (Figure 4.6) and increasing σ primarily increases the probability at the lowest r (where the experimental method is most sensitive) rather than significantly shifting the entire distribution as shown in Figure 4.9(b). From

these results I conclude that the primary driving force of collapse is the sequestration of residues of similar hydrophobicity regardless of their overall polarity.

4.6 Comparison between Two Protein L Mutants

I also find this model is sensitive to a single mutation that changes the hydrophobicity of the sequence. Figure 4.10 shows that with the mutation of Phe22 to Ala in protein L, the average number of favorable contacts ($e_{ij} \neq 0$) increased locally and decreased in several positions far from position 22. Therefore mutant F22A K23C has fewer favorable conformations that act to collapse the chain. Since the measured reaction-limited rates for K23C F22A are lower than for K23C at the same denaturant concentrations and the chain compacts less, the values of σ required to match the experimental data are approximately the same for both sequences.

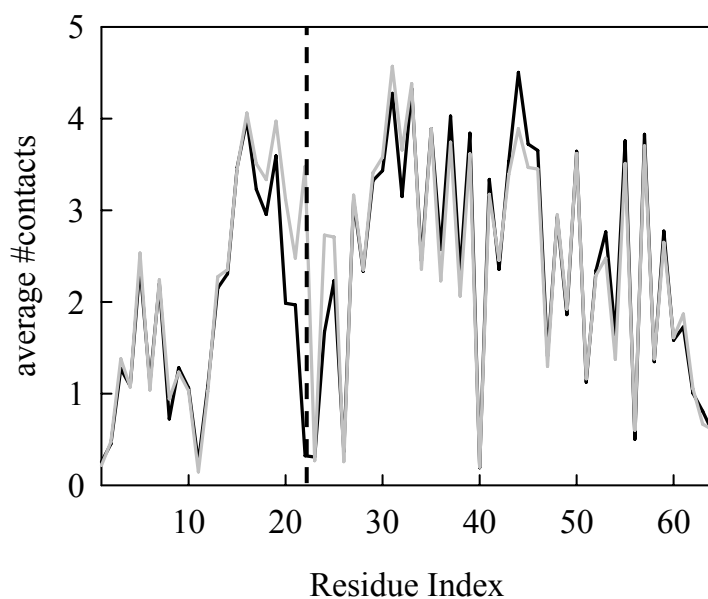


Figure 4.10 Average number of favorable contacts of each residue in protein L K23C (black) and F22A K23C (grey). The vertical dashed line shows the position of the mutation F22A.

I have also used residue contact maps generated by the WLC model before (Figure 4.11(c)) and after (Figure 4.11(a,b,d)) reweighting to analyze the conformational changes of these two protein L mutants. In Figure 4.11, the lower-right triangle region gives the average distance $\langle r_{ij} \rangle$ between residue i and j and the upper-left triangle region gives the standard deviation (spread of the distance distribution) of these two residues. This statistical model yields a trivial result at 6M GdnHCl(Figure 4.11(c)), that is, distant residues have a longer average contact distance and a wider distance distribution. Figures 4.11(a) and (b) show the contact maps for $\sigma = 2.5$. The average terminal distance decreased from 50 Å to about 30 Å and the curvature in these contour lines also indicates an overall compaction of unfolded states. In order to measure the slight difference between K23C F22A (Figure 4.11(a)) and K23C (Figure 4.11(b)), I subtract the contact map of K23C from that of F22A K23C (Figure 4.11(d)) which shows the C-terminus of F22A K23C mutant makes less contacts with the middle region of the protein than K23C mutant but the N-terminus actually makes more contacts. One may also notice from Figures 4.11(a) and (b) that those points with significant average distances also have smaller deviations, which may be due to the overweighted contribution from those conformations with the lowest energies.

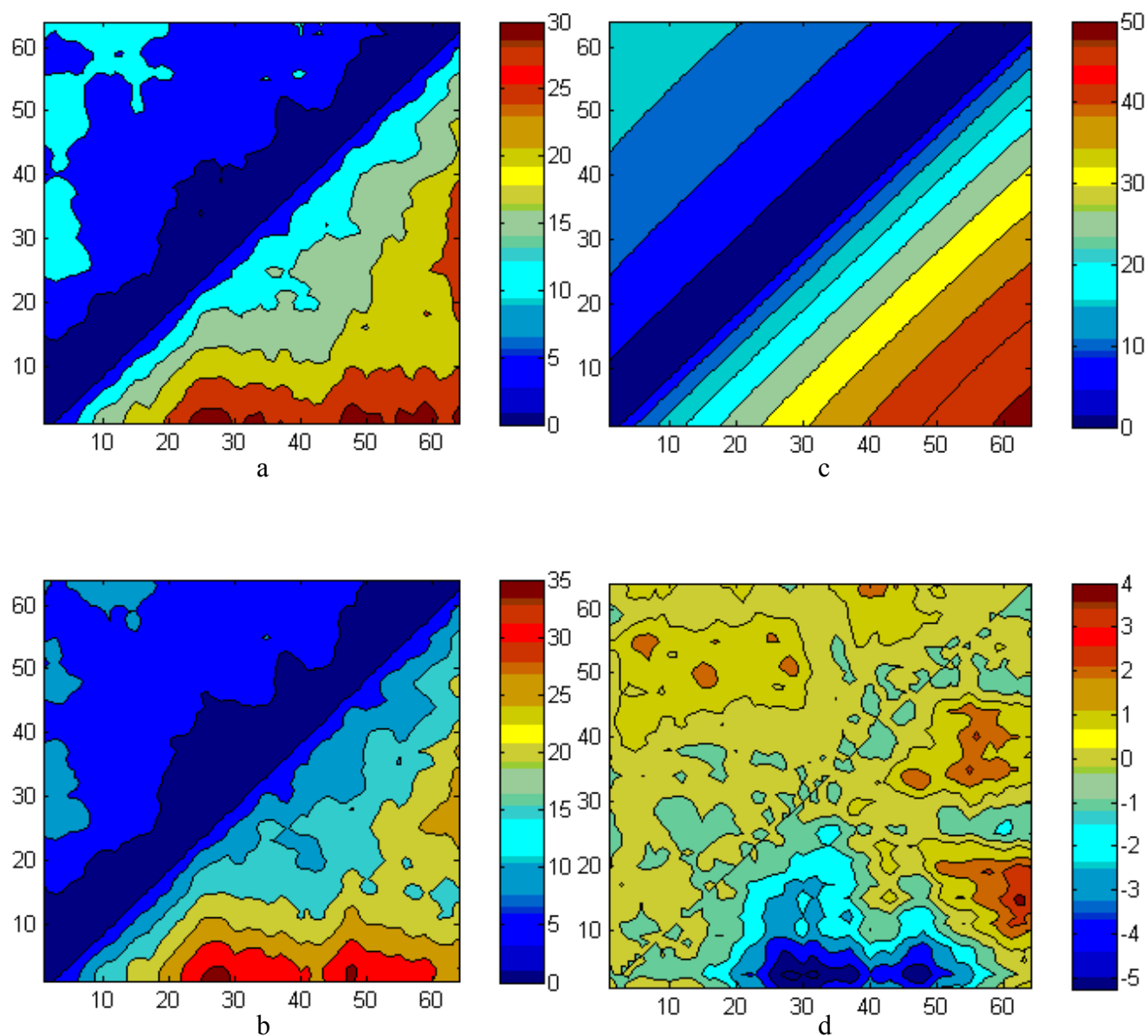


Figure 4.11 Contact maps of protein L K23C F22A (a, c) and K23C (b, c) at $\sigma = 0$ (c) and $\sigma = 2.5$ (a,b). Contact map (d) is the subtraction of (a) by (b). The lower-right triangle region shows the average distance between residues i and j and the upper-left triangle region gives the standard deviation of that average distance.

