

ABSTRACT

Title of Document: **THERMODYNAMIC PROPERTIES OF
THE UNFOLDED ENSEMBLE OF
PROTEINS**

Tanay Mahesh Desai, Ph.D., 2010

Directed By: Professor Victor Muñoz
Centro de Investigaciones Biológicas, CIB-CSIC
Madrid, Spain

Adjunct Professor
Department of Chemistry and Biochemistry
University of Maryland, College Park

A random coil, whose size is determined by its excluded volume, and net energetic interactions with its environment, has served as a representation of the unfolded ensemble of proteins. The work in this thesis involves equilibrium, nuclear magnetic resonance and time-resolved kinetics spectroscopic studies on the unfolded ensemble of BBL, a globally downhill folding 40-residue protein involved the Krebs cycle of *E. coli*, in its acid-denatured state, and on a sequence-randomized version of this protein.

The effect of variability in thermodynamic conditions, such as temperature and the presence of added chaotropes or kosmotropes, on the equilibrium properties and reconfiguration dynamics of the unfolded state, have been deduced in the absence of

competition with folding events at low pH. The unfolded ensemble experiences expansion and collapse to varying degrees in response to changes in these conditions. Individual interactions of residues of the protein with the solvent and the cosolvent (direct interactions), and the properties of the solution itself (indirect interactions) are together critical to the unfolded chain's properties and have been quantitatively estimated.

Unfolded, protonated BBL can be refolded by tuning the properties of the solvent by addition of kosmotropic salts. Electrostatic interactions turn out to be essential for folding cooperativity, while solvent-mediated changes in the hydrophobic effect can promote structure formation but cannot induce long-range thermodynamic connectivity in the protein. The effect of sequence on the properties of heteropolymers is also tested with a randomized version of BBL's sequence. Chain radii of gyration, and the degree and rate of hydrophobic collapse depend on the composition of the sequence, viz. hydrophilic versus hydrophobic content. However, the ability to maximize stabilizing interactions and adopt compact conformations is more evident in naturally selected protein sequences versus designed heteropolymers.

Chain reconfiguration of unfolded BBL takes place in $\sim 1/(100 \text{ ns})$, in agreement with theoretical estimates of homopolymer collapse rates. The refolding dynamics of salt-refolded BBL in the range of $1/(6 \text{ } \mu\text{s})$ at 320 K, emerge as being independent of the degree of folding or protonation of the chain, a result in keeping with the description of dynamics in BBL as oscillations in a single, smooth harmonic potential well, which only varies in its position and curvature with varying thermodynamic conditions.

THERMODYNAMIC PROPERTIES OF THE
UNFOLDED ENSEMBLE OF PROTEINS

By

Tanay Mahesh Desai

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Professor Victor Muñoz, Chair
Professor David Fushman
Professor George Lorimer
Assistant Professor Arthur LaPorta
Associate Professor Wolfgang Losert

© Copyright by
Tanay Mahesh Desai
2010

Acknowledgements

I would like to thank my thesis advisor Prof. Victor Muñoz for the opportunities he gave us, his patience and an insistence on never settling for anything other than an exact answer to any scientific question. During the years spent in his lab, his constant encouragement, constructive criticism and advice ensured that we looked at the smallest problem in the greatest detail.

My colleagues over the years- Abhinav, Athi, Christina, David, Eva de Alba, Fabiana, Jian Wei, Li Peng, Lorenzo, Luis, Rani, Ravi, Michele, Mourad, Susana and Urmi- have all been exceptionally helpful to me, and I've learnt a lot from them. My time in Maryland was a breeze thanks to my friends Athi, Christina, Marvourneen, Sangita, Hulo, Debo, Prateek, Nozomi and An-Ni, and Kousik and Jason from the Doyle lab. I also owe a lot to Athi, Aparna, Christina, David, Mourad and Sunita for being my pillars of strength through these years. Without them I may not have made it!

My parents Parul and Mahesh made me smile when I didn't, and made sure I never got too complacent; thank you for having been patient and encouraging. My sister Hemal and cousin Noopur have been thoroughly loving and entertaining through the years! I would like to thank my friend and wife Sunita, for her help and support throughout. A big hug to my grandmothers Veerumati and Savitaben for being the sheet anchors of the family. My aunt and uncle Nita and Rajiv, in a way, got this whole science juggernaut rolling! My granddads Ishwarlal and Balwantrai: You'd better be watching over. This is a dedication to the two of you, wherever you are.

Table of Contents

Acknowledgements	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Abbreviations	vii
1. Introduction and Research Objectives	1
1.1 Statistical Methods in Protein Folding.....	1
1.2 The Unfolded Ensemble as a Random Coil.....	6
1.3 Swelling and Collapse of an Unfolded Protein.....	8
1.4 The Effect of Kosmotropes on an Unfolded Protein	10
1.5 Research Objectives.....	11
2. Experimental Techniques: Principles and Methods	14
2.1 Förster Resonance Energy Transfer (FRET)	14
2.1.1 Fluorescence Spectroscopy.....	14
2.1.2 Resonance Energy Transfer (RET).....	16
2.2 Circular Dichroism Spectroscopy (CD).....	20
2.3 Singular Value Decomposition (SVD)	24
2.4 Infrared Spectroscopy (IR)	26
2.5 Dynamic Light Scattering (DLS).....	29
2.6 Nuclear Magnetic Resonance Spectroscopy (NMR)	32
2.7 Laser Temperature-Jump (T-jump) Instrumentation and Measurements	32
2.8 Protein and Dye Concentration Measurements.....	32
2.9 Protein Sequences	39
3. The Unfolded Ensemble: The Effect of Cosolvents	40
3.1 Introduction.....	40
3.2 The Effect of Chemical Denaturants on Unfolded Proteins	43
3.3 Analysis of Variations in the Size of Unfolded Chains	45
3.4 Extension of the Flory-Huggins Theory to 3-Component Mixtures.....	54
3.5 Features of the 3-Component Energy Equation.....	57
3.5.1 Basic Features of the Quadratic Equation.....	57
3.5.2 A Limitation of the Probability Distribution Function	61
3.5.3 Effect of the Exchange Energetic Parameters on Chain Size	62
3.5.4 Effect of Chain Length on Extracted Energetic Parameters	69
3.5.5 Exchange Constants of Cosolvent/Solvent Interaction with the Residue..	75
3.6 Application of the Quadratic 3-Component Energy Equation to the Analysis of Data from Single Molecule FRET (smFRET) Experiments.....	79
3.6.1 Protein Unfolding Studied by smFRET.....	79
3.6.2 Expansion and Collapse of the Unfolded Ensemble.....	81
3.6.3 Application of the 3-Component Quadratic Equation to smFRET Data ...	92
3.7 Conclusions.....	103
4. The Effect of Chaotropes and Kosmotropes: BBL as a Case Study	108
4.1 Introduction.....	108

4.2 BBL: A Globally Downhill Folding Protein.....	112
4.2.1 Evidence for Global Downhill Folding.....	112
4.2.2 On the Ph-Sensitivity of BBL.....	115
4.3 Expansion and Collapse of Acid-Denatured BBL.....	119
4.3.1 Hydrophobic Collapse of Acid-Denatured BBL.....	121
4.3.2 The Effect of Guanidinium Chloride and Urea on Unfolded BBL.....	123
4.3.3 Charge-Screening Effects on a Chain by a Strong Electrolyte.....	128
4.4 Refolding of Acid-Denatured BBL by Salt.....	130
4.4.1 Secondary Structure Formation in Unfolded BBL.....	130
4.4.2 The Reduced Cooperativity of Salt-Refolded BBL.....	137
4.5 Salt-Refolded BBL is not a Partially-Structured Non-Native State.....	143
4.6 BBL Refolding Dynamics at Varying Degrees of Chain Protonation.....	150
4.7 Conclusions.....	150
5. Polymeric Properties of Random versus Natural Sequences.....	161
5.1 Introduction.....	162
5.2 Equilibrium Expansion and Collapse of Ran-BBL.....	165
5.3 Conclusions.....	169
6. Future Perspectives.....	172
Bibliography.....	173

List of Tables

Table	Page Number
--------------	--------------------

3.1	55
3.2	82
3.3	96
3.4	97
3.5	143

List of Figures

Figure	Page Number	Figure	Page Number
1.1	2	4.7	119
2.1	16	4.8	122
2.2	35	4.9	123
2.3	37	4.10	124
3.1	49	4.11	124
3.2	50	4.12	125
3.3	53	4.14	126
3.4	55	4.14	127
3.5	63	4.15	128
3.6	65	4.16	131
3.7	70	4.17	133
3.8	71	4.18	133
3.9	72	4.19	133
3.10	73	4.20	134
3.11	77	4.21	137
3.12	84	4.22	142
3.13	84	4.23	147
3.14	87	4.24	148
3.15	89	4.25	149
3.16	90	4.26	149
3.17	91	4.27	153
3.18	94	4.28	154
3.19	95	4.29	155
3.20	98	4.30	157
3.21	99	4.21	161
3.22	103	5.1	166
4.1	113	5.2	166
4.2	115	5.3	167
4.3	116	5.4	167
4.4	116	5.5	168
4.5	117	5.6	168
4.6	118		

Abbreviations

N: Number of amino-acid residues in the protein
R: Universal Gas Constant
T: Temperature
3D: Three-dimensional
GdmCl: Guanidinium chloride
NMR: Nuclear Magnetic Resonance
FRET: Förster Resonance Energy Transfer
SmFRET: Single-molecule FRET
T-jump: Temperature-jump
 N_A : Avogadro Number
Naf: Naphthyl group
Dan: Dansyl group
UV: Ultraviolet
CD: Circular Dichroism
SVD: Singular Value Decomposition
IR: Infrared
FTIR: Fourier-Transform IR
DLS: Dynamic Light Scattering
ACF: Auto-correlation function
SOFAST-HSQC: Heteronuclear Single Quantum Coherence

Chapter 1. Introduction and Research Objectives

1.1 Statistical Methods in Protein Folding

Protein self-organization, or folding, is the process by which a polypeptide adopts specific three-dimensional structures in a reversible manner in the time-scale of microseconds to several seconds. An understanding of the mechanism of folding to this complex structure, and the description of the underlying physics and chemistry of the phenomenon, is a problem that has received the attention of chemists, biologists and physicists alike for over 60 years now, starting with seminal ideas on the contribution of the hydrogen-bond to helical structures in polypeptides by Pauling and Corey². Enumeration of the various interactions that drive the folding or stabilization of a protein was taken further in a review by Kauzmann where he argued in favour of the importance of the “hydrophobic interaction” to native structure, borrowing on the ideas of Irving Langmuir and J. D. Bernal³⁻⁴ on the repulsion between aliphatic molecules and water. Since then, it has been learnt that electrostatic interactions, van der Waals forces, hydrophobic interactions, and at times coordinate-covalent and covalent bonds all contribute to keep the protein in its folded state⁵, and that dihedral angles in the protein amide bond fall within specific ranges for folded polypeptide configurations in the Ramachandran plot⁶.

On the time-scales of adoption of such a structure starting from an unfolded polypeptide, a perplexing combinatorial argument known as the Levinthal’s paradox⁷ stated that a random search for the native structure by the sampling of all possible

configurations and dihedral angles by a protein's backbone, would never allow folding to proceed in biologically relevant time-scales. Thus an argument was made in favour of specific folding pathways. Anfisen's experimental identification of the reversibility of folding⁸ upon denaturation took us a step further in our understanding of protein folding.

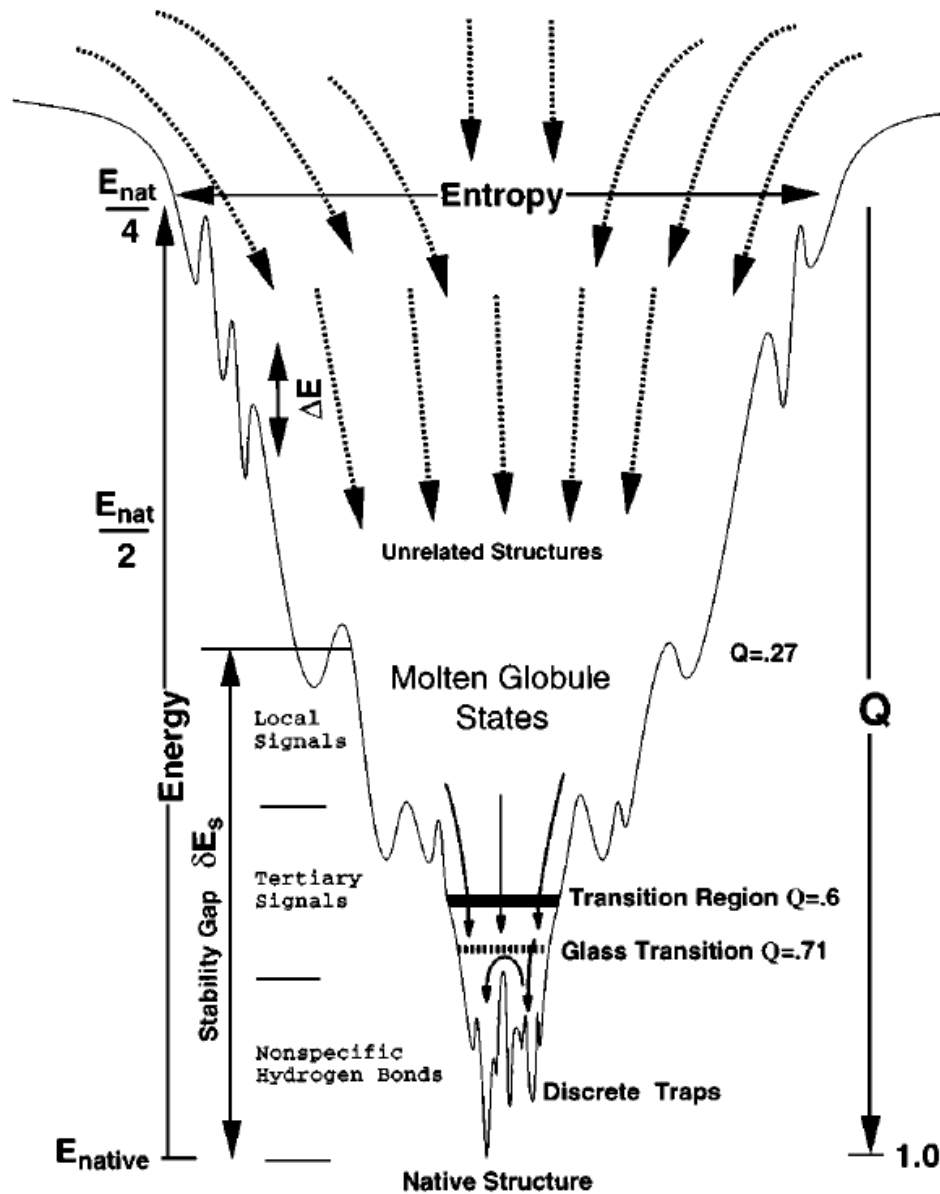


Figure 1.1 The folding funnel¹.

Since then, remarkable strides in the physical understanding of the folding mechanism have been made, ultimately culminating in our current view of protein folding- that it proceeds by an energetically biased search for a low energy, compact structure, i.e. the landscape perspective⁹. This view of protein folding stands in contrast to the ‘pathway’ idea of folding, and its genesis lay in the statistical description of the free energy landscape of a protein based on concepts from condensed matter physics, polymers and phase transitions¹⁰⁻¹¹. The free energy surface of a protein is hyperdimensional and its projection onto two order parameters, its free energy and conformational heterogeneity, resembles a funnel. The former defines the slope towards the bottom of the funnel and latter its width or circumference. The surface of this energy landscape is characterized by troughs and crests which introduce roughness on the landscape. Physically this would be equivalent to random rotations and vibrations of bonds and atomic steric clashes due to RT energy. Fluctuations with no net decrease in free energy are relatively fast compared to events where non-covalent interactions drive the protein closer to the minimum¹. When the landscape is smooth and has broad minima, first-order-like transitions take place. The experimental evidence for this is the existence of two-state proteins, whose unfolded ensemble and folded ensemble are separated by a high energy barrier. To avoid getting trapped in local minima on a rough landscape, protein sequences are naturally selected to traverse pathways of ‘minimal frustration’. Thus, protein folding does not necessarily proceed through a single progression of bond-formations, as small molecules do, but through a multitude of minimally frustrated paths from the tip of the funnel towards the bottom.

The speed or kinetics of the search for this energy minimum depends directly on the nature of the surface. In the description of effects of the landscape on folding kinetics, multiple scenarios were envisaged. A ‘Type 0’ surface is a landscape with a unimodal distribution with high bias towards the low energy folded state and no barrier separating the top and bottom of the funnel; thus, folding proceeds without energetic (or thermodynamic) bottlenecks. This is also called the downhill folding scenario. However, in this case *kinetic* bottlenecks, that affect a protein’s pause-time at any point on the funnel during its diffusion along it, do affect folding rates. The second scenario which is called “Type I” is characterized by a bimodal distribution of energy, separated by a high barrier. Type II is the third scenario when a glass transition appears at intermediate order parameters between the native and unfolded ensembles and slows down the search for the native structure.

Energy barriers between the native and the unfolded state determine the degeneracy of populations at intermediate order parameters. Thus, the ability to experimentally observe protein folding trajectories is limited by low population at intermediate degrees of nativeness when the barrier is high. This argument of course, does not hold in the case of downhill folding, since the existence of a single mode in any condition allows observation of folding events going from native to unfolded¹²⁻¹³. On the experimental side, simple two-state analyses, commonly used to analyze and interpret folding experiments, assume the existence of a barrier *a priori*. However, a model-free estimation of the height of the barrier is critical to experimentally determining the motions of the protein along the energy landscape¹⁴. Ultrafast time-resolved laser spectroscopic techniques and statistical models such as the ones used to

describe helix-coil transitions and β -hairpin formation have culminated in a deeper understanding of the free energy changes accompanying the stabilization of the backbone¹⁵⁻¹⁷ and time-scales of various protein motions. Peptide bond rotations take place on the time-scale of 1-2 ns¹⁸. α -helices form in ~ 200 ns and β -hairpins in $\sim 5\mu\text{s}$, while random hydrophobic collapse takes place in 100 ns, and contact formation between the ends of chains occurs in several microseconds¹⁸. The basis of evaluating the rates of various motions of the protein chain is firstly to analyze chain dynamics down to the fundamental diffusive motions along the free energy landscape. Secondly, an estimation of the fundamental mode of vibration of a protein is critical to the accurate estimation of barrier heights between the native and unfolded state. The global mode of such a vibration would be the fastest speed at which a protein could fold, i.e. the folding speed limit. This has led to the estimation of $N/100 \mu\text{s}$ to be the speed limit for a globular single domain protein¹⁹, where N is the number of amino acids in the protein.

These approaches have been buttressed on a parallel front, by development of methods to directly estimate barrier heights from calorimetric data²⁰. Proteins with barrier heights $>4 RT$ are two-state like in their behaviour. Proteins with marginal barrier heights $<2 RT$ would exhibit fast folding and probe-dependent observations. In the intermediate regime, the barrier height would be sensitive to thermodynamic conditions, thus switching the folding from a Type I to a Type 0 scenario.

The energy landscape view of protein folding can describe a broad range of possible behaviour that a protein can exhibit starting from its existence as an unstructured polypeptide to a fully folded structure. The dependence of these

properties on thermodynamic conditions and on the nature of the protein sequence itself has made it possible to statistically analyze and interpret experimental data on protein folding, viz. understanding a protein's thermodynamic and kinetic properties in the context of its energy landscape. The ultimate goal is to be able to define a protein's three-dimensional (3D) structure solely based on the knowledge of its sequence and environment- one of the central questions in cellular and molecular biology related to the translation of the genetic code to biological function.

1.2 The Unfolded Ensemble as a Random Coil

The statistical description of the protein in its unfolded state is the random coil. This description has served as the standard representation of the unfolded state and the starting point for protein folding simulations. Quantitative support for this representation came from intrinsic viscosity experiments by Tanford and coworkers²¹ on proteins unfolded in 6 M guanidinium chloride (GdmCl), measurements of unfolded chain hydrodynamic radii (R_h) by nuclear magnetic resonance (NMR) techniques²², spectroscopic methods measuring the radii of gyration (R_g)²³ of unfolded chains, and X-ray and neutron scattering experiments²⁴⁻²⁵. Random coils in the absence of bond-angle correlations and steric clashes follow a power-law dependence where R_g scales as $N^{1/2}$ (sequence length), based on simple Gaussian chain scaling principles. The inclusion of the chain's excluded volume changes the exponent to $\sim 3/5$. The latter power-law dependence was observed in the properties of the unfolded ensemble in the aforementioned studies within experimental error. Thus, in conditions with high denaturant concentration, proteins exist in the random coil

regime having reached their maximal degrees of expansion. This has important implications for folding reactions. The rate of chain reconfiguration in the unfolded ensemble would scale as the inverse of chain length N for a purely Gaussian chain²⁶⁻²⁷. Reconfiguration dynamics take place in the 100 ns time-scale in case of unfolded BBL during hydrophobic collapse and in the time scale of ~50 ns for the cold shock protein from *Thermatoga maritima* (CSP)²⁸. Chain dynamics in BBL were shown to depend on solvent viscosity alone up to a temperature 305 K followed by a decrease in the rate of collapse with increasing temperature and degree of collapse. This was attributed to the onset of roughness in the free energy landscape. The dynamics of CSP in high GdmCl concentrations are dependent primarily on the change in viscosity associated with added denaturant. In contrast, a study on the rate of end-to-end contact formation in a 20-residue unstructured polypeptide, by Kiefhaber and coworkers²⁹, revealed a rate of contact formation of 1/(25 ns), that significantly decreased with increasing GdmCl. This implied that the interaction of GdmCl molecules with the unfolded chain, slowed down chain dynamics, with a ~5 ns residence time of a denaturant molecule on the protein²⁹. These results together indicate that theoretical predictions of collapse rates from homopolymer theory hold good for unfolded polypeptides, though specific sequence effects such as the onset of chain ruggedness with increasing compaction can also affect collapse.

While several independent studies point to the random coil model as being a faithful description of an unfolded protein's properties, NMR spectroscopic measurements on chemically denatured states tell a different story³⁰. The observation of non-zero anisotropy by the measurement of residual dipolar coupling by NMR, in

unfolded staphylococcal nuclease at 8 M urea, pointed to the existence of native topology in the unfolded, expanded state. Native topology persists in the unfolded state of staphylococcal nuclease upon the mutation of several residues on the protein³¹, and also in short peptides³². Local structure in an unfolded protein should produce deviations from the random coil behaviour^{22,33}. Spectroscopic studies on short unfolded peptides³⁴ have shown the existence of polyproline-II helices in the absence of denaturant even for sequences with low proline content.

These diverse lines of evidence may be reconciled by the recognition of the fact that local-structure may not globally propagate to the entire sequence, and random coil-like properties can be exhibited by chains with significant local structure³⁵. Furthermore, NMR spectroscopic studies on the unfolded state could be biased by dipolar couplings in the range of several kHz, as opposed to a few Hz in the case of residual dipolar couplings, thus biasing observations in favour of those rare events where native topology is adopted.

1.3 Swelling and Collapse of an Unfolded Protein

Single molecule Förster resonance energy transfer (smFRET) experiments have allowed a direct observation of the effect of chaotropes on the unfolded ensemble in the absence of signals from the native state, as is the case in bulk experiments³⁶. The unfolded ensemble experiences an expansion in chain size with added denaturant. This is due to a change in the net properties of the solution and thereby, a change in the net interaction of the unfolded protein with its environment. According to the Tanford transfer model³⁷ and the Schellman site-exchange model³⁸,

this effect has its basis in the interactions of the denaturant with the protein. Direct interaction of denaturants such as GdmCl and urea with the protein, i.e. their binding to the protein backbone or to side chains, takes place through electrostatic and hydrogen bonding interactions³⁹⁻⁴⁵. What is not clear is whether indirect solvent-mediated effects, included in the Schellman model, also play a role in protein unfolding or expansion. These effects are a result of interactions between denaturant and water molecules, and changes in the structure of water with cosolvent addition⁴⁶⁻⁵⁴.

The acid-denatured state of BBL, a 40-residue α -helical *E. coli* protein serves as an ideal system to study the effect of thermodynamic parameters such as temperature and denaturants on chain size, in the absence of competition from folding events. At acidic pH in the absence of chemical, there is no bias towards the formation of native structure, and the effect of added denaturants on chain size and dynamics can be studied by bulk FRET measurements on a doubly, extrinsically labeled variant of the protein. The effect of chain size and thermodynamic conditions such as chemical concentration and temperature, on chain dynamics could be studied by ultrafast nanosecond laser temperature jump (T-jump) FRET measurements, which would map the rate of collapse of the protein. Collapse rates can be measured at the same degree of chain collapse, with different combinations of chemical and temperature. This would help in describing the interaction of the chemical with the protein and the dependence of this interaction on size and temperature.

The unfolded ensemble of BBL refolds with increasing pH. pH modulates the electrostatic charge, and thereby the electrostatic interactions between charged groups

on the chain. An increase in pH, for e.g. would introduce a net electrostatic repulsion on a chain with a large number of acidic amino acids, or conversely a decrease in pH would do the same for a chain with a large number of basic amino acids⁵⁵⁻⁵⁶. In the absence of refolding events, for a random heteropolymer, change in the net electrostatic interaction should therefore cause expansion or collapse. This effect can be tested on acid-denatured BBL through T-jump FRET experiments, although at progressively higher pH a significant contribution from folding can be expected.

If unfolded proteins are indeed random polymer-like in their swelling and collapse properties, then sequence effects on these properties should be self-averaging. This is to say that these properties should depend only on the net sequence composition of a protein and not on the sequence itself. The dependence of properties of the unfolded ensemble on chain sequence versus dependence on sequence composition can be tested on a randomized sequence variant of the protein BBL. Such a sequence could be obtained so as to have no propensity for structure under any condition. The effect of an additional parameter, pH, can now also be described by bulk FRET and T-jump experiments on the randomized version of BBL (Ran-BBL).

1.4 The Effect of Kosmotropes on an Unfolded Protein

While chaotropes induce chain swelling, kosmotropes, which interact strongly with bulk water, are expected to exclude the protein from the bulk solution⁵⁷⁻⁵⁸. The presence of kosmotropic cosolvents can, in principle, bring about protein compaction. Kosmotropes are known to stabilize the native state against thermal denaturation⁵⁹. The effect of added cosolvents (chaotropic or kosmotropic) on proteins is ultimately

connected with the changes in solvent quality they bring about- the so-called indirect solvent-mediated effect. The perturbation of the conformational space and stability of the unfolded protein and the stability of the folded protein in the presence of added cosolvents, stabilizing or destabilizing as they may be, could be crucial to the folding mechanism of proteins in crowded cellular milieu, where organic molecules, dissolved salts and large macromolecules interact with the protein. Based on simple polymer solution thermodynamics arguments, kosmotropes and chaotropes differ in their effects on proteins due to differences in their interactions with the solvent⁶⁰⁻⁶¹. Their effect on the size distribution of the unfolded protein, for e.g., can be studied with acid-denatured BBL. Strong, neutral kosmotropic salts such as LiCl and NaCl can be used to test the effect of kosmotropes on the acid-denatured state. What remains to be seen is if the addition of cosolvents to the denatured ensemble brings about chain compaction alone, or also restructuring in the spirit of salt-induced refolding of acid-denatured apomyoglobin illustrated in a series of articles by Fink and coworkers⁶²⁻⁶⁴ and several other proteins.

1.5 Research Objectives

The stage is set for a detailed study of the effect of a diverse range of thermodynamic conditions on the unfolded ensemble. Specifically, the acid-denatured state of the downhill folding protein BBL serves as an ideal system to study the thermodynamic properties of an unfolded ensemble in the absence of competing folding reactions, even in the absence of denaturant. While smFRET experiments can yield a wealth of data on the dependence of the size and dynamics of unfolded

proteins in isolation from signals coming from the native protein, there is a lower limit to the denaturant concentration in which an appreciable population of the denatured state can be observed. Properties of chain collapse and expansion in the absence of folding reactions and denaturants, an observation evasive to bulk experiment, have been addressed by equilibrium bulk FRET and nanosecond laser FRET T-jump studies on the expansion and collapse of acid-denatured BBL. High temperatures bring about a collapse in the acid-denatured ensemble due to an increase in the hydrophobic effect⁶⁵, and induce an initial speed-up of collapse followed by a slow-down due to stronger chain hydrophobic interactions. These studies can be extended by measuring the degree of expansion and collapse of BBL in the presence of the chaotropes GdmCl, urea and kosmotropic cosolvents such as NaCl and LiCl. The relative contributions of solvent viscosity and direct interactions of cosolvents with the protein backbone are expected to slow down the diffusive dynamics of end-to-end contact formation and collapse of a random chain. These effects, and their dependence on temperature, can be studied by measuring collapse rates in the presence of chaotropes, after having characterized their effects on the equilibrium population. If kosmotropic effects bring about protein restructuring, a novel experimental scenario can be envisaged where refolding dynamics of a downhill folding protein in a highly protonated state can be studied by IR T-jump measurements. Finally, the theoretical basis for the analysis of the equilibrium properties of unfolded chains lies in polymer physics and long-standing ideas of polymer solution thermodynamics which will be applied to the analysis of chain sizes of the unfolded states of different proteins.

The unfolded ensemble can be described as having the properties of a random coil. Thus, its properties should be a function of the net interaction of each residue with its environment. The specific effect of sequence on the properties of an unfolded protein will be tested on an unstructured peptide, with the same sequence composition as BBL, but with a completely randomized sequence. Chain size and dynamics of this randomized sequence can be measured to give a clearer picture of specific sequence effects in events such as random hydrophobic collapse.

Ultimately, the goal is to provide a deeper understanding of the folding landscape at the top of the funnel, so as to understand folding mechanisms starting from unfolded sequences, and indeed, to predict the structures of folded proteins solely based on its sequence. While several decades of research have gone into our understanding of the native state, gaining an understanding of the unfolded ensemble is the other side of the coin in the study of protein folding mechanisms. These studies are especially critical in the light of the finding that several functionally important proteins have been identified as being natively unfolded inside the cell⁶⁶⁻⁶⁹, and that a balance towards a larger fraction of charged residues, compared to hydrophobic residues, renders them unstructured in physiological conditions⁷⁰.

Chapter 2. Experimental Techniques: Principles and Methods

2.1 Förster Resonance Energy Transfer (FRET)

2.1.1 Fluorescence Spectroscopy

Electronic absorption of light refers to the interaction of a certain quanta of light of a specific energy with the electrons of a molecule, accompanied by an excitement of the molecule's electrons to higher quantized electronic energy levels such as the first singlet and second singlet excited state, from its electronic ground state. The excited electrons can return back to the ground state by several processes: traversing several vibrational and rotational energy levels in their excited singlet state, returning to the ground state by spontaneously emitting a photon, crossing over to the excited triplet state, interconversion between vibrational and rotational states in the ground state, and losing their energy by non-radiative processes. These several possibilities are summarized in the Jablonski diagram (Figure 1). Absorbing groups in proteins are the amide group which absorbs at about 220 nm owing to a $n\pi^*$ transition and at about 195 nm due to a $\pi\pi^*$ transition. Most side-chains absorb light below 200 nm. Phenylalanine, tyrosine, tryptophan, cysteine, methionine and disulphide groups absorb radiation at wavelengths just below 300 nm.

Fluorescence is a class of molecular luminescence which involves spontaneous emission of light by molecules that have excited singlet state electrons. It

is preceded by absorption of a quantum of light, following which an electron from the ground state is excited to the singlet state. Under certain conditions this excited electron can return to the ground state by spontaneously emitting a photon. This phenomenon is called fluorescence. Fluorescing molecules emit light at wavelengths typically longer than the excitation wavelength due to what is known as the Stoke's shift. This shift is seen because of rapid interconversion between various vibrational bands in the singlet and ground state, thereby reducing the amount of energy carried by the emitted photon. In practice, the probability of spontaneous emission of a photon by an excited molecule (quantum yield), the intensity of fluorescence and the final wavelength of the emitted photon can depend on the properties of the molecule, properties of the surrounding solvent and specific interactions or reactions that the fluorophore may have when in the excited state. It is for these reasons that fluorescence is widely used in the study of protein folding, since it allows the possibility of tracking changes in the local environment of specific fluorescing groups. Tryptophan, tyrosine and phenylalanine comprise the intrinsic fluorophores in proteins. It is common to chemically attach extrinsic fluorophores to proteins to monitor conformational changes, since fluorophore absorption, emission, rotational correlation time and anisotropy are sensitive to the environment.

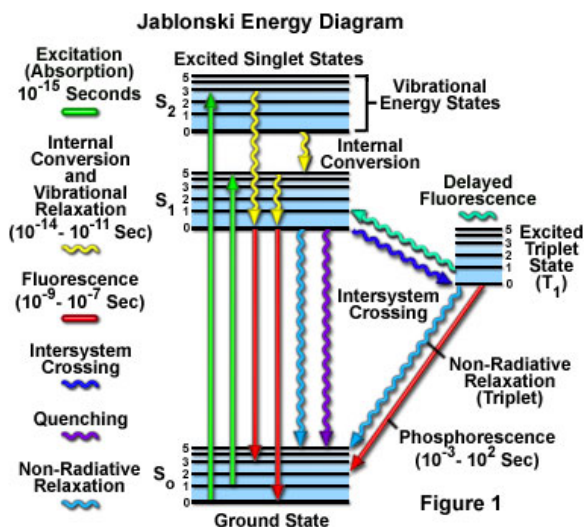


Figure 2.1 The Jablonski diagram describing the various electronic transitions upon absorption of radiation by a molecule (Source: Olympus Microscopy Resource Web-page).

2.1.2 Resonance Energy Transfer (RET)

Resonance energy transfer describes the non-radiative transfer of energy from an excited donor to an acceptor in its vicinity. The class of RET that is described here is heterotransfer, wherein the energy is transferred between two different kinds of chromophores. This energy transfer process occurs through interaction between the transition dipoles of the donor and acceptor. Förster connected the rate of this energy transfer phenomenon to the rates of fluorescence decay and non-radiative energy dissipation in the donor by the Förster rate equation and the distance of separation between the donor and acceptor.

$$k_T = (k_D + k_{Di}) \left(\frac{R_0}{R} \right)^6 = \frac{1}{\tau_D} \left(\frac{R_0}{R} \right)^6 \quad (2.1)$$

where k_T , k_D and k_{Di} are the rates of energy transfer from the donor to the acceptor, the rate of decay of fluorescence of the donor, and the rates of non-radiative dissipation of the excited electron in the donor, respectively. τ_D is the intrinsic

fluorescence lifetime of the donor, and R_0 is the characteristic Förster distance or Förster radius. The energy transfer efficiency describes the fraction of energy transferred from the donor to the acceptor in comparison with the total energy absorbed by the donor.

$$E = \frac{R_0^6}{(R_0^6 + R^6)} = 1 - \frac{\tau_D}{\tau_{DA}} = 1 - \frac{Q_D}{Q_{DA}} \quad (2.2)$$

where τ_{DA} is the lifetime of fluorescence of the donor in the presence of the acceptor, and Q_D and Q_{DA} , are donor fluorescent quantum yield (QY) in the absence and presence of the acceptor, respectively. Since the energy transfer depends on the interaction between molecular transition dipole moments, a $1/6^{\text{th}}$ distance dependence is observed. The Förster distance is a number that defines the distance at which 50% of the energy absorbed by the donor is passed on to the acceptor. It is arrived at based on a derivation by Förster, which connects this distance to the spectral properties of the donor and the acceptor.

$$R_0^6 = \frac{9000 \cdot \ln 10 \cdot \kappa^2 \cdot Q_D \cdot J}{128 \cdot \pi^5 \cdot \eta^4 \cdot N_{AV}} \quad (2.3)$$

where κ^2 is an orientation factor describing the angular orientation between the transition dipole moments of the donor and acceptor, relative to the physical angle between the two chromophores, Q_D is the intrinsic quantum yield of the donor, J is the spectral overlap integral between the donor and acceptor, η is the refractive index

of the bulk solvent, and N_A is the Avogadro number. The overlap integral J can be calculated from first principles to arrive at a wavelength dependent expression below.

$$J = \int f_D(\lambda) \cdot \epsilon_A(\lambda) \cdot \lambda^4 \cdot d\lambda \quad (2.4)$$

where $f_D(\lambda)$ is the normalized donor fluorescence signal and $\epsilon_A(\lambda)$ is the molar extinction coefficient of the acceptor at the wavelength λ . Apart from considerations concerning the orientational averaging of the donor and acceptor which affect κ^2 , the Förster theory can be readily applied to a pair of carefully selected chromophores to serve as a ‘spectroscopic ruler’ to measure distances between the donor and acceptor. On the issue of orientational averaging of the molecules, fast averaging of the donor and acceptor transition dipoles, relative to fluorescence decay lifetimes of the donor and the rate of motion of the donor relative to the acceptor, results in a value of 2/3 for complete isotropic dynamic averaging of the donor and acceptor, which is assumed to hold true for applications to protein spectroscopy. This however, is not a trivial assumption to make and care needs to be taken in considering interactions of the donor or acceptor with its environment, and a comparison of the lifetime of fluorescence decay with anisotropic motions of the donor or acceptor.

Experimental Conditions. The end-to-end distances for the protein BBL, under varying conditions, were reported by a FRET-pair comprising two fluorophores, naphthyl-alanine (the donor) and dansyl-lysine (the acceptor) at the N and C termini of the protein BBL, respectively. Naf-BBL and Naf-BBL-Dan, the singly and doubly-labeled versions, and Ran-BBL a doubly-labeled randomized sequence of BBL with

the same FRET pair, were produced by solid-state chemical synthesis by California Peptide Inc., Napa, CA. Equilibrium FRET efficiency was determined by evaluation of the quantum yield of the donor in the doubly and singly-labeled versions of the peptide. The quantum yield of the donor was measured by comparison of fluorescence from the emission band of naphthyl-alanine, spanning 300 to 450 nm, with that of N-Acetyl-L-Tryptophanamide, which has a quantum yield of 0.13 at pH 7.0 and 298 K. All fluorescence measurements were performed on a Jobin Yvon Fluorolog-3.v.2.2 (Edison, NJ), at a protein concentration of approximately 25 μM at pH 3.0, in 20 mM citrate buffer with excitation at 288 nm, a bandpass of 5 nm at the excitation and emission side, a 0.25 second integration time and 1 nm resolution, in 1 cm path length cuvettes.

Buffer reagents were of chemical grade, obtained from Merck. Salt concentrations in protein samples were measured by weight. NaCl was purchased from Merck, Protein concentrations were estimated by ultra-violet (UV) absorbance spectroscopy on a Cary100 Bio spectrophotometer (Varian Inc., Palo Alto, CA) by measurement of absorbance at 266 nm, the isosbestic point for absorbance variations of dansyl-lysine with pH, and at 280 nm, the absorbance maximum of naphthyl-alanine. The molar absorptivity of naphthyl-alanine at pH 3.0 is 3,595 $\text{M}^{-1}.\text{cm}^{-1}$ at 266 nm, and 5,526 $\text{M}^{-1}.\text{cm}^{-1}$ at 280 nm, similar to that of Naf-BBL⁶⁵. Naphthyl-alanine and Naf-BBL concentrations were estimated as the average from determinations at the two wavelengths. Naf-BBL-Dan has a molar absorptivity of 8,345 $\text{M}^{-1}.\text{cm}^{-1}$ at 266 nm and pH 3.0. These values remain invariant with salt concentration and temperature. Dansyl-lysine absorbs at 266 nm with an absorptivity of 4,528 $\text{M}^{-1}.\text{cm}^{-1}$

at 266 nm and pH 3.0. N-Acetyl-L-Tryptophanamide (molar absorptivity of $5,690 \text{ M}^{-1} \cdot \text{cm}^{-1}$ at pH 7.0) from Sigma-Aldrich was used as the standard for quantum yield determinations. The R_0 of the FRET-pair was also evaluated at varying temperatures and cosolvent concentrations, to enable calculation of average end-to-end distances from FRET measurements in a whole range of sample conditions. Naphthyl-alanine has a quantum yield of 0.13 at pH 3.0 and 298 K, which decreases with increasing GdmCl concentration and temperature. Increasing concentrations of urea were found to have no effect on the quantum yield of naphthyl-alanine. The variation in the absorbance of dansyl-lysine in the spectral overlap region with naphthyl-alanine, was also measured at varying concentrations of cosolvents and temperature. The spectral overlap integral does not change with varying cosolvent concentrations, and it increases with temperature. We assume a value of $2/3$ for κ^2 , the orientation factor since the donor and acceptor transition dipoles are freely rotating⁶⁵. The R_0 of the FRET-pair is 1.86 nm at pH 3.0 and 298 K, and decreases with increasing concentration of GdmCl.

FRET measurements on Ran-BBL, the doubly-labeled randomized BBL sequence with the same donor and acceptor pair as Naf-BBL-Dan, were performed in the same conditions, but with concentrations of the protein at 10 μM . FRET efficiencies were estimated by comparison with the free naphthyl-alanine label.

2.2 Circular Dichroism Spectroscopy (CD)

Chiral molecules are asymmetric, and such molecules are said to be optically active. Biomolecules of interest, such as DNA and proteins are chiral and hence are

optically active. What does it mean for a molecule to be optically active? When light of a certain wavelength passes through an absorbing sample, the intensity of the transmitted light is lower. Consider this incident light to be linearly polarized, with its electric field vector oscillating along a given plane, with varying amplitude. A linearly polarized light beam is effectively, a combination of two in-phase, like amplitude, oppositely-sensed circularly polarized light waves. In an optically active sample, due to molecular asymmetry, the refractive index (propagation velocity in the medium) for the two forms of circularly polarized light are different, giving rise to a change in the angle of the linearly polarized light. This is known as circular birefringence, or optical rotatory dispersion. In addition, there is also a difference in the extent of absorption of the two circularly polarized components of the linearly polarized light wave, by the optically active sample. This property is called circular dichroism, with the measured quantity being the difference in absorption between right and left circularly polarized light by the asymmetric sample. Proteins are highly asymmetric because of their secondary and tertiary structure and because of the asymmetric α -carbon centers. For the former type, also known as super-asymmetry, optical activity arises because of the interaction between transition dipoles of absorbing chromophores. These are the interactions that allow circular dichroism to measure protein secondary structure. In the far-UV wavelength region (190-250 nm), an isolated amide group has an electronic absorption peak at 222 nm due to a weak $n\pi^*$ electronic transition, and another electronic absorption band with a peak at 195 nm due to a $\pi\pi^*$ transition. In proteins, owing to the interaction between the several amide transition dipoles in an ordered α -helix, the $\pi\pi^*$ transition is predicted by the

Exciton theory to have two components. One transition occurs at a lower energy with the transition dipole parallel to the helix axis, and the other at a higher energy, with the transition dipole perpendicular to the helix axis. Hence, the absorption due to $\pi\pi^*$ transitions is split into a negative and positive couplet, the former at 208 nm due to the parallel component of the transition and the other at 190 nm due to the perpendicular component. A random coil on the other hand shows CD peaks at predominantly 195 nm, similar to an isolated amide bond, due to the $\pi\pi^*$ transition and another at 230 nm due to the $n\pi^*$ transition. β -strands have a negative absorption band at about 215 nm and a positive band at about 200 nm. Proteins with a combination of α -helices, β -sheets, and β -turns show CD spectra with a combination of absorbances from the different sources of absorbance.

An instrument measuring electronic CD is called a spectropolarimeter and its output is typically in millidegrees. The historical reason behind this form of the measured quantity lies in the fact that differential absorption of the two circularly polarized components by an optically active substance yields a transmitted light beam with unequal amplitudes of the electric field vector of the two components. These combine to give an electric field vector which traces an elliptical path. The tangent of this ellipticity is related to the different amplitudes of the two circularly polarized components.

$$\tan \theta = \frac{E_r - E_l}{E_r + E_l} = \frac{I_r^{1/2} - I_l^{1/2}}{I_r^{1/2} + I_l^{1/2}} \quad (2.5)$$

where E represents magnitude of the electric field vectors for the right and left circularly polarized light, and I , their respective intensities. Upon approximating the tangent of the ellipticity to be numerically equal to the ellipticity itself (due to the small magnitude of the ellipticity), and replacing transmitted light intensities by the absorbance of the sample from Beer's Law, one can arrive at the expression for ellipticity in degrees:

$$\theta = \frac{\Delta A \cdot (\ln 10) \cdot 180}{4 \cdot \pi} = 32.99 \cdot \Delta A \quad (2.6)$$

where ΔA is the difference in absorbance of the left and right circularly polarized light. This quantity can be normalized to give molar ellipticity $[\theta]$ in $\text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$, which accounts for the path length l in units of cm traversed by light in the optically active medium, and the concentration C of the macromolecule in units of M.

$$[\theta] = \frac{100 \cdot \theta}{C \cdot l} \text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1} \quad (2.7)$$

An additional normalization, included for proteins because of the repeating chromophoric amide bond, yields the mean residual ellipticity:

$$[\theta]_{MR} = \frac{100 \cdot \theta}{C \cdot l \cdot N} \text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1} \quad (2.8)$$

where, N is the number of amide bonds in the protein. Since the typical instrumental read-out is in millidegrees, the conversion from read-out to mean residual ellipticity is:

$$[\theta]_{MR} = \frac{\theta_{\text{read-out}}}{10.C.l.N} \text{deg.cm}^2.\text{dmol}^{-1} \quad (2.9)$$

Analysis of CD-spectra is further discussed in the following section on Singular Value Decomposition of CD data.

Experimental Conditions. Circular Dichroism (CD) experiments were performed on a Jasco J-815 spectropolarimeter. All measurements were made using Naf-BBL or Ran-BBL at a concentration of $\sim 50 \mu\text{M}$ in 20 mM citrate buffer at pH 3.0. Mean residual ellipticity was estimated based on the concentration measurements of Naf-BBL or Ran-BBL for each sample. Temperature-based unfolding was monitored by collecting CD spectra with 1 nm resolution in a temperature range starting from 268 K to 368 K, every 5 K using a cuvette of 1 mm path-length, with a bandwidth of 2 nm, a response time of 16 sec and a scan rate of 10 nm.min^{-1} in continuous scanning mode. The reversibility of each run was checked by comparing spectra at 298 K collected before and after each temperature-ramp.

2.3 Singular Value Decomposition (SVD)

Singular value decomposition refers to the theory and procedure which states that any matrix A can be decomposed into a product of three matrices:

$$A = USV^T \quad (2.10)$$

where U is a unitary and orthogonal matrix, whose columns are an orthonormal basis set for the set of vectors (columns) in A , S is a diagonal matrix with the diagonal values representing singular values for the basis vectors in U in decreasing order of magnitude along the diagonal, and V^T is another unitary and orthogonal matrix containing the least-squared coefficients that regenerate the original matrix A upon multiplication with US . This procedure finds immediate application in the field of spectroscopy, as any series of spectra which are a linear combination of a set of basis representative spectra, can be split into their individual linearly independent contributing components, i.e. a procedure similar to but more robust than traditional principal component analysis. These components are found in the columns of the first matrix U , with the first vector in U being the average representative vector or spectrum of the data-set in A . The singular values in S are numerical estimates of the degree of contribution of each of the orthonormal basis vectors or spectra in U , to the overall series of spectra in A . Finally, the matrix V charts the changes in amplitude for each basis vector in U . This procedure is advantageous for various reasons in addition to principal component analysis. It offers the possibility of eliminating or reducing spectral noise, as the noise component often turns out as a distinct vector in the matrix U . Moreover, in spectroscopy, changes at particular wavelengths are indicative of changes in a specific structural property or spectral transition being observed. The vectors in the matrix V however are calculated after taking into account the entire spectrum, and contain the varying amplitude for each component vector in U . For

example, in CD experiments this would mean looking at the changes in the amplitude at 222 nm for the alpha-helix from the corresponding vector in V for the basis set in U that most resembles the average helical spectrum, usually the first column for a helical protein, instead of simply the signal at 222 nm in the raw spectra. The second column in U can be representative of other structural signatures such as random coil, β -sheet, or β -turn or could be a complex mixture of signals. In general, red or blue shifts of a spectra, and relative correlations or anti-correlations between different structural signatures present in each spectral band measured by any procedure such as absorbance, CD, fluorescence or IR can also be judged from the orthonormal basis set of vectors produced in the matrix U by SVD. Finally, SVD allows for the possibility of reducing the effective rank of a rank-degenerate large data-set, thus reducing the computational load of analysis.

2.4 Infrared Spectroscopy (IR)

Infrared spectroscopy is a class of vibrational spectroscopy which measures the absorption of light at frequencies that activate bond vibrational modes in molecules. It turns out that light with frequencies in the infrared region have the right energy for exciting the vibrational modes of atoms in a molecule. A non-linear molecule with n atoms has $3n$ degrees of freedom, of which three are translational and 3 are rotational. The remaining $(3n-6)$ degrees of freedom correspond to the fundamental modes of vibration of atoms around bonds that connect them. Absorption occurs when the frequency of light matches the energy needed to excite an atom from its ground vibration state to higher excited vibrational state. Fundamental absorption frequencies

are responsible for such transitions to the first excited vibration state and overtones are frequencies for states above the first excited vibrational state. In addition, one can observe absorbance bands corresponding to a combination of vibration modes and also when different vibrations mutually cancel (difference bands). Typical vibrational motions include symmetric and asymmetric bond stretching, and twisting, wagging, deformation, scissoring and rocking motions which are all bending motions. Organic groups of interest in biological applications have characteristic absorptions pertaining to specific vibrations. But, the observed frequency, intensity and probability of absorption is influenced not just by the nature of the atoms and bonds involved, but also the presence of neighbouring atoms that affect vibrational motions. In proteins, IR absorption is centered around mainly three bands. The amide I region around 1650 cm^{-1} contains contributions from the amide C=O stretching, C-N stretching and N-H bending. The amide II region around 1570 cm^{-1} has contributions from C-N stretching and N-H bending. Absorption in the amide III region around 1300 cm^{-1} comes from C-N stretching, N-H bending, C=O stretching and O=C-N bending. In addition the amide A region has contributions from C-N stretching at about 3300 cm^{-1} . The amide I region is usually invoked for identification of protein secondary structure with absorption in the $1621\text{-}1640\text{ cm}^{-1}$ and $1671\text{-}1679\text{ cm}^{-1}$ region coming from β -sheets, $1651\text{-}1657\text{ cm}^{-1}$ from α -helix, $1641\text{-}1647\text{ cm}^{-1}$ for random coil, and $1658\text{-}1671\text{ cm}^{-1}$ and $1681\text{-}1696\text{ cm}^{-1}$ for turns and bends. In addition, in the amide II region α -helices absorb around 1545 cm^{-1} and β -sheets around 1525 cm^{-1} . A typical IR spectrum consists of several of these bands in any given region and they can be deconvoluted

and split into constituent bands by fitting the spectrum to a combination of representative Gaussian distributions for each absorbance band.

The instrumentation for measuring infrared absorbance at equilibrium usually involves fourier transfer infrared spectrographs (FTIR). FTIR is different from a typical scanning IR spectrograph with a monochromator, as in that it has a Michelson interferometer, which has two mirrors, one stationary and the other mobile. The different wavelengths in the IR-beam are split and sent to both mirrors, and the reflected beams undergo constructive and destructive interference at different positions of the mirror. Hence frequency becomes a function of the position of the moving mirror. Thus, absorbance at different frequencies can be collected simultaneously, with the additional step of Fourier transforming the collected absorbance to convert it from being a function of mirror position to that of frequency. This enables collection of a large range of frequencies typical of IR absorption, in very short periods of time.

Experimental Conditions. Naf-BBL samples for FTIR measurements were prepared at a concentration of 2.5 mM in buffers of 20 mM concentration, after three cycles of dissolution in D₂O, complete heat denaturation, flash freezing and lyophilization. The measurements were performed on a Jasco FT/IR-4200 Type-A spectrograph at a spectral resolution of 2 cm⁻¹, with 100 accumulations for each condition, with cosine apodization.

2.5 Dynamic Light Scattering (DLS)

Light scattering or photon correlation spectroscopy is a technique that measures scattered light from particles to estimate properties of the scattering particles such as radii of gyration, hydrodynamic radius, diffusion coefficients for translational Brownian diffusion of the particles in solvents and their molecular weights. The principle of light scattering is based on the interaction of electrons in the particle with incident light, and their ability to disperse it in directions different from that of light incidence. Rayleigh scattering is the simplest theory that can describe elastic light scattering. Such scattering would have an inverse fourth power dependence on wavelength.

Static light scattering involves the study of scattering by molecules which are either static or in practice, moving in a specific and non-random direction. The average amount of light scattered by the particle depends on the wavelength of the light used, the refractive index of the solution or suspension, the angle of observation with respect to the direction of incident light, the concentration of particles in the suspension or solution, their molecular weight(s), and their shape and size. Radii of gyration and molecular weights can be routinely measured by static light scattering for molecules such as proteins, but the radii of gyration obtained must be interpreted with caution as this measured quantity depends on the shape the molecule.

Dynamic light scattering on the other hand, measures scattering from particles that are freely diffusing in solution due to Brownian motion. Since at any given spatial point being studied, the local concentration of particles is constantly changing, here fluctuations in the scattered light are measured, as opposed to average scattering

measured in static light scattering experiments. The random motion of particles depends on their diffusion coefficient and the thermal energy in the system.

$$D = \frac{RT}{N_{AV}f} \quad (2.11)$$

where D is the diffusion coefficient in units of $\text{cm}^2.\text{s}^{-1}$, R is the ideal gas constant, T is the absolute temperature of the system, and f is the friction factor, associated with the shape of the particle. Randomly fluctuating light scattering intensity can be analyzed with an auto-correlation function which connects the decay of the autocorrelation function with correlation time, which turns out to be an exponential decay in its simplest form for monodisperse samples:

$$A(\tau) = \langle I(t).I(t + \tau) \rangle \quad (2.12)$$

$$\ln[A(\tau)] = \ln[A_0] - \left[8\pi^2.\eta^2.\sin^2\left(\frac{\theta}{2}\right) \right].D_\tau \quad (2.13)$$

where A is the autocorrelation function for intensity fluctuations with correlation time τ , I is the intensity of light scattered, η is the refractive index of the medium, θ is the angle of observation of scattered light with respect to incident light. Hydrodynamic radii (R_h) can be estimated from the diffusion coefficient through the Stokes-Einstein equation:

$$R_h = \frac{kT}{6\pi\eta D} \quad (2.14)$$

where η is the viscosity of the solution or suspension. This estimate depends on several properties of the scattering particle such as shape, surface charges, structure, interaction with the solvent to produce viscous drag, and properties of the solvent or medium itself, such as ionic strength.

Conditions for experiments performed by Dr. Begoña Monterroso at the German Rivas Laboratory at CIB-CSIC, Madrid. Quasi-elastic scattering data were acquired by a 825 nm laser light scattering detector (Dynapro Titan, Protein Solutions, Inc.) equipped with a Peltier type cell holder and using a quartz cell. The light scattering instrument reports an autocorrelation function computed from fluctuations in scattering intensity every 10 seconds. Measurements were carried out at temperatures varying between 0 and 60 °C at 5 °C intervals, and allowing the temperature to equilibrate for 5 minutes before collecting data, on a sample of Naf-BBL at a concentration of 1 mM. Raw data were acquired using Dynamics V6 (V.6.3.18, Protein Solutions, Inc.), and exported as text files for subsequent processing and modeling using user-written scripts and functions in MATLAB (Ver. 7.3, MathWorks, Natick, MA). Quasi-elastic light scattering data are reported in the form of an autocorrelation function describing the time-dependence of the correlation between scattering intensity at any given time and the intensity at a subsequent increment of time τ . For a single scattering species, the autocorrelation function (ACF) is given by:

$$ACF(\tau) = ACF_{\infty} + ACF_{\max} \exp(-2Dq^2\tau) \quad (2.15)$$

where ACF_{max} and ACF_{∞} are, respectively, the values of ACF in the short and long time limits, q is the scattering vector, which at a scattering angle of 90° is given by $q = \frac{4\pi n_0}{\lambda_0} \sin(\pi/4)$, and D is the translational diffusion coefficient of the scattering species.

All dynamic data collected for each sample at each specific temperature – typically between 9 and 20 autocorrelation functions – are averaged to yield a mean and standard deviation for $ACF(\tau)$. Fitting of equations to these combined data is performed using non-linear least-squares minimization of chi-square. The apparent hydrodynamic radius R_{app} was calculated according to the Stokes-Einstein relation for a diffusing spherical particle. Measured values of the diffusion coefficient are corrected to reference conditions (water, 20°C) according to:

$$D_{20,w} = D_{meas} \left(\frac{\eta}{\eta_{20,w}} \right) \left(\frac{293}{T(^{\circ}\text{C}) + 273} \right) \quad (2.16)$$

where η and $\eta_{20,w}$ denote the viscosities of the solution and water at 20°C , respectively, and T the temperature in Celsius.

2.6 Nuclear Magnetic Resonance Spectroscopy (NMR)

Nuclear magnetic resonance (NMR) spectroscopy is unique among the methods available for the determination of the structures of proteins at atomic resolution, since the NMR data can be recorded in solution. In NMR experiments,

solution conditions such as the temperature, pH and salt concentration can be adjusted so as to closely mimic a given physiological environment. Conversely, the solutions may also be changed to extreme non-physiological conditions, such as in studies of protein denaturation. In addition to protein structure determination, NMR applications include investigations of dynamic features of the molecular structures, as well as studies of structural, thermodynamic and kinetic aspects of interactions between proteins and other solution components, which may either be other macromolecules or low molecular weight ligands.

Spectral analysis is primarily focused on the positions of the individual NMR lines in the ^1H NMR spectrum, as given by the “chemical shift”, in parts per million (ppm) relative to a reference compound. Although the chemical shift is primarily determined by the covalent structure of the amino acid residue, it can also be significantly affected by the interactions with the environment. Therefore, the exclusion of water from the interior of a globular protein causes the chemical shifts of the core residues to be different from those of the water-exposed residues, so that even NMR lines originating from multiple residues of the same amino acid type can be distinguished. This ‘conformation-dependent chemical shift dispersion’ was found to be sufficiently large to enable ^1H NMR studies of protein denaturation. This then indicated the exciting prospect of using NMR for detailed studies of protein folding, and in particular for distinguishing between two-state and multiple-state folding and unfolding transitions.

Two-dimensional (2D) NMR spectra provide more information about a molecule than 1D spectra and are especially useful in determining the structure of a

molecule, particularly for molecules that are too complicated to work with using one-dimensional NMR. We used the 2D SOFAST-HSQC (Heteronuclear Single Quantum Coherence) experiment using the natural abundance of the ^{15}N isotope. Each residue of the protein has an amide proton attached to a nitrogen in the peptide bond. If the protein is folded, the peaks are usually well dispersed, and most of the individual peaks can be distinguished. HSQC is useful to screen candidates for structure determination by NMR.

Experimental Conditions. 1D and 2D spectra were collected for Naf-BBL at 283 K at pH 7.0 and pH 3.0 with no salt, and at pH 3.0 with 2 M LiCl, with protein concentration at 2 mM, also containing 10% D_2O and 10 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as the reference, with at least 10000 scans for good signal and well-resolved HSQC cross peaks.

2.7 Laser Temperature-Jump (T-jump) Instrumentation and Measurements

The application of laser spectroscopy to the study of protein folding stems from the need to study dynamic processes in protein folding, which can occur in the timescale of one to tens of nano-seconds to several microseconds. Laser beams have the advantage of being highly monochromatic, collimated and coherent, and can be pulsed and tuned to various degrees depending on specific requirements. The type of experimental set-up employed for the processes being studied here is the pump-probe experiment. This set-up involves the ‘pump’ step, which is basically the use of laser light to heat the sample, triggering a deviation from equilibrium for the molecules in the sample. This is the T-jump which is typically triggered by an IR-beam that is at a

wavelength that matches the IR absorption by H₂O or D₂O. The pump is then followed by the probe, which is a spectroscopic probe of the process taking place. In our experiments, we monitored either IR absorbance by the protein or fluorescence emitted by a fluorophore on the protein. Both the IR and fluorescence/FRET T-jump apparatus in the lab were set up by Dr. Michele Cerminara.

IR T-Jump Set-up

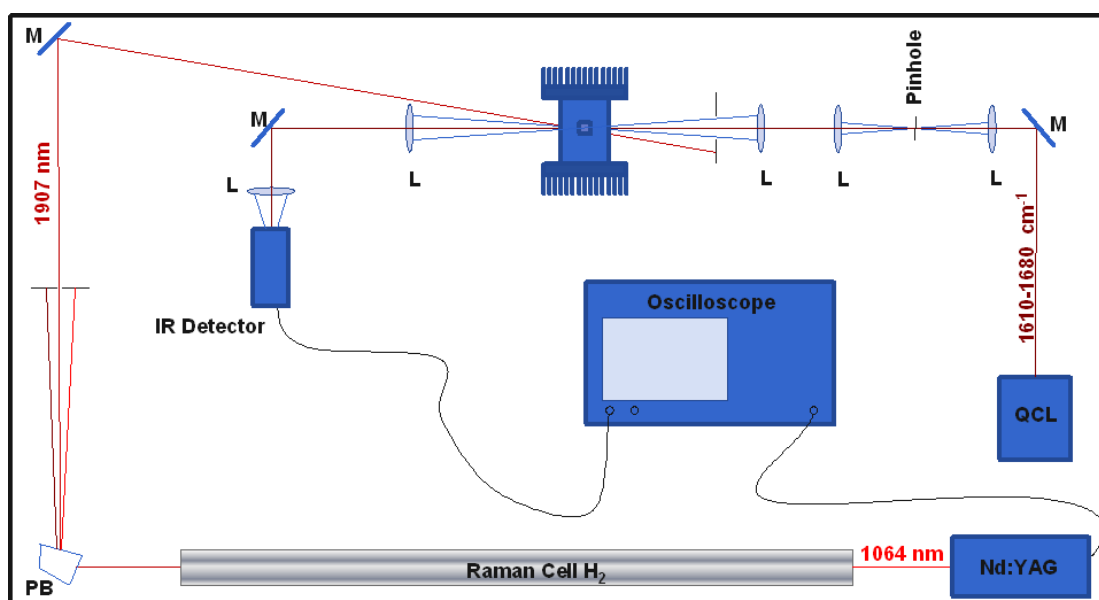


Figure 2.2 Schematic diagram of the IR laser T-Jump Spectroscopy set-up.

The IR T-jump set-up was installed to measure characteristic absorbances by a protein in the Amide I and Amide II regions. The pump beam was provided by a Nd-Yag Laser from Litron Lasers (Rugby, UK), which produces a pulsed laser beam at 1064 nm at 1, 2, 4 and 8 Hz. This beam is frequency-shifted to 1907 cm⁻¹ by a H₂-filled Raman cell from Light Age (Somerset, NJ), which is the absorbance maximum for O-D bond stretching in D₂O. Heating is carried out by 5 ns pulses of the pump

beam, from a single direction in the sample chamber at a power of about 20 mJ, heating a total volume of about 40 nl in the beam spot every 0.5 s. The heating spot is about 1 mm in diameter. The probe beam has a diameter of ~ 150 μm at the sample, corresponding to a probe volume of 0.1 nl, and is a continuous-wave IR beam aligned with the pump beam and is generated between 1610 and 1680 cm^{-1} by a Quantum Cascade Laser from Daylight Solutions (Poway, CA). Transmittance from the sample is detected by a MTC IR detector from Kolmar Technologies (Newburyport, MA). Since the probe beam is continuous, the sample can be probed starting from 10 ns after the pump beam has heated the sample to about 12 ms, when the sample begins to cool down again. Detection can be carried out at any given characteristic frequency depending on equilibrium FTIR measurements for specific structural signatures in the protein.

Experimental Conditions. The sample conditions for IR T-jump experiments were the same as those for equilibrium FTIR, i.e., 2.5 mM protein concentration, with 50 μM sample path lengths and collection every 0.2 ns for the fastest processes in the range of 20 ns (divided into 10^5 intervals), and every 4 ns for processes in the time scale of 0.4 ms (again divided into 10^5 intervals). The resolution of the instrument is determined by the IR detector, which is ~ 5 ns.

Fluorescence T-Jump Set-up

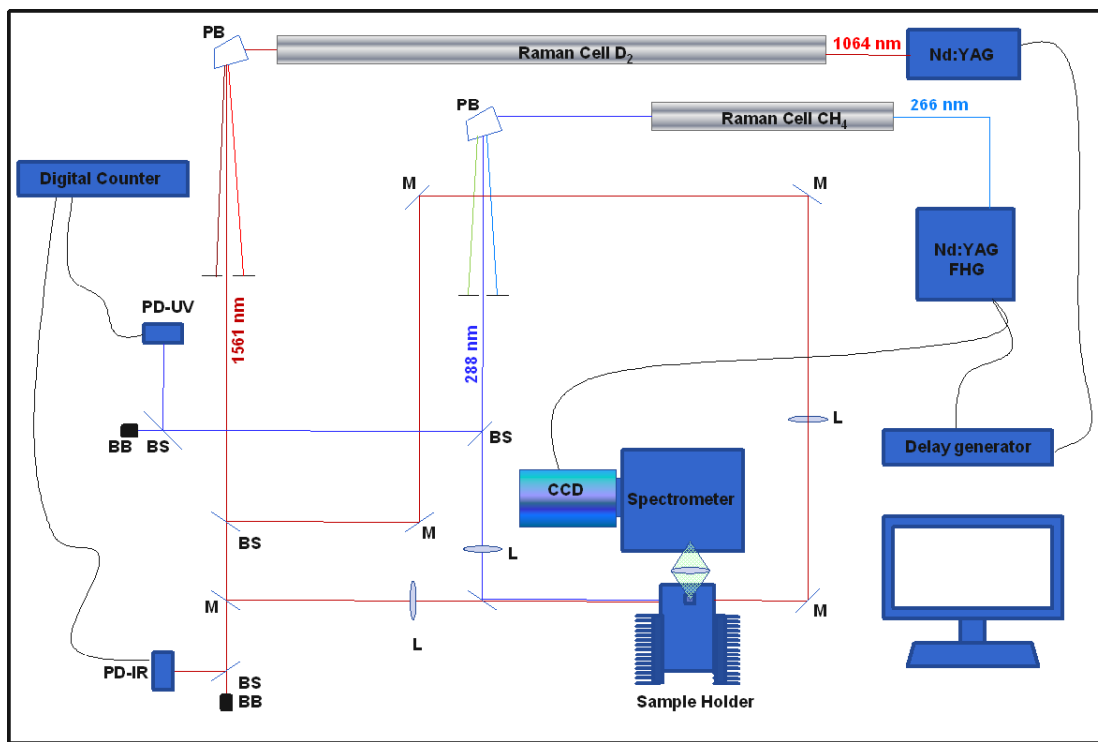


Figure 2.3 Schematic diagram of the FRET pump-probe laser T-jump spectroscopy set-up.

The set-up for fluorescence follows the same ‘pump-probe approach’, however the heating beam is Raman-shifted to 1560 cm^{-1} by a cell containing methane. This allows samples to be prepared in H_2O . Heating is carried out by two 20 mJ heating pulses of 5 ns duration separated by 125 ms (for running at 8 Hz), impinging on the sample from opposite directions. The probe beam is a UV-beam produced by frequency quadrupling to 266 nm by a fourth harmonic generator from Minilite, and Raman-shifting to 288 nm, a 1064 nm beam produced by a second Nd-Yag laser. Detection is carried out 90° to the heating beams by a CCD camera from Princeton Instruments (Trenton, NJ) that records entire spans of wavelength (a fluorescence spectrum). The delay between the pump and probe beams is generated by a delay

generator from Stanford Research Systems (Sunnyvale, CA). Heating pulses were set at a frequency of 2 Hz. The resolution of measurements in this case is not determined by the detector since the response time of the detector typically is about several fractions of a second, but is determined by the jitter of the pulse delay generator, which is typically 0.5 ns. The pump beam diameter was 1 mm corresponding to a heating volume of 400 nl for a sample path length of 0.5 mm, and the probe beam is about 100 μM in diameter, corresponding to a probe volume of 4 nl.

Experimental Conditions. Naf-BBL-Dan and Ran-BBL samples for FRET T-jump measurements were prepared in 20 mM buffer at concentrations of approximately 100 μM with CS_2 added to the samples at a concentration of 2% (v/v) to serve as a triplet quencher.

2.8 Protein and Dye Concentration Measurements

Buffer reagents were of chemical grade, obtained from Merck. Salt concentrations in protein samples where present, were measured by weight. Urea and GdmCl concentrations were estimated by refractive index measurements of their solutions. NaCl was purchased from Merck, and LiCl and CsCl from Sigma-Aldrich. Naphthyl-alanine and dansyl-lysine, used as references for donor and acceptor fluorescence, respectively, were obtained from Sigma-Aldrich. Naphthyl-alanine, dansyl-lysine, Naf-BBL, Naf-BBL-Dan, Ran-BBL and N-Acetyl-L-Tryptophanamide concentrations were determined by UV absorbance spectroscopy on a Cary100 Bio spectrophotometer from Varian. The molar absorptivity of naphthyl-alanine at pH 3.0 is $3,595 \text{ M}^{-1}.\text{cm}^{-1}$ at 266 nm, and $5,526 \text{ M}^{-1}.\text{cm}^{-1}$ at 280 nm, similar to that of Naf-

BBL⁶⁵. Naphthyl-alanine and Naf-BBL concentrations were estimated as the average from determinations at the two wavelengths. Naf-BBL-Dan has a molar absorptivity of 8,345 M⁻¹.cm⁻¹ at pH 3.0 and 266 nm⁶⁵. These values remain invariant with salt concentration and temperature. Dansyl-lysine absorbs at 266 nm with an absorptivity of 4,528 M⁻¹.cm⁻¹ at pH 3.0 and 266 nm⁶⁵. Ran-BBL concentrations at pH 3.0 were similarly estimated, as it has the same two extrinsic labels as Naf-BBL-Dan. At pH 7.0 dansyl absorbance changes although with an isosbestic point at 266 nm. Hence, the same values of molar absorptivity as those at pH 3.0 can be used at pH 7.0 for Naf-BBL-Dan and Ran-BBL. N-Acetyl-L-Tryptophanamide (molar absorptivity of 5,690 M⁻¹.cm⁻¹ at pH 7.0 and 280 nm) from Sigma-Aldrich was used as standard for quantum yield determinations.

2.9 Protein Sequences

Naf-BBL: NH₃⁺-NafA-LSPAIRLLAEHNLDASAIKGTGVGGRLTREDVEKHLAK-COO⁻

Naf-BBL-Dan: NH₃⁺-NafA-LSPAIRLLAEHNLDASAIKGTGVGGRLTREDVEKHLA-DanK-COO⁻

Ran-BBL: NH₃⁺-NafA-VKAISARLGEAHLEAESGRTLRLGDINRTLAKVPDLHG-DanK-COO⁻

Chapter 3. The Unfolded Ensemble: The Effect of Cosolvents

3.1 Introduction

The energy landscape theory of protein folding proposed by Wolynes and coworkers describes the dependence of thermodynamic properties of a protein on the various structural degrees of freedom of the protein's constituent units, i.e., its amino acid residues^{1,9}. The basis of this statistical treatment lay in seminal work by Wolynes *et al*, where concepts used to study spin glasses were applied to the statistical analysis of protein stability¹⁰. Owing to such a statistical treatment, a free energy function such as the Gibbs free energy of the ensemble and key features of a protein's energy landscape can be captured when this energy is described as a function of a few order parameters, whereas in reality such dependence is hyperdimensional. Similarity to the protein's native structure is one such coordinate or parameter, with the other one being the average size or radius of gyration of the ensemble. It is easy to see why this choice of order parameters is suitable for describing the free energy landscape. While similarity to the native ensemble turns out to be a measure of the net energetic drive towards adopting the native structure, the radius of gyration measures the conformational diversity or conformational entropy of the ensemble. The implicit assumption here is that the dynamics of variability of an ensemble at a fixed average 'distance' from the native structure are much faster than the search for native interactions that would energetically drive the folding process. This assumption greatly simplifies the treatment of a polypeptide's search for its most stable

conformation, since the energetically taxing process is the search for native contacts. The result of such an analysis reduces the free energy landscape to a funnel-shaped rugged three-dimensional surface.

Such a description treats conformational entropy at a fixed free energy, in the simplest possible manner. For example the conformational diversity of the ensemble at a particular free energy at the top of the funnel (low degree of nativeness, i.e. an unfolded ensemble) can be understood by assuming a random distribution in the ensemble, where the protein samples various structures, much like a randomly diffusing chain. A complete description of the protein folding energy landscape warrants not only an understanding of folding mechanisms leading to the native state, but also an understanding of key non-covalent interactions of a protein in its *unfolded state* (intra-protein and protein-solvent), the dynamics of such a population, and the dependence of its properties such as size and degree of nativeness on thermodynamic parameters such as temperature, solute or salt concentration, macromolecular crowding, pH and pressure. On the experimental side, studying the unfolded ensemble has proved difficult as traditional folding experiments typically monitor ensemble-averaged signals. This problem is compounded by the fact that the population of the unfolded ensemble appreciates only under conditions which destabilize the native ensemble relative to the unfolded, i.e. non-native conditions. Therefore, this precludes an estimation of an unfolded polypeptide chain's properties in native or physiological conditions. High resolution techniques such as NMR, ultra-fast laser spectroscopy and single molecule FRET (smFRET) experiments have enabled a direct observation of the unfolded ensemble^{24,35-36,71-74}. This chapter focuses

on the thermodynamic analysis of high-resolution data from smFRET measurements, which have allowed the observation of the unfolded state, in isolation from the native ensemble. In doing so several questions can be directly answered: a) In the unfolded ensemble how do the chain's interactions with the solvent or solution change observable properties such as radii of gyration and how can these effects be put on a quantitative basis. b) How do GdmCl and urea act as unfolding agents, i.e. what is the nature of their interaction with the unfolded ensemble? c) How faithfully can chemically denatured proteins be represented by the random coil model? The chapter is organized into the following sections starting with Section 3.2 which reviews the widely accepted views on the mechanism of action of unfolding agents such as GdmCl and urea on proteins, and the existing debates in the field. Section 3.3 reviews the application of polymer physics to the study of a common order parameter encountered in studies of the unfolded ensemble- the radius of gyration- and the dependence of chain size on interaction with the solvent, through the Flory-Huggins theory of polymer solution thermodynamics. Section 3.4 discusses our extension of this theory to account for solutions with 3 components instead of just two, followed by a heuristic analysis of the properties of our model in Section 3.5. Section 3.6 deals with the analysis of smFRET measurements of protein unfolding in the existing literature with our thermodynamic model, finally with concluding remarks in Section 3.7.

3.2 The Effect of Chemical Denaturants on Unfolded Proteins

The first use of cosolvents and electrolytic solutes to denature proteins goes back about eighty years to the study of the denaturation of proteins in urea. Urea and GdmCl are routinely used in protein folding experiments to destabilize proteins and alter the equilibrium between the folded and unfolded states⁷⁵. Such unfolding studies provide an estimate of the free energy difference between the folded and unfolded ensemble in native conditions by linear extrapolation. A critical question that arises from protein denaturation experiments is the mechanism of action of denaturing agents, and the broader question of the effect that osmolytes have on protein structure. While it is accepted that denaturation occurs because of the favourable interaction of proteins with denaturant molecules, compared to their interaction with water³⁷, the exact mechanism which drives this interaction is still controversial. Several different models such as the molecular transfer model^{37,76}, and the site-exchange model^{38,77-78}, point to the fact that the unfolded protein binds a larger number of denaturant molecules⁷⁹ than the folded protein. This result also finds support in the correlation of protein folding *m*-values with the change in accessible surface area of the protein upon unfolding⁸⁰. In a series of papers by Thirumalai and coworkers^{39-42,81}, the bulk of the effect of urea and GdmCl as unfolding agents has been shown to lie in their ability to modify the electrostatic interactions present on the chain. These cosolvents do so by screening electrostatic interactions, as is the case with the electrolytic cosolvent GdmCl, and by the formation of hydrogen bonds with the backbone and charged residues on the chain, i.e. a direct interaction or binding with the chain. However, several reports⁸²⁻⁸⁴, have pointed to a role that non-polar or

aromatic groups play in the interaction with urea, i.e., that urea unfolds proteins by reducing the hydrophobic effect, i.e. by facilitating the solvation of non-polar residues⁸⁵. In addition, as any kosmotropic or chaotropic cosolvent, urea and GdmCl have been implicated in altering the structure of water, and in particular as structure-breakers. Although the widely accepted view now is that urea and GdmCl unfold proteins by a direct interaction with the backbone^{37,79,86-90}, the role of these cosolvents in water restructuring is still widely debated. While it is clear that GdmCl and urea preferentially occupy sites on the protein backbone and interact with residues, do the thermodynamics of cosolvent-water interactions affect protein conformations as well? A simple thermodynamic argument would suggest that any change in the cosolvent-water interaction should necessarily influence the solvation of protein residues via the hydrophobic effect. Water restructuring by denaturants, if it does exist, may then be another key process in determining the stability of protein conformations in mixed solvents.

A wealth of data exists on transfer free energies of amino acid groups into mixed solvents, calorimetric measurements on cosolvent-protein interactions, and the various binding models used to describe protein-cosolvent interactions. The effect of cosolvents on the properties of the *unfolded state*, however, has evaded experimental investigation due to fact that the unfolded state is very low in population in native conditions, and bulk experiments have mixed signals emanating from an ensemble of structures. Now, due to the availability of information on the unfolded ensemble in isolation from the native ensemble, a systematic analysis can be performed on the dependence of their properties on solution conditions. The following sections are

devoted to the development of a theory that is specifically tuned to the analysis of the dependence of chain size of simple polymers such as the protein unfolded ensemble, on thermodynamic conditions such as temperature, and solvent quality, i.e. the presence of denaturing compounds or stabilizing cosolvents, which implicitly make the solvent ‘good’ or ‘poor’ for the unfolded protein, respectively.

3.3 Analysis of Variations in the Size of Unfolded Proteins

The unfolded ensemble, in general, experiences an increase in size when measured by techniques such as FRET and small angle X-ray scattering^{36,91}, upon the introduction of denaturing reagents such as urea and GdmCl. Although it has been argued that the random coil need not necessarily serve as a standard representation for the unfolded state^{79,92}, its properties such as a random coil-like dependence of chain size on chain length⁹¹, and the dynamics of motions in unfolded proteins which agree well with theoretical analyses^{26-27,65,93-95}, give credence to the assumption that a random coil may indeed be a reasonable representation of the unfolded state for the study of global properties such its size, radius of gyration or end-to-end distance. Given that random coil-like behaviour can be assumed to be a reasonable starting point for the description of the unfolded ensemble, a theoretical construct that is immediately applicable to the evaluation of end-to-end distances in chains is the Gaussian distribution. Seminal work by P. J. Flory summarized in his landmark textbook “Principles of Polymer Chemistry” published in 1953, on the size distributions of polymeric unbranched chains, showed that freely-jointed, random-flight chains have end-to-end distances which necessarily follow a Gaussian

distribution. Such an *ideal chain* would possess infinitesimally small joints (or beads), connected by thin links, segments or bonds, and would have no restrictions of dihedral angles between the beads. This distribution would take up the following form:

$$P(r) = 4\pi r^2 \cdot \left[\frac{3}{2\pi N b^2} \right]^{3/2} \exp \left[\frac{-3r^2}{2N b^2} \right] \quad (3.1)$$

where N is the number of bonds in the chain joining two beads (and not the number of beads or residues themselves), and b is the bond-length. A simple dependence of chain size on chain length results from this distribution.

$$\langle R^2 \rangle = N b^2 \quad (3.2)$$

Chains with a greater number of linkages (N) would thus be larger than chains with fewer linkages. Furthermore, chains with the same number of beads, but with different measured end-to-end distances, would have different effective segment lengths according to this model. Radius of gyration (R_g), another size measure that gives the position-averaged size of the chain can also be estimated readily from this model.

$$R_g = \left[\frac{N b^2}{6} \right]^{1/2} \quad (3.3)$$

Despite the simplicity of such a description, several of the simplifying assumptions made to arrive at this formalism do not actually hold true for unfolded proteins, or any polymeric chain for that matter. Amino acid residues have finite volumes and its atoms sterically clash with each other, because of their finite van der Waals' radii. Thus, such a chain would be a *self-avoiding chain*, where two beads or residues are not allowed to occupy the same position in space. Each residue would therefore have an excluded volume, and the entire chain would be more expanded than an ideal chain. Furthermore, the existence of steric clashes would also hinder hinge-movements around the bond-linkages. Thus, there would be a limited range of dihedral angles within which a given set of bonds could exist, much like amino acid residues in folded proteins. Such a chain would then have a distance distribution with an additional term because of its overall excluded volume. Such a term, as it turns out, can be arrived at through a simple combinatorial treatment using a lattice model, and has a well established algebraic form. Briefly, a *self-avoiding polymer* on a lattice is one where two segments (or residues) cannot overlap or occupy the same lattice site. The physical quantity which determines the extent of self-exclusion of the chain is the volume of each residue or bead v_c . The larger this quantity, the greater are the deviations of the self-avoiding chain from the ideal chain ensemble

$$P(r) = 4\pi r^2 \cdot \left[\frac{3}{2\pi N b^2} \right]^{3/2} \exp \left[\frac{-3r^2}{2N b^2} \right] \cdot \exp \left[-\frac{N^2 \cdot v_c}{2r^3} \right] \quad (3.4)$$

Besides the excluded volume effect, each residue has energetic interactions with other residues and solvent molecules surrounding it. For a protein, these would be non-covalent interactions such as a net hydrophobic interaction, electrostatic attractions or repulsions, hydrogen bonds, etc. Any two residues, separated by at least one other residue, exist in an overall, mutual potential field, which further affects the distance that may separate them in space. Thus, an additional term for energetic interactions is included, that would skew the distribution in Equation. 3.4 by its Boltzmann weight.

$$P(r) = 4\pi r^2 \cdot \left[\frac{3}{2\pi N b^2} \right]^{3/2} \exp \left[\frac{-3r^2}{2N b^2} \right] \cdot \exp \left[-\frac{N^2 \cdot v_c}{2r^3} \right] \cdot \exp \left[\frac{-E_{\text{interaction}}}{RT} \right] \quad (3.5)$$

Here, $E_{\text{interaction}}$ is the net energetic interaction potential of the chain, R is the Universal Gas Constant, and T the temperature. Such a potential would bias the size distribution function, depending on whether the chain is in a ‘good solvent’ or ‘poor solvent’. A good solvent is one in which the chain has relatively favourable net interactions with *dissimilar moieties* (solvent and cosolvent molecules); in a poor solvent residues on the chain have favourable net interactions with *other residues* on the chain. Figure 3.1 describes the various scenarios discussed above for chain size-distributions. Indeed a freely-jointed ideal chain occupies a lesser volume than a chain with self-avoiding moieties that interact with their environment.

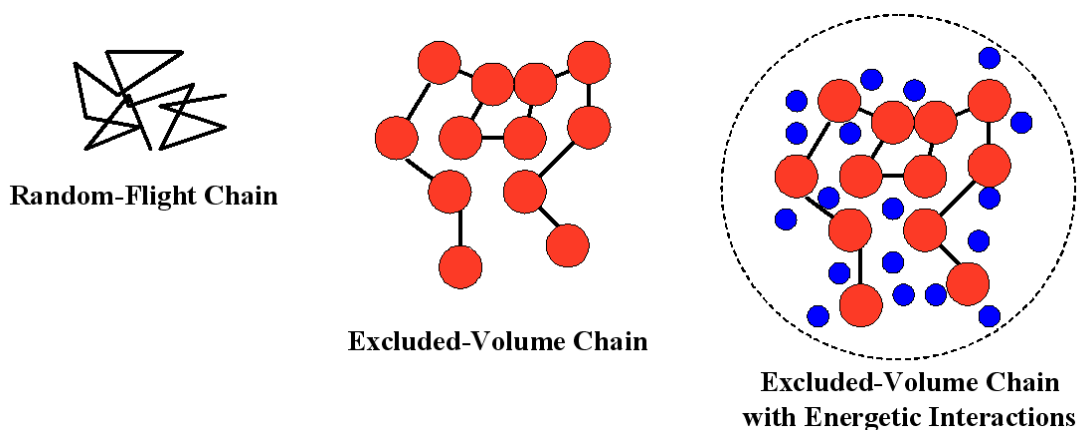


Figure 3.1 Various scenarios for chain-size distributions

One way of evaluating this potential involves a lattice method, where the chain occupies vacant positions on a lattice site with a fixed volume, with all unoccupied sites being filled with solvent moieties. Such a treatment comprises the well-known Flory-Huggins theory of polymer solution thermodynamics. A mean-field approximation known as the Bragg-Williams approximation facilitates an easy estimation of this energy. This approximation involves evaluating the overall chain energy by counting the various types of interactions and their individual energies, with the overall energy being the component-averaged energy over the lattice volume. This procedure is described below. Figure 3.2 illustrates the lattice method, with occupied sites. The equilibrium reaction on the right side of the figure indicates two like-interactions being disrupted leading to a polymer-solvent interaction. This is the central assumption in the theory, where it is assumed that any moiety on the lattice has pair-wise interactions with its neighbours. Each lattice site has z neighbours; in this case, z is assumed to be 6, with 4 neighbours on the same horizontal plane as the lattice site, and one each vertically above and below. The next important parameter is

the volume of each bead, or site, v_c . A value of $100 \text{ cm}^3 \cdot \text{mol}^{-1}$ is a reasonable estimate based on theoretical calculations of partial molar volumes of amino acid residues⁹⁶.

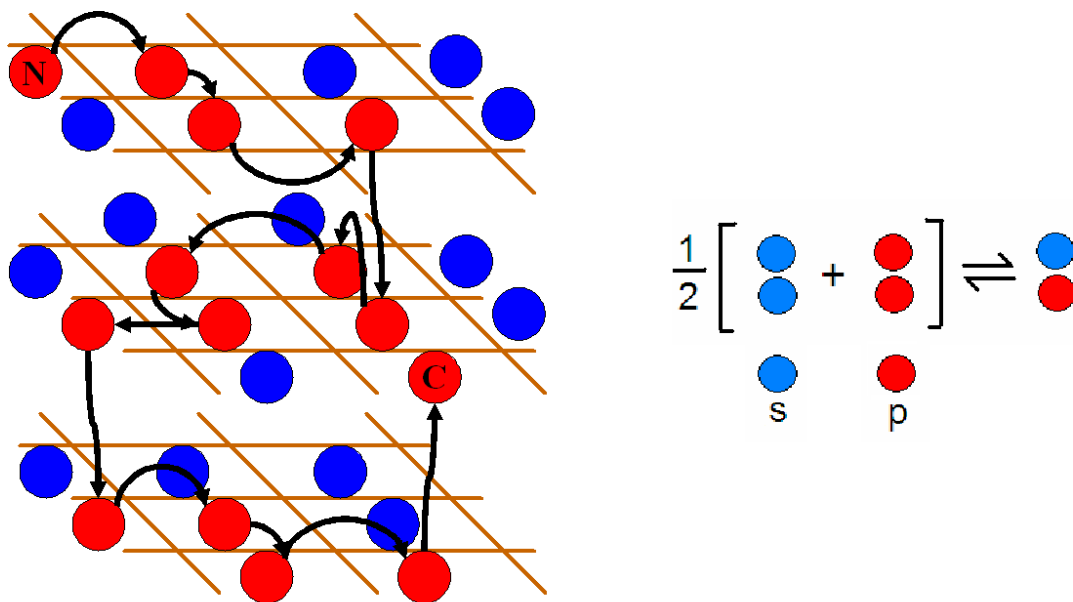


Figure 3.2 The lattice model description. Red circles indicate protein or polymer residues connected by black linkers. Blue circles are the solvent molecules that occupy the remaining lattice sites. Pair-wise interactions are indicated by the equilibrium reaction on the right, where two like-interactions ($-\epsilon_{ss}$ and $-\epsilon_{pp}$) are broken to form a polymer-solvent interaction ($-\epsilon_{ps}$).

Therefore the interaction energy would be:

$$E_{\text{interaction}} = -\sum N_{ab} \cdot \epsilon_{ab} \quad (3.6)$$

where a and b are like or unlike moieties, N_{ab} is the number of such interactions in the lattice volume under consideration, and ϵ_{ab} is the positive energy of interaction (hence the negative sign in the equation). For the lattice above:

$$E_{\text{interaction}} = -\left(N_{ss} \cdot \mathcal{E}_{ss} + N_{pp} \cdot \mathcal{E}_{pp} + N_{ps} \cdot \mathcal{E}_{ps}\right) \quad (3.7)$$

Let us assume that the lattice has N_{total} sites, with N being the number of residues in the lattice and N_s the number of solvent moieties ($N_{\text{total}} = N_s + N$). Upon counting the number of various types of interactions listed in Equation 3.7 we would arrive at:

$$E_{\text{interaction}} = -\left[\left(\frac{1}{2} z N_s \cdot N_s\right) \cdot \mathcal{E}_{ss} + \left(\frac{1}{2} z N \cdot N\right) \cdot \mathcal{E}_{pp} + z N \cdot N_s \cdot \mathcal{E}_{ps}\right] \quad (3.8)$$

$$E_{\text{interaction}} = -\left[\left(\frac{1}{2} z (N_{\text{total}} - N)^2\right) \cdot \mathcal{E}_{ss} + \left(\frac{1}{2} z N^2\right) \cdot \mathcal{E}_{pp} + z N \cdot (N_{\text{total}} - N) \cdot \mathcal{E}_{ps}\right] \quad (3.9)$$

Retaining only the N-associated terms and no terms associated with N_{total} , since we are concerned only with interactions that the residues are involved in, we get:

$$E_{\text{interaction}} = -z N^2 \left[\frac{\mathcal{E}_{ss}}{2} + \frac{\mathcal{E}_{pp}}{2} - \mathcal{E}_{ps} \right] \quad (3.10)$$

For the equilibrium reaction shown in Figure 3.2, the free energy change associated with the reaction or each lattice-site is:

$$\Delta G = \Delta \mathcal{E} = \frac{1}{2} (\mathcal{E}_{ss} + \mathcal{E}_{pp}) - \mathcal{E}_{ps} \quad (3.11)$$

Normalizing Equation 3.11 over the total number of lattice sites, we get:

$$N_{total} = \frac{r^3}{v_c} \quad (3.12)$$

$$E_{\text{net interaction}} = \frac{-zN^2 \cdot v_c \cdot \Delta \epsilon}{r^3} \quad (3.13)$$

Combining Equation 3.13 with equation with Equation 3.5,

$$P(r) = 4\pi r^2 \cdot \left[\frac{3}{2\pi N b^2} \right]^{3/2} \exp \left[\frac{-3r^2}{2N b^2} \right] \cdot \exp \left[-\frac{N^2 \cdot v_c}{2r^3} \right] \cdot \exp \left[\frac{zN^2 \cdot v_c \cdot \Delta \epsilon}{r^3 \cdot RT} \right] \quad (3.14)$$

$$\text{and, } \frac{z\Delta \epsilon}{RT} = \chi \quad (3.15)$$

where χ is called the ‘exchange parameter’, alluding to the fact that the associated free energy arises out of an exchange of interacting moieties. Thus, for a non-ideal self-avoiding chain whose components have energetic interactions with their surroundings, the chain end-to-end distance can be described by the following distribution:

$$\therefore P(r) = 4\pi r^2 \cdot \left[\frac{3}{2\pi N b^2} \right]^{3/2} \exp \left[\frac{-3r^2}{2N b^2} \right] \cdot \exp \left[-\frac{N^2 \cdot v_c}{2r^3} (1 - 2\chi) \right] \quad (3.16)$$

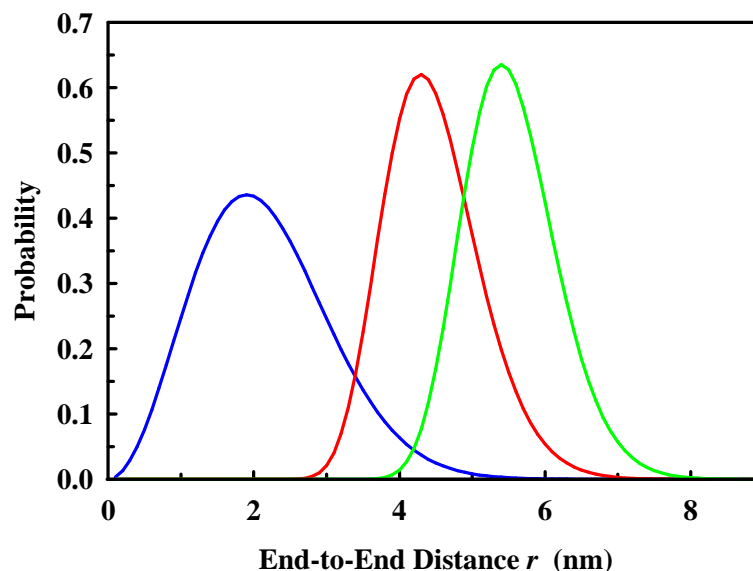


Figure 3.3 Probability distributions for the 3 scenarios in Figure 3.1. The blue curve is a Gaussian distribution with $b=0.33$ nm and $N=50$. The red curve indicates an excluded volume chain with a volume of $100 \text{ cm}^3 \cdot \text{mol}^{-1}$ per residue. The green curve indicates a distribution with an energetic interaction $\Delta\epsilon=-500 \text{ J} \cdot \text{mol}^{-1} \cdot \text{residue}^{-1}$.