4.7 PRE Prediction

The best measure of a model is its ability to predict experiments *de novo*. With good convergence of σ at low concentrations of denaturant for various sequences, I have used energy reweighting to predict another measure of the unfolded ensemble in chains outside the initial data set. Paramagnetic resonance enhancement (PRE) is a type of NMR measurement which reports long-range pairwise interactions by using a spin label attached to the polypeptide chain. Typically ^{15}N - ^1H HSQC experiments are recorded with the label in two states: oxidized (paramagnetic) and reduced (diamagnetic). Either the transverse relaxation rate of the paramagnetic sample, R_{2P} , or the ratio between the intensity of a peak in the paramagnetic sample, I_{ox} , and the diamagnetic sample, I_{red} , reflects the distance from the spin-labeled site to every backbone H - N

$$R_{2P} = K(r^i)^{-6} \left(4\tau_C + \frac{3\tau_C}{1 + \omega_H^2 \tau_C^2} \right) \quad (4.7)$$

$$\frac{I_{ox}}{I_{red}} = \frac{R_{2R} \exp(-R_{2P}t)}{R_{2R} + R_{2P}} \quad (4.8)$$

where R_{2R} is the transverse relaxation rate in the diamagnetic sample, K is $1.23 \times 10^{-32} \text{ cm}^6 \text{ s}^{-2}$ for the interaction between a single electron and a proton and ω_H is the Larmor frequency of a proton (750 MHz). τ_C is the effective correlation time, typically 2 ns. In each wormlike chain the distance r^i between the C^α of the spin-labeled residue and that of the i th residue is calculated. Individual R_{2P} and I_{ox}/I_{red} can hence be calculated for each chain using equations 4.7 and 4.8 and the average (experimental) values can be calculated as

$$\langle R_{2P} \rangle_i = \int Z(r^i) R_{2P}(r^i) dr^i \quad (4.9)$$

$$\langle I_{ox}/I_{red} \rangle_i = \int Z(r^i) I_{ox}(r^i)/I_{red}(r^i) dr^i \quad (4.10)$$

where $Z(r^i)$ comes from equation 4.3. For a random coil, one would expect deviations from a ratio of 1 only at residues nearby the spin label in sequence; deviations far in sequence indicate that the spin labels are in contact with the unfolded chain, at least transiently.

Figure 4.12 shows the calculated and measured PRE values for multiple labels of drkN SH3 domain³¹ and acyl-CoA binding protein (ACBP)³⁵. I find remarkable agreement between calculated and measured data. However the agreement between measured (black bars) and calculated PRE is much better if residual secondary structure is included in the construction of the wormlike chains (red lines) than without it (green lines). For ACBP, I find good agreement for most of the chain but for certain regions, particularly residues 24-28, 29-35 and 50-60, there is little non-local interaction, regardless of the position of the paramagnetic probe. Examination of the sequence shows that these regions are almost exclusively hydrophobic (24-28) or hydrophilic (29-35 and 50-60) so local interactions within these regions will contribute significantly to E_{TOT} . However if equation 4.1 is altered to include only interactions at least 4 residues away ($|i-j|>4$), then agreement with the data is somewhat improved (blue lines) as non-local interactions are more highly weighted (although σ increased because the total number of interactions decreased). These results suggest there are still further refinements that can be made to this model with the addition of different types of data.

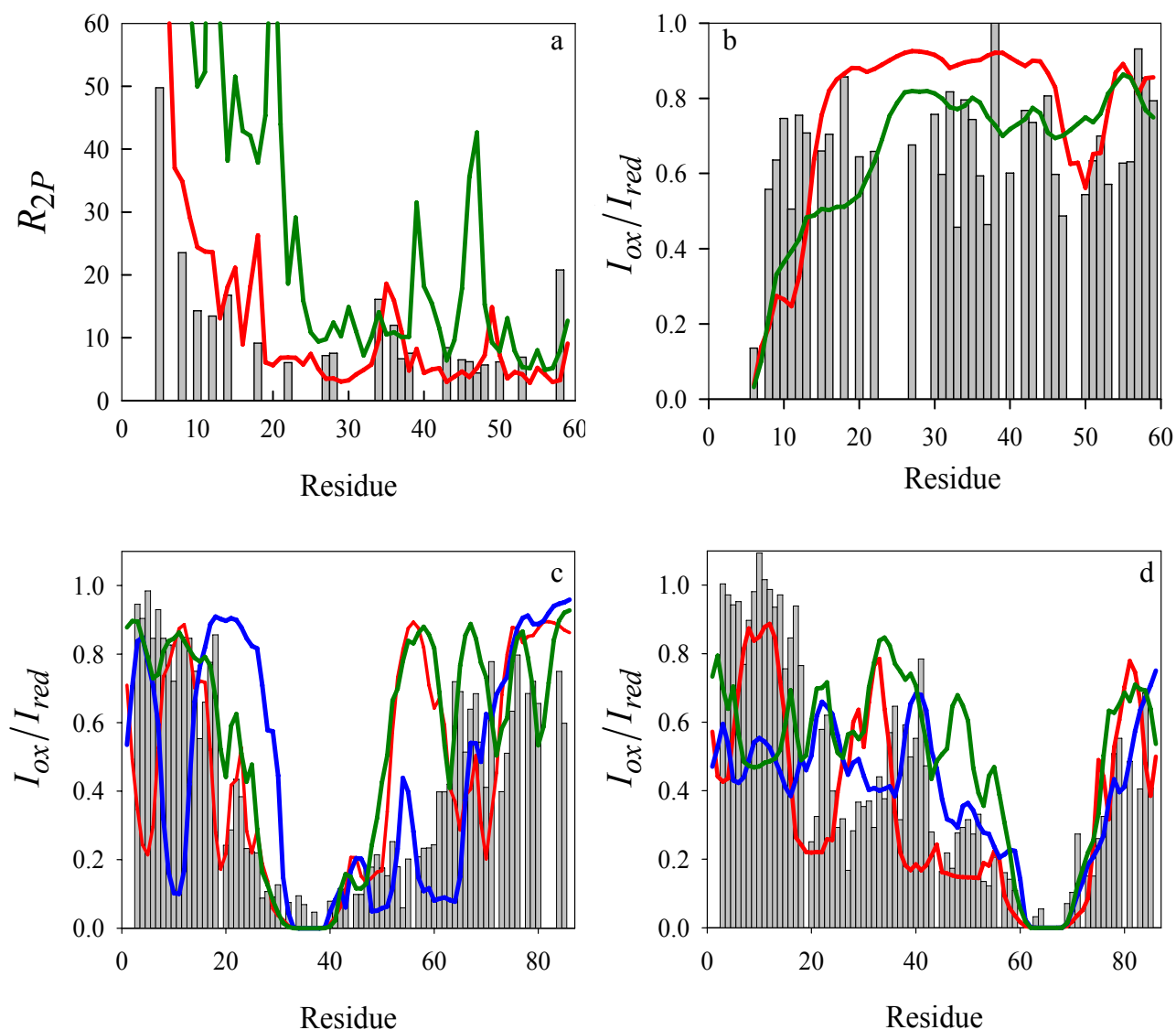


Figure 4.12 Measured (bars) and calculated (lines) PRE per residue for drkN SH3 domain (a and b) and ACBP (c and d). In each graph the green lines correspond to a completely random set of wormlike chains and the red lines correspond to an ensemble of wormlike chains with some residual secondary structure given by measured NMR chemical shifts. The blue lines correspond to the ensemble with secondary structure using a modified version of equation 4.1 in which only interactions more than 4 residues apart in sequence are included ($\sigma = 6$). The paramagnetic labels are located at residue 2 (a), the N-terminus (b), residue 36 (c) and residue 65 (d). For (a) $\tau_C = 0.5$ ns, a scaling factor in equation 4.7, is adjusted to best match the measured data. Pearson Correlation coefficient between bars and red lines equals to 0.89 (a), 0.41 (b), 0.35 (c) and 0.59 (d).

4.8 Charge Effects

Intrinsically disordered proteins generally have a high charge content and the net charge has been shown to correlate with the radius of gyration. So in order to better model the polymer dynamics of intrinsically disordered proteins, an inclusion of both hydrophobic and charge effects may be necessary. After the addition of the electrostatic energy term, the total energy (equation 4.1) of each chain becomes,

$$E_{TOT} = E_h + E_e = E_h + \gamma \sum_{|i-j|>1} \frac{q_i q_j}{|r_i - r_j|} = \sum_{|i-j|>1} \frac{\gamma q_i q_j - e_{i,j}}{|r_i - r_j|} \quad (4.11)$$
$$q_i = \begin{cases} 1, & \text{if } i \text{ is Arg or Lys} \\ -1, & \text{if } i \text{ is Asp or Glu} \\ 0, & \text{otherwise} \end{cases}$$

γ exhibits the strength of effective electrostatic interactions between two charged residues and q_i is the charge of residue i . The above expression of q_i is only valid at pH = 7. At lower pH value, one may need to consider the positive charges introduced by His. Also Arg and Lys will become mostly uncharged in basic solution and Asp and Glu will become uncharged in acidic solution. According to the Henderson-Hasselbalch equation, a charged residue possesses a certain probability of gaining or losing an electron determined by its pKa and the solution's pH. To take into account this uncertainty in our simulation, for each chain generated, a random number (0 to 1) independently assigned to each residue was compared to the probability determined by Henderson-Hasselbach equation. If this random number is smaller than the probability, this residue shall carry charge for this particular chain.

AavLEA has a large fraction of charged amino acids (~50%) and has been shown in chapter

2 to be near equally extended and diffusive over all denaturing conditions (Figure 3.11). The reaction-limited and diffusion-limited rates and corresponding WLC parameters are listed in Table 4.2. We find the range of σ values is much lower than previously determined values for other proteins (Figure 4.6), so there has to be some additional swelling forces opposing the hydrophobic effects for this protein and the high volume of charges distributed over the entire AavLEA peptide can be a good candidate.