The energetic interaction parameter $\Delta\epsilon$ is therefore the net energy per residue required to fully expose a sequestered residue to a solvent molecule. This energy is essentially enthalpic with contributions also coming from entropic changes accompanying the rearrangement of the solvent structure around the residue. Combinatorial entropic changes related to chain rearrangements are accounted for in the distribution. Good solvents therefore, would result in expanded chains, with favourable polymer-cosolvent interactions, indicated by a negative $\Delta\epsilon$. Chain compaction results when $\Delta\epsilon$ becomes higher, an indication of a poor solvent, and an energetically unfavourable residue-exposure process. Figure 3.3 indicates the resultant chain distributions when excluded volume and energetic interactions are taken into consideration over and above random fluctuations.

Equation 3.11 yields a simple result for the two-component mixtures, involving just a polymer dissolved in a solvent. However, for our purposes, it would be necessary to determine the energetic interactions accompanying residue exposure to a mixed solvent or cosolvent-solvent mixture. The section below describes the extension of the Flory-Huggins theory to 3-component mixtures involving the polymer, solvent and cosolvent.

3.4 Extension of the Flory-Huggins Theory to 3-Component Mixtures

The aim of this procedure is to determine the dependence of $\Delta\varepsilon$ on individual contributions from several interactions. Where a two-component mixture has only 3 kinds of pair-wise interactions- ε_{pp} , ε_{ss} and ε_{ps} - a 3-component mixture would have 6 different pair-wise interactions- ε_{pp} , ε_{ss} , ε_{ps} , ε_{cc} , ε_{cs} and ε_{pc} . Thus, three additional terms account for the presence of the cosolvent. An expression for $\Delta\varepsilon$ should also depend on the concentration of the cosolvent. A concentration measure that would be appropriate for our considerations would be the volume fraction of the cosolvent ψ .

$$\psi = \frac{C.V_{solute}}{V_{solution}} \quad (3.17)$$

Here, C is the concentration in molar of the cosolvent, V_{solute} is the molar volume of the solute in a solution volume of $V_{solution}$. The molar volumes of urea and GdmCl molecules are $46 \text{ cm}^3.\text{mol}^{-1}$ and $73 \text{ cm}^3.\text{mol}^{-1}$, respectively⁹⁷. The total volume of the GdmCl species is used here, instead of simply the volume of Gdm⁺ ion.

Equation 3.6 is invoked once again to evaluate all the interactions which contribute to $E_{interaction}$. The table below lists the number of pair-wise interactions of each type and their respective energies. N and N_{total} are the number of residues and the total number of lattice sites, respectively.

N_{ab}	N_{ab}	ϵ_{ab}
N_{ss}	$\frac{1}{2} z \cdot (1-\psi)^2 \cdot (N_{total}-N)^2$	ϵ_{ss}
N_{pp}	$\frac{1}{2} z \cdot N^2$	ϵ_{pp}
N_{cc}	$\frac{1}{2} z \cdot \psi^2 \cdot (N_{total}-N)^2$	ϵ_{cc}
N_{ps}	$z \cdot (1-\psi) \cdot N(N_{total}-N)$	ϵ_{ps}
N_{pc}	$z \cdot \psi \cdot N(N_{total}-N)$	ϵ_{pc}
N_{cs}	$z \cdot \psi(1-\psi) \cdot (N_{total}-N)^2$	ϵ_{cs}

Table 3.1 Enumeration of the various interactions between any two moieties on the lattice, based on the Bragg-Williams mean-field approximation.

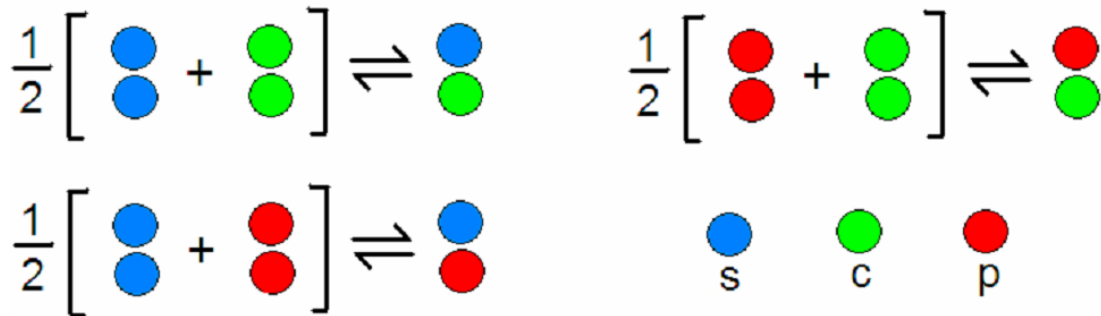


Figure 3.4 Chemical equilibria indicating the 6 different pair-wise interactions in a 3-component mixture.

Parameters from Table 3.1 are then incorporated into Equation 3.6, and again, only those terms which are N -dependent and N_{total} -independent are retained, to give:

$$E_{\text{interaction}} = -zN^2 \left[\frac{1}{2} \varepsilon_{pp} + \frac{1}{2} (1-\psi)^2 \cdot \varepsilon_{ss} + \frac{1}{2} \psi^2 \cdot \varepsilon_{cc} - (1-\psi) \cdot \varepsilon_{ps} - \psi \cdot \varepsilon_{pc} + \psi (1-\psi) \cdot \varepsilon_{cs} \right] \quad (3.18)$$

Upon clubbing together terms of the same order in ψ , we get:

$$E_{\text{interaction}} = -zN^2 \left[\left(\frac{\varepsilon_{cc}}{2} + \frac{\varepsilon_{ss}}{2} - \varepsilon_{cs} \right) \psi^2 + (\varepsilon_{ps} + \varepsilon_{cs} - \varepsilon_{pc} - \varepsilon_{ss}) \psi + \left(\frac{\varepsilon_{pp}}{2} + \frac{\varepsilon_{ss}}{2} - \varepsilon_{ps} \right) \right] \quad (3.19)$$

The second-order and zero-order coefficients conveniently reduce to $\Delta\varepsilon_{cs}$ and $\Delta\varepsilon_{ps}$, respectively, based on Equation 3.11. The terms in the first-order coefficient also reduce to a convenient form upon adding and subtracting $1/2\varepsilon_{pp}$ and $1/2\varepsilon_{cc}$.

$$E_{\text{interaction}} = -zN^2 \left[\Delta\varepsilon_{cs} \psi^2 + (\Delta\varepsilon_{pc} - \Delta\varepsilon_{ps} - \Delta\varepsilon_{cs}) \psi + \Delta\varepsilon_{ps} \right] \quad (3.20)$$

The three different exchange interaction energy terms $\Delta\varepsilon_{ab}$ represent the free energy changes accompanying each of the three equilibria shown in Figure 3.4. Therefore the free energy change that results when a sequestered residue, interacting with another residue on the chain, is exposed to a mixed solvent is given by:

$$\Delta\varepsilon = \left[\Delta\varepsilon_{cs} \psi^2 + (\Delta\varepsilon_{pc} - \Delta\varepsilon_{ps} - \Delta\varepsilon_{cs}) \psi + \Delta\varepsilon_{ps} \right] \quad (3.21)$$

This would be the energetic interaction term per lattice site, $\Delta\varepsilon$, which would appear in Equations 3.11, 3.13, 3.14 and 3.15.

3.5 Features of the 3-Component Energy Equation

A series of heuristic calculations to bring to light features of the quadratic 3-component energy equation and the effect of varying energetic interaction parameters on end-to-end distances is in order. The next three sections describe calculations that would enable a better understanding of the quadratic 3-component equation.

3.5.1 Basic Features of the Quadratic Equation

Equation 3.21 immediately points to several of the effects of mixed solvents on the size of polymer:

- 1) Since $\Delta\varepsilon$ represents the energetic change accompanying the exposure of a sequestered residue on the protein to a mixed solvent with varying concentrations of the cosolvent ψ , we notice that the energetic change is necessarily quadratic in its dependence on ψ .
- 2) The individual exchange energetic interaction parameters $\Delta\varepsilon_{cs}$, $\Delta\varepsilon_{pc}$ and $\Delta\varepsilon_{ps}$ are indeed energetic (enthalpic) changes accompanying the exchange equilibria in Figure 3.4. Each equilibrium describes an exchange reaction, where like-interactions between moieties of the same kind are replaced by cross-interactions where two different moieties interact with each other. Thus, this model is not limited by the treatment of interactions as simple binding events, manifested in the analysis of data

by the use of binding polynomials, as mentioned in Section 3.2⁹⁸. In fact, it is known that GdmCl and urea have weak binding coefficients to the protein residue or backbone, and that these solutions are highly non-ideal, pointing to interactions between the solvent and cosolvent molecules, which cannot be ignored^{38,77,97,99-100}. This is to say that the analysis of chemical unfolding experiments and the interaction of cosolvents with proteins, through binding polynomials and explicit binding equilibria, is simplistic, but insufficient. The 3-component energetic interaction model described here does not invoke the binding polynomial. An advantage of this model is that the eventual analysis would be inclusive of both types of effects: explicit binding and indirect solution driven processes. Most importantly, the model allows the explicit separation of each interaction event of the protein residue with its surroundings into its components, whereby, protein-cosolvent, cosolvent-solvent and protein-solvent interactions come out as separate terms.

3) The curvature of $\Delta\varepsilon$ with increasing cosolvent concentration depends entirely on the 2nd derivative of Equation 3.21, i.e. the interactions between the cosolvent and solvent. Thus the curvature of the variation of $\Delta\varepsilon$, and in turn R_g , with concentrations of the cosolvent, depends entirely on the physico-chemical properties of the cosolvent-water mixture, and the enthalpy of dissolution of the cosolvent in water. This is an illuminating feature of the model, which puts to rest the question of whether the bulk solution plays any role at all in the expansion and compaction of chains, as opposed to solely direct interactions of the chain residues with the cosolvent. The equation allows us to envisage a scenario where, even in the absence

of cosolvent interactions with the residue ($\Delta\varepsilon_{pc}=0$), the chain would experience an expansion or compaction.

4) Solutions of urea and GdmCl in water have positive deviations from ideality, i.e. their enthalpies of dissolution are positive due to their interactions with the solvent. Urea and GdmCl are also *chaotropic cosolvents*, which disrupt the structure of water. The free energy of their dissolution is driven by the entropic component arising from the combinatorial mixing and rearrangement of the solvent and solute species. Thus, it is easy to see, that $\Delta\varepsilon_{cs}$, the energetic change (an enthalpy change to be more precise) accompanying the dissolution of urea or GdmCl in water would be positive. This automatically sets the curvature to be positive in Equation 3.21, and thus, $\Delta\varepsilon$ and R_g variation with cosolvent concentration would be a convex quadratic curve. In case of a solute which has a negative enthalpy of dissolution in water, the scenario would be the opposite, with negative $\Delta\varepsilon$ curvature, and concave variations of $\Delta\varepsilon$ and R_g with concentrations of the cosolvent.

5) Each interaction type, whether cosolvent-solvent, polymer-solvent or polymer-cosolvent is expected to vary with cosolvent concentration, and Equation 3.21 ensures that this variation is captured by the quadratic equation. Indeed, this equation finds utility in analyzing any 3-component mixture, as no *a priori* requirement pertaining to the existence of the polymer chain is made in its derivation. The effect of the existence of the polymeric chain is captured by the zN^2 term in Equation 3.20. Thus, if the polymer-dependent terms in Equation 3.21 vanish, then the energetic term reduces to:

$$\Delta\varepsilon = -\psi(1-\psi)\cdot\Delta\varepsilon_{cs} \quad (3.22)$$

Equation 3.22 would represent the dependence of the net enthalpic change on cosolvent concentration, accompanying the dissolution of the cosolvent in water. This equation is similar to the one utilized in analyzing properties of non-ideal solutions such as surface tension, activity coefficients and dissolution enthalpies, and their dependence on solute concentration, by the Flory-Huggins theory of solution thermodynamics.

6) The slope of $\Delta\varepsilon$ versus chemical concentration ψ ,

i.e. $\frac{\partial\Delta\varepsilon}{\partial\psi} = (\Delta\varepsilon_{pc} - \Delta\varepsilon_{ps}) - (1-2\psi)\Delta\varepsilon_{cs}$, or the sensitivity of residue-exposure to the

cosolvent concentration, is a complex combination of all 3 types of cross interactions.

This further illustrates the point that, in principle, chain sensitivity to a cosolvent arises not just from interactions of the chain with the solvent or cosolvent ($\Delta\varepsilon_{ps}$ and $\Delta\varepsilon_{pc}$), but also from interactions in bulk, of the solvent with the cosolvent- an indirect effect.

Thus, Equation 3.21 could be directly applicable in scenarios where end-to-end distance or size information of an unbranched polymeric or random coil like protein chain (for e.g. an unfolded ensemble) in varying concentrations of cosolvent concentration is available.

3.5.2 A Limitation of the Probability Distribution Function

The distribution function described in Equation 3.16 (a neater form of Equation 3.14), is only valid when $(1-2\chi)>0$. When this condition is not met, the distribution tends to infinity at low distances. This condition implies that once the excluded volume term $\frac{-N^2v_c}{2r^3}$ comes into play, the energetic interactions can only produce an expansion according to the above condition, since when $(1-2\chi)>0$, it implies, $\Delta\varepsilon \leq \frac{RT}{2z}$, and this would indicate an expansion as $\Delta\varepsilon \rightarrow -\infty$. Thus, the energy of interaction per residue at the most compact state is one whose scale is set by the choice of b , the segment length of the chain. The planar projection of the bonds connecting two α -carbons is 0.38 nm. However, experimental observations based on measurements made as part of this thesis, and discussed in latter sections indicate, that a chain at its maximally compact state in conditions of high temperature without chemical denaturants would have a segment length b_{min} of 0.33 nm. This segment length has been used for evaluation of energetic parameters for all proteins under all conditions discussed in further sections and chapters. At the very least, this number sets a compaction lower bound for all unfolded ensembles, and an upper bound on extracted energetic parameters, since it underestimates the actual b_{min} that a protein can possibly access in any experimental condition at maximal possible compaction. This is not a serious limitation, except for the fact that typically smFRET measurements on the unfolded ensemble are difficult to discern at low denaturant concentrations because the unfolded ensemble is low in population, and also because its peak is merged with that of the folded ensemble at high FRET. Additionally, using

the same minimum segment-length for calculating 3-body interaction energies for all proteins renders uniformity to the analysis of different sets of data on proteins with varying sequence lengths and sizes.

3.5.3 Effect of the Exchange Energetic Parameters on Chain Size

This section contains a series of numerical calculations with varying exchange energetic interaction parameters $\Delta\epsilon_{cs}$, $\Delta\epsilon_{pc}$ and $\Delta\epsilon_{ps}$, and elucidates their effect on chain size and the average FRET read-out expected from FRET experiments. To recapitulate, these exchange interaction terms determine the energy changes accompanying an exchange reaction, such as the ones illustrated in Figure 3.4. Since cosolvent-solvent interactions lead to deviations from ideality and Raoult's laws, physico-chemical data on non-ideal properties of these solutions such as activity coefficients⁷⁷, can directly give us a handle on the cosolvent-solvent interaction parameter $\Delta\epsilon_{cs}$. The chemical potential of a non-ideal solution has a contribution from the activity coefficient as follows:

$$\mu = \mu_0 + RT \cdot \ln(\gamma\psi) \quad (3.23)$$

where, μ is the chemical potential of the solvent or cosolvent, and γ is the activity coefficient of the solvent or cosolvent, with both being on the concentration scale of volume fraction ψ . γ is unity for ideal solutions. The Flory-Huggins theory allows us to directly extract this parameter from activity coefficient data as follows:

$$\gamma_{\text{solvent}} = \exp\left[\frac{z\Delta\epsilon_{cs}}{RT} \cdot \psi^2\right] \quad (3.24)$$

$$\gamma_{\text{cosolvent}} = \exp\left[\frac{z\Delta\epsilon_{cs}}{RT} \cdot \psi(\psi - 2)\right] \quad (3.25)$$

where R is the Universal Gas Constant, T the temperature, and ψ a pertinent concentration measure of the cosolvent, in this case, its volume fraction. Detailed activity coefficient data are available for both urea and GdmCl solutions in water⁷⁷ and the activity coefficients of water on the *mole fraction scale* are plotted below in Figure 3.5 A for varying GdmCl and urea concentrations.

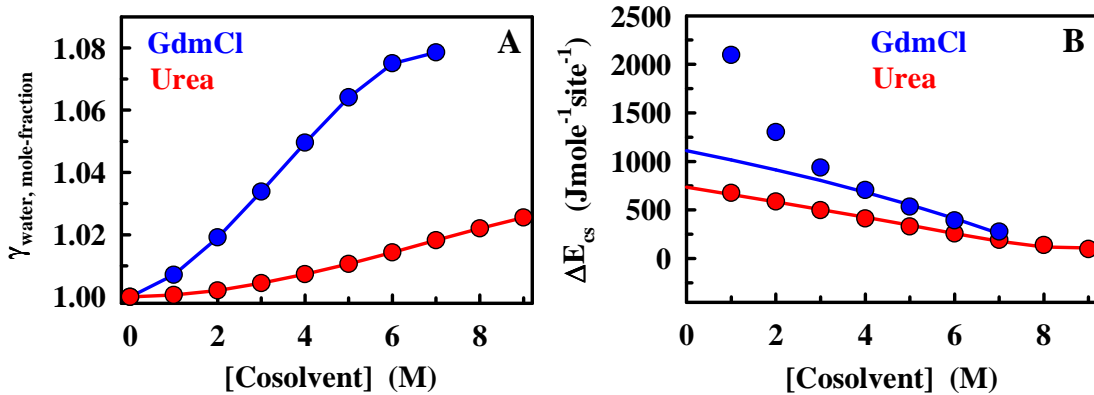


Figure 3.5 The activity coefficients of water in GdmCl and urea solutions and corresponding $\Delta\epsilon_{cs}$ at 298.16 K. Circles indicate experimental data and lines in A are to guide the eye and in B are fits. A) Activity coefficient of water in solutions with GdmCl and urea, on the mole fraction scale. B) Calculated $\Delta\epsilon_{cs}$, based on the Flory-Huggins theory of solution thermodynamics applied to the activity coefficient data. Circles indicate a direct calculation, and lines are fits, both with mole fraction as the concentration unit. Blue line is a linear fit to the high concentration range of GdmCl, and red line is a quadratic fit.

The calculated energies are plotted in Figure 3.5 B. It must be noted, that in case of GdmCl, since the solute is ionic and splits into two species upon dissolution, the mole-fraction of the cosolvent in Equations 3.24 and 3.25 would have a multiplication factor of 2, since experimental mole fraction data on GdmCl only

quotes the mole fraction of a single ionic species⁷⁷. The activity coefficient of the solute has the opposite trend, and decreases with increasing concentration. The limitation of such a calculation using Equation 3.24, is that $\Delta\epsilon_{cs}$ tends to infinity, at low solute concentrations and is indeterminate at infinite dilution. There is a larger degree of variability in $\Delta\epsilon_{cs}$ extracted from the solute activity coefficient data as ψ tends to 0, and therefore, only solvent activity coefficients are used to extract this parameter. Further, as can be seen in the values of $\Delta\epsilon_{cs}$ at low cosolvent concentration for GdmCl, in Figure 3.5 B, it is possible that the parameter tends to really large numbers, possibly because of the inaccuracy in the utilizing polynomial fits to determine activity coefficients at concentrations close to 0 M. Thus, a quadratic fit of the $\Delta\epsilon_{cs}$ data-set, for all urea concentrations from 1 M onwards and a linear fit for the high-concentration range of the GdmCl data-sets to their mole fractions, extrapolated to 0 M [Cosolvent], yields the values $\Delta\epsilon_{cs} = 1108.2 \text{ J.mol}^{-1}.\text{site}^{-1}$ for GdmCl and $\Delta\epsilon_{cs} = 762.22 \text{ J.mol}^{-1}.\text{site}^{-1}$ for urea at 298.16 K at infinite dilution. These cosolvent-solvent exchange interaction energies will change with temperature as do activity coefficients. These two numbers set the curvature for all GdmCl and urea-based expansion and compaction experiments, when analyzed by the 3-component energy equation. These cosolvent-solvent interaction energies are positive indicating a positive enthalpy of dissolution, positive deviations from ideality, and the fact that in aqueous solution of these cosolvents, like interactions are preferred over unlike interactions between the solvent and cosolvent. Thus, the overall energetic interaction parameter $\Delta\epsilon$ for a residue would always have positive curvature with increasing

concentrations of these two cosolvents, and thus the curves of $\Delta\varepsilon$ vs ψ would always be convex parabolas.

Calculations are then performed with the above $\Delta\varepsilon_{cs}$ for urea, and varying values of $\Delta\varepsilon_{ps}$ and $\Delta\varepsilon_{pc}$. The latter two range between -1600 and $200 \text{ J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$, and the first set of calculations demonstrate the effect that varying energies have on a chain with 50 residues. Figure 3.6 plots $\Delta\varepsilon$ for various values of $\Delta\varepsilon_{ps}$ and $\Delta\varepsilon_{pc}$. Energies were calculated

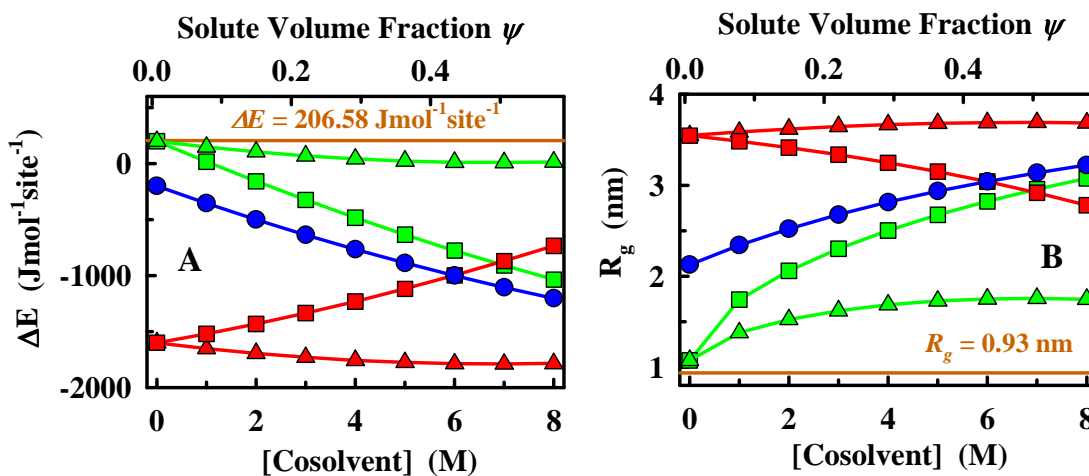


Figure 3.6 The effect of exchange energetic interactions on chain size. A) Circles indicate energies and lines are to guide the eye. Red and green symbols indicate $\Delta\varepsilon_{ps}$ equal to $-1600 \text{ J}\cdot\text{mol}^{-1}\cdot\text{site}^{-1}$ and $200 \text{ J}\cdot\text{mol}^{-1}\cdot\text{site}^{-1}$, respectively. Triangles indicate $\Delta\varepsilon_{pc} = \Delta\varepsilon_{ps}$, red squares $\Delta\varepsilon_{pc} = 200 \text{ J}\cdot\text{mol}^{-1}\cdot\text{site}^{-1}$ and green squares $\Delta\varepsilon_{pc} = -1600 \text{ J}\cdot\text{mol}^{-1}\cdot\text{site}^{-1}$. Blue circles are energies with intermediate $\Delta\varepsilon_{ps}$ and $\Delta\varepsilon_{pc}$, compared to the red and green symbols. Brown line indicates the θ -condition at 298.16 K . B) Radius of gyration of a 50-residue chain with a b_{min} of 0.33 nm at 298.16 K . Same symbol and colour scheme as in A.

using the volume fraction and $\Delta\varepsilon_{cs}$ of urea as the cosolvent, but qualitatively the results would be the same if those for GdmCl were used instead. These energies were then used as inputs into Equation 3.21, to calculate the average FRET efficiency

expected for a chain with 50 residues, through Equation 3.26, with the probability distribution defined as in Equation 3.16, with a segment length b of 0.33 nm at 298.16 K and a R_0 of 5.4 nm. The FRET efficiencies were then fit to Equation 3.26, however with probability distributions coming from Equation 3.1, a Gaussian, yielding various segment lengths b over different conditions. R_g was then determined through Equation 3.3. Though R_g could have been obtained directly from Equation 3.16, this indirect procedure was employed since it represents the typical scenario in experimental conditions where average FRET efficiencies and R_0 would be the experimentally known parameters, and R_g for bulk or single-molecule FRET experiments is typically taken to be the root-mean squared average radius of gyration for a Gaussian chain, as defined in Equation 3.3 upon fitting FRET efficiencies to the transcendental Equation 3.26.

$$\langle E \rangle = \frac{\int P(r) \cdot \frac{R_0^6}{R_0^6 + r^6} dr}{\int P(r) dr} \quad (3.26)$$

Figure 3.6 A shows energies calculated for various scenarios. The triangles indicate two extreme scenarios, where the interactions of the cosolvent and solvent with the polymer residue are equal and exactly balance out. The red triangles indicate a scenario where the solvent is ‘good’ with a $\Delta\epsilon_{ps}$ of $-1600 \text{ J.mol}^{-1}.\text{site}^{-1}$, and likewise for $\Delta\epsilon_{pc}$. The green triangles indicate the opposite scenario where the solvent is ‘poor’ with $\Delta\epsilon_{ps}$ of $200 \text{ J.mol}^{-1}.\text{site}^{-1}$ and likewise for $\Delta\epsilon_{pc}$. In these two sets of conditions where the polymer residue’s preference for solvent or cosolvent exactly match, the cosolvent-induced expansion is entirely driven by $\Delta\epsilon_{cs}$, as is obvious from Equation 3.21. The corresponding plots in Figure 3.6 B clearly show marginal chain expansion.

Thus, these two sets of calculations show that a protein can expand because of indirect solution effects, even when there is no net driving force on the chain to expand because of interactions with the solvent or the cosolvent. This is so because, as a solution becomes less poor, the chain can expand because the free energy change involved in exposing the chain to the solution is lower and more favourable. This is clear manifestation of the hydrophobic effect, and can be assessed in quantitative terms through the 3-component energetic interaction equation. The expansion is driven by the term $-\psi(1-\psi)\Delta\epsilon_{cs}$, and for solutes with positive deviations from ideality, the chain expands. It is easy to see that in a similar scenario with a cosolvent with negative deviations from ideality, we would see chain compaction. Kosmotropic salts such as NaCl and LiCl have negative enthalpies of solvation, and their presence in the solution would result in chain compaction, protein stabilization and structure formation^{42,60-61}. Thus, the effect of chaotropic or kosmotropic cosolvents on the solvent and on proteins, the effects of electrolytes in the Hofmeister series on protein stability, and the hydrophobic effect can all be explained based on the 3-component energetic interaction model, and more so, in quantitative terms. Qualitatively, this critical result can be explained as follows. Cosolvents such as urea and GdmCl prefer like-interactions over interactions with water. There is a lower energetic penalty for chain expansion in such conditions, since per site, in the hydration shell of the protein, several of the pre-existing water-water hydrogen bonds will have been broken because of the presence of cosolvents. The probability of finding a cosolvent in the vicinity of the protein in such conditions is proportional to its volume fraction, and thus with increasing cosolvent concentration, chain expansion proceeds

favourably. Any direct interaction of the cosolvent with the protein backbone or its residues is always an additional factor over this ever-present indirect, solution-driven expansion. Several studies on the role of urea and GdmCl in inducing protein unfolding implicate the loss of structure on a direct interaction of the cosolvent with the protein backbone, through hydrogen bonding and electrostatics^{39-42,81}. While it is possible that for several proteins the dominant effect for protein unfolding is a strong preference of the protein backbone to the cosolvent, one cannot rule out the possibility that depending on the net relative hydrophobic-hydrophilic properties of protein sequences, the preference of the protein backbone for cosolvent may exactly match that for the solvent.

The green squares in Figure 3.6 were produced using $\Delta\epsilon_{ps}$ equal to $200 \text{ J.mol}^{-1}.\text{site}^{-1}$ and $\Delta\epsilon_{pc}$ equal to $-1600 \text{ J.mol}^{-1}.\text{site}^{-1}$. In these conditions, the protein residue clearly favours the cosolvent leading to a large expansion in chain size. The red squares present the opposite scenario, where $\Delta\epsilon_{ps}$ is $-1600 \text{ J.mol}^{-1}.\text{site}^{-1}$, a good solvent to start off with, and $\Delta\epsilon_{pc}$ is $200 \text{ J.mol}^{-1}.\text{site}^{-1}$. In these conditions, with a preference of the protein for the solvent over cosolvent, as the volume fraction of the latter increases, the chain experiences a compaction. Experimentally this last scenario would be difficult to observe, since even highly hydrophilic chains, which interact relatively better with water than do hydrophobic chains, would have favourable interactions with cosolvents such as urea and GdmCl, precluding the possibility of observing chain compaction. The effect of cosolvents in proteins is further complicated by the existence of long range effects such as Debye-Hückel electrostatic screening of electrostatic interactions present between residues on the chain.

The blue circles indicate an intermediate scenario, where the protein has moderately favourable interactions with the solvent, due to a $\Delta\varepsilon_{ps}$ of $-200 \text{ J.mol}^{-1}.\text{site}^{-1}$ and $\Delta\varepsilon_{pc}$ equal to $-1600 \text{ J.mol}^{-1}.\text{site}^{-1}$. The result is chain expansion.

3.5.4 Effect of Chain Length on Extracted Energetic Parameters

So far, we have seen the basic features of the quadratic 3-component energy equation in Section 3.5.1, the limitations involved in applying this equation to determine the average size of an ensemble by invoking Equation 3.21 in Section 3.5.2, and heuristic calculations that pointed to the effect that these pair-wise exchange energetic interactions have on chain size in Section 3.5.3. However, for a given expansion in terms of segment length b , what effect does *chain length* N , or the number of residues in the chain, have on the exchange energetic interactions needed to produce that expansion? Ideally, for two chains with identical solvent and solute interaction properties, if segment length remains the same, then the variation in their radii of gyration is purely a function of chain length N . Their energetic interactions should necessarily be the same, and therefore, their energetic interaction parameters $\Delta\varepsilon$'s should be the same. The following calculations evaluate the net energy $\Delta\varepsilon$ needed to expand a chain through a 'range of b ' plotted below in Figure 3.7, and show that the extracted energies, in fact, vary according to chain length. This 'range of b ' corresponds to the radii of gyration plotted in Figure 3.6 B in blue circles for a 50 residue chain.

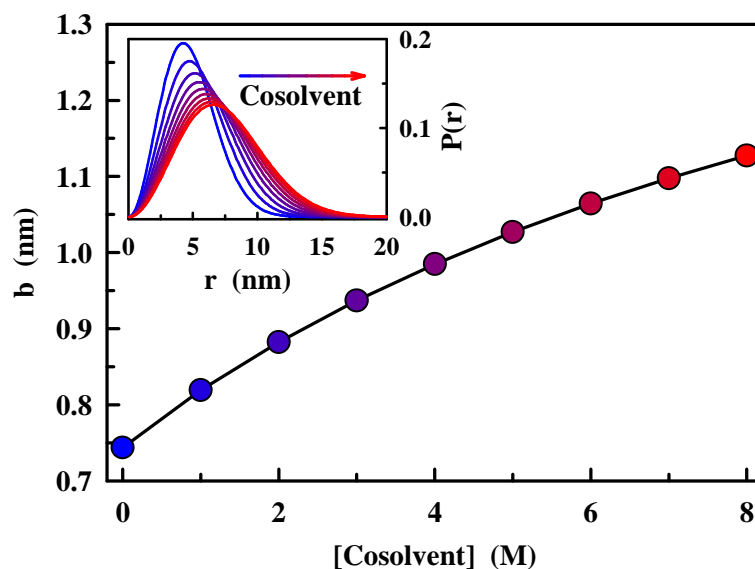


Figure 3.7 Chain expansion as a basis set for extraction of the net interaction energy $\Delta\epsilon$. (Inset) Gaussian distributions for corresponding segment lengths for a 50 residue chain.

Chains of different lengths would *not* require different energies for expanding from the same initial segment length to the same final segment length for the hypothetical range of cosolvent concentrations, as shown in Figure 3.7. However, upon inspecting Equations 3.14 or 3.16, it is apparent that different chains with different lengths N , would need different energies for the same degree of expansion based on the analytical form of the equations. Physically this makes sense, since the longer the chain, the higher its excluded volume, and the lower is the energy needed to expand chains of progressively longer lengths.

Energies corresponding to the segment length variation plotted in Figure 3.7, for chains of different lengths were obtained by the following procedure: a) expected FRET efficiencies were calculated for chains of different lengths, based on Equations 3.1 and 3.26, for the same set of segment length variations as plotted in Figure 3.7; b) these average FRET efficiencies were then plugged into Equation 3.26, however with

the distribution function from Equation 3.16, to give the expected energies, as plotted below in Figure 3.8. These energies were fit to the quadratic 3-component energetic interaction equation in Equation 3.21, with the cosolvent-solvent interaction parameter $\Delta\epsilon_{cs}$ being that for urea.

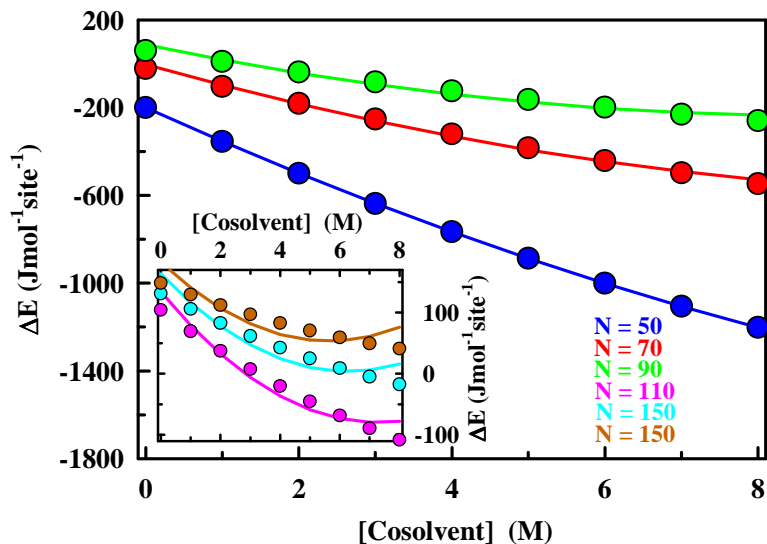


Figure 3.8 Variability with chain length of the net interaction energy for chain expansion. Circles in main figure and in the inset are energies and lines are quadratic fits of the energies.

Figure 3.8 plots the energies needed to produce the expansion within the range of segment lengths b as plotted in Figure 3.7, for chains lengths from 50 residues to 150 residues. These calculations were performed using a b_{min} of 0.33 nm assuming a temperature of 298.16 K and are represented by the circles in Figure 3.8. The lines are quadratic fits to these energies based on Equation 3.21, with urea as the cosolvent. It is important to mention that since $\Delta\epsilon_{cs}$ is fixed for a given cosolvent based on its activity coefficient data, the fits are really only linear fits with $\Delta\epsilon_{pc}$ and $\Delta\epsilon_{ps}$ being the adjustable parameters. Despite this unexpected result of the interaction energy $\Delta\epsilon$ being different for chains of different lengths, one can make the following

observations. Longer chains need lower energies to expand to the same segment length b , when compared to shorter chains. There is a large degree of variability in the energies needed to expand chains of different lengths, and the longer the chain, the lower is the net increment in interaction energy per unit increase in the volume fraction of the cosolvent. Also, the longer the chain, the higher is the interaction energy at 0 M cosolvent ($\Delta\epsilon_{ps}$). The latter result can be explained as follows. If two chains of different lengths are to have the same segment length b , then the solvent *seems to be* ‘poorer’ for the longer chain, because of its higher excluded volume. This effect is manifested in higher $\Delta\epsilon_{ps}$, indicating lower favourability of the solvent for the polymer. This result can serve as a calibration criterion described later. Figure 3.9 plots the energies at the minimum and maximum b in Figure 3.7. These energies show the trademark behaviour, i.e., lower energies needed to expand chains with increasing chain lengths.

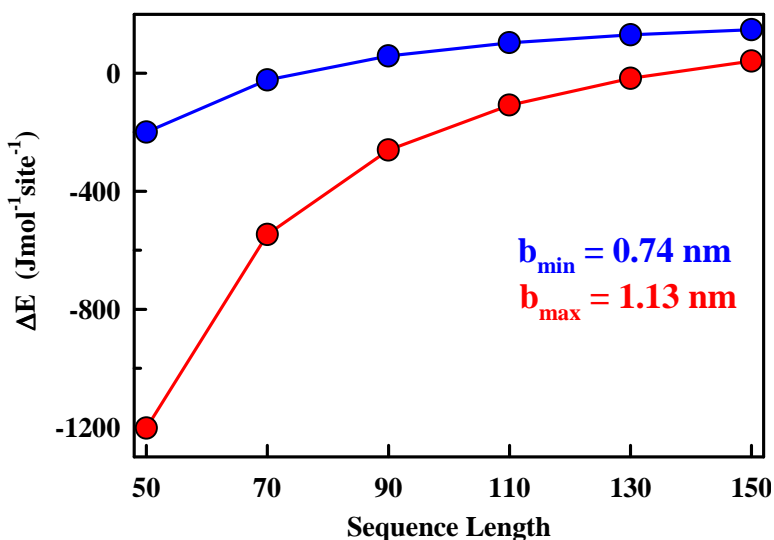


Figure 3.9 Interaction energy $\Delta\epsilon$ for different chain lengths, needed for expansions as plotted in Figure 3.7. Blue symbols indicate 0 M urea and red symbols indicate maximal expansion at 8 M urea. Lines are to guide the eye.

Evidently, another illustrative calculation would be to evaluate the amount of energy needed to expand chains of different lengths between the same ranges of segment lengths b , however this time, with the minimum and maximum b 's being extreme values. In keeping with our definition of b_{min} to be 0.33 nm for the most compact chains, we can also define an upper bound for segment length b_{max} to be 1.3 nm, based on experimental observations, described in forthcoming chapters. The utility of such a calculation would be to evaluate energies needed to expand chains to a maximal segment length b_{max} , and also to illustrate that the above observations persist even for a different set of b 's.

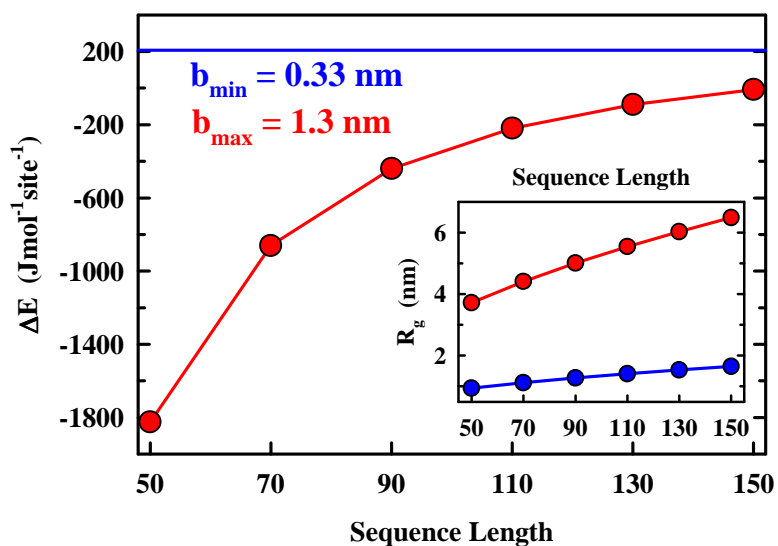


Figure 3.10 Interaction energy $\Delta\epsilon$ for different chain lengths, needed for expansion between the same range of segment lengths b . Circles indicate calculated energies. Lines to guide the eye. Blue line indicates energy at θ -condition. (Inset) Corresponding radii of gyration for b_{min} and b_{max} . Lines are to guide the eye

For the same b_{min} and b_{max} of 0.33 nm and 1.3 nm respectively, the required interaction energies are plotted versus sequence length in Figure 3.10. Since b_{min} represents the θ -condition, where the excluded volume effect and interaction energies cancel out to make the chain a random-walk chain, the chain size now is purely a function of chain length, as can be seen in the blue symbols in the inset of Figure 3.10. The red symbols in the inset also show an increase in chain size with chain length, for the same segment length b of 1.3 nm. Ideally this should also be a purely chain length effect, however the interaction energies in Figure 3.10 reflect similar observations as made earlier. The blue line in the main figure indicates a θ -condition energy of $RT/2z$ which is $206.58 \text{ J}\cdot\text{mol}^{-1}\cdot\text{site}^{-1}$. How can we explain this observation where different energies are obtained for the same physical process? And can we correct for it, so that energetic interaction parameters reflect the true physical process?

To answer the first question, if we were to differentiate Equation 3.16 with respect to r , and equate the derivative to zero, to determine the end-to-end distance where the distribution has its maximum, we get the following result and notice the dependence of the exchange factor χ on chain length N :

$$\text{If } \frac{\partial P(r)}{\partial r} = 0, \text{ then } (1 - 2\chi) : \frac{(b^5 - b^3)}{N^2} \quad (3.27)$$

In simple terms, the longer the chain the higher the *number* of likely interactions of the cosolvent with the chain (scaling as N^2) and the lower the extracted

energies we get from the fitting procedure. This explains the dependence of interaction energies $\Delta\varepsilon$ on chain length.

3.5.5 Exchange Constants of Cosolvent/Solvent Interaction with the Residue

To further inspect the quadratic 3-component equation, and its constituent coefficients, consider two equilibria as described in Figure 3.4.

$$\frac{1}{2}[p.p + s.s] f \quad p.s \quad (3.28)$$

$$\frac{1}{2}[p.p + c.c] f \quad p.c \quad (3.29)$$

Equations 3.28 and 3.29 would have free energy changes as in Equation 3.11:

$$\Delta G_{ps} = \Delta\varepsilon_{ps} = \frac{1}{2}(\varepsilon_{pp} + \varepsilon_{ss}) - \varepsilon_{ps} \quad (3.30)$$

$$\Delta G_{pc} = \Delta\varepsilon_{pc} = \frac{1}{2}(\varepsilon_{pp} + \varepsilon_{cc}) - \varepsilon_{pc} \quad (3.31)$$

Upon subtracting Equation 3.28 from Equation 3.29, we get:

$$\frac{1}{2}c.c + p.s \quad f \quad \frac{1}{2}s.s + p.c \quad (3.32)$$

Likewise, subtraction of Equation 3.30 from Equation 3.31 yields the energy difference ($\Delta\varepsilon_{pc}-\Delta\varepsilon_{ps}$), which is the free energy for the site exchange reaction in Equation 3.32. This is the classical Schellman site-exchange process where a single cosolvent moiety replaces a single solvent moiety on a residue site⁹⁷. One could thus define an exchange constant here in the spirit of Schellman's site exchange constant:

$$K_{exchange} = \frac{[pc]}{[ps]} = \exp\left[\frac{-(\Delta\varepsilon_{pc} - \Delta\varepsilon_{ps})}{R.T}\right] \quad (3.33)$$

Note that solvent or cosolvent concentrations do not appear in Equation 3.33; thus the exchange constant would simply be a ratio of the occupation probability of the cosolvent over the solvent at the site of a given residue. In other words, $K_{exchange}$ indicates the relative preference of the residue for the cosolvent moiety over that for the solvent moiety.

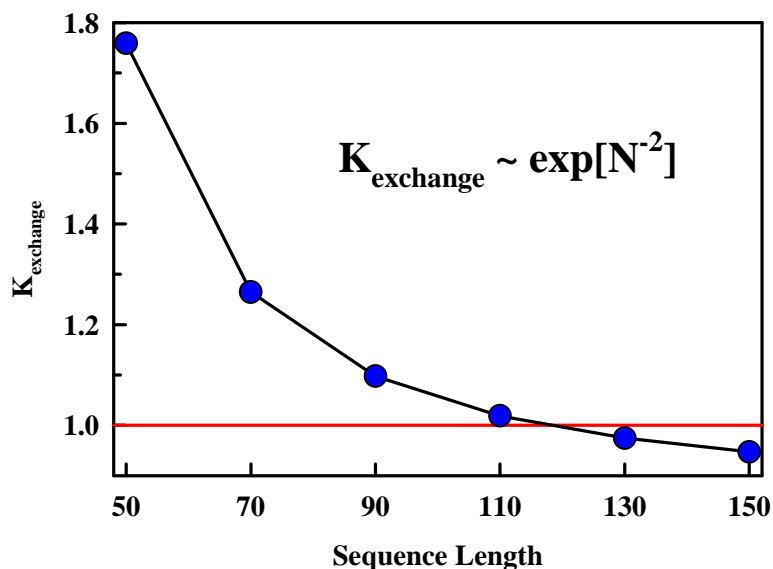


Figure 3.11 Site exchange constants or preference of a residue for cosolvent over that for the solvent. Blue circles represent K_{exchange} for different chain lengths. Black line is to guide the eye. Red line represents unity or equal preference of residue for cosolvent and solvent.