Table 4.2 Energy reweighted worm-like chain parameters and effective diffusion coefficients for AavLEA without charge

[GdnHCl]	$k_R (s^{-1})$	$k_{D+}(s^{-1})$	$l_p(\text{\AA})$	$d_\alpha(\text{\AA})$	$\langle r \rangle (\text{\AA})$	$\sigma(1/T)$	$D (x10^{-6} \text{ cm}^2 \text{ s}^{-1})$
S2C							
0	256000	830000	4	4	32.1233	0.6	0.55
1	260000	1024000	4	4	32.1233	0.6	0.65
6	223000	696000	4	4	32.6823	0.45	0.5
T58C							
0	267000	950000	4	4	32.0618	0.6	0.6
1	250000	643000	4	4	32.2638	0.55	0.4
6	132000	811000	4	4	33.9852	0.1	0.8

Table 4.3 reweighting with both σ and γ for AavLEA T58C

[GdnHCl]	assumed $\sigma(1/T)$	$\gamma(1/T)$
0	2.5	31
1	1.8	19
6	0.5	5

One way to take into account the charge effects is to first assume an identical or different γ value for each [GdnHCl] and then adjust σ to match measured reaction-limited rates. However, because we have little knowledge of this newly introduced parameter, γ , we don't know what to choose for its value. So instead, I take a complimentary approach and assume a similar set of σ values for AavLEA at different [GdnHCl] to those values shown in Figure 4.6. And then fit γ to experimental data. The Trp-Cys quenching experiments of AavLEA were carried out in pH=7, so I didn't consider partial charges discussed above.

I find the addition of the charge effects to the WLC model yields very meaningful results for the AavLEA T58C mutant. The fitted γ values to σ values are listed in Table 4.3. We can see γ is consistently about 10 times larger than σ at all three GdnHCl concentrations, suggesting that the strength of charge effects is homogenously greater than the hydrophobic effects at all [GdnHCl] (note that q_i in equation 4.11 only indicates the sign of the charge, so σ and γ have the same dimension), which makes sense if one considers the charge effects as a representative of the electrostatic forces and the hydrophobic effects as a result from the Van der Waals forces. The consistent ratios of magnitude between 0M and 6M GdnHCl for both σ and γ also indicate that these two parameters might represent well the solvent conditions. As the GdnHCl concentration decreases, the ionic strength decreases accordingly causing a weaker charge shielding. Hence, the increase of γ reflects a longer Debye length at low denaturant concentration. Similarly, the increase of σ as [GdnHCl] decreases reflects a weaker polar shielding of the water molecules from the guanidinium ion. Figure 4.13 shows the Trp/Cys distance distributions before and after reweighting by σ and (or) γ . We can see an opposing effect of σ and γ on the distance distribution

$P(r)$ and a combination of them results in a squeezed distribution with a decreasing in the population of both compact and stretched conformations.

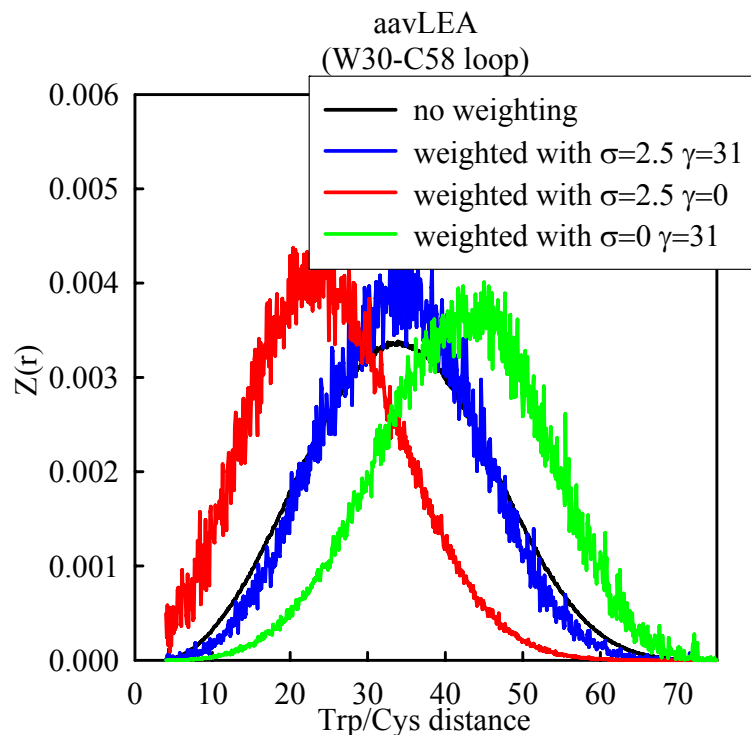


Figure 4.13 Probability distribution of Trp-Cys distances in AavLEA T58C before (black) and after reweighting for $\sigma = 2.5$ (red), $\gamma = 31$ (green) and $\sigma = 2.5$, $\gamma = 31$ (blue)

Table 4.4 reweighting with only γ for AavLEA S2C

[GdnHCl]	$k_R (s^{-1})$	$k_{D+}(s^{-1})$	$l_p(\text{\AA})$	$d_\alpha(\text{\AA})$	$\langle r \rangle (\text{\AA})$	$\gamma(1/T)$	$D (x10^{-6} \text{ cm}^2 \text{ s}^{-1})$
S2C							
0	256000	830000	4	4	31.4681	10	0.5
1	260000	1024000	4	4	31.4681	10	0.6
6	223000	696000	4	4	32.0442	8	0.45

The charge effects on the S2C mutant seem to work on the opposite way: they further compact the chains (Table 4.4). While it is good to see the heterogeneity between different loops

of the same protein, the homogenous effects of σ and γ make it impossible to fit both to the experimental data. This suggests there are other effects than charge preventing the C2-W30 loop from collapsing.

4.9 Conclusion

The main feature of this model is that a distribution of random conformations is reweighted to favor those with close interactions between residues of similar hydrophobicity. This includes residues that are relatively hydrophilic because I assume that in a random ensemble of collapsed conformations, not all hydrophilic residues will be accessible to solvent and intramolecular interactions between hydrophilics will also help collapse the chain. Figure 4.14 shows the cumulative probability of buried surface area for hydrophobic (Phe to Tyr in Figure 4.1) and hydrophilic residues of protein L. At $\sigma = 0$, they are trivially identical and remain fairly similar at $\sigma = 2.5$ even though there are twice as many hydrophilic residues in the sequence than hydrophobic residues. Therefore, the typical unfolded conformation does not have a hydrophobic core, but in water-like solvent conditions hydrophobes are significantly more buried from water than in highly denatured conditions. Interestingly, I also found that binary HP models (Figure 4.8), in which all residues were grouped into only two categories of hydrophobicity, produced more variance in σ than a more graduated model based on similarity of amino acids on the Miyazawa-Jernigan scale (Figure 4.6) and that only including interactions of

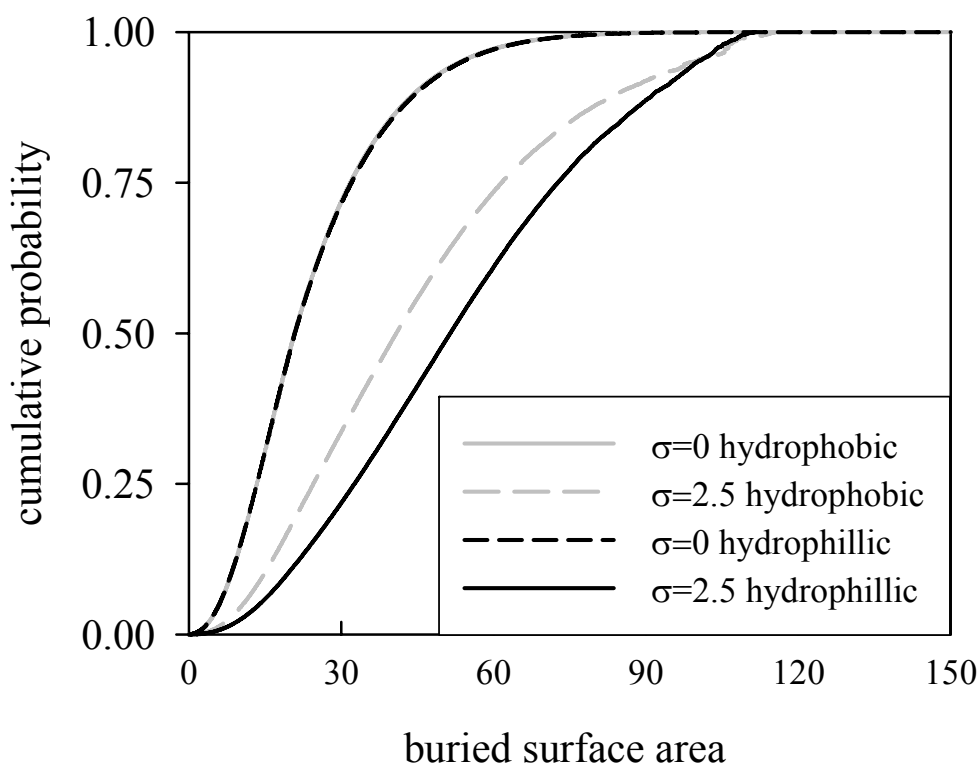


Figure 4.14 The cumulative probability ($C(a) = N \int_0^a P(x) dx$, N is a normalization constant) of buried surface area for hydrophobic and hydrophilic residues of protein L at $\sigma = 0$ and $\sigma = 2.5$. Buried surface area is defined as the cross-section of overlap between two links on the wormlike chain assuming each link has spherical volume of diameter 6.5 \AA . Since no two links can be closer than $d_\alpha = 4 \text{ \AA}$, the maximum buried surface area is 20.6 \AA^2 per residue.