Figure 3.11 plots the site exchange constant K_{exchange} as obtained from quadratic fits to the interaction energies plotted in Figure 3.8, followed by evaluation of exchange constants, based on Equation 3.33. We again see the trademark sequence length dependence as was observed for $\Delta\varepsilon$.

To recapitulate the series of calculations in Section 3.5.4, interaction energies were evaluated for expansion between the same range of segment lengths for chains of different lengths. The basis set was taken to be a $\Delta\varepsilon_{ps}$ of $-200 \text{ J}\cdot\text{mol}^{-1}\cdot\text{site}^{-1}$, and a $\Delta\varepsilon_{pc}$ of $-1600 \text{ J}\cdot\text{mol}^{-1}\cdot\text{site}^{-1}$. The difference between the two energies exactly matches the energy difference expected for the 50 residue chain, i.e., $-1400 \text{ J}\cdot\text{mol}^{-1}\cdot\text{site}^{-1}$ or an exchange constant of 1.76, since the segment lengths used to perform this calculation were based on that for the 50 residue chain. Therefore, the variation in K_{exchange} seen in Figure 3.11 is again an effect of the length of the chain on extracted energetic

parameters. However, there is no reason for exchange coefficients to be dependent on chain length. Thus, the chain length effect on interaction energies and exchange coefficients is an effect that needs to be taken into account before further meaningful inferences are made about a chain's preference for interactions with cosolvent or solvent.

To summarize, the quadratic 3-component energetic interaction equation and the resultant exchange coefficients can quantify the effect that cosolvents have on simple polymeric chains. Expansion and compaction are a direct result of the balance between three different interaction energies- $\Delta\epsilon_{cs}$, $\Delta\epsilon_{pc}$ and $\Delta\epsilon_{ps}$. Chain expansion and compaction can be driven solely by the interactions of the solvent and cosolvent, and any direct interaction of the cosolvent with the residue is an additional effect over this solution-driven phenomenon. Furthermore, due to the non-ideality and high concentration regimes of chemical denaturant solutions and mixed cosolvents typically utilized in protein folding studies, no binding theory is invoked in this analysis. Instead interaction preferences in the form of exchange coefficients are arrived at based on a first principle evaluation of a residue's interaction with its environment. Although parameters such $\Delta\epsilon$ and $K_{exchange}$ as extracted from FRET data would have a dependence on sequence length, this can be corrected for, to reflect the true physical process where site-wise interaction between residue, cosolvent and solvent drive expansion and compaction.

3.6 Application of the Quadratic 3-Component Energy Equation to the Analysis of Data from Single Molecule FRET (smFRET) Experiments

3.6.1 Protein Unfolding Studied by smFRET

High resolution techniques such as force spectroscopy¹⁰¹ and single molecule fluorescence spectroscopy on freely diffusing and immobilized single molecules^{36,102-103} have enabled a high resolution study of protein folding and unfolding reactions compared to experiments in bulk. Such experiments have facilitated the prospect of studying folding trajectories at resolutions which were hitherto confined to atomistic simulations¹⁰⁴, and have eliminated the need to synchronize molecular reactions and to work around the problem of signals emanating from a mixed ensemble in bulk experiments. SmFRET experiments in particular are ideally suited to the study of protein folding, with the observable reaction coordinate being FRET efficiency, and the size or end-to-end distance of a protein chain¹⁰⁵. Such experiments trace their origin back to seminal studies on fluorescence from single molecules¹⁰⁶⁻¹⁰⁷, and found applications in the study of FRET between extrinsically labeled ends of proteins in conditions of varying native stability such as chemical denaturant concentrations¹⁰⁸. SmFRET experiments can also unequivocally demonstrate the simultaneous existence of different sets of molecular ensembles under given equilibrium conditions, as seen by the existence of FRET-separated populations, with the higher FRET population being the folded state and the lower FRET population being the unfolded ensemble³⁶. Thus, smFRET allows us to slice the observation of an equilibrium population into its minimum number of representative components or ensembles. Upon changing thermodynamic conditions such as GdmCl or urea concentration, the relative

populations of the folded and unfolded peaks change; this is indeed the expected observation for a classical two-state protein^{75,105,109}. SmFRET, thus serves as a critical assessment of a central assumption in the analysis of several protein folding experiments- the two-state model. Accurate estimates of FRET efficiencies depends critically on the careful determination of donor and acceptor quantum yields, the Förster radius, determination of and correction for instrumental detector efficiencies, corrections for erroneous leakage of signals into alternate channels, control of laser excitation power to minimize photo-bleaching and triplet formation, and optimization of signal-to-noise. Another observation that this technique allows is the determination of chain dynamics in the unfolded state⁹⁵. The observation of most relevance to this work is that of the unfolded ensemble in coexistence with the native state, all the way down to conditions close to those for maximal stability of the native state, i.e., low denaturant concentrations of 1-3 M. An immediate question arises of whether the unfolded ensemble itself responds to chemical denaturants by changes in the average size of the ensemble. So far smFRET experiments on several proteins indicate that in addition to a decrease in the relative population of the folded ensemble with increasing denaturant, there is an accompanying expansion of the unfolded ensemble^{36,103,105,110-111}. Qualitatively, these results point to the favourable interaction of chemical denaturants with the unfolded chain, which drives the unfolding reaction and the expansion of the unfolded state as well^{40,42,77,81,99-100}. It has therefore been possible to perform a detailed analysis on the expansion and compaction (or collapse) of the unfolded ensemble of proteins studied by smFRET, by monitoring changes in the FRET efficiencies of the unfolded ensemble^{103,112-117}. FRET efficiency

information from such experiments, on a diverse set of proteins with a range of sizes and lengths, when analyzed with the 3-component quadratic energetic interaction model, would allow the determination of the dependence of the degree of collapse on chain length, a purely polymeric property. Further, for proteins with the same lengths, the differences in the degrees of chain expansion and compaction, may point to the dependence of these properties on sequence. Finally, such an analysis would shed light on the mechanism of action of denaturants, since the 3-component quadratic equation allows us to dissect the expansion and compaction of the unfolded ensemble into contributions from cosolvent-solvent, polymer-solvent and polymer-cosolvent interactions.

3.6.2 Expansion and Collapse of the Unfolded Ensemble

SmFRET data is available on the following proteins, as of the present time, and have been used here for an analysis by the 3-component energy equation: In each reference, the first moments of the FRET efficiencies of the unfolded ensemble for each protein were typically plotted against the concentration of the corresponding denaturant used in the study. Experiments were typically performed at 25° C, unless otherwise mentioned. FRET data-sets were digitized from each reference using the DigitizeIt software (DigitizeIt, Köln, Germany). FRET data were first analyzed by a simplistic Gaussian chain model as in Equation 3.1 and radii of gyration (R_g) were determined from Equation 3.3, based on the assignment of all changes in FRET efficiency and size to the varying segment length b - a reasonable assumption for the purposes of R_g calculation, based on Equation 3.26. The more serious errors in such

an analysis can of course, arise from inadequately accurate determinations of the Förster radii (R_0), and whether this property of the FRET pair changes with chemical concentration due to changes in the absorption spectrum, molar absorptivities of either dye, or changes in their quantum yields, with changing conditions. In most of the cases below, a standard R_0 was used for all cosolvent or chemical concentrations when no changes in dye quantum yields or lifetimes were observed. Furthermore, γ -correction is necessary to account for differences in detector efficiencies for the two dyes and is thus important for accurate determination of FRET Efficiencies¹⁰².

Table 3.2 Properties of proteins studied by smFRET. Dye-pairs in main text.

Protein	N	R_0 (nm)	m_{eq} (kJ.mol ⁻¹ .M ⁻¹)	C_m (M)
CI2	54	6.2	7.49 ⁷⁵	3.92 ⁷⁵
ACBP	61	6.2	11.30 ¹¹⁸	2.36 ¹¹⁸
Protein L	65	5.4	7.95 ¹¹⁵	2.60 ¹¹⁵
CspTm	67	5.4	7.90 ¹¹⁹	2.0 ¹¹⁵
RNase H	154	7.1	21.9 ¹²⁰	1.81 ¹²⁰
Im9 (Urea)	85	5.4	4.20 ¹¹⁶	
Staph. Nuclease	71	5.13	6.83 ¹²¹	
Staph. Nuclease	98	5.13	6.83 ¹²¹	
Barstar (GdmCl)	79	5.4	7.70	1.24
Barstar (Urea)	79	5.4	5.23 ¹²²	3.87 ¹²²
SH3 (GdmCl)	69	5.4	7.31	1.90
SH3 (Urea)	69	5.4	2.566	5.43

1) Chymotrypsin Inhibitor 2 (CI2) and Acetyl-Coenzyme A Binding Protein (ACBP)¹²³:

CI2 is a serine protease inhibitor of the potato inhibitor I family, is found in the albumin of seeds from the *Hiproly* strain of barley, and has been shown to be a two-state folder by equilibrium and kinetics experiments⁷⁵. ACBP is a single-chain α -helical protein found in the bovine liver, and like CI2 has no disulphide bonds, prosthetic groups, metal ions or cofactors, and is a two-state protein, as indicated by cooperative folding¹¹⁸. SmFRET measurements were performed on freely diffusing molecules by nanosecond alternating laser excitation of two extrinsically chemically attached dyes, Alexa Fluor 532 and Alexa Fluor 647 as donor and acceptor, respectively, in the study by Weiss and coworkers¹²³. Being two-state folders, two separate ensembles coexisting at equilibrium are observed by smFRET, as is the denaturant dependent expansion of the unfolded ensemble. The ‘sequence length’ that should be plugged into Equations 3.1, 3.3 and 3.14, corresponds to the number of bonds linking the amino acid residues in the protein ($N_{\text{amino-acid residues}} - 1$) plus two bonds for the linkers of the dyes connecting them to the protein backbone by the terminal residues. In cases where chemical labeling was performed at non-terminal sites, segment lengths were calculated with the above rule, and R_g were calculated based on the number of linkers in the full length protein. The tabulated value of N in Table 3.2 is the figure which is used to calculate R_g . The digitized data from the reference by Laurence and coworkers¹²³ is plotted here in Figure 3.12 A. Corresponding end-to-end distances and R_g were calculated based on the Gaussian chain assumption. CI2 and ACBP have linker lengths of 54 and 61, respectively, and

ACBP is slightly larger in size than CI2. This difference in size has to do with sequence length. Furthermore, as can be seen from Figure 3.12 B, ACBP seems a lot more sensitive to GdmCl, than CI2, from the GdmCl concentration needed for similar degrees of expansion in terms of R_g . This can also be inferred from their denaturation midpoints which are 2.36 M and 3.92 M GdmCl for ACBP and CI2, respectively.

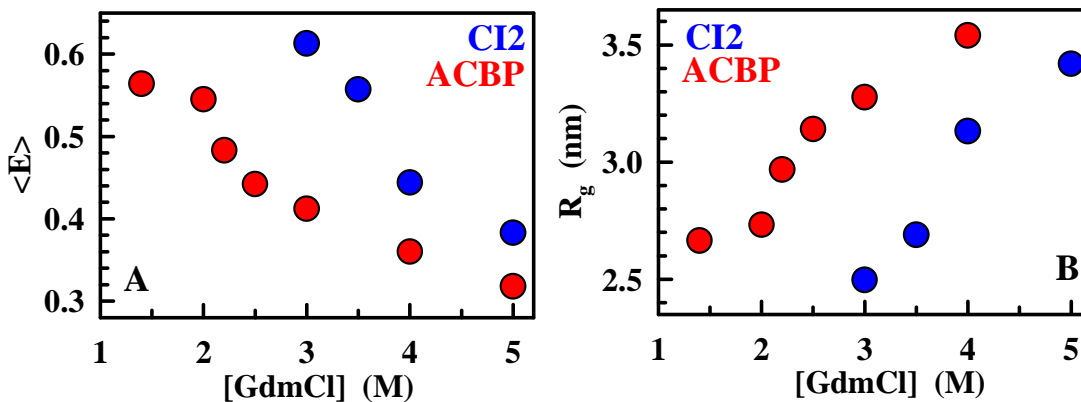


Figure 3.12 FRET efficiency and R_g of the unfolded ensembles of CI2 and ACBP. A) SmFRET efficiencies measured by nanosecond alternating laser excitation with Alexa Fluor 532 and Alexa Fluor 647 as the donor and acceptor dyes respectively. B) R_g calculated by the Gaussian chain approximation.

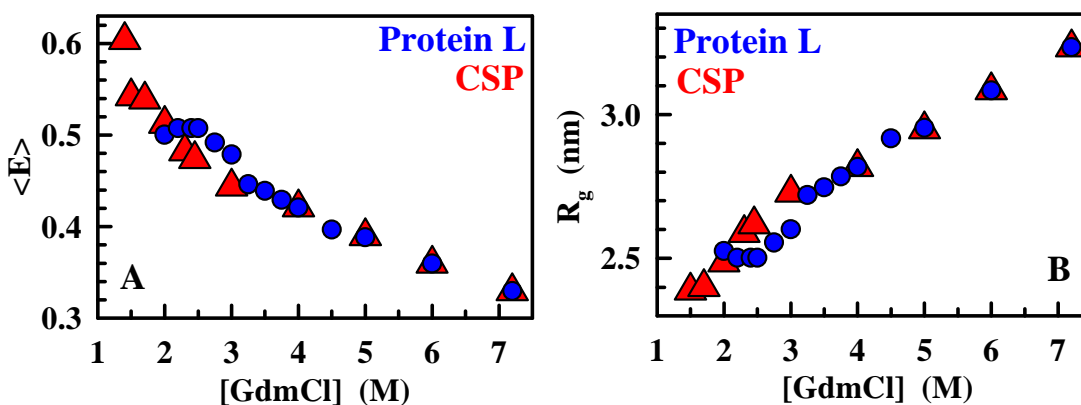


Figure 3.13 FRET efficiency and R_g of the unfolded ensembles of Protein L and CSP. A) SmFRET efficiencies measured by intensity and lifetime measurements on freely diffusing molecules with Alexa Fluor 488 and Alexa Fluor 594 as the donor and acceptor dyes respectively. B) R_g calculated by the Gaussian chain approximation.

2) Protein L and Cold Shock Protein (CSP) from *Thermatoga maritima*¹¹⁵:

The B1 IgG binding domain of the peptostreptococcal protein L, or simply protein L, is a 62 residue IgG binding α/β protein, which had been modified for the smFRET study, by the addition of cysteines and C-terminal residues^{115,124}, bringing the length to 65. The cold shock protein from the thermophile *Thermatoga maritima* is a β protein, with a length of 67. Both proteins have been extensively studied in equilibrium^{36,119,124-126} and are known to be two-state folding proteins. Being similar in length, both proteins exhibit FRET Efficiencies in similar ranges with Alexa Fluor 488 and Alexa Fluor 594 as the donor and acceptor dyes, respectively, in Figure 3.13 A, based on both lifetime and intensity measurements on freely diffusing molecules, as opposed to simply intensity measurements on protein L by Sherman and Haran¹²⁷. In the protein L data-set, we observe a rollover in R_g at low GdmCl concentration where the chain is seen to ‘collapse marginally’. However, at the concentration where this is observed, i.e. 2-3 M GdmCl, the Debye-Hückel charge-screening effect will have saturated. Hence, this effect is most probably because of the low population of the unfolded ensemble in those given conditions, making it difficult to accurately estimate mean FRET efficiencies. This point is of critical importance, since in conditions of native protein stability, at low chemical concentration, depending on the threshold of counts in the single molecule histograms, the unfolded ensemble might be difficult to resolve. Both proteins experience an expansion with increasing GdmCl concentration well up to the highest concentration range of 7 M. These observations are in contrast with those made by Plaxco *et al.*¹²⁸, where time-resolved small angle X-ray scattering measurements on protein L yielded radii of gyration which matched

well with the R_g measured by smFRET at 4 M GdmCl, but not at a low concentration of 1.2 M, where scattering data indicated a similar R_g as that at 4 M GdmCl, i.e., ~ 2.6 nm. The source for this conflicting result was purported to lie in the high protein concentrations used in time-resolved X-ray scattering experiments, which could skew the measured R_g at low GdmCl, to higher values¹¹⁵. Also, the R_g of the two proteins as plotted in Figure 3.13 B, are similar at high GdmCl concentrations above 3 M. However, at lower concentrations, protein L seems to be more collapsed than CSP, and this difference could be attributed to their different hydrophobicity indices, as protein L is slightly more hydrophobic than CSP. This difference could also arise because of the presence of transient structure in the unfolded ensemble of protein L¹²⁹⁻¹³⁰, whereas no such residual unfolded state structure was found in the case of CSP¹³¹⁻¹³². Either way, these results were posited to be consistent with the demonstration by Kohn *et al.*⁹¹, that at high denaturant concentrations unfolded polypeptide chains behave like random homopolymers.

3) Ribonuclease H1 (RNase H)¹⁰³:

The 155 residue protein RNase H from *Escherichia coli* is an α/β protein, and refolds by the population of a transient intermediate that is observed as a ‘burst-phase intermediate’^{103,120}. In the smFRET unfolding study by Kuzmenkina *et al.*, surface-immobilized RNase H molecules, labeled with Alexa Fluor 546 and Alexa Fluor 647 through maleimide chemistry, were observed by confocal microscopy, while solution conditions were changed *in situ*¹⁰³. The labels were attached on residue 3 and 135; therefore, end-to-end distance information was reported for an effective length of 134 linkers. In the study by Kuzmenkina *et al.* two well-separated populations were

observed, with the trademark GdmCl-dependent expansion of the unfolded ensemble. The study further posits that the unfolded ensemble being sensitive to denaturant stress, populates a ‘continuum of substates’, with GdmCl stabilizing expanded conformations at high concentrations. This unfolded ensemble could, in principle, coexist with the native state at low denaturant concentrations and would be compact in size. Direct observation of this compact denatured state was precluded by the absence of a well defined FRET histogram because of low signal¹⁰³. The smFRET efficiencies in Figure 3.14 A in blue circles were converted to R_g based on the Gaussian chain assumption with the number of sequence linkers being 154 residues for the full length protein with 155 residues.

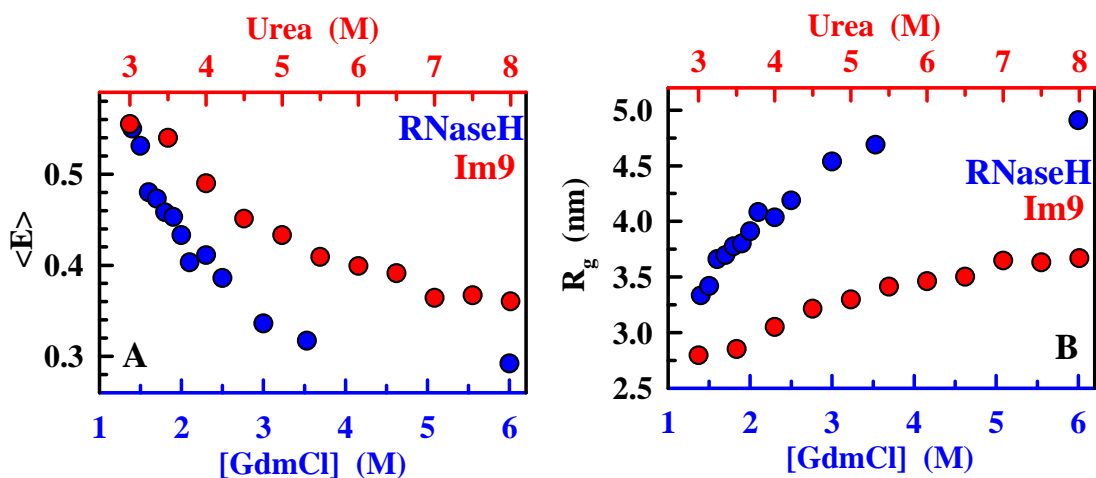


Figure 3.14 FRET Efficiency and R_g of the unfolded ensemble of RNase and Im9. A) SmFRET efficiencies. B) R_g calculated by the Gaussian chain approximation for 155 residue chain for RNase H and 86 residue chain for Im9.

4) E colicin binding immunity protein Im9¹¹⁶:

Im9 is an 86-residue 4-helical protein which folds by the two-state mechanism at neutral pH. In the smFRET study by Tezuka-Kawakami *et al.* a freely-diffusing 6 His-tagged version of the protein was used. The dyes Alexa Fluor 488 and Alexa

Fluor 594 were introduced at a S81C mutation site, and at a naturally occurring C23. Thus the effective length for Gaussian analysis in this case is 60 linkers. The value of $(N_{\text{amino-acid}}-1)$ for calculating the R_g is 85. Unfolding was initiated by urea and performed at 10° C. The choice of denaturant and temperature was made to compare smFRET unfolding data to those at equilibrium¹³³.

5) Staphylococcal Nuclease (Staph. Nuclease)¹¹³: Staphylococcal nuclease is a 149 residue protein with a N-terminal five-stranded β barrel subdomain, and a C-terminal 3 α -helical α subdomain. There is an ongoing debate regarding the folding mechanism of this protein, as to whether it folds via a two-state mechanism or a three-state mechanism with a collapsed intermediate in between the unfolded and native states¹³⁴⁻¹³⁶. SmFRET experiments were performed on two variants, where the dyes Alexa Fluor 555 and Alexa Fluor 647, were placed on residues with the N-terminal β domain at positions K28C and K97C, and another between the two domains at positions K28C and H124C. The smFRET efficiencies of the variable unfolded ensemble were converted to segment lengths b in this case to allow a comparison of the degrees of collapse within the β domain and between domains, corrected for the difference in the number of residues separating them, in Figures 3.15 A and B.

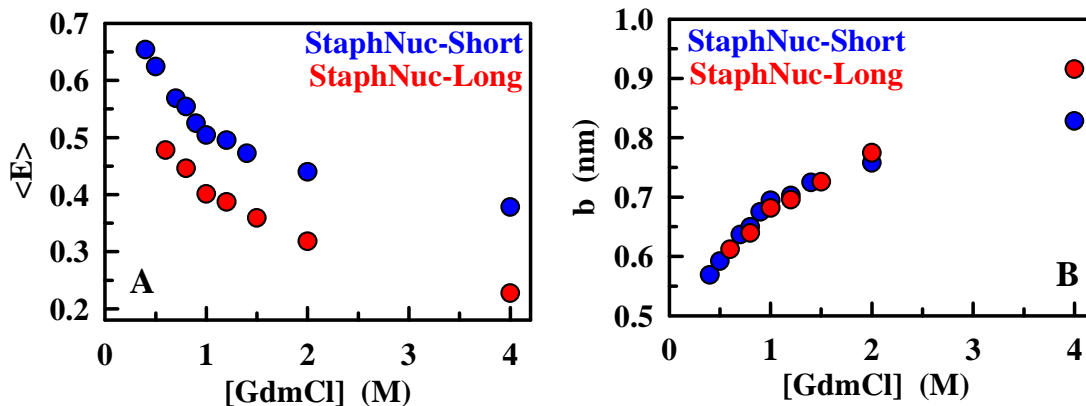


Figure 3.15 A) SmFRET Efficiencies of unfolded Staph. Nuclease versions. B) b , the effective segment length.

The smFRET efficiencies and segment lengths in Figure 3.15 A and B respectively, argue against the presence of a low denaturant compact intermediate. This follows from the fact that no intermediate species was observed at low GdmCl concentrations, with the caveat that if low in population, this species would be hard to detect by smFRET. Furthermore, segment lengths, upon correction for chain length dependence show very similar response of the sequences to chemical denaturant, indicating uniform domain collapse and expansion, precluding the possibility of any domain-specific collapse within the β domain. The β domain does happen to be less expanded at 4 M GdmCl than as indicated by the interdomain segment lengths for the longer Staph. Nuclease variant. However, at such high concentrations of chemical, proteins are expected to behave like random coil heteropolymers, despite claims being made about Staph. Nuclease possessing significant long-range residual structure even at 8 M urea³⁰. Thus, based on the smFRET data, the β domain may have some intra-domain residual structure with native-like topology that is manifested as a lower segment length at 4 M GdmCl.

6) Barstar¹¹⁴: Barstar is a 90 amino acid inhibitor of barnase, a ribonuclease from *Bacillus amyloliquefaciens*. Cysteine residues were introduced at positions 12 and 89 to facilitate labeling with the fluorophores Alexa Fluor 488 and Alexa Fluor 594. The Barstar variant used here is known as the pseudo-wild-type or pWT. Results from smFRET measurements with urea and GdmCl, indicate that low concentrations of GdmCl induce chain collapse because of the electrostatic interaction of guanidinium ions with negative charges on the protein at pH 7-8, since Barstar has 16 acidic and 10 basic residues. Debye-Hückel charge-screening of the electrostatic repulsions on the chain, can also induce chain collapse at low ionic strengths of GdmCl. An extrapolation of the unfolded ensemble R_g 's measured by GdmCl or urea denaturation do not have the same values at 0 M chemical, indicating the possibility that screening and modulation of electrostatics on the chain can induce different secondary structures in the presence of different denaturants¹¹⁴.

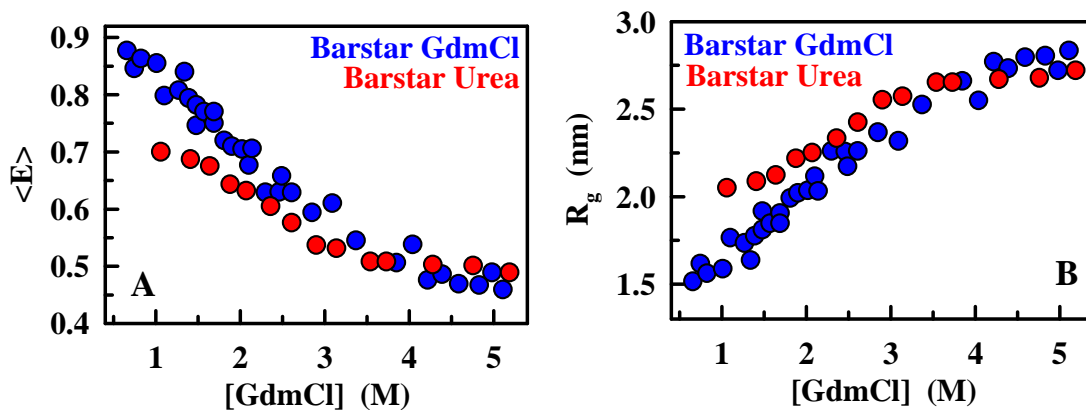


Figure 3.16 SmFRET Efficiencies and R_g of the unfolded population of Barstar, induced by GdmCl and urea, based on measurements of freely diffusing molecules.

7) SRC Homology 3 Domain (SH3; Experiments performed in the Muñoz group by Dr. Jian Wei Liu and Dr. Luis Alberto Campos-Prieto): The SH3 domain is a β -barrel protein with 5 β -strands packed to form to antiparallel β -sheets. The variant studied here is 70 amino acids in length with Alexa Fluor 488 and Alexa Fluor 594, the donor and acceptor dyes, respectively, labeled at positions C3 and C70.

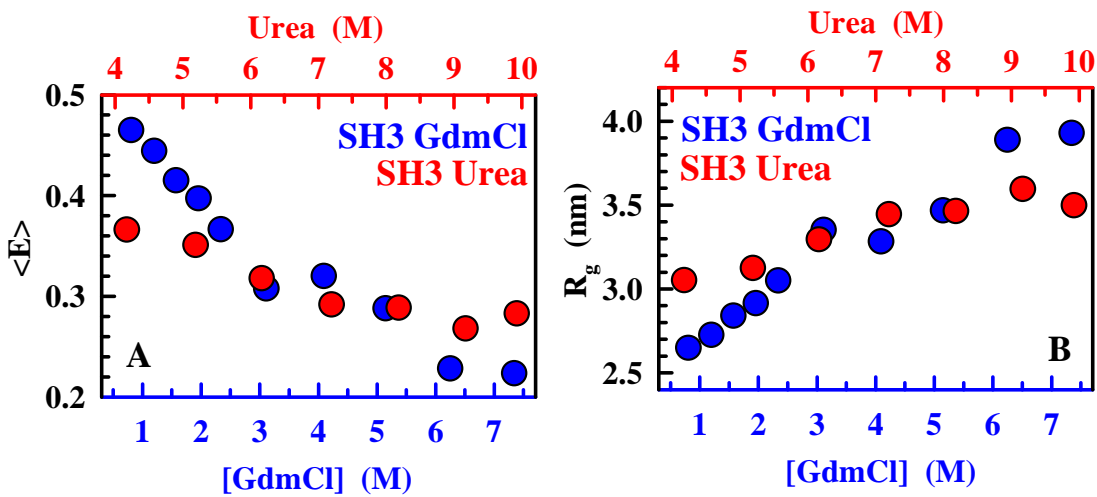


Figure 3.17 SmFRET and R_g of the unfolded ensemble of SH3.

Among all the proteins analyzed above by smFRET, several of them such as CI2, ACBP, Protein L, CSP and SH3 are highly sensitive to GdmCl and expand well into the high concentration ranges of 5-6 M GdmCl. Cosolvent-driven expansion in proteins such as Barstar and RNase H though, seem to taper off at concentrations of ~ 5 M, though they are two among four of the longest proteins in these studies (with Im9 and Staph. Nuclease-Long being the other two). Thus, other than a simple saturation of binding sites which can be expected with increasing GdmCl concentrations¹³⁷, sequence and protein-specific effects are also observed which need to be investigated in further detail here. These effects are even more pronounced at low chemical concentrations. Protein L and CSP being very similar in length, have

similar R_g at high GdmCl concentrations but are markedly different at low concentrations, where Protein L seems to be more collapsed. The properties of unfolded Barstar are particularly interesting since, with two different chemical denaturants its collapse-expansion behaviour in the low-concentration regime is completely different. Being a protein with a large number of charged amino acids, it is sensitive to the electrolytic cosolvent GdmCl, and in its presence, Barstar's enhanced degree of collapse in comparison to urea, may suggest the formation of structures which are specific to the property of the mixed solvent involved. The analysis with the 3-component quadratic equation follows in the next section.

3.6.3 Application of the 3-Component Quadratic Equation to smFRET Data

In this section, a detailed analysis of the smFRET data is performed by fitting the interaction energies obtained from the probability distribution in Equation 3.14 to the 3-component quadratic equation. Such a fit is performed by fixing values of the cosolvent-solvent interaction energy $\Delta\epsilon_{cs}$ for GdmCl and urea from their activity coefficient data, and fitting the $\Delta\epsilon_{ps}$ and $\Delta\epsilon_{pc}$. The procedure is therefore the following:

- 1) smFRET efficiencies are fit to the transcendental Equation 3.26, with the probability distribution coming from Equation 3.14, with b being fixed at 0.33 nm, and R_0 coming from experimental determination, to yield $\Delta\epsilon$ as the fitted parameter. Since Im9 experiments were performed in urea at 283.16 K instead of 298.16 K, $\Delta\epsilon_{cs}$ at 283.16 K was calculated assuming a linear dependence of $\Delta\epsilon_{cs}$ on temperature.
- 2) $\Delta\epsilon$'s for a given protein are then fit to the 3-component Equation 3.21.

- 3) The fitted parameters $\Delta\epsilon_{pc}$ and $\Delta\epsilon_{ps}$ are then used to recalculate $\Delta\epsilon$, which is then plugged back into Equation 3.14, to recalculate 'fitted' $\langle E \rangle$ from Equation 3.26 to assess the fits.
- 4) In case of the Barstar and SH3 data-sets $\Delta\epsilon_{ps}$ were fitted to be the same for both data-sets, and were also separately allowed to freely float, yielding different sets of results.
- 5) Exchange factors $K_{exchange}$ were calculated based on Equation 3.33 to arrive at relative preferences of the residue for the cosolvent over that for the solvent. Exchange factors were also calculated to evaluate preference of the cosolvent for the protein residue over H₂O.

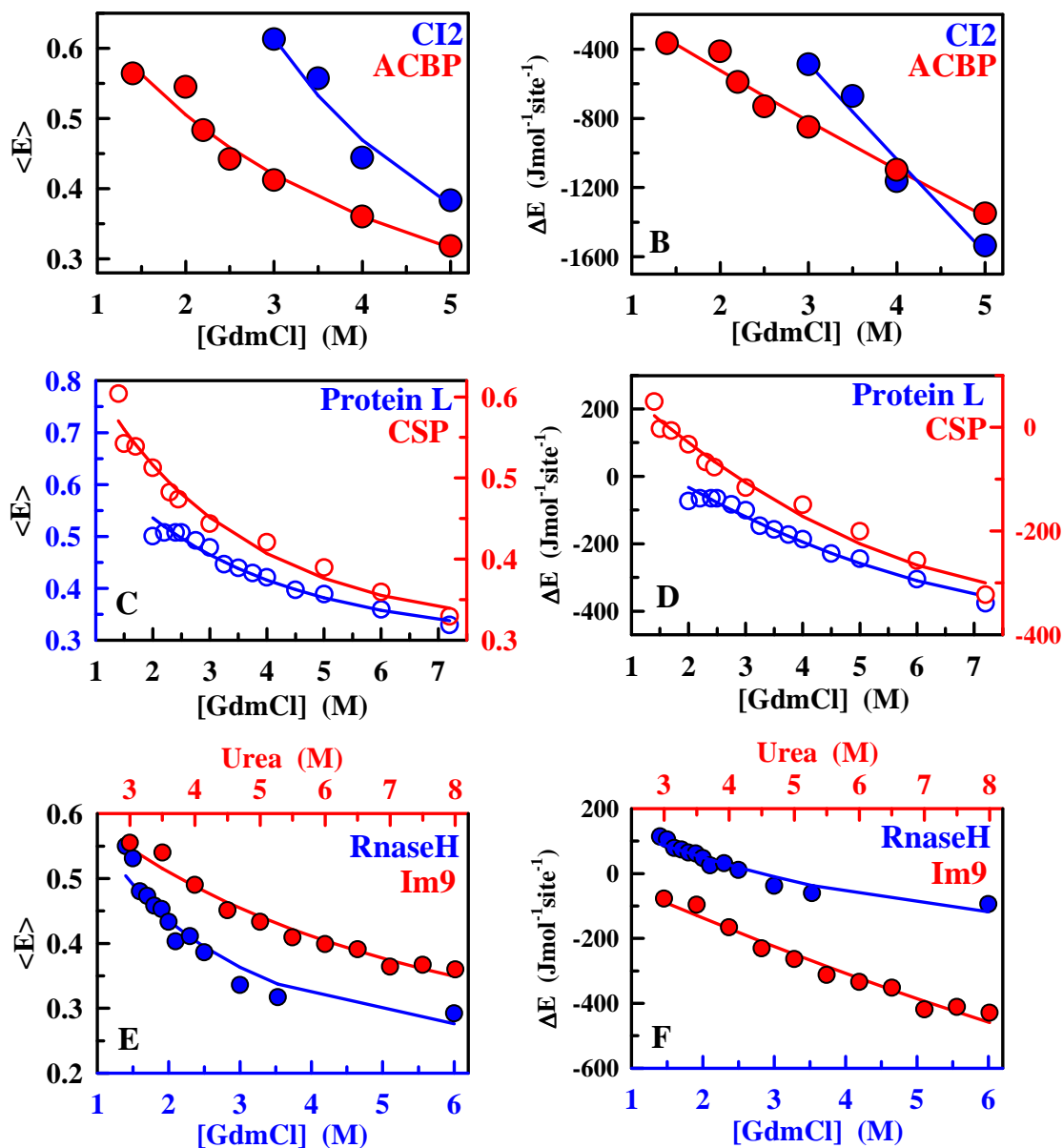


Figure 3.18 Fits to smFRET Efficiencies and $\Delta\epsilon$ by the 3-component quadratic equation. Circles indicate experimental data ($\langle E \rangle$) and calculated $\Delta\epsilon$. The lines indicate fits to $\Delta\epsilon$ by the 3-component equation and recalculated $\langle E \rangle$ with the fitted $\Delta\epsilon$.

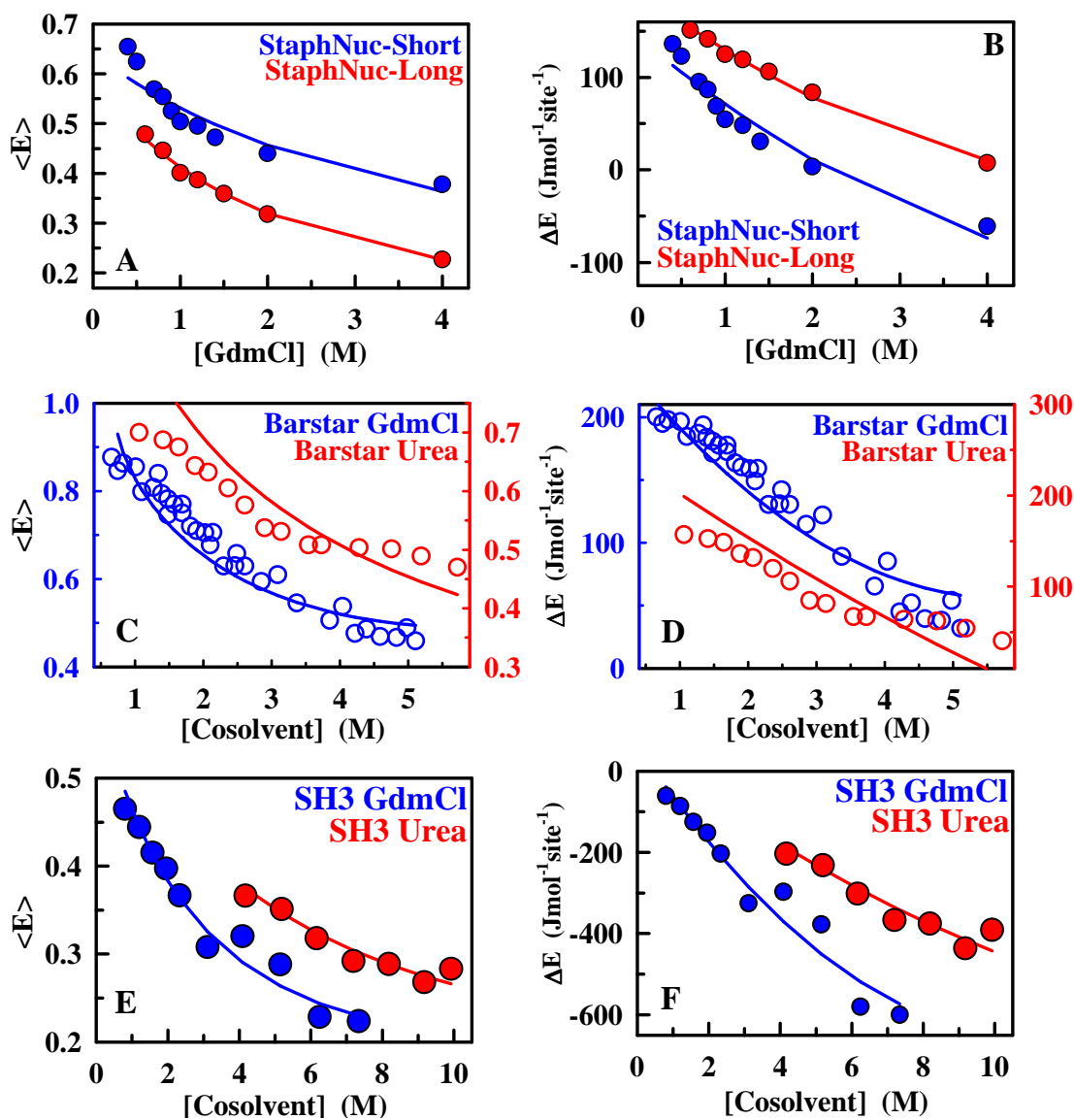


Figure 3.19 Fits to smFRET Efficiencies and ΔE by the 3-component quadratic equation. Circles indicate experimental data ($\langle E \rangle$) and calculated ΔE . The lines indicate fits of ΔE by the 3-component equation and recalculated $\langle E \rangle$ with the fitted ΔE .

The fitting procedure above produces the results in Table 3.3. Fits for RNase H, Staph. Nuclease-Short and Barstar deviate from experimental data. In case of RNase H, fits deviate towards lower values compared to FRET efficiencies at low chemical concentration, and is most evident at the lowest two concentrations. This

could have to do with errors in determination of the cosolvent-solvent interaction energy $\Delta\epsilon_{cs}$ by our procedure, with a resultant curvature that is not high enough to reproduce the curvature seen in the RNase data. This deviation could arise from the fact the curvature is dominated by the first two data points and also from the absence of data-points between 3.5 M and 6 M. Low chemical concentration deviations could arise from the fact that it is difficult to discern the unfolded population signal from that of the folded population at low chemicals, where the two peaks are often merged.

Table 3.3 Fits of smFRET Efficiencies to 3-Component Energetic Equation.

Protein	N	$\Delta\epsilon_{cs}$ (J.mol ⁻¹ .K ⁻¹)	$\Delta\epsilon_{pc}$ (J.mol ⁻¹ .K ⁻¹)	$\Delta\epsilon_{ps}$ (J.mol ⁻¹ .K ⁻¹)	$K_{exchange,p-cs}$
CI2	54	1108.2	-5780.5±1105	1243.9 ± 306	17.0
ACBP	61	110.8.2	-3212.1 ± 274	98.0 ± 60	3.8
Protein L	65	1108.2	-310.1 ± 44	176.0 ± 12	1.2
CSP	67	1108.2	-188.4 ± 42	159.1 ± 11	1.2
RNase H	154	1108.2	89.8 ± 251	204.4 ± 48	1.0
Im9 (Urea)	85	1013.7	-929.5 ± 22	187.5 ± 24	1.6
Staph. Nuclease	71	1108.2	184.9 ± 174	143.1 ± 21	1.0
Staph. Nuclease	98	1108.2	352.0 ± 20	193.0 ± 3	0.9
Barstar (GdmCl)	79	1108.2	424.0 ± 30	253.9 ± 6	0.9
Barstar (Urea)	79	762.2	-143.3 ± 1082	253.9 ± 166	1.2
SH3 (GdmCl)	69	1108.2	-609.3 ± 100	62.5 ± 28	1.3
SH3 (Urea)	69	762.2	-628.9 ± 135	62.5 ± 43	1.3

The fits to Staph. Nuclease-Short suffer from the same shortcomings as the RNase H data-set. In general both RNase H and Staph. Nuclease-Short FRET efficiencies have higher curvature with chemical concentration than that expected for GdmCl. Fits could be recalculated by allowing the cosolvent-solvent terms for these two data-sets to float freely.

Barstar on the other hand has poor fits because both data-sets were fit with a common polymer-solvent energy at 298.16 K. A polymer should, in principle, collapse to the same average ensemble at decreasing concentrations of the cosolvent, irrespective of which cosolvent is used. However, it is known that Barstar collapses to a different extrapolated R_g in the presence of urea when compared to GdmCl, which has been attributed to the possible existence of secondary structure in the unfolded ensemble in native conditions in the presence of an electrolytic cosolvent such as GdmCl¹¹⁴. Thus, one could arrive at better fits if the two polymer-solvent terms for the two-data sets were allowed to float independently. The results are plotted in Figure 3.20 and the fitted parameters are listed in Table 3.4.

Table 3.4 Fits of smFRET Efficiencies with freely-floating parameters.

Protein	N	$\Delta\epsilon_{cs}$ (J.mol ⁻¹ .K ⁻¹)	$\Delta\epsilon_{pc}$ (J.mol ⁻¹ .K ⁻¹)	$\Delta\epsilon_{ps}$ (J.mol ⁻¹ .K ⁻¹)	$K_{exchange,p-cs}$
RNase H	154	2755.7	955.2 ± 23	294.3 ± 5	0.8
Staph. Nuclease	71	3161.1	1607.8 ± 33	176.3 ± 4	0.6
Barstar (GdmCl)	79	1108.2	358.6 ± 24	273.7 ± 5	1.0
Barstar (Urea)	79	762.22	141.7 ± 40	195.4 ± 6	1.0

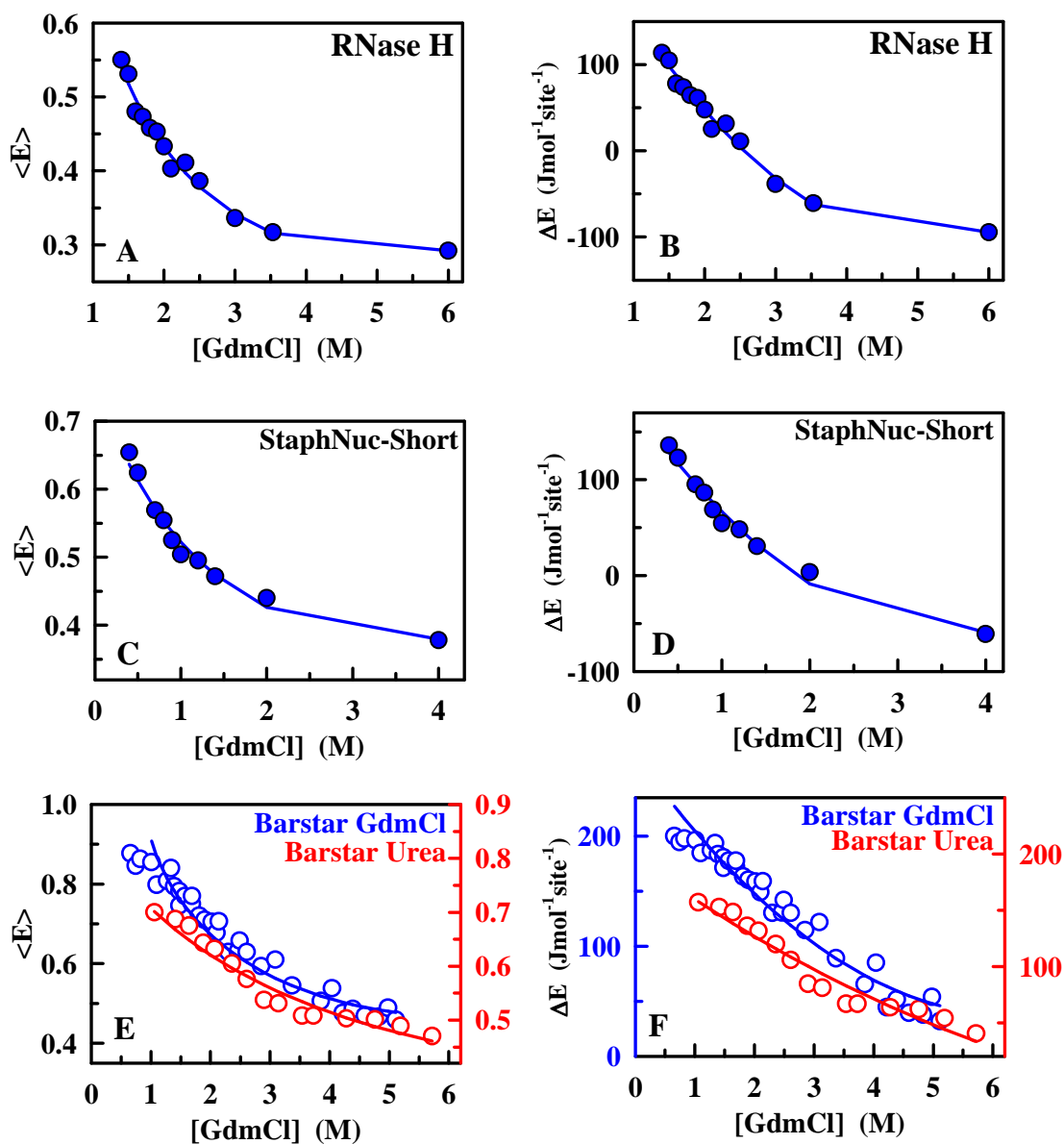


Figure 3.20 Fits with independent $\Delta\epsilon_{cs}$ for RNase H and Staph. Nucl.-Short, and independent $\Delta\epsilon_{ps}$ for the Barstar GdmCl and urea data-sets.