hydrophobic residues did not significantly collapse the chains at all (Figure 4.9). This suggests that collapse of random chains in low denaturant conditions is driven by interactions of all amino acids of similar polarity. I also found that convergence of σ required only very short range r -dependence (less than 6.5 \AA) but the form of the distance-dependence did not matter. This suggests that only the closest intramolecular interactions drive the collapse of random chains.

As described in chapter 2, an intramolecular diffusion coefficient, D , which reflects the time for unfolded conformations to reconfigure, can be determined from Trp-Cys quenching

measurements (see equation 4.5). Figure 4.15 shows D for various sequences and [GdnHCl] using $Z(r)$ from the model and measured k_{D+} . There is some variance in D at high denaturant which increases as denaturant decreases, with unstructured or unstable sequences (apocytochrome c and protein L F22A) remaining fairly diffusive and well folded proteins (proteins L and G) slowing dramatically. A recent measurement of Trp/Cys contact quenching of protein L after rapid dilution (250 μ s) of denaturant in a microfluidic mixer determined $k_{D+} \sim 10^5 \text{ s}^{-1}$ which gives $D = 2 \times 10^{-9} \text{ cm}^2 \text{ s}^{-1}$ ³⁸, more than two orders of magnitude slower than apocytochrome c. Importantly, in this measurement diffusion was so slow that $k_{obs} \sim k_{D+}$ so k_R could not be determined directly. Therefore a predictive model of $Z(r)$ is critical to determine D .

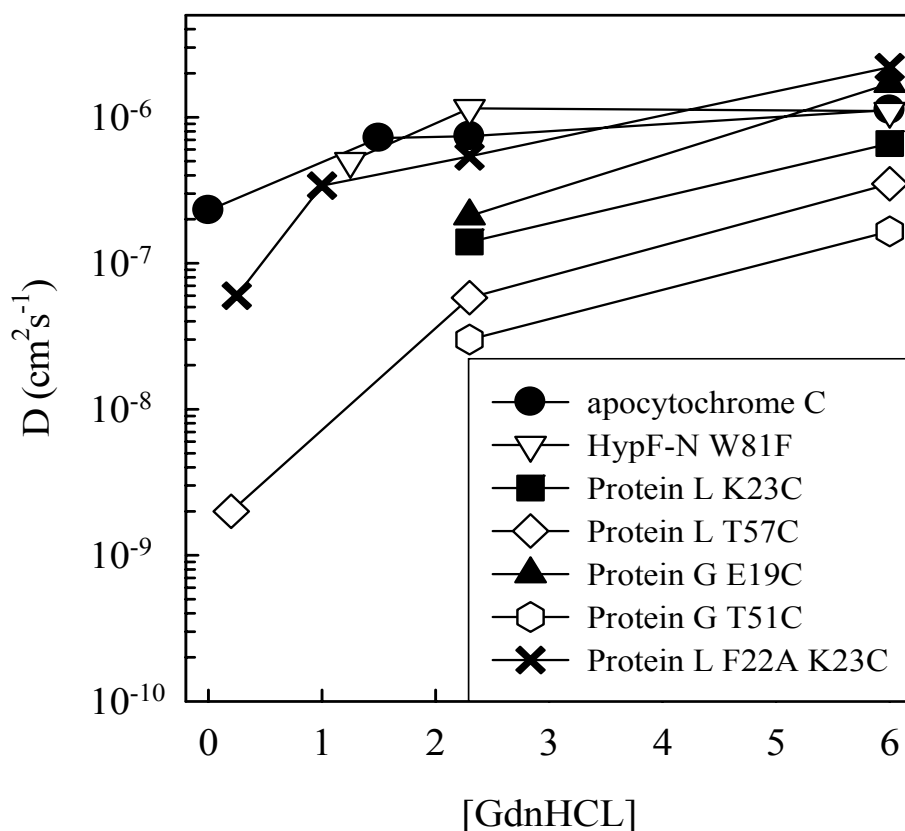


Figure 4.15 The values of D which best reproduce the measured diffusion-limited rates using equation 4.5 with values of σ shown in Figure 4.6.

Chapter 5

Conclusions

In the end, I would like to summarize the conformational and diffusional properties of several typical proteins with very different folding profiles. Figure 5.1 shows the measured values of reaction-limited and diffusion-limited rates for various sequences shown in Table 5.1. There is little difference in rates at 6 M GdnHCl, but the reaction-limited rates at low concentrations of GdnHCl (determined either from experiment or molecular dynamics simulation) vary by over an order of magnitude. The slowest k_R , which corresponds to a more expanded chain, are observed in apocytochrome c, which is intrinsically disordered, and the amyloidogenic protein HypF-N. Proteins L and ACBP are the fastest. Significantly, the mutation of a single phenylalanine to alanine in protein L decreases k_R significantly at all [GdnHCl]. Therefore sequence plays a significant role in determining intramolecular contact formation under conditions that favor folding.

Table 5.1 Several protein sequences studied in our lab

protein	sequence length	probe position (mutation)	loop length
apocytochrome c	104	C17, W59	42
HypF-N	91	W27, C65(C7S/C40S/W81F)	38
G	56	C19, W43 (E19C) W47, C51 (T51C)	23 8
L	64	C23, W47 (K23C) C23, W47 (F22A/K23C) W47, C57 (T57C)	23 23 10
ACBP	86	C17, W58 (T17C/W55A) W58, C86 (W55A/I86C)	28 28

Figure 5.1 Experimental reaction-limited (black) and diffusion-limited (white) rates measured at various concentrations of denaturant for (a) ACBP T17C (b) ACBP I86C (c) HypF-N W81F, (d) apocytochrome c, (e) Protein L K23C and (f) Protein G E19C. The errors in these rates are typically 10%. The triangular points in (a) and (b) are measured values from molecular dynamics simulations. The triangular points in (e) are the destabilizing mutation K23C F22A in which the unfolded state is observable in as little as 0.25 M GdnHCl. The reaction-limited rate in (e) (black circle) at 0 M GdnHCl was calculated from a Trp-Cys distribution derived from a molecular dynamics simulation³⁷ using equation 2.9. All other points were derived from measured tryptophan triplet decay rates at various temperatures and viscosities using equations 2.5 and 2.6.