The CI2 and ACBP data-sets show particularly steep sensitivity to chemical, as seen by the slopes of their FRET efficiencies and the resultant $\Delta\epsilon$'s versus chemical concentration. This can be attributed to the fact that FRET efficiencies were not γ -corrected for variations in detector efficiencies in the study by Laurence *et al*¹²³. This explains the high exchange coefficients $K_{exchange\ p-cs}$ and $K_{exchange\ c-ps}$ obtained for CI2

and ACBP, which physically represent the preference of the polymer residue for the cosolvent over that for the solvent H₂O, and the preference of cosolvent molecules for polymer residues over that for solvent molecules, respectively. In general $K_{exchange,p-cs}$, the microscopic equilibrium constant for polymer interactions with cosolvent over solvent moieties, is greater than unity indicating favourable interactions between the polymer and the cosolvent with respect to those with the solvent.

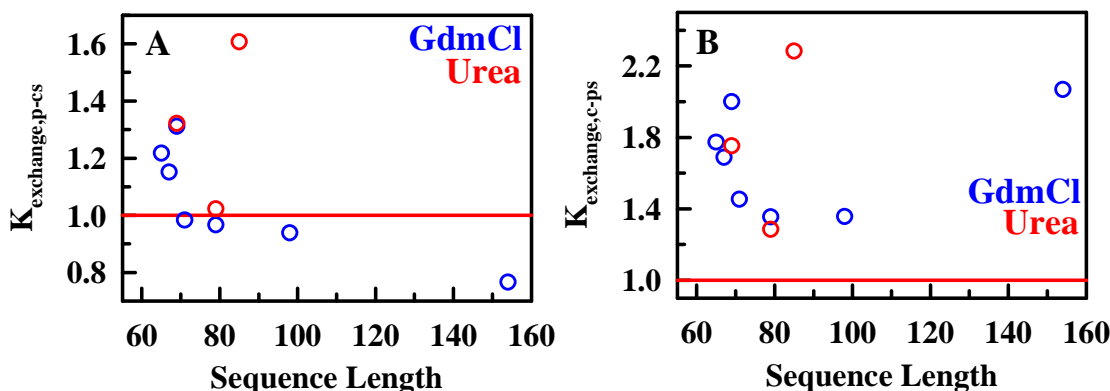


Figure 3.21 Exchange coefficients from 3-component fits. A) Residue preference for cosolvent over that for H₂O. B) Cosolvent preference for residue over that for H₂O. Red lines in both panels indicate equal relative preferences, i. e. $K_{ex}=1$.

The same holds for $K_{exchange,c-ps}$ as well, and the two sets of exchange coefficients are plotted above in Figure 3.21. It is to be noted that in this figure, except for RNase H, and Staph. Nuclease-Short, all data were fit with common $\Delta\epsilon_{cs}$ coming from cosolvent-water activity coefficients, and with a common $\Delta\epsilon_{ps}$ for the SH3 data-set. RNase H data is fitted with a freely-floating $\Delta\epsilon_{cs}$ indicating a higher curvature, which implies stronger relative like-interactions in GdmCl-water solutions if the data are taken as is. Freely floating $\Delta\epsilon_{cs}$ fits for Staph. Nuclease-Short are discarded since the standard $\Delta\epsilon_{cs}$ from activity coefficient data of GdmCl seems to fit the Staph-

Nuclease-Long data well. This difference could arise, as mentioned previously, from errors in determination of FRET efficiencies at low GdmCl concentration, the paucity of data points in the high concentration limit skewing the fits, or specific interactions of GdmCl within the N-terminal β domain. Another source for error could be interactions between the dyes which would yield higher FRET efficiencies than expected, and would skew curvature to higher values. Therefore, $\Delta\varepsilon_{cs}$ from activity coefficient data is retained for Staph. Nuclease. Barstar data are fit with independent $\Delta\varepsilon_{ps}$ as justified earlier. Taken together, these results are all plotted as the exchange coefficients in Figure 3.21 A and B, except those for CI2 and ACBP. The immediate observation one can make is the spread of points around unity. Unit exchange coefficients imply equal interaction propensity for the pair of molecules being compared. Thus, all cosolvent based effects when $K_{exchange,p-cs}$ is unity arise from cosolvent driven effects. In general the protein residues have a higher preference for cosolvents than for water, except in the cases of Staph. Nuclease (both variants), RNase H and the GdmCl data-set of Barstar in Figure 3.21 A. The corresponding $K_{exchange,c-ps}$ in Figure 3.21 B show stronger interactions of the cosolvent with the protein residue, than with water for all data-sets. Thus a combination of the residue's preference for cosolvent, and the cosolvent's more frequent occupation of the protein's interaction shell compared to bulk solution (direct and indirect cosolvent effects, respectively), drive the expansion of the unfolded chain. The latter indirect effect is clearly seen in the cases of RNase H and Staph. Nuclease, where strong indirect preference of the cosolvent GdmCl to the residue, compared to that for water is what drives the expansion process, rather than direct cosolvent-protein interaction.

$K_{exchange,p-cs}$ for both these data-sets fall below unity, but their $K_{exchange,c-ps}$ are well above unity.

Barstar and SH3 are the only two data-sets with smFRET measurements with both urea and GdmCl as cosolvents. SH3 has higher preference for either cosolvent compared to water, and a relatively higher preference for urea than GdmCl, by a faint margin (Figure 3.21 A). Likewise, either cosolvent molecule prefers interactions with the residue than with water, and this is seen from Figure 3.21 B. However, stronger indirect solution effects in the presence of GdmCl, drive chain expansion when compared to urea as the $K_{exchange,c-ps}$ is much higher for GdmCl (2.0) than urea (1.8). Barstar, on the other hand, has stronger interactions with urea than GdmCl (Figure 3.21 A), and this is the reason for collapse seen in the presence of GdmCl, but not in the presence of urea, despite the marginally higher solvent-relative preference of GdmCl for the residue than urea (Figure 3.21 B).

Another feature that emerges from Figure 3.21 is the clear chain-length dependence of exchange coefficients described in Sections 3.5.4 and 3.5.5. Despite this model-dependent artifact, the main conclusions drawn above still hold. To compare these results with data from protein folding experiments, a correlation can be sought between the exchange coefficients and protein folding m -values. The underlying physical process for both unfolded chain expansion and protein unfolding is the same, i.e. interaction of the chemical denaturant with the protein residues. The sensitivity of a protein to chemical denaturation can be obtained by the linear extrapolation method where protein folding free energies are fit to a linear equation with ΔG_{water} as the intercept on the y-axis and $-m$ as the slope⁸⁰. Protein folding m -

values are known to correlate linearly with the change in accessible surface area (ΔASA) upon protein unfolding, viz. the area of the protein that is exposed to the solvent upon unfolding. The expansion of an unfolded ensemble upon the introduction of chemical denaturants is akin to the unfolding of a protein in the presence of chemical and indeed is connected by a simple thermodynamic cycle¹³⁷. The change in accessible surface area upon protein unfolding was shown to vary linearly with the number of residues on the protein with a correlation coefficient of 0.994⁸⁰. It may be useful to also consider variation of m -values with $N^{6/5}$, since:

$$m : \Delta ASA, \Delta ASA : R^2, R : N^{3/5} \Rightarrow m : N^{6/5} \quad (3.34)$$

Furthermore, the exchange coefficients also vary as $\exp(N^{-2})$ (Section 3.5.5). Thus, m -values and $K_{exchange,p-cs}$ should have a positive correlation with an increase in m -value correlating with an increase in $K_{expansion,p-cs}$. Unfortunately, because of the length-dependent feature of the model, as the chain grows longer, $K_{expansion,p-cs}$ decreases because of the $\exp(N^{-2})$ dependence. This an unphysical result because one would expect $K_{expansion,p-cs}$ to increase as the sensitivity to chemical gets stronger, as reflected in the magnitude of the length-corrected m -value. As a result the corrected m -values and exchange coefficients have a negative correlation (Figure 3.22). The length dependence of the exchange coefficients is not trivial to correct for. Even so, the correlation is poor with a correlation coefficient of 0.18. This particular comparison would require further analysis where the hydrophobic content of the proteins are measured from amino acid transfer free energy data, as well as a

reasonable correction which accounts for the chain length dependence of exchange coefficients is made.

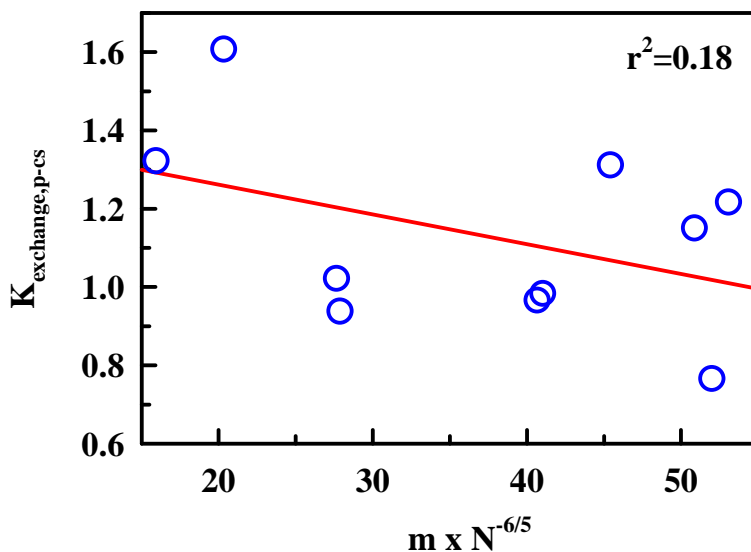


Figure 3.22 Correlation between experimental protein folding m -values corrected for chain length dependence and exchange coefficients from smFRET data on the unfolded ensemble for protein residue's preference for cosolvent over water.

3.7 Conclusions

Proteins unfold due to their interaction with chemical denaturants. The interaction that drives the unfolding process has been analyzed by several experimental and theoretical studies that point to the direct interaction of these denaturants with the peptide backbone and residues through hydrogen bonding. Thus predominantly electrostatic interactions on the protein are perturbed by the denaturing molecules. It has also been shown in various studies that GdmCl and urea do not perturb interactions between nonpolar residues⁴²⁻⁴³ and therefore, do not perturb the structure of water significantly, except at high concentrations. Thus, the weakened

hydrophobic effect in the presence of denaturants may not be a dominant effect for producing chain unfolding⁴³.

We have developed an analytical theory as an extension of the Flory-Huggins theory of polymer solution thermodynamics to incorporate an additional third component in solutions, i.e. the cosolvent or added solute, in addition to the solvent and the protein, to analyze the effects of denaturants on the size of unfolded proteins. The model makes no *a priori* assumption except for the Bragg-Williams mean-field approximation. Thereby, no comment can be made about the number of bound molecules to the unfolded protein, since the model does not explicitly incorporate this number, or even consider explicit binding events. The absence of the number of bound molecules to the protein in our analysis would be equivalent to the interaction of one covalent molecule per residue or lattice site, with the thermodynamics of the interaction being governed by the probability of occurrence of a cosolvent molecule in the solution, i.e. its concentration. This may not be such a disadvantage since, a thermodynamic analysis of denaturant effects on chain size by Haran and coworkers yielded a similar result^{112,137}. The other feature, which is the treatment of interaction between moieties on a thermodynamic basis, without the explicit incorporation of a binding event is in fact an advantage, since aqueous solutions of denaturing molecules such as GdmCl and urea show significant deviations from ideality and there is an explicit difference between binding and preferential interaction, both of which drive unfolding reactions as described in several papers by Schellman,. The model predicts a quadratic dependence of interaction energies of a protein residue on the volume fraction or mole fraction of the cosolvent, with the curvature of this

dependence coming entirely from cosolvent-solvent interactions alone. The model puts observations of chain collapse and expansion on a thermodynamic footing, where good mixed solvents result in favourable net thermodynamic interactions of the protein with the solution.

Our results also point to two important effects which bring about protein unfolding. Firstly, the interaction of the cosolvents GdmCl and urea with the proteins is stronger than the interaction of the protein with solvent. Thus, in the presence of these cosolvents, there is a preferential interaction of these molecules with the protein over that of the protein with water, with this thermodynamic ratio of preference, or preferential interaction or exchange coefficient in the range 0.9 to 1.6 depending on the protein. These results are similar to those obtained for preferential concentration independent interactions of these denaturants with the protein based on the studies of Schellman and Record^{86,97}. The second important factor in determining the expansion of proteins by denaturants is the cosolvent-solvent interaction. The model predicts quite clearly, that the non-ideal endothermic nature of the cosolvents GdmCl and urea should, by itself, drive a polymeric chain to expand, even if the chain has an equal preference for the cosolvent or the solvent. Physically, this result implies that because of the poor interaction between the solvent and cosolvent, as is the case for endothermic solutes, the protein chain experiences an increasingly favourable environment for residue exposure, i.e. a weakened hydrophobic effect with increasing chemical concentration. This result is clearly seen in the case of proteins RNase H, Staph. Nuclease and Barstar's interactions with GdmCl, where the relative interaction of the protein with the solvent is stronger, but the preference of the cosolvent for the

protein's residues over water, thermodynamically drives the unfolded chain's expansion. In general the protein chain is seen to have a relatively stronger preference for urea than for GdmCl as observed from the Barstar and SH3 data on urea and GdmCl, where the $K_{exchange,p-cs}$ is marginally higher for urea than for GdmCl. In fact, in the case of Barstar, this effect has been documented to induce a collapse of the chain in the presence of GdmCl, but not in the presence of urea. This stronger preference for urea may emerge from the polar nature of urea molecules. However the more pronounced unfolding of in the presence of GdmCl, as is seen from lower C_m values is due to the stronger relative preference of GdmCl for the peptide over water, when compared with urea. This is again corroborated by the fact that among the two GdmCl is the more endothermic solute, and it has weaker interactions with water.

In summary, smFRET experiments have unequivocally demonstrated the expansion of the unfolded chain in the presence of denaturants at increasing concentrations. This effect is universal in proteins seen so far, except for cases such as Barstar or Protein L where a high degree of charged or polar amino acids on the chain, or the presence of structure in the unfolded ensemble, could perturb the purely polymeric dependence of chain properties on chemical denaturants and cosolvents. GdmCl being an electrolyte is especially prone to stabilizing structures with high degrees of charged interactions, thereby bringing about collapse. Although proteins in high denaturant concentrations have properties similar to those of a random coil, where $R_g \sim N^{3/5}$, there have been studies to show that denaturants do not produce unfolded state expansion by X-ray scattering measurements⁹¹. However, the use of small angle X-ray scattering measurements to determine the size of a random coil

may not be the ideal technique for assessing R_g (as in the study by Kohn and coworkers), due to possible errors coming from the high protein concentration needed for the experiment, which could interfere with accurate size measurements¹¹⁵. The stage is now set for further single molecule unfolding measurements on proteins with varying sequence lengths, and simultaneously on proteins with similar lengths to address the issue of chain length dependence of some of the exchange coefficients obtained from our analysis. In the interim, a correction needs to be incorporated in the model to account for this effect. Further, a comparison between microscopic energies of interactions between amino acid residues and mixed solvents and the thermodynamics of the transfer of amino acid groups from water to mixed solvents is in order. This is ongoing work, and would allow a direct comparison of the extracted pair-wise interaction parameters from our model and free energy estimates from transfer experiments. Temperature dependence of the expansion and collapse of the unfolded state in the presence of denaturants can also be analyzed with our model. In a study by Schuler and coworkers, unfolded state collapse and expansion was observed under varying conditions of temperature, on the cold shock protein from *Thermatoga maritima*, by smFRET. Such measurements in combination with our theory would further explain the thermodynamics of protein denaturation by chaotropic cosolvents, and the polymeric properties of unfolded chains. These studies would contribute immensely to a deeper characterization of the unfolded ensemble enabling the possibility of improving computational simulations on protein folding reactions.

Chapter 4. The Effects of Chaotropes and Kosmotropes: BBL as a Case Study

4.1 Introduction

While the previous chapter focused on the effect of the chaotropes GdmCl and urea on protein conformations, and specifically the unfolded ensemble, this chapter attempts to determine the effect that chaotropes and kosmotropes would have on protein structure in acid-denatured BBL. A formal definition of the terms ‘kosmotrope’ and ‘chaotrope’ is in order. Kosmotropes, or kosmotropic solutes are those, which upon dissolution in water, bring about a decrease in the entropy of the first shell of dissolved water molecules surrounding the ions or the non-electrolytic solute⁶⁰⁻⁶¹, compared to the bulk solvent. Neutral salts with ions of high charge density such as Li^+ or F^- appear in the category of kosmotropes. Thus kosmotropic solutes, or kosmotropic cosolutes or cosolvents (terms used interchangeably in this thesis) are ‘structure-makers’. Chaotropic solutes on the other hand bring about a relative increase in the entropy of the first shell of water molecules surrounding the solute, when compared to the bulk solvent. Thus, they are characterized as water structure-breakers, in at least the first shell of water molecules. These definitions, whereby dissolved cosolvents are split into two types or groups, and their relative ranking or ability to change the stability of biopolymers, stems from the classic paper by Hofmeister (published in German, in the Archives of Experimental Pathology and Pharmacology in 1888) on the ability of dissolved neutral salts to precipitate proteins out of their aqueous solutions *in vitro*, i.e. the salting-out of a protein. The salts and

their component ions seemed to follow a rank-ordering, whereby there were large differences in the minimum concentrations of salt required to salt-out proteins, as measured by their precipitation cloud-points. This rank-ordering of salts and ionic species was termed the Hofmeister effect, whereby a correlation was observed between the cloud-point and ionic properties of the solute or constituent ions. Small ions with high charge densities or ions with multiple charges such as SO_4^{2-} or Li^+ salt-out proteins the earliest, where as larger ions such as Br^- and Rb^+ do so at higher concentrations. Indeed low concentrations of the latter group salted-in proteins whereby between 0.1 and 1 M, they increased the solubility of proteins in water. This series rank-ordering has been observed time and again on several other kinds of experimental observables, such as the relative viscosities of solutions, surface tension measurements, solute enthalpies of hydration, change in the melting temperature of protein unfolding, and the order in which these ions elute out of Sephadex columns⁶¹. Kosmotropes have relatively higher heats of hydration (they are exothermic), are strongly solvated by water molecules, and hydrogen-bonding at the solute-water interface drives the relative entropy reduction over bulk water.

The scenario is the opposite for chaotropes, where several potential hydrogen bonds are compromised at the ion-water interface, leaving the water molecules free to rotate or translate, resulting in higher entropies relative to the bulk. Kosmotropes stabilize proteins (in terms of the stability of the native structure relative to other ensembles) due to their ability to bind water strongly, which excludes them from the protein surface¹³⁸⁻¹³⁹. This forces the protein structure to minimize exposed surface area in the presence of what is effectively a poorer solvent; one where the

hydrophobic interaction becomes stronger, bringing about compaction and structure formation. It is this increase in hydrophobic interaction and drive to minimize exposed surface area that promotes protein-protein association, and eventually precipitation or salting-out. Chaotropes on the other hand, destabilize protein structures relative to kosmotropes, allow for a greater degree of exposure to mixed-solvents bringing about a decrease in the hydrophobic effect. They salt-out proteins at much higher concentrations than kosmotropes. They are characterized by lower heats of hydration compared to chaotropes, and in the case of GdmCl and urea, their solutions are endothermic. They are not excluded from the first hydration shell of proteins to the same extent as kosmotropes allowing for a greater degree of occupation of the volume surrounding protein residues, often resulting in binding to the backbone or amino acid residues. It is important to mention that apart from this solvent-driven effect, often direct interaction with charged residues, and a Debye-Hückel screening of electrostatic interactions also affects the extent and the manner in which electrolytic solutes or polar non-ionic solutes affect protein stability.

The relevance to the present study comes from this very ability to affect protein stability and modulate interactions of the protein with its mixed-solvent, and the affect solutes have on intra-protein interactions. The following critical questions can then be raised about the effect of kosmotropes on proteins: How would dissolved solutes which are kosmotropic relative to GdmCl and urea, affect protein size distributions in the unfolded ensemble? Would a certain degree of compaction be observed in the unfolded ensemble in the presence of salts such as NaCl? Further, would any degree of structure formation, i.e., refolding, be observed in the presence

of such salts due to the tuning of the hydrophobic effect and electrostatic interactions? If so, how would these refolded ensembles compare with the natively folded state in terms of stability, degree of structure formation and melting temperature? The questions may be immediately addressed with the protein BBL, which is acid-denatured at pH 3.0^{13,65}. The sections in this chapter are organized in the following order. Section 4.2 reviews the folding properties of the protein BBL and touches upon its pH sensitivity. Section 4.3 describes bulk FRET experiments on unfolded BBL at pH 3.0 in the presence of urea, GdmCl and NaCl as cosolvents, and discusses their effects on protein size. Section 4.4 then describes a series of CD experiments on acid-denatured BBL in the presence of these cosolvents to assess the possibility of salt refolding or chain compaction. It also discusses the effect of other solutes such as LiCl and CsCl on unfolded BBL at pH 3.0, and the relative differences between the effect of these solutes, which are different in terms of the size of the cation involved, and also their relative solubilities. Section 4.5 delves further into high-resolution NMR analyses of the ensembles observed in the presence of salt. Section 4.6 describes IR T-jump experiments to study the dynamics of chain reconfiguration in the presence of salts, and describes the combined effect that pH and salt concentration have on BBL reconfiguration, resulting in a critical reassessment of the manifestations of BBL's pH sensitivity on its structure. Section 4.7 summarizes these results in the overall context of BBL as a globally downhill folding protein.

4.2 BBL: A Globally Downhill Folding Protein

4.2.1 Evidence for Global Downhill Folding

BBL (Protein Databank ID 1BBL) is a 40 residue variant of a 51 residue peptide which comprises the peripheral subunit binding domain (dihydrolipoamide dehydrogenase (E3) binding domain), and is part of the dihydrolipoamide succinyltransferase (E2) core, of the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli*¹⁴⁰. 2-oxoglutarate dehydrogenase is responsible for oxidative decarboxylation of 2-oxoglutarate resulting in succinyl-CoA as a product along with NADH. This multienzyme, which has 3 different subunits, E1, E2 and E3, is involved in the citric-acid cycle in an energy producing step during glycolysis. The E2 subunit shuttles a covalently bound succinyl substrate (arising from 2-oxoglutarate decarboxylation, by the thiaminodiphosphate group on the 2-oxoglutarate decarboxylase- E1 subunit), between the E1 subunit and the E3 subunit (the dihydrolipoamide dehydrogenase).

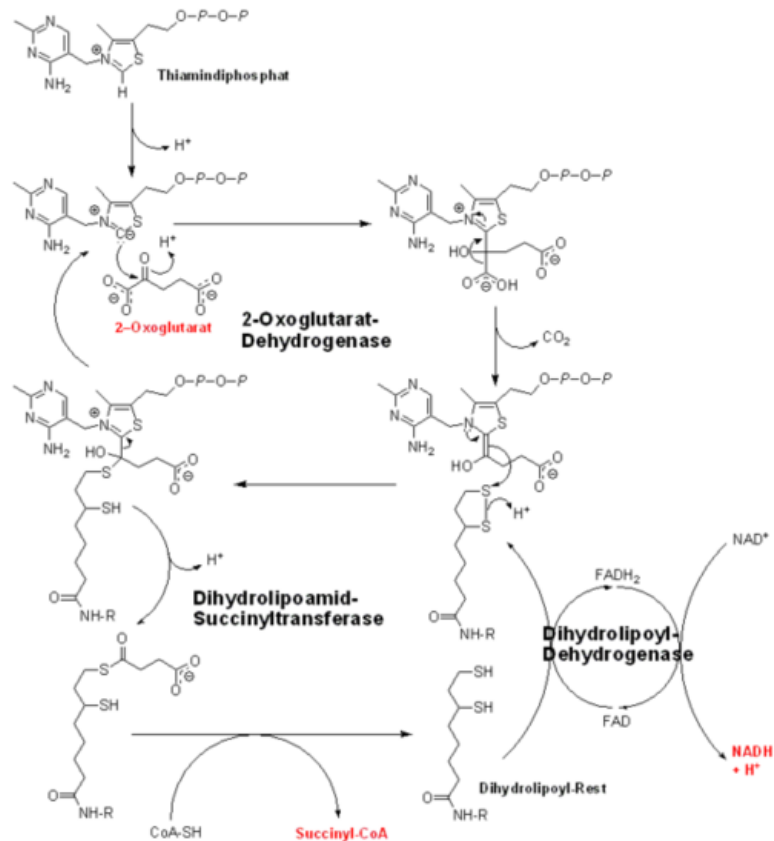


Figure 4.1 Enzymatic reactions carried out by the 2-oxoglutarate dehydrogenase multienzyme complex. (Source: Wikipedia).

The E1 subunit decarboxylates the 2-oxoglutarate group, which is then transferred to a ~14 Å lipoylated lysine residue on the 80 residue N-terminal domain of E2. This is called the ‘swinging arm’ which shuttles between the E1 domain and the dihydrolipoamide succinyl transferase domain at the 300 residue long C-terminus of the E2 subunit. It catalyzes the transfer of the succinyl group out of the lipoylated arm to produce succinyl-CoA. Finally, the E3 subunit, which is the dihydrolipoamide dehydrogenase, regenerates the lipoyl group on the N-terminus of the E2 subunit. Between the N and C-terminal domains of the E2 subunit is a small and folded peripheral subunit binding domain which contains a site that promotes binding of the E2 subunit to the E3 subunit. This binding domain on E2 is flanked by flexible linkers

which are important for the shuttling function of the E2 swinging arm. Overall the 2-oxoglutarate dehydrogenase complex contains 24 copies of the E2 subunit as an octahedral core, around which homodimers of E1 and E3 aggregate peripherally, with at least 6 homodimers of each being required for retaining full enzymatic activity¹⁴⁰.

The 40 residue variant of the 51 residue subunit binding domain, termed BBL, has been identified as a globally downhill folding protein with a single population existing in all thermodynamic conditions, varying in the quantity of native structure. There is no barrier to folding in any thermodynamic condition, and this finding was the first experimental identification of a downhill folder, the existence of which was predicted by the analytical funnel landscape theory of Wolynes and coworkers as the Type 0 scenario⁹ of protein folding. The population shifts gradually from fully native to fully unfolded by progressive shortening of the helices and gradual melting of structure^{13,141}. The thermal melting of tertiary structure, the unraveling of secondary structure and melting of the helices are thermodynamically decoupled, indicating a temperature unfolding process with a low degree of cooperativity. This low thermodynamic cooperativity results in decoupled unfolding events, broad thermal transitions as studied by low resolution techniques such as CD and differential scanning calorimetry, changes in signal in the folded baseline, and a distribution of melting temperatures (T_m 's) spanning 60 K. These results together rule out the possibility of an all-or-none transition. Analysis of the thermal unfolding data with a structure-based statistical mechanical approach developed by Muñoz and Eaton¹⁴², yielded no barrier heights to folding under any condition- a globally downhill folding scenario. This finding has been confirmed by a wide range of techniques over several

reports^{12,14,143-146} despite the supposed controversial description by Fersht and coworkers of BBL as a two-state folding protein¹⁴⁷⁻¹⁵⁵.

4.2.2 On the pH-Sensitivity of BBL

BBL is acid-denatured at pH 3.0. Two histidines on BBL with pKa's of ~6.13 and 5.37 are responsible for the unfolding of BBL with decreasing pH due to their protonation^{13,65,141,146,152} (unpublished data, Sadqi and Muñoz). A calculation of the net charge on the protein considering a simple scenario of invariant pKa's of the ionizable amino acids on the protein (of which there are 14), indicates that the net charge of BBL should increase from 2.3 at pH 7.0 to 8.5 at pH 3.0, as shown in Figure 4.2. Electrostatic repulsion at lower pH could also be a source for long-range interactions that destabilize the protein. Acid-denatured BBL is completely unfolded as indicated by the CD thermal melts at 222 nm in Figures 4.3 A and B.

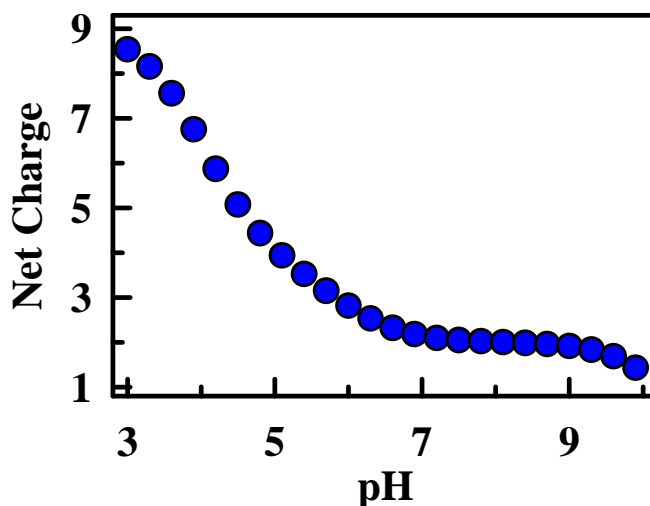


Figure 4.2 Theoretical net charge on BBL as a function of pH.

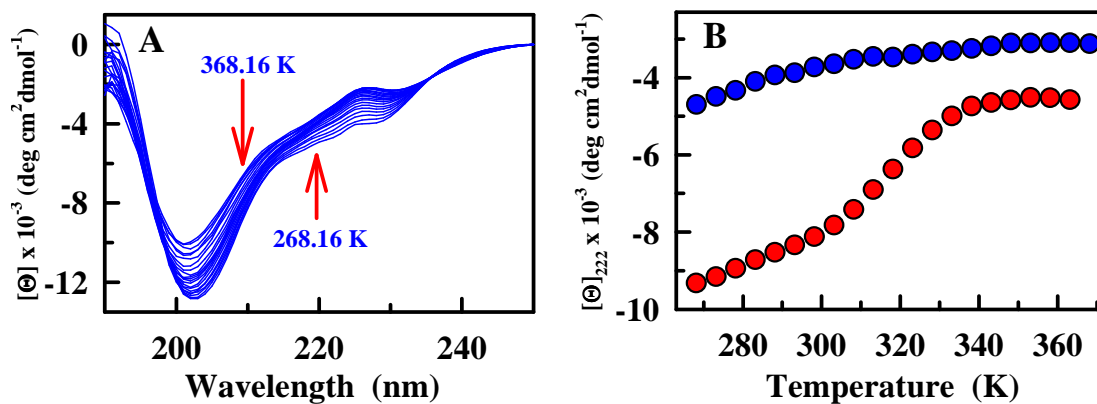


Figure 4.3 CD melts of Naf-BBL at pH 3.0 and pH 7.0. A) Lines indicate molar residual ellipticities measured between 268.16 and 368.16 K. B) CD signal at 222 nm at pH 7.0 (red circles) and pH 3.0 blue circles in 20 mM phosphate and citrate buffers, respectively

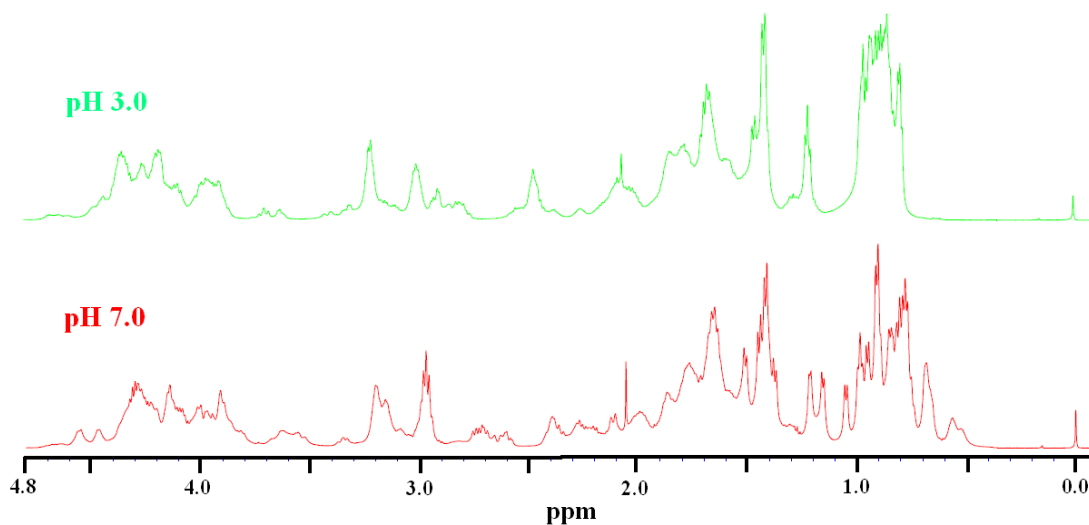


Figure 4.4 NMR 1D proton chemical shifts in the aliphatic region for folded Naf-BBL at pH 7.0, and acid-denatured at pH 3.0, at 283.16 K

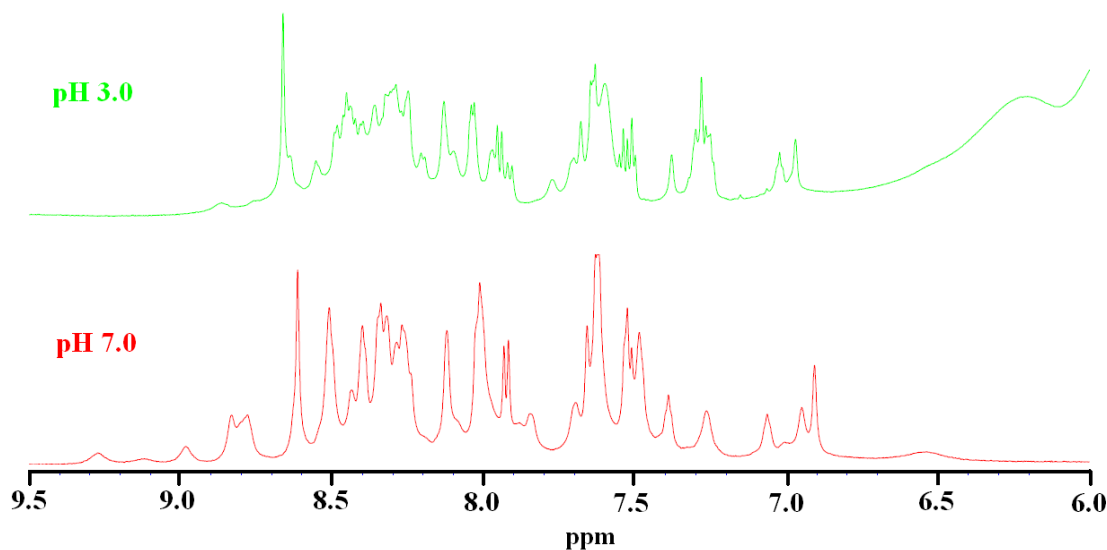


Figure 4.5 NMR 1D proton chemical shifts in the amide region for folded Naf-BBL at pH 7.0, and acid-denatured at pH 3.0, at 283.16 K

Figure 4.3 B shows the CD signal at 222 nm which indicates a loss of helical structure with increasing temperature. There is still some signal at pH 7.0 at 360 K, indicating some residual structure which disappears as the pH decreases. NMR 1D chemical shifts of the methyl protons compared at pH 3.0 and 7.0 in Figures 4.4 at 283.16 K show the disappearance of signals from the methyl protons in addition to line broadening. This result indicates loss of structure, as does Figure 4.5, with chemical shifts from the amide protons. Figures 4.6 and 4.7 are the NMR SOFAST HSQC spectra with ^{15}N natural abundance, for native Naf-BBL at pH 7.0 at 283.16 K and acid-denatured Naf-BBL at pH 3.0 at 283.16 K, respectively. The peaks cluster between 8 and 8.5 ppm in acidic conditions indicating a loss of structure in the protein.

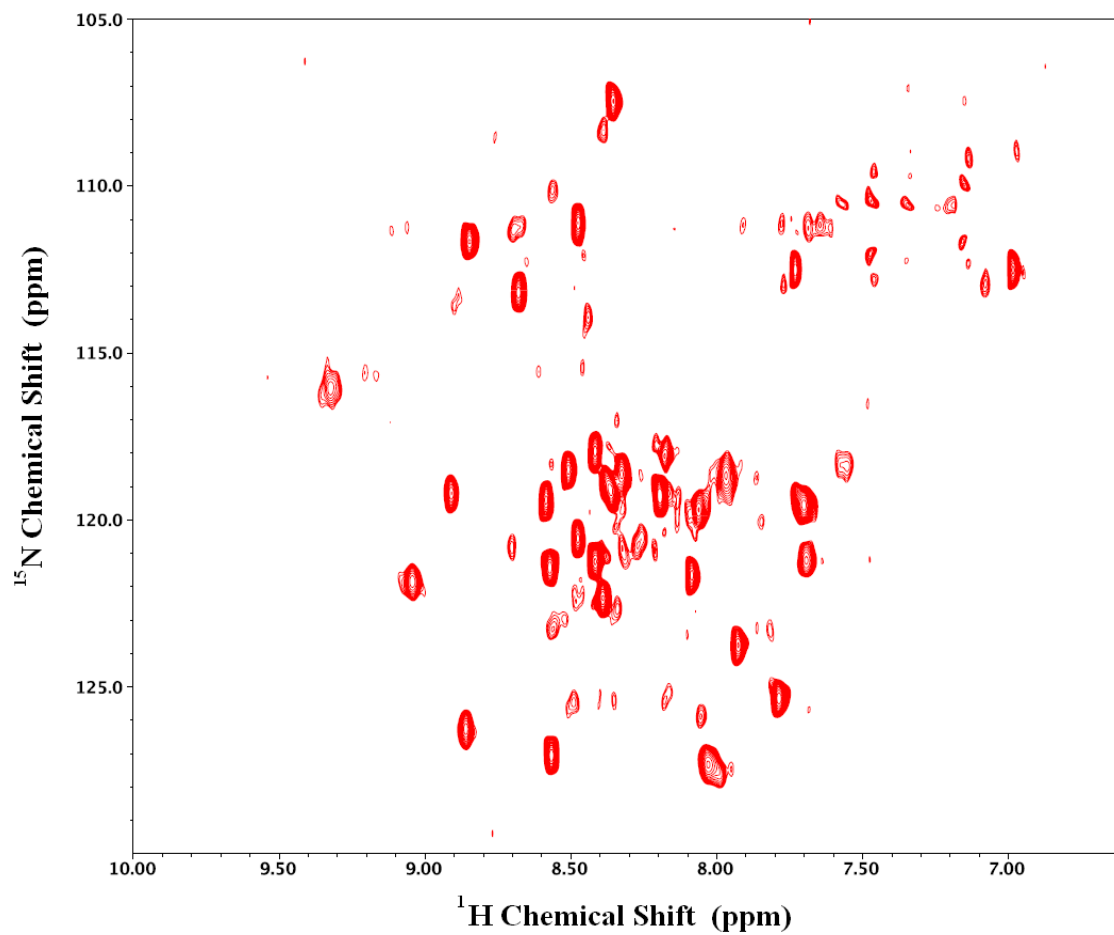


Figure 4.6 NMR SOFAST HSQC spectra of Naf-BBL with ^{15}N natural abundance at pH 7.0 and 283.16 K.

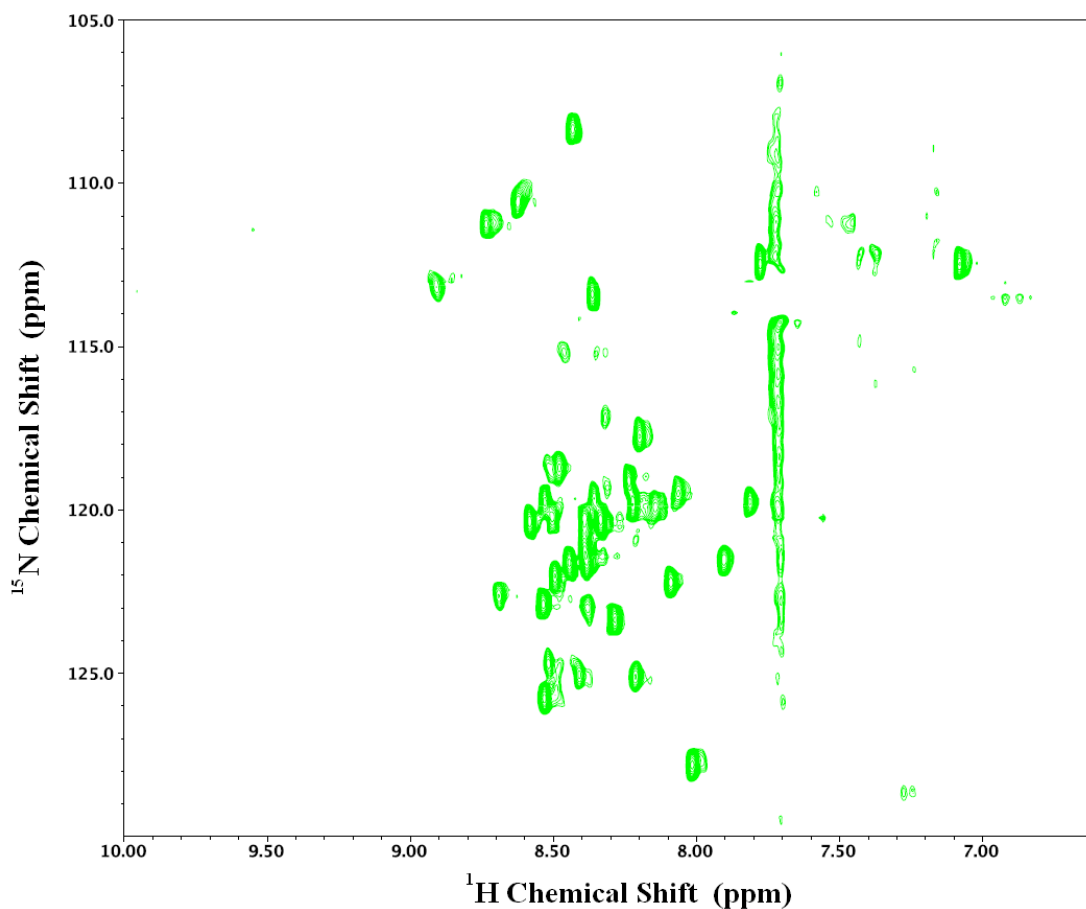


Figure 4.7 NMR SOFAST HSQC spectra of Naf-BBL with ^{15}N natural abundance under acid-denaturation at pH 3.0 and 283.16 K.

The chemical shifts at pH 3.0 are invariant with temperature⁶⁵, and thus no structural evolution takes place with changing temperature in the acid-denatured ensemble, thus ruling out the possibility of the existence of some acid-denatured molten-globule like state at low pH.

4.3 Expansion and Collapse of Acid-Denatured BBL

BBL unfolds by acid denaturation and is completely unfolded at pH 3.0. This sensitivity to pH allows for the possibility of tuning native structure in the protein

without changing solution conditions by the addition of chaotropes or by changing the temperature. The latter two can in fact serve as orthogonal thermodynamic variables in unfolding experiment. This scenario is the unfolded state equivalent of double-perturbation experiments described by Oliva and Muñoz¹⁵⁶. Furthermore, in the context of the unfolded ensemble, this property of pH-based denaturation can be exploited to study the effect of unfolding agents such as GdmCl and urea, and the effect of temperature on the *unfolded ensemble alone* in bulk solutions. Such an experiment would typically be hard to perform in bulk due to the competing processes of protein unfolding with changes in GdmCl or urea concentrations or changes in temperature, and the existence of ensemble-averaged signals coming from an entire slew of coexisting populations in bulk. A technique such smFRET confocal microscopy allows unprecedented access to the observation of the properties of the unfolded ensemble. However, even in such studies, the unfolded ensemble appears to have an accurately determinable and sizeable population above the background only in moderate to large concentrations of denaturant. Thus, the acid-denatured BBL system could be useful in the study of chain expansion and collapse in varying conditions by FRET experiments in bulk. Such experiments could be performed on the doubly-labeled Naf-BBL-Dan variant of BBL, previously described in the work identifying BBL as a downhill folder, and would report on the size of the protein¹³.

4.3.1 Hydrophobic Collapse of Acid-Denatured BBL

Protein thermal unfolding occurs because the free energy of unfolding is curved with respect to temperature. This curvature arises from slight imbalances in the enthalpic-entropic compensation when they change with temperature^{12,145}. In a structural sense, temperature unfolds proteins because of the dependence of its stabilizing non-covalent interactions such as the hydrophobic effect and hydrogen-bonding on temperature. High temperature perturbs these interactions and an increase in entropy drives the unfolding process. However, how does the unfolded ensemble itself respond to temperature changes? For large polymers in water, a temperature increase is accompanied by a large increase in the radius of gyration of the polymer—the globule to coil transition¹⁵⁷. Temperature is also known to cause collapse in the cold-shock protein and the intrinsically disordered prothymosin A¹⁵⁸. Such a process of contraction in the average size of the unfolded ensemble upon change in thermodynamic conditions, without the formation of native structure, has been termed collapse.

Equilibrium FRET studies on unfolded BBL, using the doubly-labeled Naf-BBL-Dan variant indicated a decrease in FRET efficiency with increasing temperature^{13,65}, which are reproduced here in experiments performed by the author in Figure 4.8 A. Root mean-squared end-to-end distances ($\langle R^2 \rangle^{1/2}$) and radii of gyration are calculated based on the Gaussian chain assumption and plotted in Figure 4.8 B, after incorporating R_0 corrections (Figures 4.9 and 4.10).

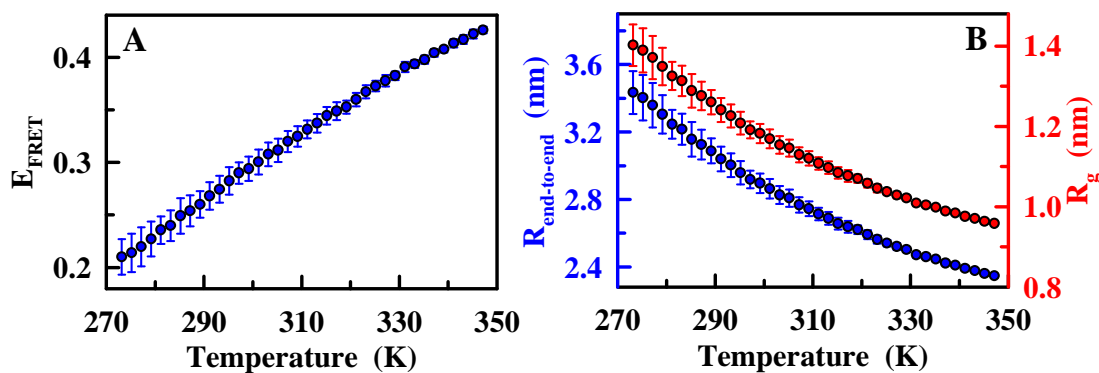


Figure 4.8 Temperature collapse of acid-denatured BBL. A) FRET efficiencies measured by comparison of naphthyl (donor) quantum yield in the presence and absence of dansyl (acceptor). Errorbars represent 3 repeats. B) Root mean-squared end-to-end distance (blue) and R_g (red) calculated based on Gaussian chain model.

Unfolded BBL experiences a collapse with increasing temperature due to the hydrophobic effect^{3,65} whereby, with increasing temperature the free energy of transfer of hydrophobic or aliphatic groups on the chain, from water to a sequestered environment decreases (becomes more favourable)¹⁵⁹⁻¹⁶⁰. This collapse with increasing temperature is akin to hydrophobic collapse seen in proteins, upon their transfer from concentrated to dilute solutions of denaturants, an observation now confirmed by smFRET experiments^{36,161}. Hydrophobic collapse in BBL was found to be diffusive and in the range of $\sim 1/(100 \text{ ns})$ by FRET T-jump experiments⁶⁵. The rate of collapse increases with increases temperature, following the decreases in solvent viscosity, and then experiences a roll-over around 308 K. Since there is no activation energy for collapse in these conditions, the rollover was posited to have a dynamic rather than a thermodynamic origin. The decrease in rate arises from an increase in the roughness of the landscape at high temperature, when the protein is collapsed and several hydrophobic interactions must be broken to reconfigure the protein chain⁶⁵.

4.3.2 The Effect of Guanidinium Chloride and Urea on Unfolded BBL

The effect of the chaotropic cosolvents GdmCl and urea on unfolded BBL at pH 3.0 was determined by bulk FRET experiments on Naf-BBL-Dan, the doubly-labeled version of BBL. Estimates of the size variation with FRET as the reporter were made based on the careful determination of R_0 as a function of cosolvent concentration and temperature. The variation in R_0 comes about owing to a change in the quantum yield of the donor with temperature and GdmCl, an increase in the overlap integral with temperature, and an increase in refractive indices of cosolvent aqueous solutions with cosolvent concentration. Naphthyl-alanine quantum yield changes with increasing GdmCl concentration but not with urea (Figure 4.10).

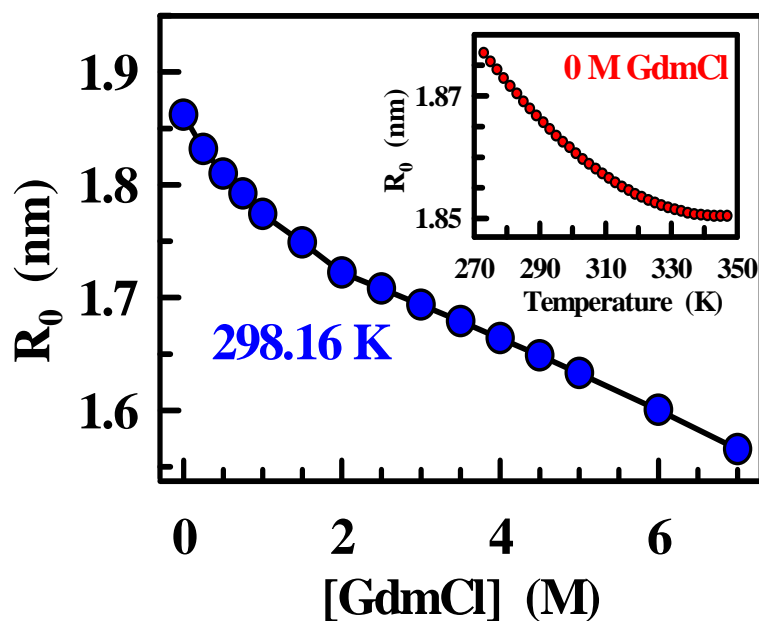


Figure 4.9 Förster radius of FRET pair. A) Variation in R_0 with [GdmCl] at 298.16 K. (Inset) Variation with temperature at 0 M GdmCl.

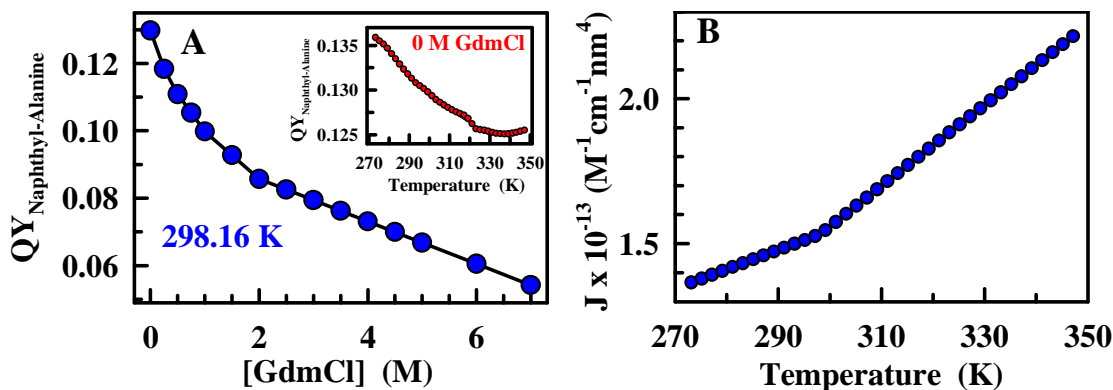


Figure 4.10 A) Quantum yield of Naphthyl-Alanine (donor) as a function of [GdmCl] at 298.16 K. Line to guide the eye. (Inset) Donor QY changes with temperature at 0 M GdmCl. B) Variation in overlap integral of the FRET pair with temperature.

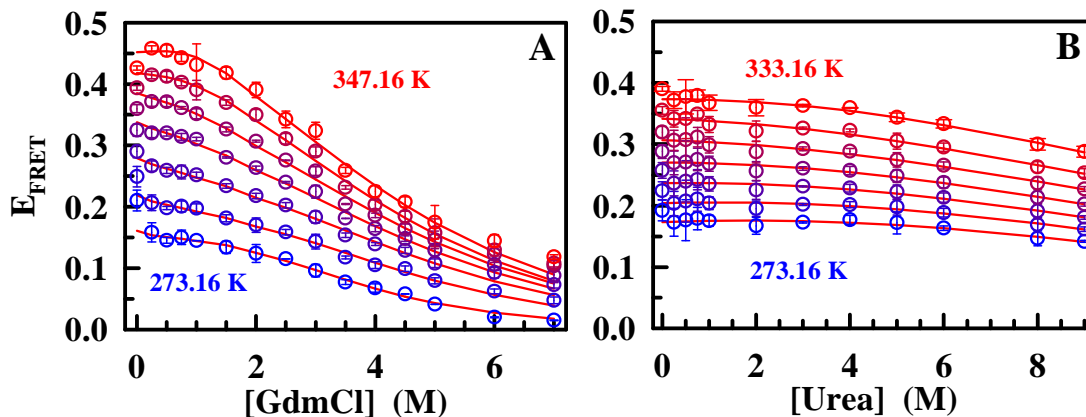


Figure 4.11 Expansion of unfolded BBL with cosolvent at pH 3.0. Circles indicate experimental data with error bars from 3 repeats. Lines indicate fits with the 3-component interaction model to guide the eye. A) GdmCl expansion and fits for data from 0.25 M to 5 M. B) Urea expansion and fits for data from 0 M to 9 M.

Unfolded BBL at pH 3.0 undergoes an expansion as seen from the decreasing FRET efficiency, with increasing GdmCl and urea concentration in Figure 4.11 A and B. FRET efficiency decreases from 15% to ~0 at 273.16 K, corresponding to an increase in R_g from ~1.5 nm to ~2.4 nm at 5 M, and from 50% to 10% at 347.16 K, corresponding to an expansion from ~1.0 nm to ~1.5 nm (Figure 4.12 A). The expansion in the presence of urea is less prominent as seen from the FRET efficiency decrease in Figure 4.11 B and from the R_g in Figure 4.12 B, where the R_g increases by

approximately 0.15 nm at 2743.16 K and by 0.1 nm at 333.16 K from 0 M to 9 M urea. The property of hydrophobic collapse with increasing temperature is preserved through the entire concentration range of either chemical. However, the degree of collapse decreases with increasing GdmCl, but remains the same in relative terms, with increasing urea.

The expansion with chemical is not monotonous, as a compaction occurs upon transfer from 0 M to 0.25 M GdmCl, as seen in Figure 4.12 A. This is unexpected based on a purely chaotropic effect of GdmCl, since the curvature of FRET versus chemical should necessarily be convex as discussed in the previous chapter. A fit of the energies as plotted in Figure 4.13, with 3 freely floating parameters, i.e. $\Delta\epsilon_{cs}$, $\Delta\epsilon_{ps}$ and $\Delta\epsilon_{pc}$, to serve as visual aids, clearly yield negative curvature (and negative $\Delta\epsilon_{cs}$'s) contrary to the positive energies expected from the positive deviations from ideality of aqueous solutions of GdmCl. The increase in R_g with urea is almost linear, indicating a critical difference in the strengths of interaction of the two chemical as seen from the higher degree of expansion in the presence of GdmCl, and a lower degree of hydrophobic collapse with increasing temperature.

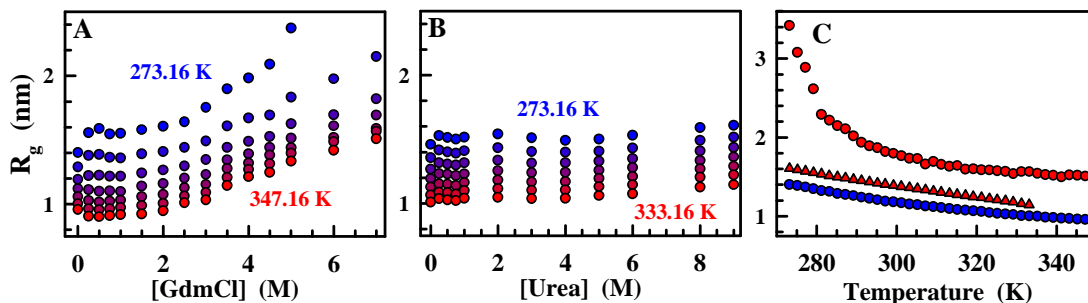


Figure 4.12 Expansion and Collapse of unfolded BBL with chaotrope addition and temperature change at pH 3.0. A) Circles indicate R_g calculated from FRET efficiencies by the Gaussian chain model. B) Same as A. C) Blue circles are BBL R_g in 20 mM citrate buffer, red circles in 7 M GdmCl, and red triangles in 9 M urea.

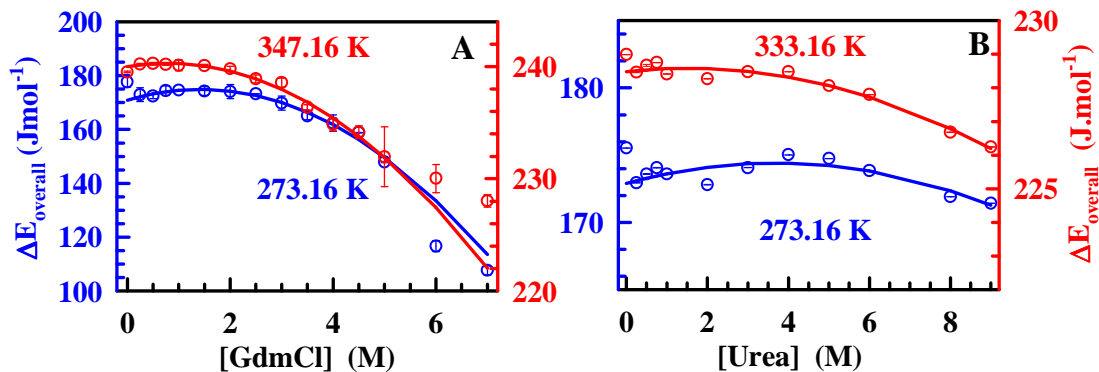


Figure 4.13 BBL residues' net interaction energy ($\Delta\varepsilon$) in different conditions at pH 3.0. A) Circles represent calculated $\Delta\varepsilon$'s, and lines indicate fits of $\Delta\varepsilon$ from 0.25 M to 5 M GdmCl to the 3-component energy equation. B) Same as A, with fits for data from 0.25 M to 9 M urea.

The $\Delta\varepsilon$'s in Figure 4.13 clearly indicate opposite curvature with chemical at all experimental temperatures, to what is expected for GdmCl and urea based on their solution properties. Furthermore in GdmCl, upon going from 0 M to 0.25 M, chain expansion is observed followed by compaction between 0.25 M and 0.75 M, after which chain expansion resumes (blue circles, Figure 4.12 A). The high temperature counterpart of this behaviour is chain compaction from 0 M to 0.75 M, after which the expected chain expansion occurs (red circles, Figure 4.12 A). This peculiar rollover at low GdmCl concentrations, which is not observed in the presence of urea, could be attributed to the fact that GdmCl is an electrolyte and its ionization in water shields the electrostatic repulsions on the chain arising from the net charge of ~ 8.5 at pH 3.0 (Figure 4.2). This screening effect on a random polymer can be determined by weighting the Gaussian distribution in Equation 3.1 with the electrostatic potential in the presence of screening ions, i.e., the Poisson screening equation:

$$V_{\text{electrostatic}} = 1.3865 \times 10^5 \cdot \frac{\exp(-\kappa r)}{\epsilon r} \quad (4.1)$$

with κ being the screening constant:

$$\kappa = \sqrt{\frac{2 \cdot e^2 \cdot N_A^2 \cdot 10^3 \cdot I}{8.85 \times 10^{-12} \cdot \epsilon \cdot R \cdot T}} \quad (4.2)$$

where e is the elementary electric charge of 1.6×10^{-19} C, N_A is the Avogadro number, I is the ionic strength of the solute, ϵ is the dielectric constant of the medium, R is the universal gas constant, T the temperature, and r in Equation 4.1, the distance separating two charged particles. The net electrostatic potential of the protein calculated for the several charged residues on the chain at pH 3.0, with the distance between any pair of charges being weighted by the Gaussian distribution is plotted below in Figure 4.14. Although the calculation is only approximate with water's dielectric constant used for the calculation, one observes a decrease in the electrostatic interaction on the chain with increasing concentration (ionic strength), and tapers off by a concentration of 2 M.

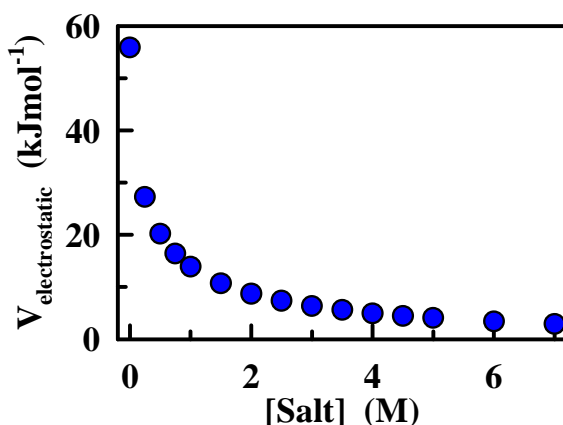


Figure 4.14 Electrostatic potential of a charged 40 residue random polymer, with charge-screening due to dissolved ions.

4.3.3 Charge-Screening Effects on a Chain by a Strong Electrolyte

The charge-screening effect could account for the low concentration rollover in chain size of unfolded BBL in the presence of GdmCl. At low concentrations, the electrolytic properties of GdmCl could dominate over its chaotropic properties. This effect is known to bring about a stabilization of the native state at low concentrations of GdmCl. If so, NaCl should also be able to produce the same effects as GdmCl. Bulk FRET measurements on Naf-BBL-Dan at pH 3.0 in the presence of NaCl as the solute, yielded the efficiencies plotted in Figure 4.15.

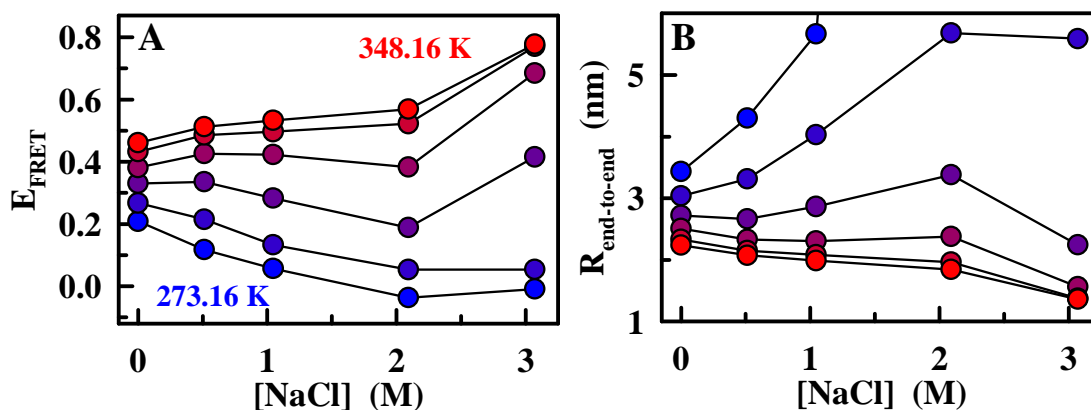


Figure 4.15 Effect of NaCl on unfolded BBL's chain size at pH 3.0. A) Circles represent experimentally determined FRET efficiencies. Lines are to guide the eye. B) R_g calculated by the Gaussian chain assumption. Lines are to guide the eye.

FRET measurements in the presence of NaCl corroborate the rollover observed in low GdmCl concentrations since the FRET efficiency decreases at low temperature. This result qualitatively seems the same as that in the presence of GdmCl. However, in case of GdmCl, the decrease in FRET efficiency at low concentrations was sharp followed by a second phase starting at 1 M GdmCl with modest decreases in FRET thereafter. This corresponded to an increase in R_g at 0.25

M, followed by a compaction up to 1 M GdmCl and then an expansion, indicating the point where the electrostatic screening effect would have saturated, and the chaotropic effect took over. No such biphasic behaviour is observed at low temperature (273.16 K) in the presence of NaCl, indicating a persistence of the same effect present at low concentrations all the way to the highest concentration of 3 M. The FRET efficiency at high temperature also continues increasing indicating chain compaction, and this trend continues through the entire range of concentrations from low to high salt. Electrostatic screening of charged interactions on the protein should bring about chain compaction at all temperatures with increasing concentration of NaCl. However, the trend of compaction is reversed with temperature compared to the chain size variation with temperature in buffer. This is to say, that with increasing salt, the chain seems to have a larger range of sizes to sample. This unexpected result can only be explained by the presence of specific interaction of the salt with the labels (a possibility ruled out by control experiments on the labels), or due to specific effects of NaCl on chain structure. The screening of electrostatic interactions on the chain at pH 3.0 by NaCl, could bring hitherto distant areas on the chain closer into contact, due to the lower energetic penalty for approach, as the electrostatic repulsions are screened. This effect could promote structure formation. The advent of secondary structure in BBL, in the presence of salt at low pH alone can explain the large range of end-to-end distances explored by the protein, so as to bring about expansion at low temperature, and compaction at high temperature in salt. Secondary structure at low temperature could force the two labels apart, due to increasing chain stiffness- an effect that would disappear at high temperature, when the salt-refolded protein has

been unfolded by the increase in temperature. This hypothesis can only be ascertained by further spectroscopic studies on BBL at pH 3.0 in the presence of salt.

4.4 Refolding of Acid-Denatured BBL by Salt

4.4.1 Secondary Structure Formation in Unfolded BBL

Near-UV CD spectroscopic measurements on the singly-labeled variant Naf-BBL at pH 3.0 in the presence of NaCl indicate secondary structure formation. A representative series of spectra at 3 M NaCl in Figure 4.16 A show a trademark decrease in the peak at ~222 nm, the wavelength for the amide bond $n\pi^*$ transition. This effect weakens with loss of α -helical structure, where the peak increases in value indicated helix melting with increasing temperature. Singular value decomposition on the spectra produces three components, the first indicating the average signal, and the second indicating the anti-correlation between peaks at 225 nm and 235 nm in Figure 4.16 B. These are the signals from α -helix and random coil. Amplitudes of the two components in Figure 4.16 C indicate a decrease for the first component, and an anti-correlation in the change in signals of the peaks in the second component due to the amplitude sign change.

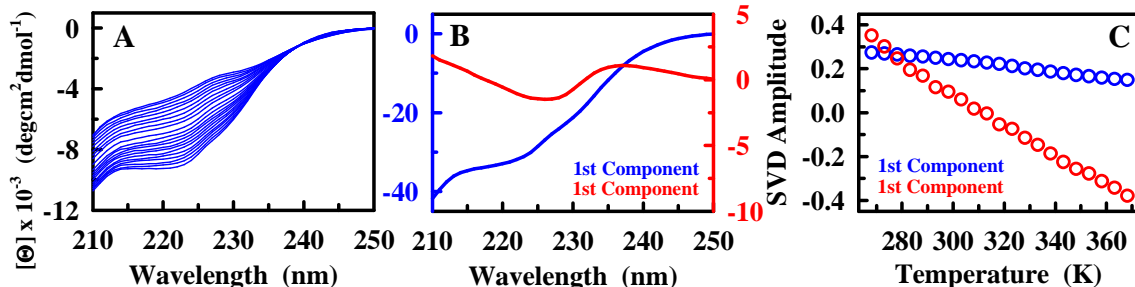


Figure 4.16 Far-UV CD of Naf-BBL at pH 3.0 in 3 M NaCl. A) CD melts of Naf-BBL. B) First two components from SVD analysis, multiplied by their singular values. C) Amplitudes of the first two components.

The third component not shown here contains peaks at 222 nm and 230 nm, which are opposite in sign, again indicating signals from the α -helix and random coil, with amplitudes which are convex with temperature, and have a minimum at 305 K. This component represents the non-concerted manner in which helix length decreases, as this third component is a function the number of helix-coil junctions on the peptide, which increase and then decrease with temperature¹⁴⁶. This simple analysis points to the non-cooperative thermal unfolding of salt-refolded acid-denatured BBL. Figures 4.17, 4.18 and 4.19 plot the helical signal at 222 nm during the thermal unfolding of Naf-BBL in the presence of increasing quantities of NaCl, CsCl and LiCl. These three salts have different cations, but the same anion, and the key set of differences in the behaviour of these salts is the radii of the cations and the solubility of the salts in water. Three different salts were employed to determine the effect that varying cations would have on the refolding of BBL with salt. The upper limit of salt concentration employed to perform CD measurements, was determined by the maximum concentration up to which reversible thermal unfolding could be performed. These were 3 M, 4 M and 6 M for NaCl, CsCl and LiCl, respectively. As

such, it is clear that all three salts induce refolding in acid-denatured BBL at pH 3.0. Their ability to do so stems from the screening of electrostatic repulsions on the unfolded BBL chain at pH 3.0 and also from increases in the hydrophobic effect due to strong interactions with the solvent (water) and exclusion from the hydration shell of the protein. Their ability to induce protein restructuring can be determined by a more detailed analysis of the folded and unfolded protein's mean residual ellipticities at low and high temperatures, respectively. These results are summarized in Figures 4.20 and 4.21.

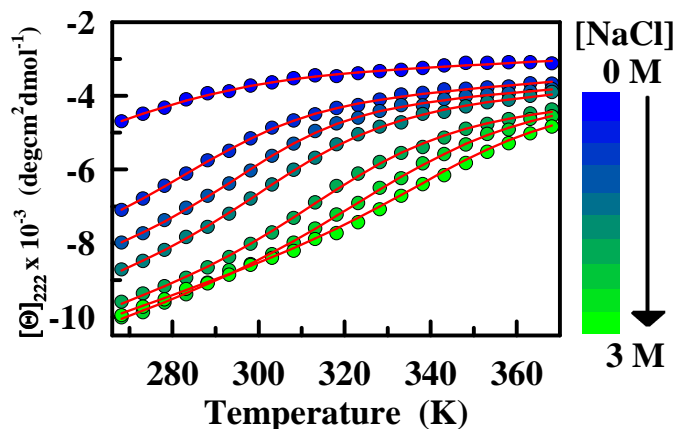


Figure 4.17 Thermal melts of Naf-BBL in the presence of NaCl. Circles and lines represent data and phenomenological two-state fits, respectively.

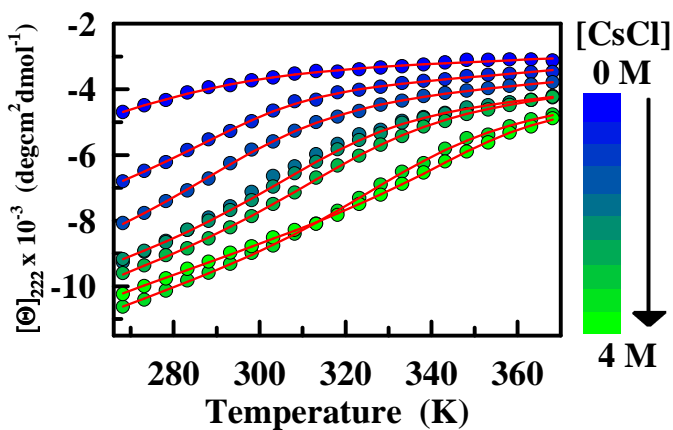


Figure 4.18 Thermal melts of Naf-BBL in the presence of CsCl. Circles and lines represent data and phenomenological two-state fits, respectively.

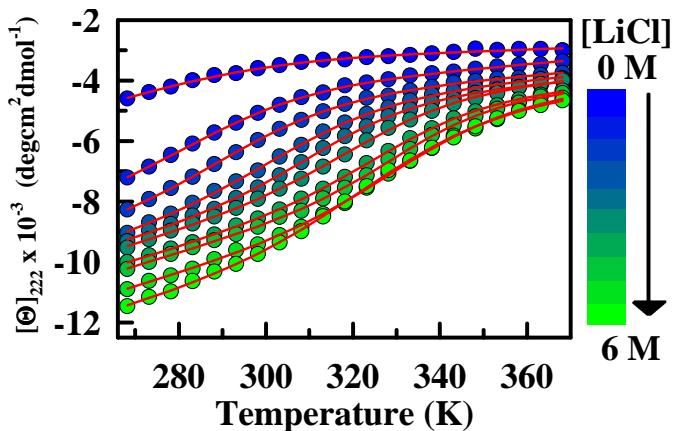


Figure 4.19 Thermal melts of Naf-BBL in the presence of LiCl. Circles and lines represent data and phenomenological two-state fits, respectively.

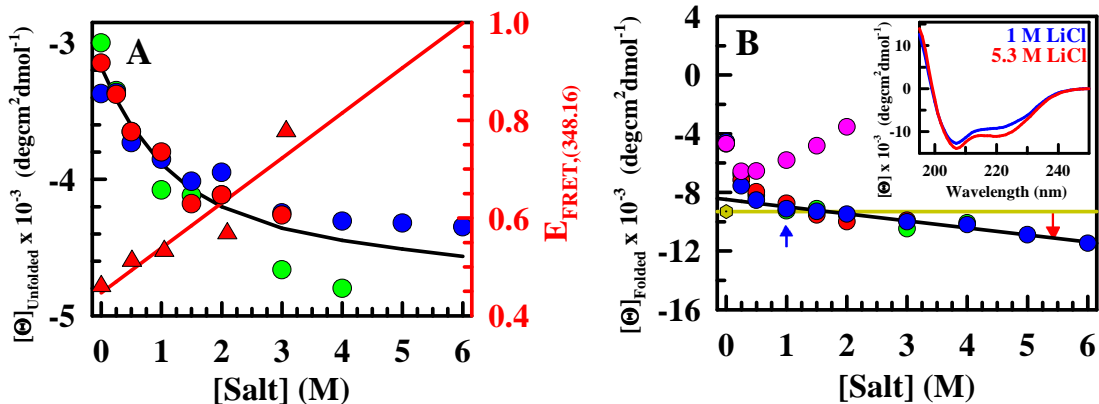


Figure 4.20 The unfolded and folded baselines of BBL in the presence of salt. A) Unfolded baseline of BBL in the presence of salt at pH 3.0 on the left axis. Mean residual ellipticity at 368.16 K determined from the phenomenological two-state fits of the signal at 222 nm. Red, green and blue circles indicate NaCl, CsCl and LiCl, respectively. Black line is a polynomial fit to guide the eye. Red triangles are bulk FRET efficiencies of Naf-BBL-Dan on the right axis at pH 3.0 and 348.16 K with NaCl. Red line is to guide the eye. B) Folded baseline of BBL in the presence of salt at pH 3.0 and 268.16 K obtained from phenomenological two-state fits. Circle colour scheme same as in A. Pink circles are low temperature CD signals in the presence of GdmCl. (Inset) CD spectra of Naf-BBL at 1 and 5.3 M LiCl at 268.16 K., indicated by blue and red arrows, respectively, in the main figure. The brown line is CD signal at pH 7.0 with no salt.

Figure 4.20 A plots the unfolded ‘baseline’ signal at high temperature, when the protein is unfolded. The change in signal is biphasic with a sharp drop in mean residual ellipticity between 0 and 2 M for all salts, followed by a flatter, linear region. The three salts induce structure even in the thermally unfolded ensemble, as can be observed from the decreasing CD signal with increasing salt concentration. The CD signals for all three salts overlay in the entire concentration range, except for CsCl, which seems to have lower signal between 3-4 M than NaCl and LiCl. The FRET efficiency change in Naf-BBL-Dan, previously discussed in Figure 4.15, in the presence of NaCl is directly related to the increase in structure with increasing salt. The existence of some residual structure or stiffness in the thermally unfolded ensemble in the presence of salt drives the FRET donor and acceptor further apart

upon restructuring in salt, causing a decrease in FRET efficiency at low temperature. At high temperature the *increase* in FRET efficiency with salt concentration can be explained based on the rationale that upon thermal melting, chain stiffness decreases and the distance between the FRET pair decreases due to decreased chain stiffness. It is also plausible that in the salt-refolded structure the donor or acceptor specifically interact with certain side-chains (e.g. the C-terminal histidine could quench the dansyl-lysine acceptor), leading to a change in its local environment¹³. As mentioned earlier, these results with NaCl-induced FRET changes are similar to those in low GdmCl concentrations, with the exception that instead of the rollover observed in the latter, the FRET and end-to-end distance changes monotonically with increasing concentration of NaCl.

Figure 4.20 B shows the folded ‘baseline’ at 268.16 K in the presence of various salts. These were determined from phenomenological two-state fits to the CD melts at 222 nm. All three salts induce structure in the protein, and can be said to do so to the same degree, since the ellipticities match in analogous conditions, within experimental error. Similar to the salt-induced restructuring of the protein observed at 368.16 K, at low temperature a drastic strengthening of signal at low concentrations up to 1 M is observed, followed by a linear increase. In fact by a concentration of 0.25 M of either salt, ~50% of the native signal, indicated by the brown line, is regained. The most striking result is the change in signal with GdmCl, where it decreases (strengthens) with concentration up to a 1 M, after which the CD signal increases. The electrolytic properties of GdmCl cause the electrostatic repulsions to be screened, which results in a BBL restructuring or compaction, followed by the

regime beyond 1 M, where the charge-screening property will have saturated and the chaotropic property of GdmCl takes over. In the case of the neutral salts, the sharp increase in structure at low concentrations followed by a regime of linear change has its origin in the same properties, with the difference that these salts are kosmotropic in comparison with GdmCl and restructure the protein beyond the 1 M regime. By a concentration of 2 M, the CD low temperature baseline of BBL in the presence of either salt, matches that under native conditions indicating a full recovery of the same degree of structure as at neutral pH.

At pH 7.0, the net charge on the protein is ~ 2 , contributing relatively favourably to the stabilization of the protein, in comparison to the contribution at pH 3.0, when a net charge of ~ 8.5 produces a higher degree of electrostatic repulsion between charged moieties on the protein. Thus, the repulsive electrostatics at pH 3.0 replace the relatively less unfavourable electrostatics present at neutral pH. This difference has a direct influence on structure, since BBL unfolds with decreasing pH. At pH 3.0, in the presence of a far greater electrostatic potential (similar to the effect described by Stigter and Dill¹⁶²), solvent restructuring and an increase in the hydrophobic interaction in the presence of 2 M salt, almost exactly compensate the destabilizing effect of long-range electrostatic repulsion. Thus, the landscape of unfolded BBL can be tuned towards higher structure by simply modulating pH and salt concentration.

4.4.2 The Reduced Folding Cooperativity of Salt-Refolded BBL

The first derivatives of low resolution thermal melts are a first approximation of the cooperativity of thermal unfolding, and conversely, folding, since they are essentially the rates of change of the signal being measured with the independent thermodynamic variable, such as denaturant concentration, pH or in this case temperature. The maxima in such curves indicate the temperature at which the signal undergoes an inflection.

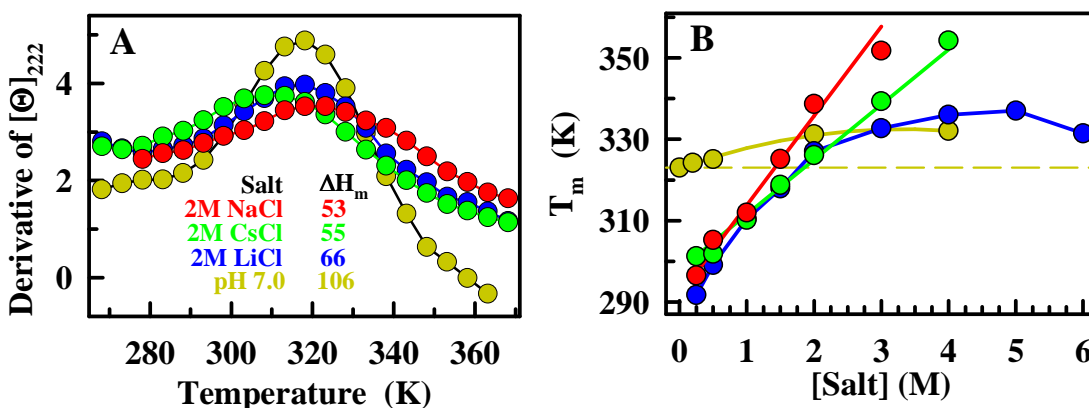


Figure 4.21 Thermal melting temperature and broadness of CD thermal melts of Naf-BBL. A) Smoothed first derivatives of CD signal, with red, green, blue and brown circles indicating Naf-BBL at pH 3.0 with 2 M NaCl, CsCl and LiCl, and Naf-BBL at pH 7.0 with no salt, respectively. Lines are to guide the eye. The table indicates ΔH_m 's of Naf-BBL in the three conditions with that at pH 7.0 and no salt, to indicate relative cooperativity of refolding with respect to that under native conditions in units of kJmol^{-1} . B) The T_m 's of thermal unfolding determined from phenomenological two-state fits to thermal unfolding CD data, with same colour scheme as in A, but with LiCl as the added salt in the pH 7.0 condition. Dashed brown line indicates T_m of Naf-BBL at pH 7.0 and no salt. Lines are polynomial fits to guide the eye.

The point of inflection is the condition at which variation in signal is maximum. Mathematically, this occurs when the signal has undergone a 50% change compared to its upper and lower limits. This is the T_m in thermal unfolding

experiments. The slope at the T_m , i.e. the value of first derivative of signal versus temperature (such as those plotted in Figure 4.21 A) and the broadness of the transition both indicate the degree of cooperativity of the transition. Theoretically, an all-or-none transition should be infinitely sharp around the T_m with little or no change in signal in the temperatures below and above the melting temperature. However, even for two-state folding proteins, folding curves have finite slope and broadness^{75,109}. For a globally downhill folding protein such as BBL with no energy barrier to folding and even for proteins with marginal barriers, thermal unfolding transitions in experiments such as differential scanning calorimetry, or spectroscopic measurements are broader than that expected for established two-state folders^{12-14,141,144-146,156,163}. Thus, the broadness of a thermal melts is an immediate indicator of folding cooperativity in low resolution experiments.

The first derivatives of the CD thermal melts at 2 M salt are plotted in Figure 4.21 A. The T_m 's in the three curves almost exactly match the T_m of BBL under native conditions at neutral pH (~318 K). The continual restructuring of BBL with salt to produce the native baseline of Figure 4.20 B, and the unfolded baseline at high temperature in Figure 4.20 A, the coincidence of structural content (determined from the signal), matching T_m 's in 2 M salt and at neutral pH, and concomitant increase of T_m and structure formation with salt (Figure 4.21 B), all indicate the direct correlation between structure and stability, as expected for a downhill folding protein¹⁵⁶. It also points to the fact that salt-refolded protonated BBL at pH 3.0 adopts the same structure as native BBL in neutral conditions. This is further confirmed by NMR experiments described later.

A critical difference between the thermal unfolding of salt-refolded protonated BBL and native BBL is the lower cooperativity of thermal melts as judged from the increased broadness of first derivative curves in Figure 4.21 A. In a phenomenological two-state analysis, ΔH_m defines the sharpness of the thermal transition and the table in Figure 4.21 A lists the mid-point folding enthalpy for all four conditions. Salts with different cations follow the rank-ordering of $\text{Na}^+ > \text{Cs}^+ > \text{Li}^+$ in their ability to reduce the mid-point folding enthalpy with respect to that at neutral pH. A ratio of the ΔH_m 's in salt to that at neutral pH is an indication of the relative cooperativity or sharpness of the salt-refolded thermal melting curves in comparison to that at neutral pH, and yield values of 50% (Na^+), 52% (Cs^+) and 62% (Li^+). These ratios are relative cooperativities, and the thermal denaturing curves in 2 M salt are at least twice as broad as the thermal unfolding curves of the downhill folding protein BBL in neutral pH; thus, the transition is at least half as cooperative. Similarly, another estimate of relative cooperativity can be obtained from the derivative curves themselves, as follows. The first derivative curves are interpolated mathematically to 0.1 K increments, and the values of the first derivative at 15 K below and above the thermal denaturation midpoint are summed and normalized by the value of the first derivative at the T_m for each condition of salt concentration or pH. This ratio when compared to the ratio obtained for BBL at neutral pH also gives an estimate of the relative sharpness of the thermal unfolding curves. The relative cooperativities arrived at are 47% (Na^+), 44% (Cs^+) and 69% (Li^+), similar to the relative cooperativities obtained by comparison of the midpoint unfolding enthalpies.

The same rank ordering is observed in the effect of salt on the T_m 's of salt-refolded protonated BBL at pH 3.0, with the largest increase in T_m for unit increase in salt concentration being produced by NaCl, followed by CsCl and LiCl. This rank-ordering does not follow the Hofmeister series or the cationic radii, but in fact follows the aqueous solubilities of the three salts in increasing order NaCl (6.2 M)<CsCl (11.0 M)<LiCl (19.6 M). This is a direct indication of the fact that protein restructuring is caused by solvent effects in the presence of salt, where strong kosmotropic interactions between the salt's ions and water, exclude the ions from the protein's surface, and also induce structure. Lower solubilities are a result of a greater degree of water sequestration, and the more a protein is excluded from the bulk solvent, the more stable it becomes⁶⁰⁻⁶¹.

Table 4.1 has the midpoint unfolding enthalpies for all conditions determined by phenomenological two state fits, and the relative cooperativities in all conditions are less than unity, indicating broader thermal transitions compared to thermal unfolding of BBL at neutral pH. Thus, although solvent restructuring and an increase in the hydrophobic effect at pH 3.0 in the presence of salt can induce protein refolding in a pH-sensitive protein such as BBL, the presence of highly unfavourable electrostatics on the protonated chain, screened nevertheless due to the presence of salt, precludes the possibility of cooperative folding. Electrostatic interactions are long-range forces and act beyond simple immediate-neighbour two-body interactions such as hydrogen-bonding or van der Waals interactions. The source for folding cooperativity is this non-local thermodynamic coupling or interaction mediated between a pair of moieties by the presence of at least one other moiety in between¹⁴¹.

Structural connectivity or the thermodynamic coupling between sequence non-local portions of a protein arise from such long-range interactions, and electrostatic interactions can contribute to this effect⁵⁴. Thus, in acidic pH, the presence of unfavourable electrostatics prevents thermodynamic coupling between sequence-distant groups on the chain, and prevents the chain from folding as cooperatively as BBL at neutral pH. Although interactions such as the hydrophobic effect are sufficient to induce folding of a protein, folding cooperativity, i.e. thermodynamic and structural coupling between distant regions of a protein cannot be induced. Therefore, the presence of favourable electrostatic interactions is determined to be critical to folding cooperativity in addition to the presence of hydrophobic interactions⁵. BBL has a well-defined hydrophobic core¹³, but even so has a low degree of folding cooperativity¹⁴¹. Figure 4.22 shows the first derivatives of the CD signal of Naf-BBL in 4 M LiCl at two different pH's- 3.0 and 7.0. BBL has similar T_m 's in either condition and similar degrees of relative cooperativities (Table 4.1) and broadness of the thermal transition, based on a visual inspection of the curves. BBL in these conditions is stabilized with respect to BBL at neutral pH, due to structuring of the solvent. This shows that the same physical forces which are at play at low pH and a high degree of protonation of BBL, are also present at neutral pH, and the same forces, i.e., electrostatic interactions and strengthening of the hydrophobic effect due to solvent structuring contribute to protein stability and cooperativity. Figure 4.22 also serves as an important control experiment which shows that the structuring effect is not an artifact which emerges from the effect of the protonation of histidines on the nature of the folding transition, as has been claimed by Fersht and coworkers¹⁵² and

Wolynes and coworkers¹⁶⁴. Rather, the stabilizing effect of salt, and the rendering of lower folding cooperativity in its presence is preserved in both sets of conditions at pH 3.0 and pH 7.0, with overlaying T_m and similar relative cooperativities (Fig. 4.22, 4 M LiCl condition).

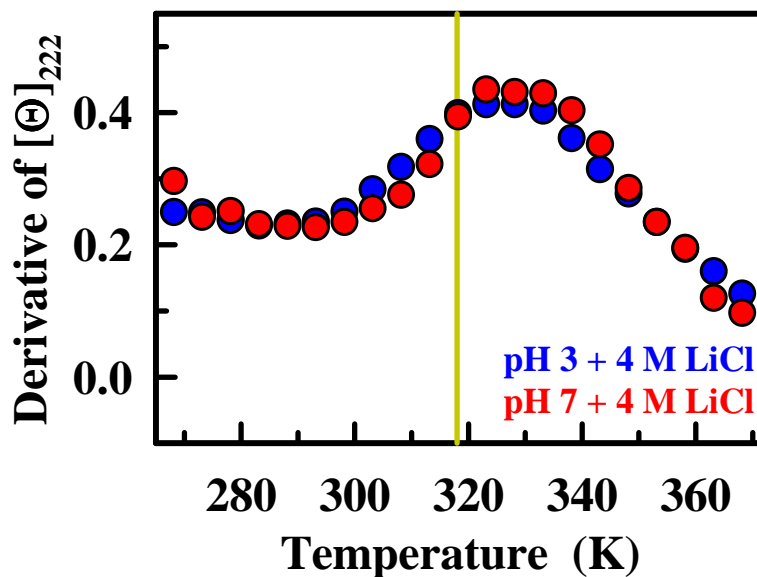


Figure 4.22 Reduced folding cooperativity and increased stability in the presence of salt. Circles represent first derivatives of mean residual ellipticity of Naf-BBL with 4 M LiCl at pH 3.0 (blue) and pH 7.0 (red). Brown line is T_m of BBL at pH 7.0

Table 4.1 Unfolding midpoint enthalpies (ΔH_m) in $\text{kJ}\cdot\text{mol}^{-1}$ in different salt concentrations at pH 3.0 with NaCl, CsCl, and LiCl as the salts employed, and at pH 7.0 with LiCl as the salt. Asterisks indicate enthalpies normalized to that at pH 7.0 in citrate buffer. The numbers in parentheses are relative cooperativities calculated by a ratio of ΔH_m in the given condition by that of BBL at pH 7.0 from two-state fits.