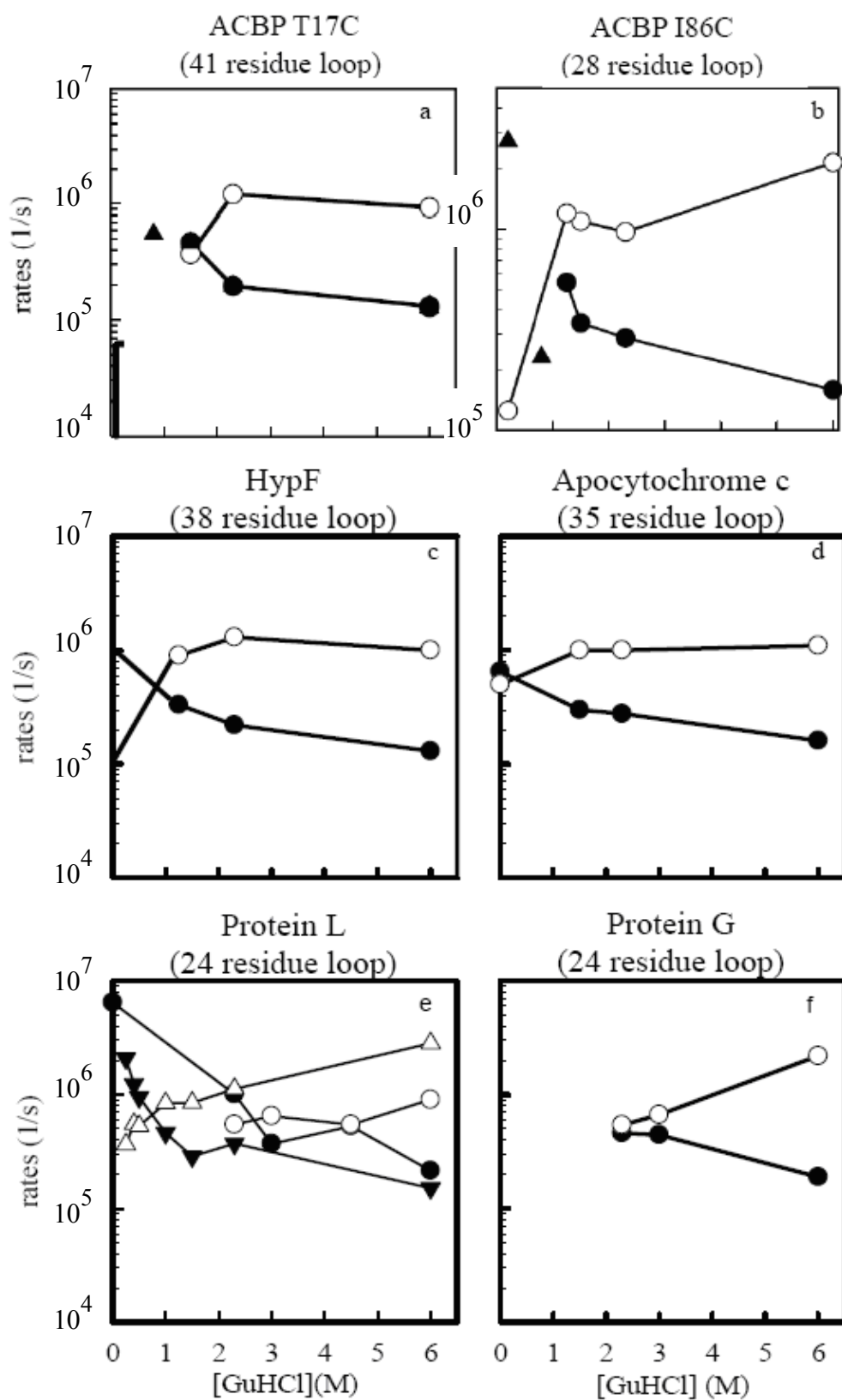


Figure 5.1

Theoretically, I developed a sequence specific worm-like chain model to compare to the experimental results of those unfolded proteins, especially under folding conditions. I believe the method of re-weighting worm-like chain distributions to account for intrachain interactions is robust, realistic and predictive and may be a useful starting point for further models of protein folding. For example, a recent study of molecular dynamics folding trajectories found that randomly generated chains folded more slowly and less efficiently than chains generated with native-like conformations determined by a knowledge-based potential¹⁰⁶. This indicates that not all conformations in the unfolded state are equally likely to proceed to the folded state and that there may in fact be specific bottlenecks for leaving the unfolded basin. However, that study did not examine how realistic or probabilistic the productive unfolded conformations are. That study also shows the native-like chains are much more extended. If extension is required for folding then finding the conformation that leads to folding may be extremely unlikely and knowing how quickly a random unfolded conformation can access the productive conformation defines the folding rate.

Figure 5.2 shows the intramolecular diffusion coefficients of several proteins in various folding categories. We can see nonfolding peptides and intrinsically disordered proteins (aavLEA and apocytochrome C) are highly diffusive in water while well-behaved proteins (protein L and ACBP) diffuse very slowly. The diffusivities of aggregation-prone proteins (long polyQ, L F22A and HypF-N) lie right in the middle between these two groups. Therefore I propose that there is a dangerous dynamic range of diffusivity in which unfolded proteins or polypeptides are more likely to aggregate.

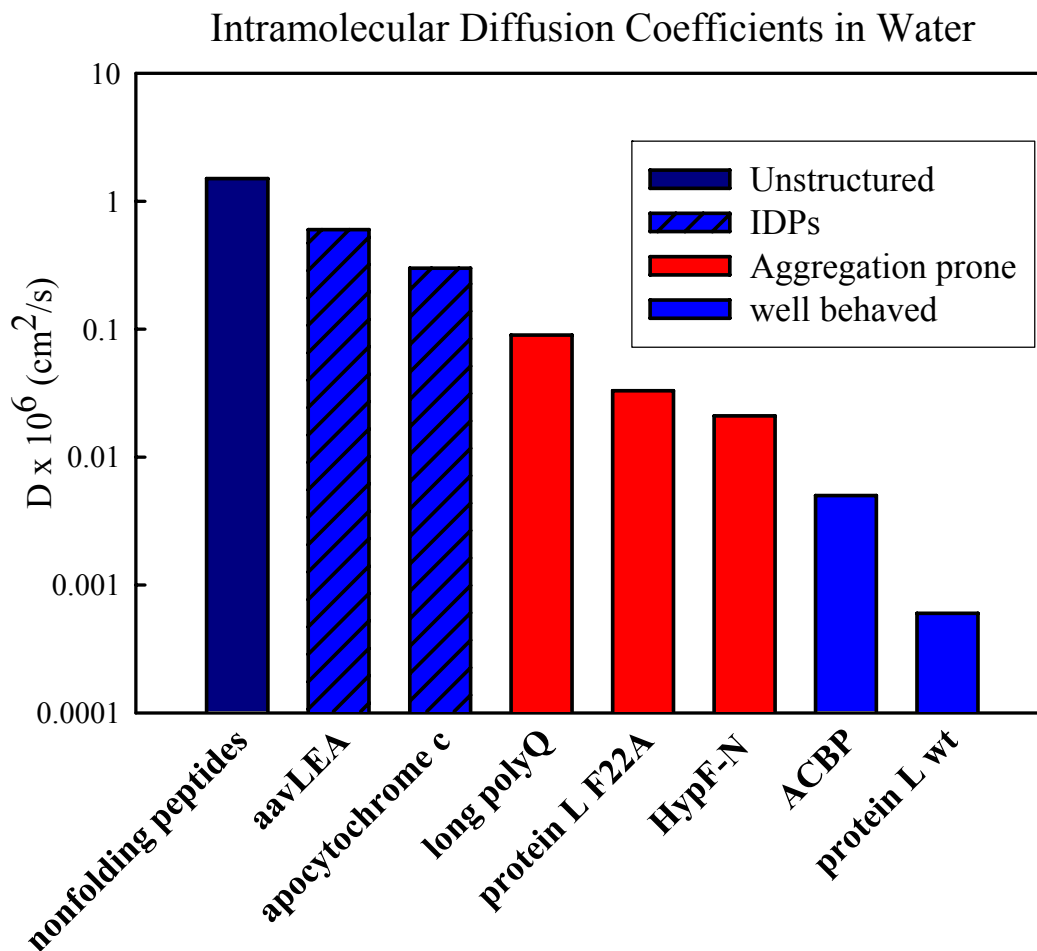
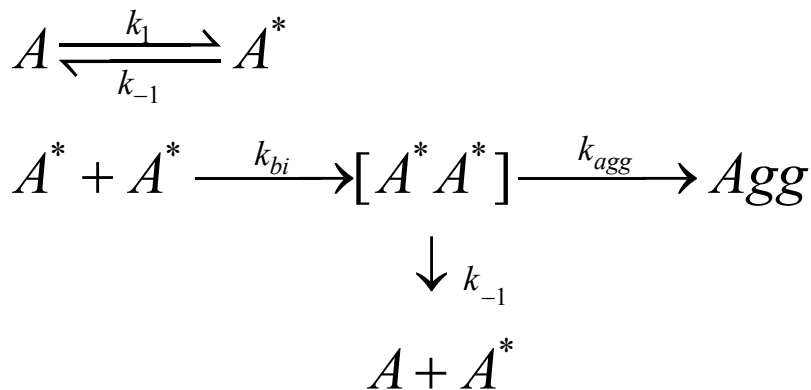


Figure 5.2 Intramolecular diffusion coefficients of several different proteins. Long polyQ is an aggregation-prone protein like HypF-N. One single mutation F22A destabilizes well-behaved wildtype protein L tremendously and gives it a tendency to aggregate.

In this dynamic range, a freely diffusing chain reconfigures just fast enough to expose hydrophobic residues to solvent long enough to form bimolecular associations. Polypeptides that diffuse faster do not leave hydrophobes exposed for long, destabilizing bimolecular complexes, and well behaved proteins diffuse slow enough that most hydrophobes are exposed to solvent infrequently. This is illustrated by a simple kinetic model



where A is the unfolded ensemble of conformations and A^* is the conformational subset that can make bimolecular associations. $[A^* A^*]$ is the bimolecular encounter complex that can lead to irreversible aggregation, Agg , but can also be returned to monomers by conversion of A^* into A . The intramolecular reconfiguration rates, k_1 and k_{-1} , are related to the intramolecular diffusion coefficient, D . k_{agg} consists of a sequential steps of the aggregation process from the formation of oligamers to fiberizations. Keeping k_{bi} , k_{agg} and $K \equiv k_1/k_{-1}$ constant, Figure 5.3 shows the formation of Agg is substantially slower when k_1 is much higher or much lower than k_{bi} than when they are the about the same. This is not a realistic model but the simplest model I can extract from my experimental results. Although there is a big difference in the timescale between intramolecular diffusion and aggregation, this model should still be considered intuitive as long as k_{agg} is irreversible. Proof of this hypothesis will certainly require more measurements of aggregation-prone sequences, but a more complete understanding of hydrophobic encounter complexes between proteins will also help to define the limits of this dynamical regime of aggregation.