[Salt] (M)	pH 7.0+LiCl	pH 3.0+NaCl	pH3.0+CsCl	pH3.0+LiCl
0	105.65 (1.00)	–	–	–
0.2	96.00* (0.91)	–	–	–
0.25	–	50.67 (0.48)	62.38 (0.59)	49.08 (0.46)
0.5	106.12* (1.00)	60.18 (0.57)	49.64 (0.47)	52.890 (0.50)
1.0	–	54.11 (0.51)	52.82 (0.5)	53.16 (0.50)
1.5	–	57.17 (0.54)	58.55 (0.55)	65.36 (0.62)
2.0	70.76* (0.67)	52.53 (0.50)	54.54 (0.52)	65.86 (0.62)
3.0	–	66.23 (0.63)	61.93 (0.59)	72.04 (0.68)
4.0	74.65* (0.71)	–	86.89 (0.82)	80.28 (0.76)
5.0	–	–	–	67.20 (0.64)
6.0	–	–	–	55.31 (0.52)

4.5 Salt-Refolded BBL is not a Partially-Structured Non-Native State

A compact folding intermediate that has well-preserved secondary structure, but compromised tertiary structure^{62-64,165}, with exposed hydrophobic patches and larger radius of gyration, higher enthalpy and entropy than the native state, among other spectroscopic and hydrodynamic properties in which it is different from the native protein, is called the ‘molten globule state’. It is a generic term and can refer to the existence of this state in conditions where it is populated either by temperature or

denaturant. The term is often interchangeably used with the term ‘A state’, whereas strictly speaking, the latter refers to those denatured states produced by acidic pH, which do not have the same degree of loss of structure at low pH as in the presence of denaturing agents, and can refold upon addition of salt. A seminal study on salt-refolding of the acid-denatured state of apomyoglobin posited that anion binding to protonated sites on the protein at low pH induced refolding in highly acidic conditions, where $[H^+] < 10^{-2} \text{ M}$ ⁶⁴. Since our observations on protein refolding by addition of salt are made in conditions similar to those that are employed in populating molten globule or A states, it is critical to determine whether salt-refolded protonated BBL is truly native-like in its structural and dynamic properties.

It has already been established that in terms of spectroscopic signal and thermal melting temperature, salt-refolded BBL adopts the same properties as those of native BBL, hinting at the readoption of native structure merely as a simple function of the balance between destabilizing protonation and solvent-induced stabilization. However, caution must be exercised here, since we do not make the observation with a low resolution technique such as CD that the acid-denatured state does retain some residual signal which decreases further upon addition of GdmCl (Figure 4.20 B). However, the proton chemical shift and SOFAST HSQC spectra at pH 3.0 in comparison with those at pH 7.0 in Figures 4.4, 4.5, 4.6 and 4.7 clearly indicate that the onset of acidic pH brings about complete loss of structure. Furthermore, the chemical shifts at pH 3.0 do not show any variation with temperature, whereas the ones at pH 7.0 do⁶⁵. Nevertheless, it is crucial to verify that

salt-refolded BBL is indeed not a molten globule state or an A state, which if it were, would give rise to the observed low cooperativity in thermal unfolding transitions.

Upon comparison of chemical shifts in the amide region of their proton 1D spectra, at pH 3.0 and pH 3.0 with 2 M LiCl (conditions in which BBL refolds) in Figure 4.23 A, we see that the chemical shifts at pH 3.0 show significant line broadening, along with the absence of peaks at several positions, when compared with the signal from the salt-refolded sample. Furthermore, the chemical shifts for a molten globule or A state with secondary structure, but no tertiary structure should have peaks bunched in the region of 8-9 ppm. Clearly this is not the case. Figure 4.23 B plots the chemical shifts for the same two conditions but in the methyl region, where again we notice line broadening and the disappearance of peaks. This indicates that at a structural level, the protein in pH 3.0 and pH 3.0 with 2 M LiCl are distinct. The protein is indeed unfolded at pH 3.0 and structure is consolidated in the presence of salt. Most importantly, the tell-tale signs of molten globules in the 1D amide spectra of BBL at pH 3.0 and 2 M LiCl are absent. Another critical comparison would be that between the salt-refolded protein at pH 3.0 and the native protein at neutral pH. Figures 4.24 A and B plot the overlaid amide and methyl region 1D spectra of salt-refolded protonated BBL, and BBL at neutral pH. The existence of peaks indicating structure in the spectra of refolded BBL and the absence of line-broadening effects, and correspondence between the peaks in refolded and native BBL indicate that the salt-refolded structure is indeed native-like and not a non-native molten globule. The overlaid SOFAST HSQC spectra in Figures 4.25 and 4.26 also corroborate these observations. While a difference in pH between salt-refolded and

native BBL would shift the chemical peaks, the appearance of shifted but corresponding peaks between the two proteins in Figure 4.26 indicate that the salt-refolded BBL is truly native in structure. 1D spectra are plotted on the following pages.

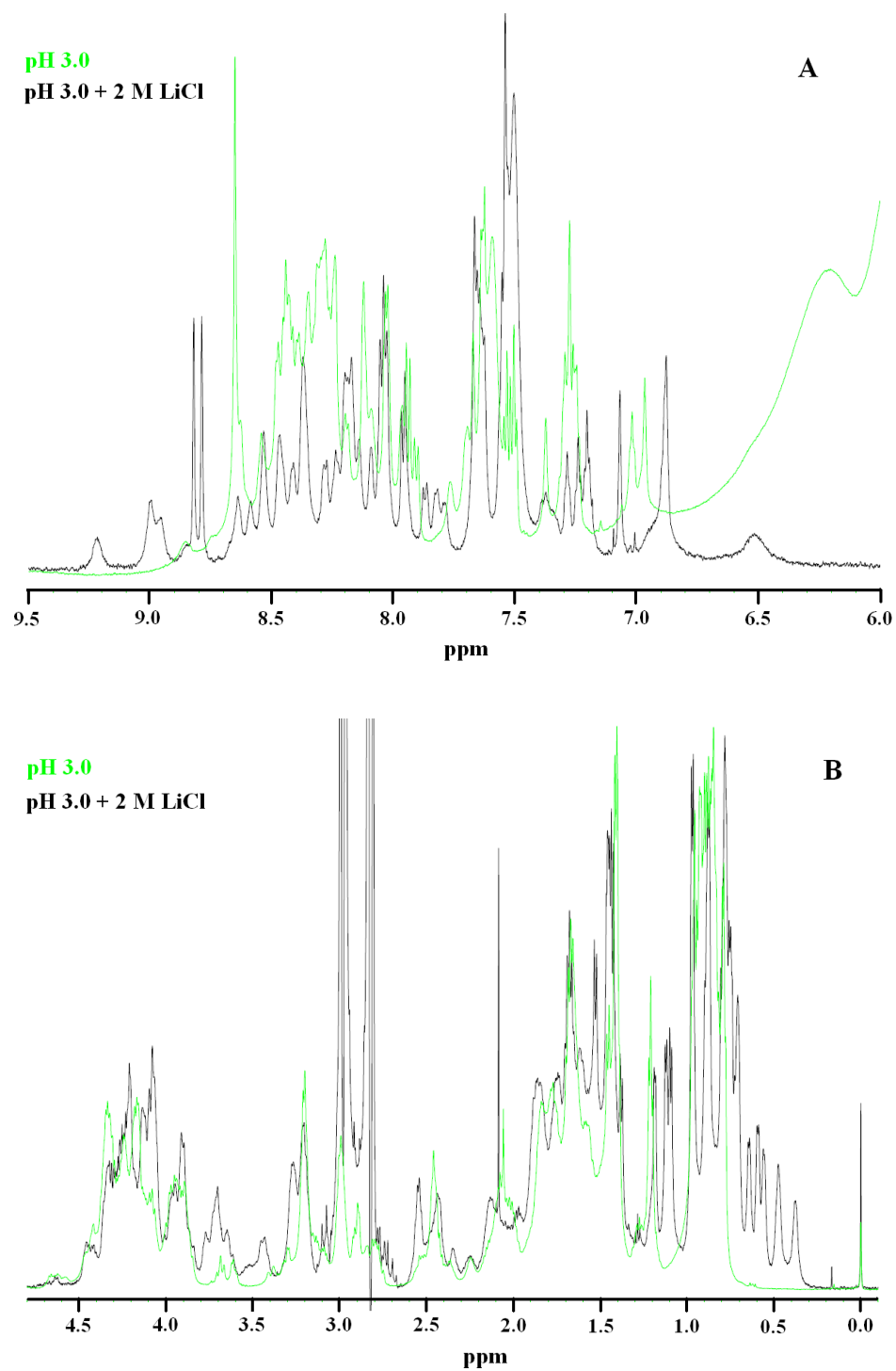


Figure 4.23 1D proton NMR chemical shifts for acid-denatured (green) and salt-refolded (black) BBL in the amide (A) and methyl (B) regions.

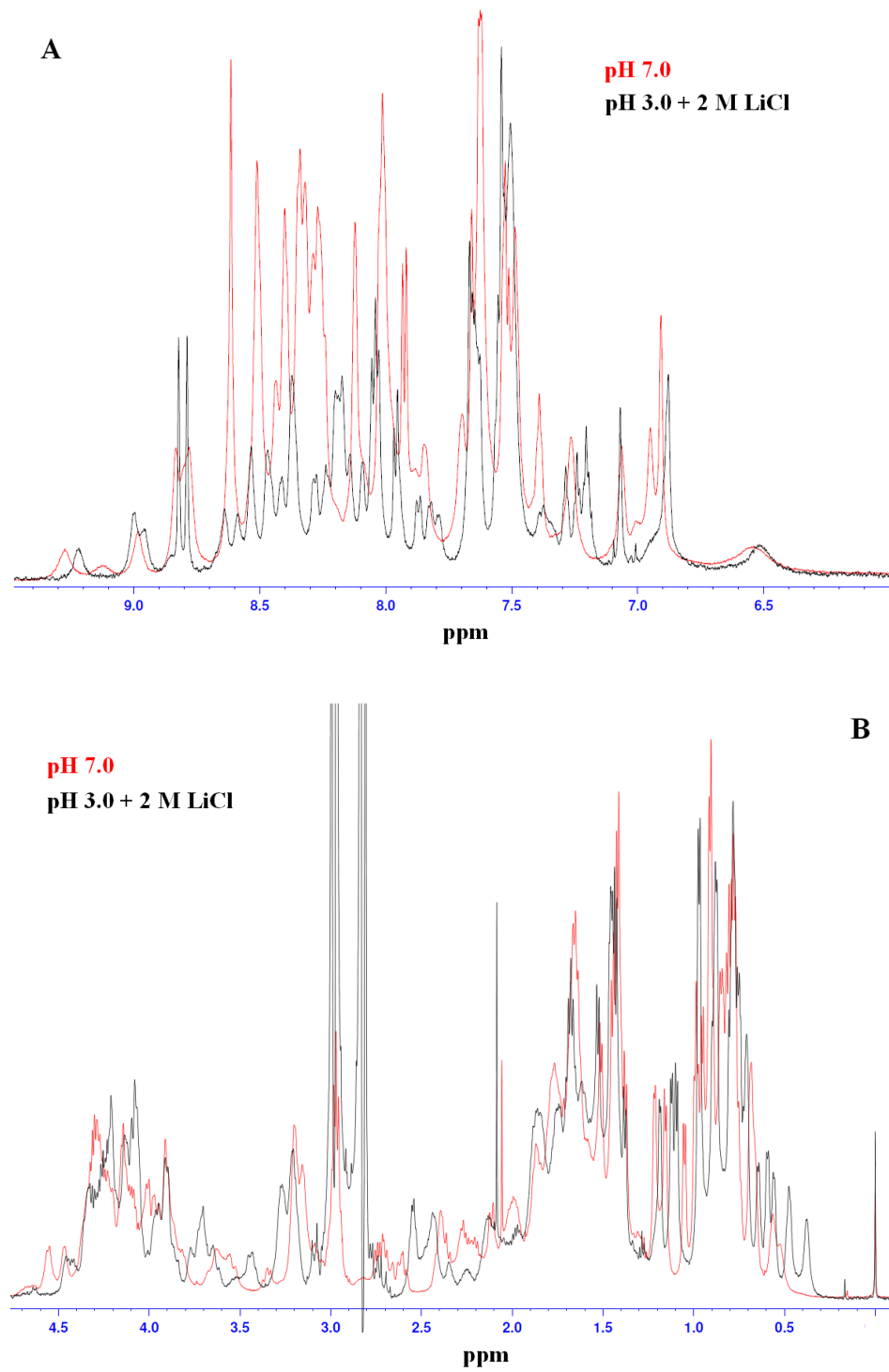


Figure 4.24 1D proton NMR chemical shifts for native (red) and salt-refolded (black) BBL in the amide (A) and methyl (B) regions.

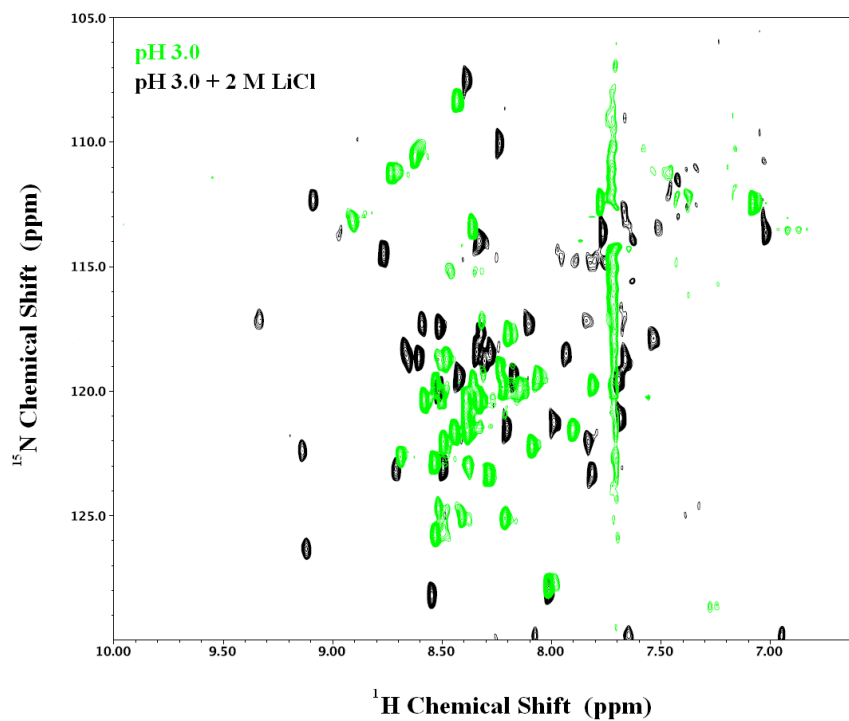


Figure 4.25 NMR SOFAST-HSQC spectra with ^{15}N natural abundance for acid-denatured (green) and salt-refolded (black) BBL.

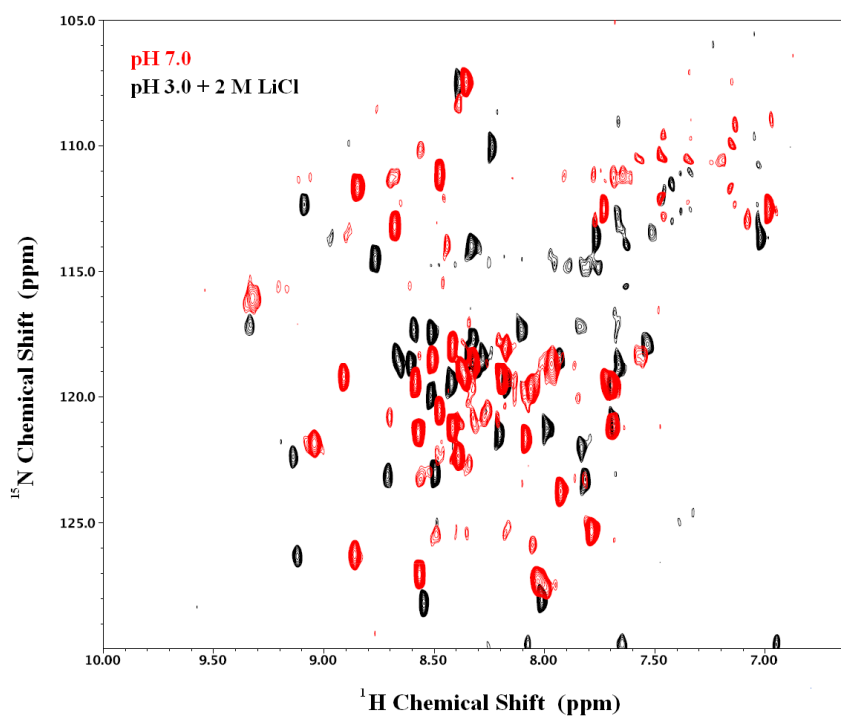


Figure 4.26 NMR SOFAST-HSQC spectra with ^{15}N natural abundance for native BBL at neutral pH (red) and salt-refolded (black) BBL.

4.6 BBL Refolding Dynamics at Varying Degrees of Chain Protonation

BBL is a globally downhill folding protein characterized by the absence of a free energy barrier to folding under all thermodynamic conditions^{12-14,141,143-146,156}. However, a recent series of articles by Fersht and coworkers¹⁵⁰⁻¹⁵² and a report by Wolynes and coworkers¹⁶⁴ seem to suggest that: a) The protonation of the chain, and in particular the two histidines in the sequence at reduced pH introduces instability in the structure, especially since the second histidine lies in the second α -helix and has a perturbed pKa (5.37). This is claimed to bring about a dispersion in the T_m observed in the atom-by-atom thermal unfolding of BBL at pH 5.3¹⁴¹; and b) The differences in the degree of protonation between pH 7.0 and pH 5.3 bring about a change in the folding mechanism from barrier-limited at neutral pH to downhill at lower pH's.

While it is known that a decrease in pH destabilizes and unfolds BBL, the structure of BBL was solved at pH 5.3 and does not show any dispersion, or conformational degeneracy or heterogeneity in the overall structure or in the second α -helix¹⁴¹. Thus, the structure of BBL at pH 5.3 is fully folded and native as determined by NMR. The effect of pH on the folding mechanism of BBL can be tested by performing nanosecond T-jump kinetics experiments on the BBL at different degrees of protonation of the chain and the histidines. Time-resolved IR T-jump kinetics measurements on BBL, measuring dynamics of the backbone structural melting, showed the existence of a single exponential relaxation upon temperature perturbation of equilibrium, with relaxation time in the order of tens of microsecond and a maximum in amplitude at ~325 K. FRET T-jump kinetics measurements on BBL showed a relaxation with contributions from two components, reporting on the

dynamics of the end-to-end distance variation (decrease in size with thermal unfolding) in BBL. The faster component is in the time-scale of ~100 ns and its amplitude increases linearly with temperature, and corresponds to the hydrophobic collapse of acid-denatured BBL. The second slower component has relaxation times in the range of tens of microseconds and amplitude with a maximum at ~290 K¹⁴³. The folding dynamics were shown to be incompatible with two-state and three-state folding models, and were explained by the dynamics of the protein on a one-dimensional harmonic potential well with a smooth surface, and where the relaxation rates (k_{obs}) depend on the curvature of the well (ω), the diffusion coefficient of motions on the surface $D(T,n,probe)$ (a function of temperature, solvent viscosity, folding order parameter n) and temperature (T), according to Equation 4.3:

$$k_{obs} = \frac{\omega^2 \cdot D(T, n, probe)}{RT} \quad (4.3)$$

The dynamics of BBL refolding at various degrees of protonation are measured here by nanosecond IR T-jump experiments. The following protocol is employed. pH decreases the T_m of BBL. LiCl at precise concentrations so as to match the T_m at two pH's 3 and 5.3 (and additionally 4.5 and 6 not studied in the kinetics experiments) is added to acidified BBL. This ensures that temperature dependent differences folding stability do not affect refolding dynamics. Two scenarios can be envisaged. If the hypothesis that BBL's folding mechanism changes from barrier-limited to downhill upon reduction in pH from 7.0 is true, and that the a molten-globule or A state is populated in the acidic pH with a barrier separating this state,

and the true unfolded state, then relaxation in the low pH regime should necessarily be faster than that at pH 7.0. This is because the barrier at neutral pH would decrease in height but still exist, enough to affect folding relaxation dynamics. A small decrease of the energy barrier by even $1RT$ should speed up the relaxation by 2.7-fold. Further we could expect differences in the relaxation dynamics at the two acidic pH's 5.3 and 3, however the stabilization introduced by the salt could have complicated effects on relaxation dynamics. This would not only confirm the existence of the A state, but would also indicate the existence of a barrier between the two purported states that BBL adopts in its folding mechanism.

In the alternate scenario, if BBL's folding relaxation dynamics, as shown by Li and coworkers in a study in the Muñoz lab, is diffusional over a single harmonic potential well, then dynamics at matched T_m 's should only be a function of the curvature of the well. This curvature is determined by the degree of structural connectivity or cooperativity of folding at the given condition. It is known that salt-refolded structures refold less cooperatively than in the absence of salt, producing a broadening of the potential well, and a greater spread in order parameters that define the potential well. This reduced curvature of the potential should slow down the relaxation according to Equation 4.3. IR T-jump measures the relaxation of the backbone during the structural transitions and unfolding accompanying the ~ 10 K temperature jumps. Figure 4.27 shows the mean residual ellipticities at 222 nm measured by CD at three different pH's with LiCl added so as to match the melting temperature and eliminate order parameter effects on the diffusion coefficient of relaxation, making it dependent solely on curvature.

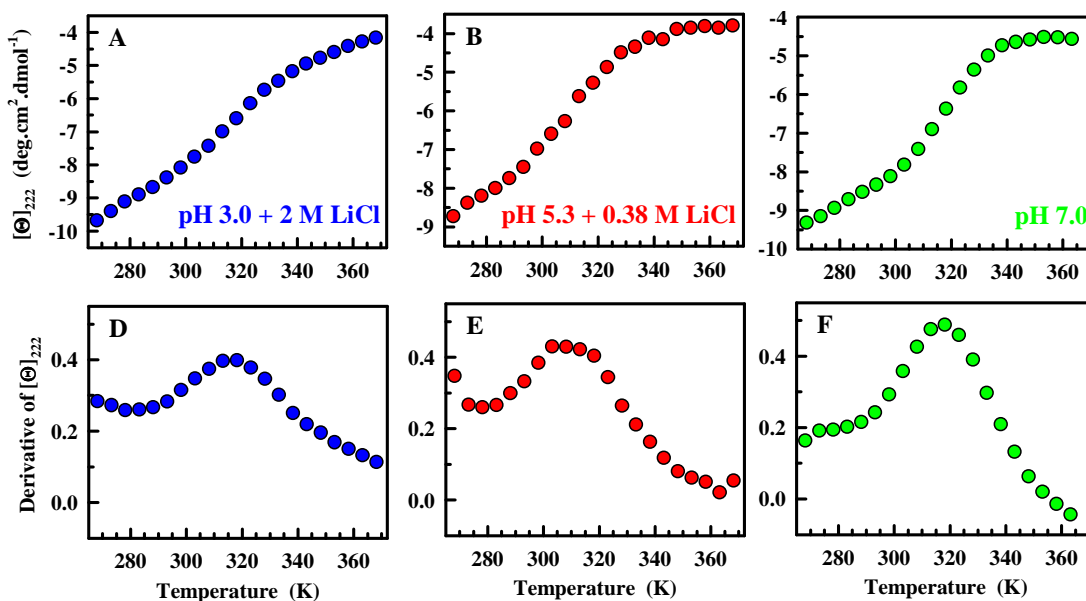


Figure 4.27 Matching T_m 's at varying degrees of chain and histidine protonation with LiCl. A, B and C are molar residual ellipticities at 222 nm determined by CD, at pH 3.0 with 2 M LiCl, pH 5.3 with 0.38 M LiCl and pH 7.0 with no salt, respectively. D, E and F show the corresponding first derivatives of the signals in A, B and C.

[LiCl] for intermediate pH's between 3.0 and 7.0 was determined by iterative interpolation of [LiCl], considering that 2 M LiCl is needed at pH 3.0 to match the T_m to that pH 7.0. This was confirmed by CD measurements in the presence of the corresponding [LiCl]. T_m measured by finding the maxima of the first derivatives of the CD signal at 222 nm with temperature, at pH 4.5 no salt, 5.3 no salt, 6 no salt, and 7.0 with no salt yielded values of ~298 K, ~305 K, 318 K and ~318 K, respectively, although the thermal melt of pH 6.0 had a lower slope at T_m . T_m measured likewise at pH 4.5 with 0.93 M LiCl, pH 5.3 with 0.38 M LiCl and pH 6 with 0.23 M LiCl, yielded similar T_m 's (as seen above by visual inspection in Figures 4.27 D, E and F) of 318 K, although that for the pH 5.3 with 0.38 M LiCl sample was slightly lower at ~313-315 K, thus prompting a recalculation of the LiCl concentration to be 0.42 M at

pH 5.3 to match the T_m of pH 7.0. A phenomenological two-state analysis on these thermal melts yields similar T_m 's of 323 K for all conditions where T_m 's were supposed to be matched by the combination of pH and salt, and relative cooperativities of 0.51, 0.66 and 0.73 at pH 4.5, 5.3 and 6 with added salt (compared to 0.65, 0.61 and 0.9 at these pH's with no salt added). This again shows that even at matched T_m 's, the folding cooperativity lowers as more salt is added

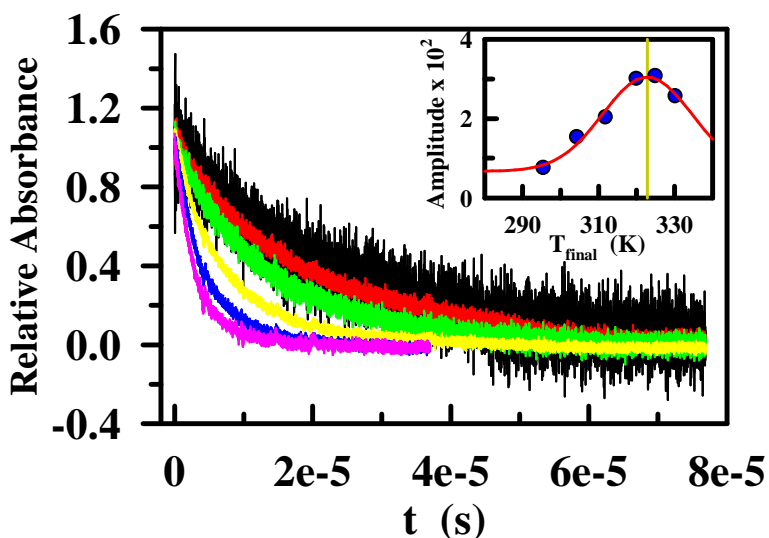


Figure 4.28 Exponential decays of the absorbance at 1632 cm^{-1} by Naf-BBL at pH 7.0 with $\sim 11.5\text{ K}$ jumps to final temperatures of 295.34 K (black), 304.26 K (red), 311.77 (green), 319.91 (yellow), 324.83 (blue) and 330.09 (pink). The corresponding relaxation times (τ_{obs}) are 22.58 μs , 20.30 μs , 14.15 μs , 7.97 μs , 4.55 μs and 3.13 μs , respectively. (Inset) Circles are amplitudes of the relaxation and the red line a Gaussian fit to the amplitudes with a mean of 322.8 K (the T_m) indicated by the brown line.

The normalized relaxation of the absorbance at 1632 cm^{-1} after $\sim 11.5\text{ K}$ T-jumps are plotted in Figure 4.28 for a few sample temperatures around 315 K for Naf-BBL at pH 7.0. The decay speeds up with increasing temperature (see figure caption).

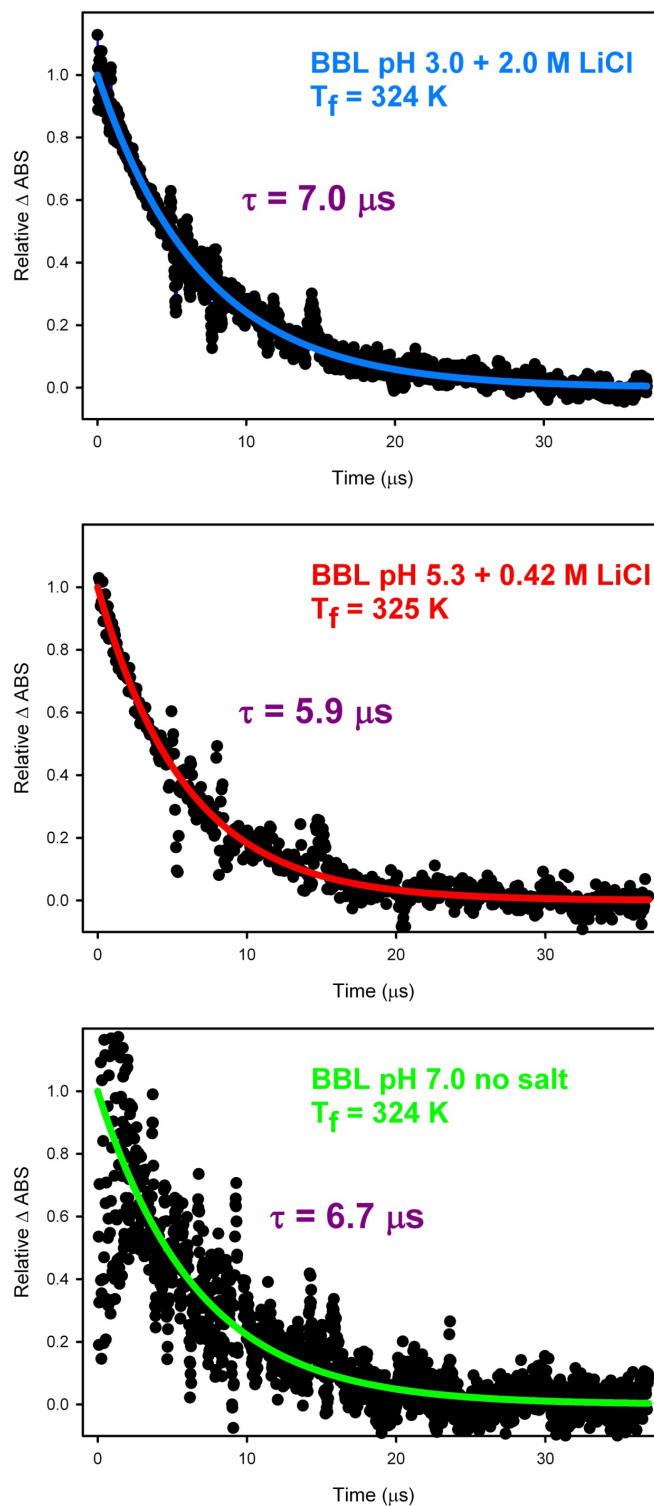


Figure 4.29 T-jump relaxation at the T_m 's in the three matched conditions.

The relaxation time at 333 K is predicted to be 0.59 μs from a linear fit of the relaxation times in Figure 4.29 needed to fit the decays. This value is remarkably close to that of 0.4 μs , predicted by Eaton and coworkers to be the protein folding ‘speed limit’,¹⁹ ($N/100 \mu\text{s}$, where N is the sequence length). The relaxation times are in the order of a few microseconds at the T_m of ~ 324 K for BBL at all the pH’s, as shown in the legends in Figure 4.29. This result shows striking similarity between the rates at acidic pH’s and pH 7.0. A barrier-limited unfolding process, with reduced barrier-heights at low pH, would result in much faster relaxation (a 2.7-fold speed-up for $1RT$ increase in barrier-height) when compared to pH 7.0. However, the τ_{obs} are similar ($\sim 6.5 \mu\text{s}$) at T_m for all pH’s within experimental error, indicating that the second scenario holds, i.e. that BBL folds by a globally downhill folding mechanism, and there is no mechanistic difference between folding at neutral pH and acidic pH, as claimed by Wolynes and coworkers. Furthermore, the coincidence of rates in the entire range of temperatures from 295 K to 350 K for all 3 pH’s, 3.0, 5.3 and 7.0 in Figure 4.29, indicate a unified folding mechanism that is preserved in all conditions as previously demonstrated^{12-14,141,143-146,156,163}.

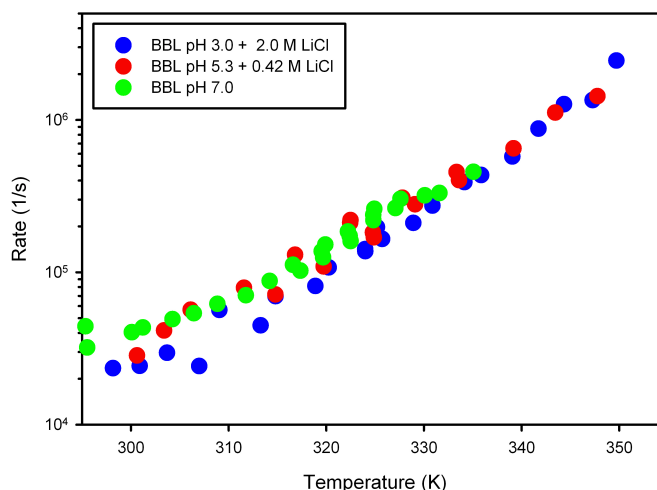


Figure 4.30 Relaxation rates of BBL folding monitored by nanosecond IR T-jump.

Thus, the possibility of the existence of a barrier to folding in BBL is unequivocally precluded by the coincidence of folding relaxation rates at all conditions of pH and temperature within experimental error, irrespective of the protonation state of the protein and in particular, the histidines. The rates at low pH should have been faster than those at pH 7.0 in such a scenario since the existence of such a state would involve energy barriers between the existing population at any given condition, and these barriers would decrease with decreasing pH, speeding up refolding rates. Furthermore, the wealth of CD and NMR spectroscopic data on the salt-refolded protonated state of BBL confirms that this ‘state’ is indeed native-like and not a molten globule or ‘A state’. However, no such speed-up is seen in the case of BBL. The destabilization of BBL with pH is indeed a manifestation of the protonation of ionizable moieties on the protein, and a single conformational ensemble is populated under all conditions¹³. The one-dimensional projection of BBL’s hyperdimensional free energy surface onto a single order parameter yields a

surface that is smooth (with respect to the RT fluctuations). Reconfiguration or folding dynamics in this case can therefore be described as diffusion along a smooth harmonic potential well, with rates being proportional to the square of the curvature. Although folding at low pH with added salt proceeds with lower cooperativity than at neutral pH, indicating a broader population distribution and a broader corresponding potential well with lower curvature, no slow down of relaxation rates with pH is observed. It is possible that at matching T_m 's the effect of lower curvature at low pH is compensated by a higher effective diffusion coefficient D along the surface, due to looser thermodynamic connectivity between portions on the chain¹⁴¹.

4.7 Conclusions

The effect of the chaotropic cosolvents urea and GdmCl on the unfolded ensemble of various two-state proteins was determined by applying the 3-component quadratic interaction energy equation to the net energy of a protein residue, to yield subtle differences in the relative effects of the cosolvent and solvent on protein expansion and compaction. This chapter dealt with the effect of the chaotropes GdmCl and urea, and additionally kosmotropes NaCl, CsCl and LiCl on the unfolded ensemble of BBL as the starting point. BBL is unfolded at pH 3.0 due to a pH-dependent protonation of the ionizable amino acids on the protein. In particular two histidines play an important role in modulating the pI and stability of the protein. This system is ideal for the study of the effect of thermodynamic conditions on the unfolded ensemble in the absence of interference from the folding reaction. Bulk FRET measurements indicated that the unfolded state is more compact than the native ensemble, and that increasing temperatures result in a compaction of the acid-

denatured ensemble due to the stronger hydrophobic effect at elevated temperatures. However, a peculiar low concentration ‘rollover’ in the presence of GdmCl was observed, which has its origins in the compaction of acid-denatured BBL at low GdmCl concentrations. This occurs due to the Debye-Hückel charge-screening of the electrostatic repulsions by the Gdm^+ and Cl^- ions. However, at concentrations above 1-1.5 M when the charge-screening effect saturates for monovalent neutral salts, the chaotropic property of GdmCl takes over and produces the expected expansion for a chaotrope. GdmCl has a stronger effect on chain size bringing about a two-fold increase in R_g . This is in stark contrast to the effect of urea, which does not increase the R_g appreciably (<0.2 nm) at all temperatures even at a concentration of 9 M. A detailed analysis on the 3-component-like effect on chain expansion, similar to that of the second chapter cannot be performed on acid-denatured BBL, except for heuristic purposes, since the denaturant driven expansion commences only after 1 M, and is in competition with prior folding events in the presence of GdmCl. Further, the presence of some residual structure in the acid-denatured state renders negative curvature to the interaction energy versus chemical curves, i.e. an unphysical result.

The effect of kosmotropes, on the other hand, is interesting, since the addition of a salt such as NaCl, causes refolding of acid-denatured BBL. CD spectroscopy measurements with different salts NaCl, CsCl and LiCl varying in the size of their cation, showed sigmoidal thermal unfolding curves. These experiments are similar in spirit to double-perturbation experiments¹⁵⁶, but a kosmotrope stabilizes the protein whereas a chaotrope unfolds it. Refolding in the presence of salt takes place less cooperatively than the already broad thermal transition of BBL at neutral pH. Charge-

screening of electrostatic repulsions and solvent-driven strengthening of the hydrophobic effect in the presence of salt (due to solvent sequestration and ion exclusion from the protein's surface), are sufficient to stabilize the protein and refold it, but cannot replace the favourable (or relatively less unfavourable) electrostatics present at neutral pH. The thermal transitions get sharper with salt concentration as structure is initially formed at low concentration and then rolls over at higher concentration to higher broadness, corroborating the previous conclusions. Thermal melting temperatures increase with increasing adoption of structure, and the rank-order of salts in their ability to increase T_m varies inversely with their respective solubilities (NaCl<CsCl<LiCl in increasing order of solubility), confirming the role of stabilization and alteration of folding cooperativity by solvent-driven effects.

Refolded BBL is native-like and is not a molten-globule as seen from CD and NMR 1D and HSQC spectroscopy. Further, nanosecond IR T-jump kinetics experiments yield the striking result that irrespective of the degree of protonation of the protein (controlled by pH) the protein folds with the same relaxation rate at all temperatures. This is exactly what is expected of a downhill folding protein, since for a barrier-limited scenario lower pH's with lower folding energy barriers should significantly speed up folding relaxation rates. Thus, the degree of protonation or the pH does not change the mechanism of BBL folding, viz. downhill folding is a global property of BBL. These experiments again rule out the possibility of the existence of the molten globule, since similar speed-ups of the relaxation rates would be expected with decreasing pH's where the barrier between the supposed 'unfolded state' and the 'A state' would progressively decrease.

In summary, the unfolded ensemble of BBL has a diverse phase phase diagram as indicated in Figure 4.31, with a slew of varying responses to thermodynamic variability, such as change in pH, chaotrope concentration, kosmotrope concentration and finally temperature. Temperature changes can induce chain compaction, chaotropes chain expansion, and kosmotropes chain compaction and even refolding. Results from the experiments on this system confirm the role of electrostatic interactions in producing folding cooperativity.

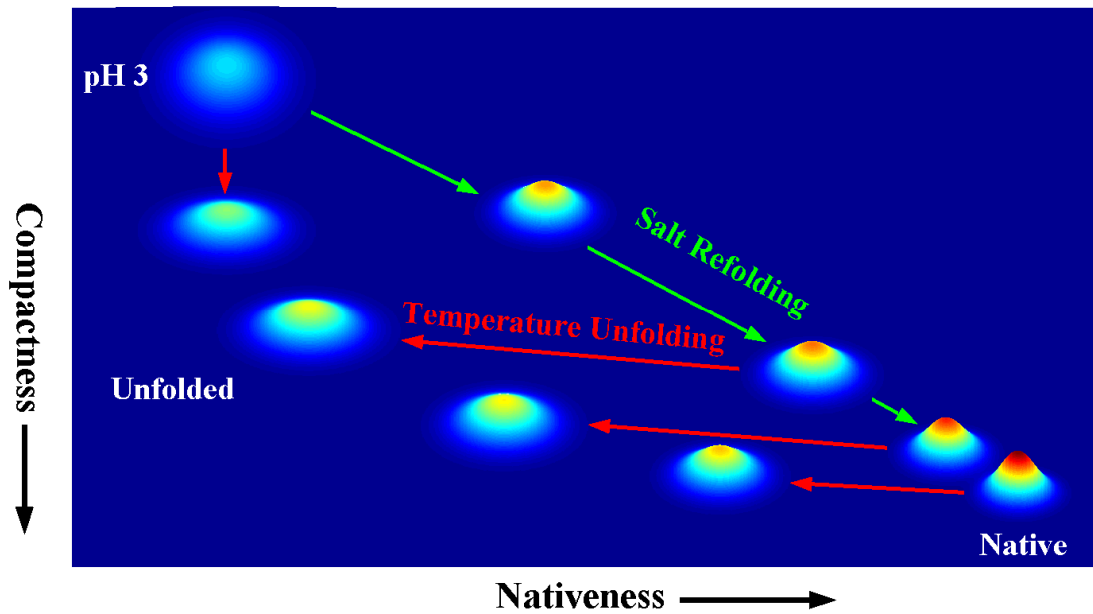


Figure 4.31 Phase diagram for properties of the unfolded state modulated by a combination of temperature, pH and added salt. Green arrows indicate refolding, and red arrows indicate thermal melting or hydrophobic collapse.

Chapter 5. Polymeric Properties of Random versus Natural Sequences

5.1 Introduction

Protein sequences are polymers naturally selected by evolution in geological time-scales that perform a vast array of biological functions such as molecular and electronic transport, catalysis, maintenance of structural integrity, cellular response to antagonists, cell adhesion, signaling reactions, homeostasis and macromolecular processing such as replication and translation. They fold in biologically relevant time-scales of the order of a few a seconds. Stable proteins are energetically biased towards the native state in physiological conditions by the presence of moderate to high energy barriers ($>2RT$)¹⁴. The absence of folding barriers¹³ and the presence of natively unfolded proteins is also posited to be critical for several cellular functions⁷⁰. The central quest in the folding field has been to understand the physico-chemical mechanism behind such rapid folding to stable structures, and to predict protein structure from sequence. De novo designed polypeptides that fold to stable structures represent a step in this direction¹⁶⁶. Designed polypeptides have also been pushed into the limit of fast folding in the order of $\sim 0.5 \mu\text{s}$, as in the case of a mutated version of the λ repressor¹⁹. However, a central question to the field of protein design and reconfiguration dynamics concerns the nature of the folding energy landscape of a random polypeptide. Theoretical predictions on the folding of random heteropolymers predict that this energy landscape should be rough with several crests and troughs that represent traps on the gradient towards the polypeptide's search for its compact free

energy minimum⁹. The existence of roughness and the absence of evolutionary selected minimally frustrated paths towards the energy minimum could be characterized by the onset of the glass-transition phase where the roughness on the landscape would severely restrict chain motion and slow down its dynamics. Such a phase would yield stretched exponential decays of chain reconfiguration, and super-Arrhenius temperature dependence or rates. This was illustrated in the case of a designed polypeptide FSS-1ss¹⁶⁶ marked by the switch in dynamics from fast folding to glassy in the regime of biological temperatures. This was brought about by change in a single residue into a non-natural amino acid. Thus minor alterations in protein sequence can have profound effects on its search for its global energetic minimum and on the speed at which it does so.

This chapter focuses on the description of a fundamental property of proteins-hydrophobic collapse- and the effect that specific chain sequence can have on size and reconfiguration dynamics. Hydrophobic collapse emerges as the response of a protein to decrease in denaturant stress and the onset of native-like conditions. However the interplay between collapse and secondary structure formation is difficult to discern in the presence of denaturants¹⁶⁷⁻¹⁶⁹. Does specific collapse in natural proteins impede or aid secondary structure formation, or vice versa? To discern this effect, collapse can be studied in a protein with no tendency to form folded structures. This can be achieved by studying the effect of sequence on collapse states and rates of a random heteropolymer. To this end, a randomized version of the protein BBL (termed Ran-BBL), was designed to have the same sequence composition as BBL, but a completely random sequence connectivity, with no propensity to form

secondary structure (q. v.). This system offers several exciting prospects relating to the study of folding dynamics and response to thermodynamic conditions such as temperature and cosolvent. Firstly, polymer physics ideas suggest that collapse in random heteropolymers should be a self-averaged function of sequence content. This is to say that the ratio of hydrophilic-hydrophobic amino acids should determine a random heteropolymer's properties. However, can sequence-specific effects be expected for such a polymer, i.e. would Ran-BBL expand and collapse to the same degree and size as unfolded BBL, in response to temperature change? Would there be a difference in the dependence of collapse rates with temperature when comparing the two? Would chain interactions with added cosolvents such as urea or GdmCl be different between the two proteins, as judged by the cosolvent-dependent expansion and collapse? Would variations in the charged state of Ran-BBL affect its degree and rate of collapse, due to changes in electrostatic interactions? If the collapsed degrees and rates of BBL and Ran-BBL are the same, it would appear that although sequences which are naturally designed to fold may adopt structure depending on specific conditions (change in pH in case of BBL, for e.g.), hydrophobic collapse is mainly controlled by self-averaging interactions.

Ran-BBL's sequence was obtained in the following manner. The Random Sequence Generator on the ExPASy Proteomics Server of the Swiss Institute of Bioinformatics was used to generate several (15-20) random sequences with the composition of BBL. The secondary structure forming propensities of these sequences were determined by the secondary structure prediction program AGADIR¹⁷⁰⁻¹⁷⁴ and one with the least propensity was chosen. The same pair of

extrinsic fluorescent labels as Naf-BBL-Dan was attached to the N and C-terminii of the peptide, produced by solid-state synthesis. Equilibrium bulk FRET experiments were performed in the presence of GdmCl and urea with varying temperature and at pH 3.0 and 7.0, to measure changes in the end-to-end distance of the peptide. DLS experiments were performed to compare R_h of Naf-BBL and Ran-BBL at various temperatures.

5.2 Equilibrium Expansion and Collapse of Ran-BBL

Ran-BBL is characterized by the complete absence of structure in the entire span of pH's from 3.0 to 7.0 (deduced by CD measurements, results not shown). Therefore, the effect of pH on chain collapse and expansion can also be tested in this system. FRET efficiencies measured at pH 7.0 and pH 3.0 at increasing concentrations of GdmCl and urea are plotted in Figure 5.1 A and B, and Figures 5.2 A and B, respectively. At pH 7.0 GdmCl brings about a pronounced decrease in the FRET efficiency, which decays to ~10% at 5 M and 283.16 K. FRET efficiency decrease in the presence of urea takes place to a lower extent, with almost 15% FRET at 7 M urea and 283.16 K.