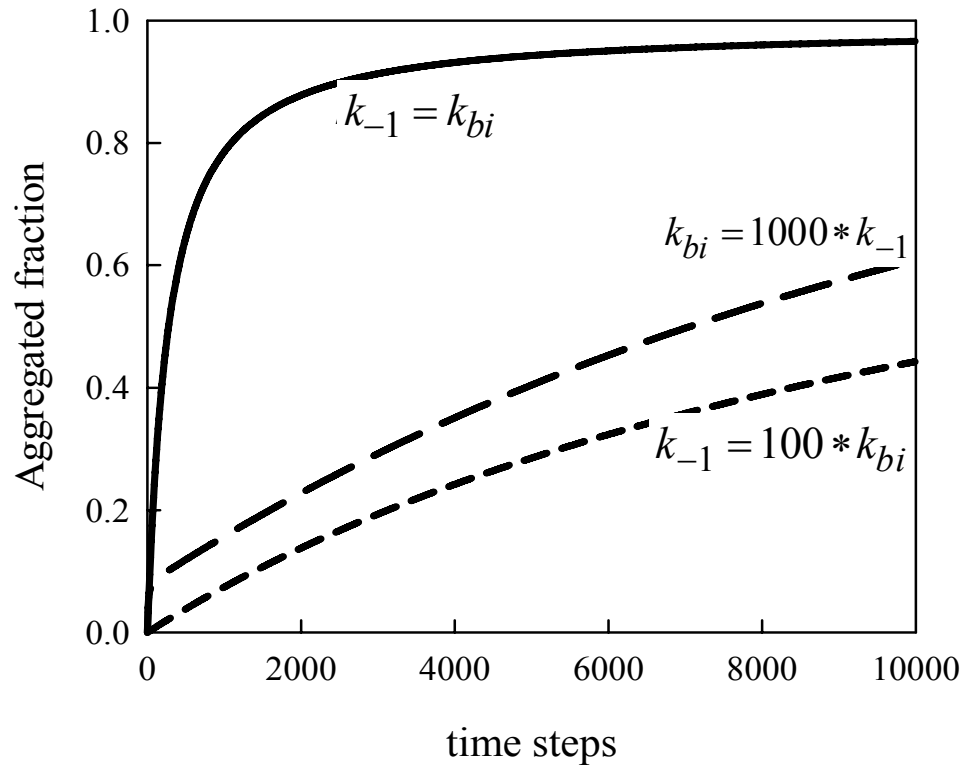


Figure 5.3 Formation of aggregated species, *Agg*, according to the model outlined in the text.

For these simulations, $k_{bi} = k_{agg} = 1$, $K = 0.1$ and k_{-1} varies as indicated on the plot.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Murray, R. K., Daryl K. Granner, Peter A. Mayes, and Victor W. Rodwell, *Harper's Illustrated Biochemistry*. 26th ed.; New York: Lange Medical Books/McGraw-Hill: 2000.
2. Spillantini, M. G.; Schmidt, M. L.; Lee, V. M. Y.; Trojanowski, J. Q.; Jakes, R.; Goedert, M., alpha-synuclein in Lewy bodies. *Nature* **1997**, 388, (6645), 839-840.
3. Baba, M.; Nakajo, S.; Tu, P. H.; Tomita, T.; Nakaya, K.; Lee, V. M. Y.; Trojanowski, J. Q.; Iwatsubo, T., Aggregation of alpha-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with lewy bodies. *American Journal of Pathology* **1998**, 152, (4), 879-884.
4. Southall, N. T.; Dill, K. A.; Haymet, A. D. J., A view of the hydrophobic effect. *Journal of Physical Chemistry B* **2002**, 106, (3), 521-533.
5. Anfinsen, C. B., Principles That Govern Folding of Protein Chains. *Science* **1973**, 181, (4096), 223-230.
6. Levinthal, C. In *How to fold gracefully*, Mossbauer Spectroscopy in Biological Systems, Monticello, Illinois, 1969; De Brunner, P.; Tsibris, J.; Munck, E., Eds. University of Illinois Press: Monticello, Illinois, 1969; pp 22-24.
7. Baker, D.; Agard, D. A., Kinetics Versus Thermodynamics in Protein-Folding. *Biochemistry* **1994**, 33, (24), 7505-7509.
8. Govindarajan, S.; Goldstein, R. A., On the thermodynamic hypothesis of protein folding. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, 95, (10), 5545-5549.
9. Viguera, A. R.; Serrano, L., Loop length, intramolecular diffusion and protein folding. *Nature Structural Biology* **1997**, 4, (11), 939-946.
10. Schindler, T.; Schmid, F. X., Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry* **1996**, 35, (51), 16833-16842.
11. Chen, B. L.; Baase, W. A.; Schellman, J. A., Low-Temperature Unfolding of a Mutant of Phage-T4 Lysozyme .2. Kinetic Investigations. *Biochemistry* **1989**, 28, (2), 691-699.
12. Burton, R. E.; Huang, G. S.; Daugherty, M. A.; Fullbright, P. W.; Oas, T. G., Microsecond

protein folding through a compact transition state. *Journal of Molecular Biology* **1996**, 263, (2), 311-322.

13. Fersht, A. R., Characterizing Transition-States in Protein-Folding - an Essential Step in the Puzzle. *Current Opinion in Structural Biology* **1995**, 5, (1), 79-84.

14. Fersht, A. R., Protein-Folding and Stability - the Pathway of Folding of Barnase. *Febs Letters* **1993**, 325, (1-2), 5-16.

15. Kramers, H. A., Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940**, 7, 284-304.

16. Hynes, J. T., Chemical-Reaction Dynamics in Solution. *Annual Review of Physical Chemistry* **1985**, 36, 573-597.

17. Hanggi, P.; Talkner, P.; Borkovec, M., Reaction-Rate Theory - 50 Years after Kramers. *Reviews of Modern Physics* **1990**, 62, (2), 251-341.

18. Thirumalai, D.; Klimov, D. K.; Dima, R. I., Insights into specific problems in protein folding using simple concepts. In *Computational Methods for Protein Folding*, 2002; Vol. 120, pp 35-76.

19. Jacob, M.; Geeves, M.; Holtermann, G.; Schmid, F. X., Diffusional barrier crossing in a two-state protein folding reaction. *Nature Structural Biology* **1999**, 6, (10), 923-926.

20. Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A., Diffusion-limited contact formation in unfolded cytochrome c: Estimating the maximum rate of protein folding. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, 93, (21), 11615-11617.

21. Hagen, S. J.; Hofrichter, J.; Eaton, W. A., Rate of intrachain diffusion of unfolded cytochrome c. *Journal of Physical Chemistry B* **1997**, 101, (13), 2352-2365.

22. Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T., The speed limit for protein folding measured by triplet-triplet energy transfer. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, 96, (17), 9597-9601.

23. Lapidus, L. J.; Eaton, W. A.; Hofrichter, J., Measuring the rate of intramolecular contact formation in polypeptides. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, 97, (13), 7220-7225.

24. Krieger, F.; Fierz, B.; Bieri, O.; Drewello, M.; Kiefhaber, T., Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *Journal of Molecular*

Biology **2003**, 332, (1), 265-274.

25. DeMaeyer, M. E. a. L., *Technique of Organic Chemistry*. Interscience, New York: 1963.

26. Hoffman, G. W., Nanosecond Temperature-Jump Apparatus. *Review of Scientific Instruments* **1971**, 42, (11), 1643-&.

27. Rounsevell, R.; Forman, J. R.; Clarke, J., Atomic force microscopy: mechanical unfolding of proteins. *Methods* **2004**, 34, (1), 100-111.

28. Cecconi, C.; Shank, E. A.; Bustamante, C.; Marqusee, S., Direct observation of the three-state folding of a single protein molecule. *Science* **2005**, 309, (5743), 2057-2060.

29. Brody, J. P.; Yager, P.; Goldstein, R. E.; Austin, R. H., Biotechnology at low Reynolds numbers. *Biophysical Journal* **1996**, 71, (6), 3430-3441.

30. Lipman, E. A.; Schuler, B.; Bakajin, O.; Eaton, W. A., Single-molecule measurement of protein folding kinetics. *Science* **2003**, 301, (5637), 1233-1235.

31. Mittag, T.; Forman-Kay, J. D., Atomic-level characterization of disordered protein ensembles. *Current Opinion In Structural Biology* **2007**, 17, (1), 3-14.

32. McCarney, E. R.; Kohn, J. E.; Plaxco, K. W., Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. *Critical Reviews In Biochemistry And Molecular Biology* **2005**, 40, (4), 181-189.

33. Kammerer, R. A.; Kostrewa, D.; Zurdo, J.; Detken, A.; Garcia-Echeverria, C.; Green, J. D.; Muller, S. A.; Meier, B. H.; Winkler, F. K.; Dobson, C. M.; Steinmetz, M. O., Exploring amyloid formation by a de novo design. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, 101, (13), 4435-4440.

34. Bruun, S. W.; Iesmantavicius, V.; Danielsson, J.; Poulsen, F. M., Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein. *Proceedings of the National Academy of Sciences of the United States of America* 107, (30), 13306-13311.

35. Teilum, K.; Kragelund, B. B.; Poulsen, F. M., Transient structure formation in unfolded acyl-coenzyme A-binding protein observed by site-directed spin labelling. *Journal of Molecular Biology* **2002**, 324, (2), 349-357.

36. Waldauer, S. A.; Bakajin, O.; Ball, T.; Chen, Y.; DeCamp, S. J.; Kopka, M.; Jager, M.; Singh,

V. R.; Wedemeyer, W. J.; Weiss, S.; Yao, S.; Lapidus, L. J., Ruggedness in the folding landscape of protein L. *Hfsp Journal* **2008**, 2, (6), 388-395.

37. Voelz, V. A.; Singh, V. R.; Wedemeyer, W. J.; Lapidus, L. J.; Pande, V. S., Unfolded-State Dynamics and Structure of Protein L Characterized by Simulation and Experiment. *Journal of the American Chemical Society* **2010**, 132, (13), 4702-4709.

38. Waldauer, S. A.; Bakajin, O.; Lapidus, L. J., Extremely slow intramolecular diffusion in unfolded protein L. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, 107, (31), 13713-13717.

39. Buscaglia, M.; Schuler, B.; Lapidus, L. J.; Eaton, W. A.; Hofrichter, J., Kinetics of Intramolecular Contact Formation in a Denatured Protein. *Journal of Molecular Biology* **2003**, 332, 9-12.

40. Lapidus, L. J.; Eaton, W. A.; Hofrichter, J., Dynamics of intramolecular contact formation in polypeptides: Distance dependence of quenching rates in a room-temperature glass. *Physical Review Letters* **2001**, 87, (25).

41. Singh, V. R.; Kopka, M.; Chen, Y.; Wedemeyer, W. J.; Lapidus, L. J., Dynamic Similarity of the Unfolded States of Proteins L and G. *Biochemistry* **2007**, 46, 10046-10054.

42. Chan, H. S.; Dill, K. A., Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. *Proteins-Structure Function and Bioinformatics* **1998**, 30, (1), 2-33.

43. Onuchic, J. N.; LutheySchulten, Z.; Wolynes, P. G., Theory of protein folding: The energy landscape perspective. *Annual Review of Physical Chemistry* **1997**, 48, 545-600.

44. Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiagarajan, P.; Sosnick, T. R.; Hasan, M. Z.; Pande, V. S.; Ruczinski, I.; Doniach, S.; Plaxco, K. W., Random-coil behavior and the dimensions of chemically unfolded proteins. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **2004**, 101, (34), 12491-12496.

45. Bicout, D. J.; Szabo, A., Entropic barriers, transition states, funnels, and exponential protein folding kinetics: A simple model. *Protein Science* **2000**, 9, (3), 452-465.

46. Dinner, A. R.; Sali, A.; Smith, L. J.; Dobson, C. M.; Karplus, M., Understanding protein folding via free-energy surfaces from theory and experiment. *Trends in Biochemical Sciences* **2000**, 25, (7), 331-339.

47. Finke, J. M.; Cheung, M. S.; Onuchic, J. N., A structural model of polyglutamine determined from a host-guest method combining experiments and landscape theory. *Biophysical Journal* **2004**, 87, (3), 1900-1918.
48. Oliveberg, M.; Wolynes, P. G., The experimental survey of protein-folding energy landscapes. *Quarterly Reviews of Biophysics* **2005**, 38, (3), 245-288.
49. Mitsutake, A.; Sugita, Y.; Okamoto, Y., Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* **2001**, 60, (2), 96-123.
50. Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S., Accelerating Molecular Dynamic Simulation on Graphics Processing Units. *Journal of Computational Chemistry* **2009**, 30, (6), 864-872.
51. Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S. P.; Khaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; Zagrovic, B., Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers* **2003**, 68, (1), 91-109.
52. Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*. Plenum: 1999.
53. Vanderkooi, J. M., Tryptophan Phosphorescence from Proteins at Room Temperature. In *Topics in Fluorescence Spectroscopy: Biochemical Applications*, Lakowicz, J. R., Ed. Plenum Press: New York, 1992; Vol. 3.
54. Lapidus, L. J.; Steinbach, P. J.; Eaton, W. A.; Szabo, A.; Hofrichter, J., Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: Testing a 1-dimensional diffusion model for peptide dynamics. *Journal of Physical Chemistry B* **2002**, 106, (44), 11628-11640.
55. Eaton, W. A.; Munoz, V.; Hagen, S. J.; Jas, G. S.; Lapidus, L. J.; Henry, E. R.; Hofrichter, J., Fast kinetics and mechanisms in protein folding. *Annual Review of Biophysics and Biomolecular Structure* **2000**, 29, 327-359.
56. Amouyal, E.; Bernas, A.; Grand, D., Photo-Ionization Energy Threshold of Tryptophan in Aqueous-Solutions. *Photochemistry and Photobiology* **1979**, 29, (6), 1071-1077.
57. Bent, D. V.; Hayon, E., Excited-State Chemistry of Aromatic Amino-Acids and Related Peptides .3. Tryptophan. *Journal of the American Chemical Society* **1975**, 97, (10), 2612-2619.
58. Cheng, R. R.; Uzawa, T.; Plaxco, K. W.; Makarov, D. E., The Rate of Intramolecular Loop

Formation in DNA and Polypeptides: The Absence of the Diffusion-Controlled Limit and Fractional Power-Law Viscosity Dependence. *Journal of Physical Chemistry B* **2009**, 113, (42), 14026-14034.

59. Szabo, A.; Schulten, K.; Schulten, Z., 1st Passage Time Approach to Diffusion Controlled Reactions. *Journal of Chemical Physics* **1980**, 72, (8), 4350-4357.

60. Singh, V. R.; Lapidus, L. J., The Intrinsic Stiffness of Polyglutamine Peptides. *Journal of Physical Chemistry B* **2008**, 112, (42), 13172-13176.

61. Hagerman, P. J.; Zimm, B. H., Monte-Carlo Approach To The Analysis Of The Rotational Diffusion Of Wormlike Chains. *Biopolymers* **1981**, 20, (7), 1481-1502.

62. Aref, H., Stirring by Chaotic Advection. *Journal of Fluid Mechanics* **1984**, 143, (JUN), 1-21.

63. Waldauer, S. A. Early Events in Protein Folding Investigated through Ultrarapid Microfluidic Mixing. DISSERTATION, Michigan State University, East Lansing, 2009.

64. Rosano, C.; Zuccotti, S.; Bucciantini, M.; Stefani, M.; Ramponi, G.; Bolognesi, M., Crystal structure and anion binding in the prokaryotic hydrogenase maturation factor HypF acylphosphatase-like domain. *Journal of Molecular Biology* **2002**, 321, (5), 785-796.

65. Colbeau, A.; Elsen, S.; Tomiyama, M.; Zorin, N. A.; Dimon, B.; Vignais, P. M., Rhodobacter capsulatus HypF is involved in regulation of hydrogenase synthesis through the HupUV proteins. *European Journal of Biochemistry* **1998**, 251, (1-2), 65-71.

66. Dubay, K. F.; Pawar, A. P.; Chiti, F.; Zurdo, J.; Dobson, C. M.; Vendruscolo, M., Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *Journal of Molecular Biology* **2004**, 341, (5), 1317-1326.

67. Ferreira, S. T.; Vieira, M. N. N.; De Felice, F. G., Soluble protein oligomers as emerging toxins in Alzheimer's and other amyloid diseases. *Iubmb Life* **2007**, 59, (4-5), 332-345.

68. Demuro, A.; Mina, E.; Kaye, R.; Milton, S. C.; Parker, I.; Glabe, C. G., Calcium dysregulation and membrane disruption as a ubiquitous neurotoxic mechanism of soluble amyloid oligomers. *Journal of Biological Chemistry* **2005**, 280, (17), 17294-17300.

69. Dobson, C. M., Principles of protein folding, misfolding and aggregation. *Seminars in Cell & Developmental Biology* **2004**, 15, (1), 3-16.

70. Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G.; Dobson, C. M., Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **2003**, 424, (6950), 805-808.
71. Calamai, M.; Taddei, N.; Stefani, M.; Ramponi, G.; Chiti, F., Relative influence of hydrophobicity and net charge in the aggregation of two homologous proteins. *Biochemistry* **2003**, 42, (51), 15078-15083.
72. Schmittschmitt, J. P.; Scholtz, J. M., The role of protein stability, solubility, and net charge in amyloid fibril formation. *Protein Science* **2003**, 12, (10), 2374-2378.
73. Zbilut, J. P.; Mitchell, J. C.; Giuliani, A.; Colosimo, A.; Marwan, N.; Webber, C. L., Singular hydrophobicity patterns and net charge: a mesoscopic principle for protein aggregation/folding. *Physica A-Statistical Mechanics And Its Applications* **2004**, 343, 348-358.
74. Chiti, F.; Mangione, P.; Andreola, A.; Giorgetti, S.; Stefani, M.; Dobson, C. M.; Bellotti, V.; Taddei, N., Detection of two partially structured species in the folding process of the amyloidogenic protein beta 2-microglobulin. *Journal of Molecular Biology* **2001**, 307, (1), 379-391.
75. Tjernberg, L.; Hsiao, W.; Bark, N.; Thyberg, J.; Johansson, J., Charge attraction and beta propensity are necessary for amyloid fibril formation from tetrapeptides. *Journal of Biological Chemistry* **2002**, 277, (45), 43243-43246.
76. Hsiao, W.; Bark, N.; Liepinsh, E.; Tjernberg, A.; Persson, B.; Hallen, D.; Thyberg, J.; Johansson, J.; Tjernberg, L., Folding into a beta-hairpin can prevent amyloid fibril formation. *Biochemistry* **2004**, 43, (16), 4655-4661.
77. Calamai, M.; Chiti, F.; Dobson, C. M., Amyloid fibril formation can proceed from different conformations of a partially unfolded protein. *Biophysical Journal* **2005**, 89, (6), 4201-4210.
78. Campioni, S.; Mossuto, M. F.; Torrassa, S.; Calloni, G.; de Laureto, P. P.; Relini, A.; Fontana, A.; Chiti, F., Conformational properties of the aggregation precursor state of HypF-N. *Journal of Molecular Biology* **2008**, 379, (3), 554-567.
79. Calloni, G.; Zoffoli, S.; Stefani, M.; Dobson, C. M.; Chiti, F., Investigating the effects of mutations on protein aggregation in the cell. *Journal Of Biological Chemistry* **2005**, 280, (11), 10607-10613.
80. Marcon, G.; Plakoutsi, G.; Canale, C.; Relini, A.; Taddei, N.; Dobson, C. M.; Ramponi, G.; Chiti, F., Amyloid formation from HypF-N under conditions in which the protein is initially in its native state. *Journal of Molecular Biology* **2005**, 347, (2), 323-335.

81. Chiti, F.; Taddei, N.; White, P. M.; Bucciantini, M.; Magherini, F.; Stefani, M.; Dobson, C. M., Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Structural Biology* **1999**, 6, (11), 1005-1009.
82. Chiti, F.; Taddei, N.; Bucciantini, M.; White, P.; Ramponi, G.; Dobson, C. M., Mutational analysis of the propensity for amyloid formation by a globular protein. *Embo Journal* **2000**, 19, (7), 1441-1449.
83. Kyte, J.; Doolittle, R. F., A Simple Method for Displaying the Hydropathic Character of a Protein. *Journal of Molecular Biology* **1982**, 157, (1), 105-132.
84. Fisher, W. R.; Taniuchi, H.; Anfinsen, C. B., Role of Heme in Formation of Structure of Cytochrome-C. *Journal of Biological Chemistry* **1973**, 248, (9), 3188-3195.
85. Buscaglia, M.; Lapidus, L. J.; Eaton, W. A.; Hofrichter, J., Effects of Denaturants on the Dynamics of Loop Formation in Polypeptides *Biophysical Journal* **2006**, 91, 276-288.
86. Bezsonova, I.; Evanics, F.; Marsh, J. A.; Forman-Kay, J. D.; Prosser, R. S., Oxygen as a paramagnetic probe of clustering and solvent exposure in folded and unfolded states of an SH3 domain. *Journal Of The American Chemical Society* **2007**, 129, (6), 1826-1835.
87. Bezsonova, I.; Singer, A.; Choy, W. Y.; Tollinger, M.; Forman-Kay, J. D., Structural comparison of the unstable drkN SH3 domain and a stable mutant. *Biochemistry* **2005**, 44, (47), 15550-15560.
88. Chakrabortee, S.; Boschetti, C.; Walton, L. J.; Sarkar, S.; Rubinsztein, D. C.; Tunnacliffe, A., Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, 104, (46), 18073-18078.
89. Goyal, K.; Tisi, L.; Basran, A.; Browne, J.; Burnell, A.; Zurdo, J.; Tunnacliffe, A., Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. *Journal of Biological Chemistry* **2003**, 278, (15), 12977-12984.
90. Kragelund, B. B.; Robinson, C. V.; Knudsen, J.; Dobson, C. M.; Poulsen, F. M., Folding of a 4-Helix Bundle - Studies of Acyl-Coenzyme-a Binding-Protein. *Biochemistry* **1995**, 34, (21), 7217-7224.
91. Kragelund, B. B.; Osmark, P.; Neergaard, T. B.; Schiodt, J.; Kristiansen, K.; Knudsen, J.; Poulsen, F. M., The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP. *Nature Structural Biology* **1999**, 6,

(6), 594-601.

92. Teilum, K.; Maki, K.; Kragelund, B. B.; Poulsen, F. M.; Roder, H., Early kinetic intermediate in the folding of acyl-CoA binding protein detected by fluorescence labeling and ultrarapid mixing. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, 99, (15), 9807-9812.

93. Teilum, K.; Poulsen, F. M.; Akke, M., The inverted chevron plot measured by NMR relaxation reveals a native-like unfolding intermediate in acyl-CoA binding protein. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, 103, (18), 6877-6882.

94. Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M., Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *Journal Of The American Chemical Society* **2005**, 127, (2), 476-477.

95. Alonso, D. O. V.; Dill, K. A., Solvent Denaturation and Stabilization of Globular-Proteins. *Biochemistry* **1991**, 30, (24), 5974-5985.

96. Bennion, B. J.; Daggett, V., The molecular basis for the chemical denaturation of proteins by urea. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, 100, (9), 5142-5147.

97. Choi, H. S.; Huh, J.; Jo, W. H., Comparison between denaturant- and temperature-induced unfolding pathways of protein: A lattice Monte Carlo simulation. *Biomacromolecules* **2004**, 5, (6), 2289-2296.

98. Myers, J. K.; Pace, C. N.; Scholtz, J. M., DENATURANT M-VALUES AND HEAT-CAPACITY CHANGES - RELATION TO CHANGES IN ACCESSIBLE SURFACE-AREAS OF PROTEIN UNFOLDING. *Protein Science* **1995**, 4, (10), 2138-2148.

99. Stumpe, M. C.; Grubmuller, H., Urea Impedes the Hydrophobic Collapse of Partially Unfolded Proteins. *Biophysical Journal* **2009**, 96, (9), 3744-3752.

100. O'Brien, E. P.; Brooks, B. R.; Thirumalai, D., Molecular Origin of Constant m-Values, Denatured State Collapse, and Residue-Dependent Transition Midpoints in Globular Proteins†. *Biochemistry* **2009**, 48, (17), 3743-3754.

101. O'Brien, E. P.; Morrison, G.; Brooks, B. R.; Thirumalai, D., How accurate are polymer models in the analysis of Forster resonance energy transfer experiments on proteins? *Journal of Chemical Physics* **2009**, 130, (12), -.

102. O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D., Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, 105, (36), 13403-13408.
103. Kellermayer, M. S. Z.; Smith, S. B.; Granzier, H. L.; Bustamante, C., Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science* **1997**, 276, (5315), 1112-1116.
104. Czaplewski, C.; Rodziewicz-Motowidlo, S.; Liwo, A.; Ripoll, D. R.; Wawak, R. J.; Scheraga, H. A., Molecular simulation study of cooperativity in hydrophobic association. *Protein Science* **2000**, 9, (6), 1235-1245.
105. Zhang, H. Y.; Neal, S.; Wishart, D. S., RefDB: A database of uniformly referenced protein chemical shifts. *Journal of Biomolecular Nmr* **2003**, 25, (3), 173-195.
106. Gursoy, A.; Keskin, O.; Turkay, M.; Erman, B., Relationships between unfolded configurations of proteins and dynamics of folding to the native state. *Journal of Polymer Science Part B-Polymer Physics* **2006**, 44, (24), 3667-3678.