The FRET efficiency is further reduced at pH 3.0 (Figures 5.2 A and B). A part of this reduction is due to the lower R_0 of the FRET pair at pH 3.0 (2.50 nm at pH 7.0 versus 1.85 nm at pH 3.0). GdmCl concentrations above 3 M have negligible FRET. FRET efficiencies are outside the dynamic range of the FRET pair, and accurate estimation of end-to-end distance even at 0.25 M GdmCl and low

temperatures cannot be made. Similar to the observation at pH 7.0, urea has a far weaker effect on chain size, with significant retention of FRET up to a concentration

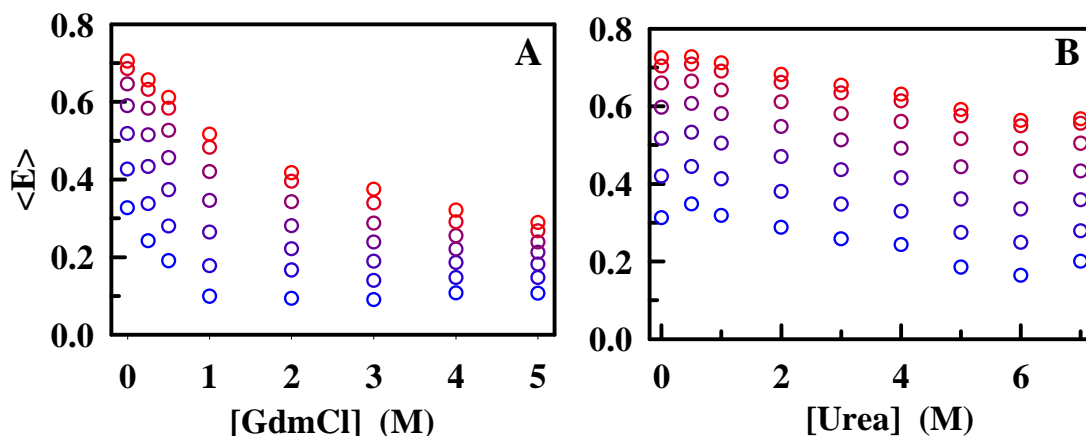


Figure 5.1 Chain expansion and collapse in Ran-BBL at pH 7.0. A) Bulk FRET efficiency variation with GdmCl. Colours go from blue to red indicating temperatures from 283.16 K to 368.16 K. B) Bulk FRET efficiency variation with urea. Same colour scheme as A.

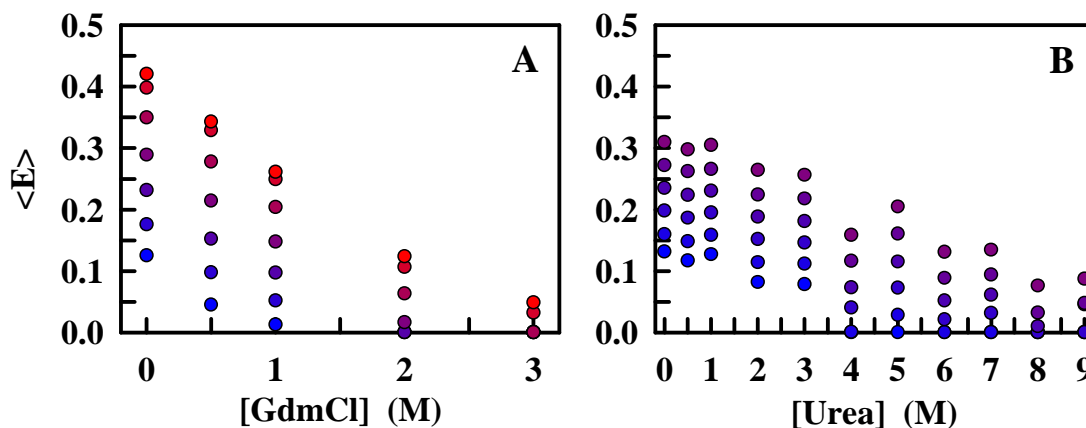


Figure 5.2 Chain expansion and collapse in Ran-BBL at pH 3.0. A) Bulk FRET efficiency variation with GdmCl. Colours go from blue to red indicating temperatures from 283.16 K to 368.16 K. B) Bulk FRET efficiency variation with urea. Colours go from blue to purple indicating temperatures from 283.16 K to 333.16 K.

of 7 M. Hydrophobic collapse with increasing temperature is observed in all conditions. However, at pH 3.0 it is not possible to determine end-to-end distance at

low temperatures and high GdmCl, due to a much reduced R_0 at pH 3.0 (1.85 nm compared to 2.5 nm at pH 7.0, (Figure 5.2 A). This effect is seen upwards of 4 M urea at low temperatures. The temperature range of FRET measurement at pH 3.0 in the presence of urea was limited by the carboxy-methylation of the C-terminus resulting in large apparent FRET efficiencies at temperatures >333 K.

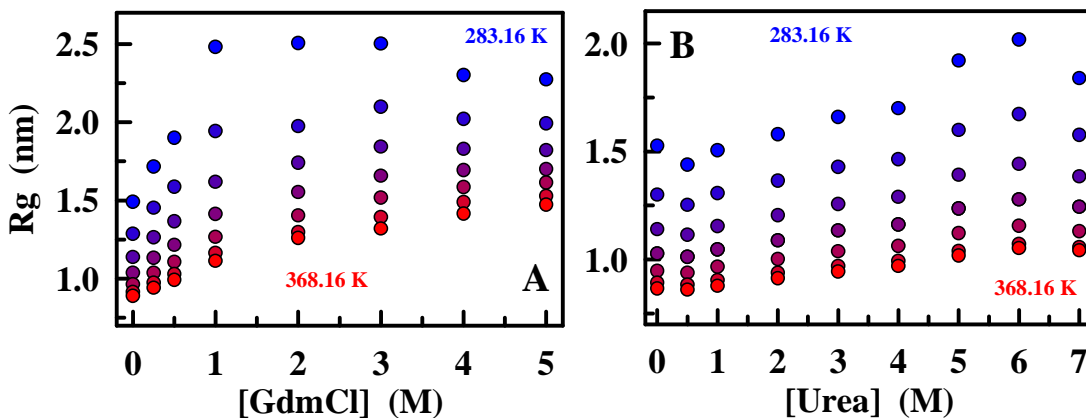


Figure 5.3 R_g of Ran-BBL at pH 7.0 calculated by the Gaussian chain assumption after R_0 correction.

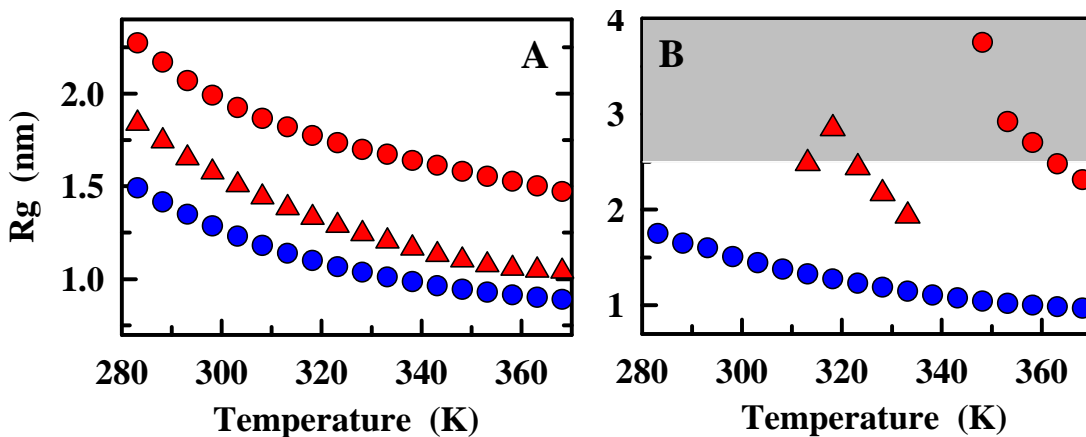


Figure 5.4 Hydrophobic collapse of Ran-BBL at the two pH's in buffer and limiting chemical concentration. A) Blue and red circles are R_g 's at 0 M and 5 M GdmCl at pH 7.0, respectively. Red triangles are R_g 's at 7 M urea at pH 7.0. B) Blue and red circles are R_g 's at 0 M and 3 M GdmCl at pH 3.0, respectively. Red triangles are R_g 's at 9 M urea at pH 3.0. The shaded region represents the upper limit of dynamic range for distance measurement of the FRET pair at pH 3.0, at ~10% FRET.

R_g calculated by the Gaussian chain assumption for the pH 7.0 data-set, after accounting for R_0 changes, show that at 368.16 K the chain is ~0.8 nm in size (R_g) in the absence of denaturant. The introduction of GdmCl expands the chain to a size of ~1.5 nm at 5 M GdmCl at 368.16 K. At 283.16 K, the expansion plateaus at 1 M GdmCl indicating the limit of expansion of the chain at 2.5 nm. Urea on the other hand brings about a net expansion by ~0.5 nm in R_g at 7 M. This relative expansion decreases with increasing temperature (Figure 5.4 A). Radii of gyration calculated for the pH 3.0 condition shows dramatic expansion of the chain size and at a moderate concentration of 3 M, the dynamic range of the FRET pair is lost (Figure 5.4 B). However, at 333.16 K, an expansion can still be discerned in the case of urea (red triangles). On an average Ran-BBL is far more sensitive to chaotropes than BBL. The congruent observations for BBL are plotted in Figures 5.5 A, B and C.

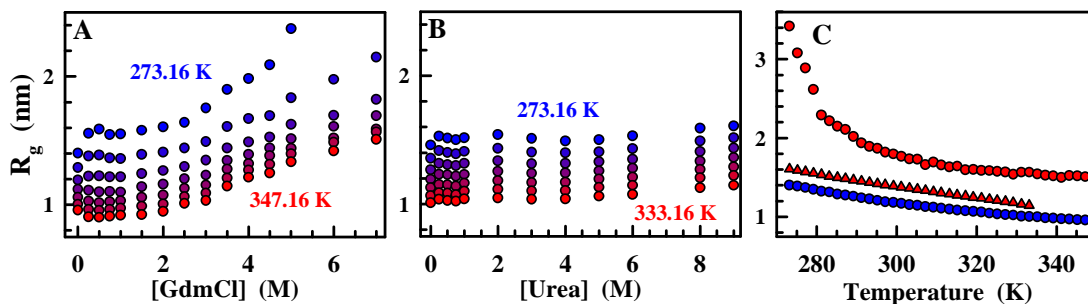


Figure 5.5 Response of BBL to chaotrope addition and temperature. A) Expansion of acid-denatured BBL with GdmCl at pH 3.0, with colours from blue to red indicate temperatures from 273.16 K to 347.16 K. B) Same as A, but in the presence of urea, up to a temperature of 333.16 K for the red circles. C) Blue circles are R_g 's in the absence of chemical, red circles in 7 M GdmCl and red triangles in 9 M urea

At 283.16 (10 K higher than the lowest temperature for measurements on BBL), Ran-BBL reaches its maximal expansion of 2.5 nm with the addition of just 0.25 M GdmCl, whereas even 7 M GdmCl does not bring about this saturation in

BBL (Figures 5.3 A and 5.5 A). Ran-BBL is also marginally more sensitive to urea than BBL. Any given concentration of urea does not cause an expansion of more than 0.2 nm in BBL (Figure 5.5 C). In case of Ran-BBL, this limit of 0.2 nm expansion is reached only at the highest temperature probed, i.e. 368.16 K. Thus, while the interaction of urea with BBL over the entire temperature range is constant, in case of Ran-BBL, the interaction becomes stronger resulting a marginally higher expansion by 0.5 nm at low temperature. It would thus appear, that the strength of the peptide-urea interaction has a stronger temperature dependence in Ran-BBL than in BBL.

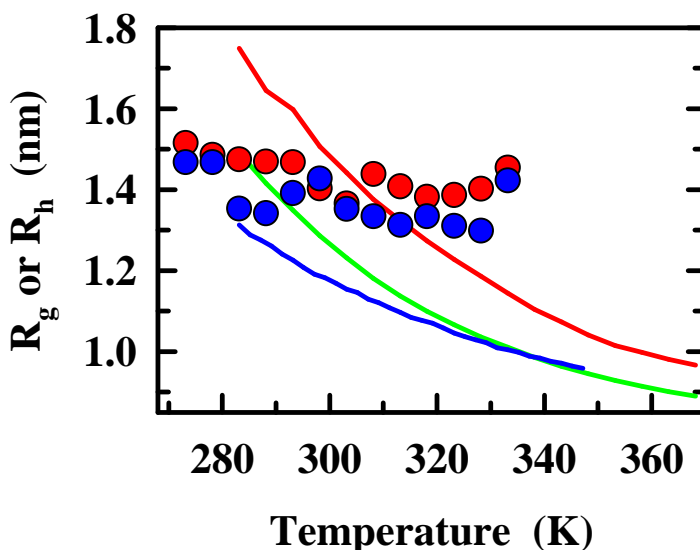


Figure 5.6 Hydrophobic collapse measured by DLS. Circles indicate hydrodynamic radii from DLS measurements. Blue and red circles are BBL at pH 3.0 and Ran-BBL at pH 3.0, respectively. Lines represent estimates from bulk FRET measurements. Blue, red and green lines are BBL at pH 3.0, Ran-BBL at pH 3.0 and Ran-BBL at pH 7.0, respectively.

Radii of gyration estimated from FRET measurements, and hydrodynamic radii measured by DLS for Ran-BBL and BBL at pH 3.0 (Figure 5.6) show good agreement. Finally, the degrees of collapse are compared of in Figure 5.6 with R_g from FRET measurements plotted as lines. Quite remarkably, the radii of gyration of

both BBL at pH 3.0 and Ran-BBL at pH 7.0 decay to identical values at high degrees of collapse (blue and green lines in Figure 5.6). Therefore, hydrophobic collapse by sequestration of protein residues from the solvent appears to be a self-averaging property in the case of BBL and Ran-BBL, and depends only on the sequence composition. At low temperatures, however, Ran-BBL is more open than BBL. CD spectra show that BBL retains residual structure at pH 3.0. This could explain the higher compaction observed for BBL at low temperatures. With increasing temperature, as this structure melts, the chain enters a self-averaging regime. Hydrophobic collapse is therefore, governed by the relative hydrophilic-hydrophobic content of the chain, rather than by sequence specific effects.

Ran-BBL at pH 3.0 is slightly more expanded than at pH 7.0 possibly due to electrostatic repulsions owing to a net charge of ~ 8.5 on the chain at pH 3.0. If a higher temperature regime were accessible, it is possible that Ran-BBL at pH 3.0 would collapse to the same size as BBL and Ran-BBL at pH 7.0. Thus, the hydrophobic effect, governed by temperature, and modulation of the net interactions of the protein with the solvent, is strong enough to overcome the high positive electrostatic potential on the chain, and charged repulsions.

5.3 Conclusions

Ran-BBL, a 40-residue polypeptide with a randomized sequence, and the same sequence composition as BBL, experiences chain expansion and collapse in much the same way as a natural sequence, in response to chemical denaturants and temperature increase. Thus, specific clustering of hydrophobic groups on the chain, do not seem to

be essential for causing a global collapse in the chain. Ran-BBL at pH 7.0 is more expanded than BBL at low temperature in the absence of cosolvent. However, Ran-BBL collapses to an identical R_g at high temperatures as BBL, indicating that hydrophobic collapse is controlled only by the content of a sequence. The relative ratio between hydrophilic and hydrophobic content, therefore modulates the net polymer-solvent interaction. Ran-BBL is more expanded at pH 3.0 than at pH 7.0 due to ~8.5 units of positive charge (~6 units higher than at pH 7.0). It is tempting to hypothesize that Ran-BBL at pH 3.0 may well collapse to the same R_g as its low-charge counterpart and BBL, if a higher temperature regime were accessible.

The stage is therefore set for measuring the rates of hydrophobic collapse of Ran-BBL in different scenarios. This is work in progress. The dynamics of chain reconfiguration during collapse can be measured by ultrafast laser techniques following T-jumps to identical temperature and to identical degrees of chain collapse (equivalent scenarios on at $T > \sim 330$ K). Furthermore, there exists the interesting possibility of deducing the effect of charge on chain collapse rates. We know now that a higher positive charge on the chain cause Ran-BBL to be further expanded. Would this electrostatic repulsion impede the dynamics of collapse as well? This could be determined after correcting for the differences in collapse rates due to different average sizes of the polymer based on a simple scaling law. These experiments will further elucidate the subtle balance between non-covalent interactions that leads to native structure and stability. Finally the generality of these results will be tested by similar experiments on other random sequences.

Chapter 6. Future Perspectives

The equilibrium effects of chaotropes, kosmotropes, temperature and pH on an unfolded protein were discerned by studies on acid-denatured BBL. Kinetics measurements, where the dynamics of hydrophobic collapse in the presence of chaotropes and kosmotropes can be studied by laser T-jump experiments with FRET as the experimental reporter. Protein chains with bound cosolvents, or with strong interactions with the solution, could collapse slower than in the absence of such interactions. Such a slow-down of collapse rates is bound to occur due to increased viscosities of solutions with added cosolvents. However, apart from this trivial effect, interactions of the chain with bound molecules could dramatically decrease its effective diffusion coefficient. Hydrophobic collapse in Ran-BBL can be studied in a varied range of thermodynamic conditions with pH, temperature, chaotrope and kosmotrope concentration as the variables. This would allow a detailed estimation of unfolded protein chain's properties in the absence of folding events. Such studies would provide experimental benchmarks that would improve theoretical simulations of the folding of polypeptides. Another area of research that has seen tremendous advances on the theoretical side is the study of folding events in conditions that mimic cellular milieu, such as the presence of macromolecular crowding agents or proteins in confined spaces. Studies on the fundamental interactions of a protein with its environment are thus critical for quantitatively understanding these processes.

Bibliography

1. Onuchic, J. N., LutheySchulten, Z. & Wolynes, P. G. Theory of protein folding: The energy landscape perspective. *Annual Review of Physical Chemistry* **48**, 545-600 (1997).
2. Pauling, L. & Corey, R. B. Configuration of polypeptide chains. *Nature* **168**, 550-551 (1951).
3. Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv Protein Chem* **14**, 1-63 (1959).
4. Tanford, C. How protein chemists learned about the hydrophobic factor. *Protein Sci* **6**, 1358-1366 (1997).
5. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29**, 7133-7155 (1990).
6. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* **7**, 95-& (1963).
7. Levintha.C. Are there pathways for protein folding. *Journal De Chimie Physique Et De Physico-Chimie Biologique* **65**, 44-& (1968).
8. Anfinsen, C. B. Principles That Govern Folding of Protein Chains. *Science* **181**, 223-230 (1973).
9. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, Pathways, and the Energy Landscape of Protein-Folding - a Synthesis. *Proteins-Structure Function and Genetics* **21**, 167-195 (1995).
- 1.0 Bryngelson, J. D. & Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* **84**, 7524-7528 (1987).
11. Bryngelson, J. D. & Wolynes, P. G. Intermediates and Barrier Crossing in a Random Energy-Model (with Applications to Protein Folding). *Journal of Physical Chemistry* **93**, 6902-6915 (1989).
12. Naganathan, A. N., Doshi, U., Fung, A., Sadqi, M. & Munoz, V. Dynamics, energetics, and structure in protein folding. *Biochemistry* **45**, 8466-8475 (2006).
13. Garcia-Mira, M. M., Sadqi, M., Fischer, N., Sanchez-Ruiz, J. M. & Munoz, V. Experimental identification of downhill protein folding. *Science* **298**, 2191-2195 (2002).
14. Naganathan, A. N., Sanchez-Ruiz, J. M. & Munoz, V. Direct measurement of barrier heights in protein folding. *J Am Chem Soc* **127**, 17970-17971 (2005).
15. Eaton, W. A. *et al.* Fast kinetics and mechanisms in protein folding. *Annual Review of Biophysics and Biomolecular Structure* **29**, 327-359 (2000).
16. Munoz, V. & Eaton, W. A. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 11311-11316 (1999).

17. Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. Folding dynamics and mechanism of beta-hairpin formation. *Nature* **390**, 196-199 (1997).
18. Munoz, V. Conformational dynamics and ensembles in protein folding. *Annu Rev Biophys Biomol Struct* **36**, 395-412 (2007).
19. Kubelka, J., Hofrichter, J. & Eaton, W. A. The protein folding 'speed limit'. *Curr Opin Struct Biol* **14**, 76-88 (2004).
20. Munoz, V. & Sanchez-Ruiz, J. M. Exploring protein-folding ensembles: A variable-barrier model for the analysis of equilibrium unfolding experiments. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17646-17651 (2004).
21. Tanford, C., Kawahara, K. & Lapanje, S. Proteins in 6-M guanidine hydrochloride. Demonstration of random coil behavior. *J Biol Chem* **241**, 1921-1923 (1966).
22. Wilkins, D. K. *et al.* Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424-16431 (1999).
23. Tcherkasskaya, O. & Uversky, V. N. Denatured collapsed states in protein folding: example of apomyoglobin. *Proteins* **44**, 244-254 (2001).
24. Kohn, J. E. *et al.* Random-coil behavior and the dimensions of chemically unfolded proteins. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12491-12496 (2004).
25. Millet, I. S., Doniach, S. & Plaxco, K. W. Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins. *Unfolded Proteins* **62**, 241-262 (2002).
26. Pitard, E. Influence of hydrodynamics on the dynamics of a homopolymer. *European Physical Journal B* **7**, 665-673 (1999).
27. Pitard, E. & Orland, H. Dynamics of the swelling or collapse of a homopolymer. *Europhysics Letters* **41**, 467-472 (1998).
28. Nettels, D., Gopich, I. V., Hoffmann, A. & Schuler, B. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2655-2660 (2007).
29. Moglich, A., Krieger, F. & Kiefhaber, T. Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains. *Journal of Molecular Biology* **345**, 153-162 (2005).
30. Shortle, D. & Ackerman, M. S. Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **293**, 487-489 (2001).
31. Ackerman, M. S. & Shortle, D. Robustness of the long-range structure in denatured staphylococcal nuclease to changes in amino acid sequence. *Biochemistry* **41**, 13791-13797 (2002).
32. Ohnishi, S. & Shortle, D. Observation of residual dipolar couplings in short peptides. *Proteins* **50**, 546-551 (2003).
33. Fiebig, K. M., Schwalbe, H., Buck, M., Smith, L. J. & Dobson, C. M. Toward a description of the conformations of denatured states of proteins. Comparison of a random coil model with NMR measurements. *Journal of Physical Chemistry* **100**, 2661-2666 (1996).

34. Barron, L. D., Blanch, E. W. & Hecht, L. Unfolded proteins studied by Raman optical activity. *Adv Protein Chem* **62**, 51-90 (2002).
35. 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. *Crit Rev Biochem Mol Biol* **40**, 181-189 (2005).
36. Schuler, B., Lipman, E. A. & Eaton, W. A. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743-747 (2002).
37. Nozaki, Y. & Tanford, C. The Solubility of Amino Acids and Related Compounds in Aqueous Urea Solutions. *J Biol Chem* **238**, 4074-4081 (1963).
38. Schellman, J. A. A Simple-Model for Solvation in Mixed-Solvents - Applications to the Stabilization and Destabilization of Macromolecular Structures. *Biophysical Chemistry* **37**, 121-140 (1990).
39. 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* **48**, 3743-3754 (2009).
40. 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. *Proc Natl Acad Sci U S A* **105**, 13403-13408 (2008).
41. Hua, L., Zhou, R., Thirumalai, D. & Berne, B. J. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc Natl Acad Sci U S A* **105**, 16928-16933 (2008).
42. O'Brien, E. P., Dima, R. I., Brooks, B. & Thirumalai, D. Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism. *J Am Chem Soc* **129**, 7346-7353 (2007).
43. Mountain, R. D. & Thirumalai, D. Alterations in water structure induced by guanidinium and sodium ions. *Journal of Physical Chemistry B* **108**, 19711-19716 (2004).
44. Mountain, R. D. & Thirumalai, D. Importance of excluded volume on the solvation of urea in water. *Journal of Physical Chemistry B* **108**, 6826-6831 (2004).
45. Tobi, D., Elber, R. & Thirumalai, D. The dominant interaction between peptide and urea is electrostatic in nature: A molecular dynamics simulation study. *Biopolymers* **68**, 359-369 (2003).
46. Xu, H. F. & Dill, K. A. Water's hydrogen bonds in the hydrophobic effect: A simple model. *Journal of Physical Chemistry B* **109**, 23611-23617 (2005).
47. Dill, K. A., Truskett, T. M., Vlachy, V. & Hribar-Lee, B. Modeling water, the hydrophobic effect, and ion solvation. *Annual Review of Biophysics and Biomolecular Structure* **34**, 173-199 (2005).
48. Urbic, T., Vlachy, V., Kalyuzhnyi, Y. V., Southall, N. T. & Dill, K. A. A two-dimensional model of water: Solvation of nonpolar solutes. *Journal of Chemical Physics* **116**, 723-729 (2002).
49. Southall, N. T., Dill, K. A. & Haymet, A. D. J. A view of the hydrophobic effect. *Journal of Physical Chemistry B* **106**, 521-533 (2002).

50. Hribar, B., Southall, N. T., Vlachy, V. & Dill, K. A. How ions affect the structure of water. *Journal of the American Chemical Society* **124**, 12302-12311 (2002).
51. Silverstein, K. A. T., Dill, K. A. & Haymet, A. D. J. Hydrophobicity in a simple model of water: solvation and hydrogen bond energies. *Fluid Phase Equilibria* **151**, 83-90 (1998).
52. Chan, H. S. & Dill, K. A. Solvation: How to obtain microscopic energies from partitioning and solvation experiments. *Annual Review of Biophysics and Biomolecular Structure* **26**, 425-459 (1997).
53. Chan, H. S., Krukowski, A. E. & Dill, K. A. Extracting hydrophobic parameters: Solvation effects of molecular size and shape. *Biophysical Journal* **70**, MP422-MP422 (1996).
54. Alonso, D. O. V. & Dill, K. A. Solvent Denaturation and Stabilization of Globular-Proteins. *Biochemistry* **30**, 5974-5985 (1991).
55. Stigter, D., Alonso, D. O. V. & Dill, K. A. Protein Stability - Electrostatics and Compact Denatured States. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 4176-4180 (1991).
56. Stigter, D. & Dill, K. A. Charge Effects on Folded and Unfolded Proteins. *Biochemistry* **29**, 1262-1271 (1990).
57. Auton, M. & Bolen, D. W. Application of the transfer model to understand how naturally occurring osmolytes affect protein stability. *Methods Enzymol* **428**, 397-418 (2007).
58. Ferreon, A. C. M., Rosgen, J., Ferreon, J. C. & Bolen, D. W. The complex dependence of protein conformational stability on salts. *Biophysical Journal* **86**, 345A-345A (2004).
59. Shi, Z. S., Chen, K., Liu, Z. G. & Kallenbach, N. R. Conformation of the backbone in unfolded proteins. *Chemical Reviews* **106**, 1877-1897 (2006).
60. Cacace, M. G., Landau, E. M. & Ramsden, J. J. The Hofmeister series: salt and solvent effects on interfacial phenomena. *Quarterly Reviews of Biophysics* **30**, 241-277 (1997).
61. Collins, K. D. & Washabaugh, M. W. The Hofmeister effect and the behaviour of water at interfaces. *Q Rev Biophys* **18**, 323-422 (1985).
62. Goto, Y., Takahashi, N. & Fink, A. L. Mechanism of acid-induced folding of proteins. *Biochemistry* **29**, 3480-3488 (1990).
63. Goto, Y. & Fink, A. L. Phase-diagram for acidic conformational states of proteins. *Journal of Molecular Biology* **214**, 803-805 (1990).
64. Goto, Y., Calciano, L. J. & Fink, A. L. Acid-induced folding of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 573-577 (1990).
65. Sadqi, M., Lapidus, L. J. & Munoz, V. How fast is protein hydrophobic collapse? *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12117-12122 (2003).
66. Sanchez-Puig, N., Veprintsev, D. B. & Fersht, A. R. Binding of natively unfolded HIF-1 alpha ODD domain to p53. *Molecular Cell* **17**, 11-21 (2005).
67. Sanchez-Puig, N., Veprintsev, D. B. & Fersht, A. R. Human full-length Securin is a natively unfolded protein. *Protein Science* **14**, 1410-1418 (2005).

68. Uversky, V. N. What does it mean to be natively unfolded? *European Journal of Biochemistry* **269**, 2-12 (2002).
69. Uversky, V. N. Natively unfolded proteins: A point where biology waits for physics. *Protein Science* **11**, 739-756 (2002).
70. Uversky, V. N., Gillespie, J. R. & Fink, A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**, 415-427 (2000).
71. Merchant, K. A., Best, R. B., Louis, J. M., Gopich, I. V. & Eaton, W. A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 1528-1533 (2007).
72. Jha, A. K., Colubri, A., Freed, K. F. & Sosnick, T. R. Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13099-13104 (2005).
73. Sinha, K. K. & Udgaonkar, J. B. Dependence of the size of the initially collapsed form during the refolding of barstar on denaturant concentration: Evidence for a continuous transition. *Journal of Molecular Biology* **353**, 704-718 (2005).
74. Hoffmann, A. *et al.* Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 105-110 (2007).
75. Jackson, S. E. & Fersht, A. R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428-10435 (1991).
76. 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? *J Chem Phys* **130**, 124903 (2009).
77. Schellman, J. A. & Gassner, N. C. The enthalpy of transfer of unfolded proteins into solutions of urea and guanidinium chloride. *Biophysical Chemistry* **59**, 259-275 (1996).
78. Schellman, J. A. Fifty years of solvent denaturation. *Biophysical Chemistry* **96**, 91-101 (2002).
79. Makhatadze, G. I. & Privalov, P. L. Protein interactions with urea and guanidinium chloride. A calorimetric study. *J Mol Biol* **226**, 491-505 (1992).
80. 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 Sci* **4**, 2138-2148 (1995).
81. Tobi, D., Elber, R. & Thirumalai, D. The dominant interaction between peptide and urea is electrostatic in nature: a molecular dynamics simulation study. *Biopolymers* **68**, 359-369 (2003).
82. Stumpe, M. C. & Grubmuller, H. Aqueous urea solutions: structure, energetics, and urea aggregation. *J Phys Chem B* **111**, 6220-6228 (2007).
83. Stumpe, M. C. & Grubmuller, H. Interaction of urea with amino acids: implications for urea-induced protein denaturation. *J Am Chem Soc* **129**, 16126-16131 (2007).

84. Stumpe, M. C. & Grubmuller, H. Polar or apolar--the role of polarity for urea-induced protein denaturation. *PLoS Comput Biol* **4**, e1000221 (2008).
85. Bennion, B. J. & Daggett, V. The molecular basis for the chemical denaturation of proteins by urea. *Proc Natl Acad Sci U S A* **100**, 5142-5147 (2003).
86. Courtenay, E. S., Capp, M. W. & Record, M. T., Jr. Thermodynamics of interactions of urea and guanidinium salts with protein surface: relationship between solute effects on protein processes and changes in water-accessible surface area. *Protein Sci* **10**, 2485-2497 (2001).
87. Wang, A. & Bolen, D. W. A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation. *Biochemistry* **36**, 9101-9108 (1997).
88. Robinson, D. R. & Jencks, W. P. The Effect of Compounds of the Urea-Guanidinium Class on the Activity Coefficient of Acetyltetraglycine Ethyl Ester and Related Compounds. *J Am Chem Soc* **87**, 2462-2470 (1965).
89. Moglich, A., Krieger, F. & Kiefhaber, T. Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains. *J Mol Biol* **345**, 153-162 (2005).
90. Timasheff, S. N. & Xie, G. Preferential interactions of urea with lysozyme and their linkage to protein denaturation. *Biophys Chem* **105**, 421-448 (2003).
91. Kohn, J. E. *et al.* Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci U S A* **101**, 12491-12496 (2004).
92. Shortle, D. R. Structural analysis of non-native states of proteins by NMR methods. *Curr Opin Struct Biol* **6**, 24-30 (1996).
93. Ostrovsky, B. & Baryam, Y. Motion of Polymer Ends in Homopolymer and Heteropolymer Collapse. *Biophysical Journal* **68**, 1694-1698 (1995).
94. Kuznetsov, Y. A., Timoshenko, E. G. & Dawson, K. A. Kinetics at the Collapse Transition of Homopolymers and Random Copolymers. *Journal of Chemical Physics* **103**, 4807-4818 (1995).
95. Gopich, I. V., Nettels, D., Schuler, B. & Szabo, A. Protein dynamics from single-molecule fluorescence intensity correlation functions. *J Chem Phys* **131**, 095102 (2009).
96. Imai, T., Kinoshita, M. & Hirata, F. Theoretical study for partial molar volume of amino acids in aqueous solution: Implication of ideal fluctuation volume. *Journal of Chemical Physics* **112**, 9469-9478 (2000).
97. Schellman, J. A. Protein stability in mixed solvents: A balance of contact interaction and excluded volume. *Biophysical Journal* **85**, 108-125 (2003).
98. Makhatadze, G. I. & Privalov, P. L. Protein Interactions with Urea and Guanidinium Chloride - a Calorimetric Study. *Journal of Molecular Biology* **226**, 491-505 (1992).
99. Schellman, J. A. The Thermodynamics of Solvent Exchange. *Biopolymers* **34**, 1015-1026 (1994).
100. Schellman, J. A. The Relation between the Free-Energy of Interaction and Binding. *Biophysical Chemistry* **45**, 273-279 (1993).

101. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109-1112 (1997).
102. Michalet, X., Weiss, S. & Jager, M. Single-molecule fluorescence studies of protein folding and conformational dynamics. *Chem Rev* **106**, 1785-1813 (2006).
103. Kuzmenkina, E. V., Heyes, C. D. & Nienhaus, G. U. Single-molecule FRET study of denaturant induced unfolding of RNase H. *J Mol Biol* **357**, 313-324 (2006).
104. Snow, C. D., Sorin, E. J., Rhee, Y. M. & Pande, V. S. How well can simulation predict protein folding kinetics and thermodynamics? *Annu Rev Biophys Biomol Struct* **34**, 43-69 (2005).
105. Deniz, A. A. *et al.* Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc Natl Acad Sci U S A* **97**, 5179-5184 (2000).
106. Ha, T. *et al.* Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc Natl Acad Sci U S A* **93**, 6264-6268 (1996).
107. Deniz, A. A. *et al.* Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of Forster distance dependence and subpopulations. *Proc Natl Acad Sci U S A* **96**, 3670-3675 (1999).
108. Talaga, D. S. *et al.* Dynamics and folding of single two-stranded coiled-coil peptides studied by fluorescent energy transfer confocal microscopy. *Proc Natl Acad Sci U S A* **97**, 13021-13026 (2000).
109. Jackson, S. E. & Fersht, A. R. Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry* **30**, 10436-10443 (1991).
110. Schuler, B. Single-molecule fluorescence spectroscopy of protein folding. *Chemphyschem* **6**, 1206-1220 (2005).
111. Rhoades, E., Cohen, M., Schuler, B. & Haran, G. Two-state folding observed in individual protein molecules. *J Am Chem Soc* **126**, 14686-14687 (2004).
112. Ziv, G., Thirumalai, D. & Haran, G. Collapse transition in proteins. *Phys Chem Chem Phys* **11**, 83-93 (2009).
113. Liu, P., Meng, X., Qu, P., Zhao, X. S. & Wang, C. C. Subdomain-specific collapse of denatured staphylococcal nuclease revealed by single molecule fluorescence resonance energy transfer measurements. *J Phys Chem B* **113**, 12030-12036 (2009).
114. Hofmann, H., Golbik, R. P., Ott, M., Hubner, C. G. & Ulbrich-Hofmann, R. Coulomb forces control the density of the collapsed unfolded state of barstar. *J Mol Biol* **376**, 597-605 (2008).
115. Merchant, K. A., Best, R. B., Louis, J. M., Gopich, I. V. & Eaton, W. A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc Natl Acad Sci U S A* **104**, 1528-1533 (2007).

116. Tezuka-Kawakami, T., Gell, C., Brockwell, D. J., Radford, S. E. & Smith, D. A. Urea-induced unfolding of the immunity protein Im9 monitored by spFRET. *Biophys J* **91**, L42-44 (2006).
117. Nienhaus, G. U. Exploring protein structure and dynamics under denaturing conditions by single-molecule FRET analysis. *Macromolecular Bioscience* **6**, 907-922 (2006).
118. Kragelund, B. B., Robinson, C. V., Knudsen, J., Dobson, C. M. & Poulsen, F. M. Folding of a four-helix bundle: studies of acyl-coenzyme A binding protein. *Biochemistry* **34**, 7217-7224 (1995).
119. Perl, D. *et al.* Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat Struct Biol* **5**, 229-235 (1998).
120. Yamasaki, K., Ogasahara, K., Yutani, K., Oobatake, M. & Kanaya, S. Folding pathway of Escherichia coli ribonuclease HI: a circular dichroism, fluorescence, and NMR study. *Biochemistry* **34**, 16552-16562 (1995).
121. Shortle, D. & Meeker, A. K. Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. *Proteins* **1**, 81-89 (1986).
122. Schreiber, G. & Fersht, A. R. The refolding of cis- and trans-peptidylprolyl isomers of barstar. *Biochemistry* **32**, 11195-11203 (1993).
123. Laurence, T. A., Kong, X., Jager, M. & Weiss, S. Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc Natl Acad Sci U S A* **102**, 17348-17353 (2005).
124. Kim, D. E., Fisher, C. & Baker, D. A breakdown of symmetry in the folding transition state of protein L. *J Mol Biol* **298**, 971-984 (2000).
125. Kremer, W. *et al.* Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium Thermotoga maritima. *Eur J Biochem* **268**, 2527-2539 (2001).
126. Schuler, B., Kremer, W., Kalbitzer, H. R. & Jaenicke, R. Role of entropy in protein thermostability: folding kinetics of a hyperthermophilic cold shock protein at high temperatures using 19F NMR. *Biochemistry* **41**, 11670-11680 (2002).
127. Sherman, E. & Haran, G. Coil-globule transition in the denatured state of a small protein. *Proc Natl Acad Sci U S A* **103**, 11539-11543 (2006).
128. Plaxco, K. W., Millett, I. S., Segel, D. J., Doniach, S. & Baker, D. Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nat Struct Biol* **6**, 554-556 (1999).
129. Scalley, M. L., Nauli, S., Gladwin, S. T. & Baker, D. Structural transitions in the protein L denatured state ensemble. *Biochemistry* **38**, 15927-15935 (1999).
130. Yi, Q., Scalley-Kim, M. L., Alm, E. J. & Baker, D. NMR characterization of residual structure in the denatured state of protein L. *J Mol Biol* **299**, 1341-1351 (2000).
131. Magg, C. & Schmid, F. X. Rapid collapse precedes the fast two-state folding of the cold shock protein. *J Mol Biol* **335**, 1309-1323 (2004).

132. Magg, C., Kubelka, J., Holtermann, G., Haas, E. & Schmid, F. X. Specificity of the initial collapse in the folding of the cold shock protein. *J Mol Biol* **360**, 1067-1080 (2006).
133. Friel, C. T., Beddard, G. S. & Radford, S. E. Switching two-state to three-state kinetics in the helical protein Im9 via the optimisation of stabilising non-native interactions by design. *J Mol Biol* **342**, 261-273 (2004).
134. Ferguson, N. & Fersht, A. R. Early events in protein folding. *Curr Opin Struct Biol* **13**, 75-81 (2003).
135. Shortle, D. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J* **10**, 27-34 (1996).
136. Carra, J. H. & Privalov, P. L. Thermodynamics of denaturation of staphylococcal nuclease mutants: an intermediate state in protein folding. *FASEB J* **10**, 67-74 (1996).
137. Ziv, G. & Haran, G. Protein Folding, Protein Collapse, and Tanford's Transfer Model: Lessons from Single-Molecule FRET. *J Am Chem Soc* **131**, 2942-2947 (2009).
138. Collins, K. D. Charge density-dependent strength of hydration and biological structure. *Biophys J* **72**, 65-76 (1997).
139. Von Hippel, P. H. & Wong, K. Y. On the conformational stability of globular proteins. The effects of various electrolytes and nonelectrolytes on the thermal ribonuclease transition. *J Biol Chem* **240**, 3909-3923 (1965).
140. Robien, M. A. *et al.* 3-Dimensional Solution Structure of the E3-Binding Domain of the Dihydrolipoamide Succinyl Transferase Core from the 2-Oxoglutarate Dehydrogenase Multienzyme Complex of Escherichia-Coli. *Biochemistry* **31**, 3463-3471 (1992).
141. Sadqi, M., Fushman, D. & Munoz, V. Atom-by-atom analysis of global downhill protein folding. *Nature* **442**, 317-321 (2006).
142. Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. Folding dynamics and mechanism of beta-hairpin formation. *Nature* **390**, 196-199 (1997).
143. Li, P., Oliva, F. Y., Naganathan, A. N. & Munoz, V. Dynamics of one-state downhill protein folding. *Proc Natl Acad Sci U S A* **106**, 103-108 (2009).
144. Munoz, V., Sadqi, M., Naganathan, A. N. & de Sancho, D. Exploiting the downhill folding regime via experiment. *HFSP J* **2**, 342-353 (2008).
145. Naganathan, A. N., Doshi, U. & Munoz, V. Protein folding kinetics: barrier effects in chemical and thermal denaturation experiments. *J Am Chem Soc* **129**, 5673-5682 (2007).
146. Naganathan, A. N., Perez-Jimenez, R., Sanchez-Ruiz, J. M. & Munoz, V. Robustness of downhill folding: guidelines for the analysis of equilibrium folding experiments on small proteins. *Biochemistry* **44**, 7435-7449 (2005).
147. Neuweiler, H., Johnson, C. M. & Fersht, A. R. Direct observation of ultrafast folding and denatured state dynamics in single protein molecules. *Proc Natl Acad Sci U S A* **106**, 18569-18574 (2009).
148. Huang, F., Ying, L. & Fersht, A. R. Direct observation of barrier-limited folding of BBL by single-molecule fluorescence resonance energy transfer. *Proc Natl Acad Sci U S A* **106**, 16239-16244 (2009).

149. Neuweiler, H. *et al.* The folding mechanism of BBL: Plasticity of transition-state structure observed within an ultrafast folding protein family. *J Mol Biol* **390**, 1060-1073 (2009).
150. Settanni, G. & Fersht, A. R. Downhill versus barrier-limited folding of BBL 3. Heterogeneity of the native state of the BBL peripheral subunit binding domain and its implications for folding mechanisms. *J Mol Biol* **387**, 993-1001 (2009).
151. Neuweiler, H. *et al.* Downhill versus barrier-limited folding of BBL 2: mechanistic insights from kinetics of folding monitored by independent tryptophan probes. *J Mol Biol* **387**, 975-985 (2009).
152. Arbely, E., Rutherford, T. J., Sharpe, T. D., Ferguson, N. & Fersht, A. R. Downhill versus barrier-limited folding of BBL 1: energetic and structural perturbation effects upon protonation of a histidine of unusually low pKa. *J Mol Biol* **387**, 986-992 (2009).
153. Ferguson, N., Sharpe, T. D., Johnson, C. M., Schartau, P. J. & Fersht, A. R. Structural biology: analysis of 'downhill' protein folding. *Nature* **445**, E14-15; discussion E17-18 (2007).
154. Ferguson, N. *et al.* Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family. *J Mol Biol* **353**, 427-446 (2005).
155. Ferguson, N., Schartau, P. J., Sharpe, T. D., Sato, S. & Fersht, A. R. One-state downhill versus conventional protein folding. *J Mol Biol* **344**, 295-301 (2004).
156. Oliva, F. Y. & Munoz, V. A simple thermodynamic test to discriminate between two-state and downhill folding. *J Am Chem Soc* **126**, 8596-8597 (2004).
157. Sun, S. T., Nishio, I., Swislow, G. & Tanaka, T. The coil-globule transition - Radius of gyration of polystyrene in cyclohexane. *Journal of Chemical Physics* **73**, 5971-5975 (1980).
158. Nettels, D. *et al.* Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc Natl Acad Sci U S A* (2009).
159. Lee, B. Solvent reorganization contribution to the transfer thermodynamics of small nonpolar molecules. *Biopolymers* **31**, 993-1008 (1991).
160. Makhatadze, G. I. & Privalov, P. L. Energetics of interactions of aromatic hydrocarbons with water. *Biophys Chem* **50**, 285-291 (1994).
161. Dill, K. A. Theory for the folding and stability of globular proteins. *Biochemistry* **24**, 1501-1509 (1985).
162. Stigter, D. & Dill, K. A. Charge effects on folded and unfolded proteins. *Biochemistry* **29**, 1262-1271 (1990).
163. Naganathan, A. N. & Munoz, V. Determining denaturation midpoints in multiprobe equilibrium protein folding experiments. *Biochemistry* **47**, 6752-6761 (2008).
164. Cho, S. S., Weinkam, P. & Wolynes, P. G. Origins of barriers and barrierless folding in BBL. *Proc Natl Acad Sci U S A* **105**, 118-123 (2008).
165. Dill, K. A. & Shortle, D. Denatured states of proteins. *Annu Rev Biochem* **60**, 795-825 (1991).

166. Sadqi, M., de Alba, E., Perez-Jimenez, R., Sanchez-Ruiz, J. M. & Munoz, V. A designed protein as experimental model of primordial folding. *Proc Natl Acad Sci U S A* **106**, 4127-4132 (2009).
167. Pabit, S. A., Roder, H. & Hagen, S. J. Internal friction controls the speed of protein folding from a compact configuration. *Biochemistry* **43**, 12532-12538 (2004).
168. Qiu, L. L., Zachariah, C. & Hagen, S. J. Fast chain contraction during protein folding: "Foldability" and collapse dynamics. *Physical Review Letters* **90** (2003).
169. Jacob, J., Krantz, B., Dothager, R. S., Thiyagarajan, P. & Sosnick, T. R. Early collapse is not an obligate step in protein folding. *J Mol Biol* **338**, 369-382 (2004).
170. Munoz, V. & Serrano, L. Helix design, prediction and stability. *Curr Opin Biotechnol* **6**, 382-386 (1995).
171. Munoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. *J Mol Biol* **245**, 297-308 (1995).
172. Munoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J Mol Biol* **245**, 275-296 (1995).
173. Munoz, V. & Serrano, L. Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices: comparison with experimental scales. *Proteins* **20**, 301-311 (1994).
174. Munoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. *Nat Struct Biol* **1**, 399-409 (1994).