ON THE FOLDING AND CONFORMATION OF PEPTIDES AND THE DEVELOPMENT OF NOVEL METHODS FOR THEIR STUDY

Matthias M. Waegele

A DISSERTATION

in

CHEMISTRY

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2011

Supervisor of Dissertation
Feng Gai,
Professor of Chemistry

Graduate Group Chairperson
Gary A. Molander,
Professor of Chemistry

Dissertation Committee Members:

Tobias Baumgart,
Assistant Professor of Chemistry

Jeffery G. Saven,
Associate Professor of Chemistry

DEDICATION

To the Memory of my Mother, Brigitte Angelika Waegele

ACKNOWLEDGMENTS

The past five years at the University of Pennsylvania have been a remarkable time. Over the years, I have learned a great deal and have matured as a scientist. My doctoral advisor, Professor Feng Gai, deserves most credit for this achievement by giving me both freedom and guidance in my scientific studies, by providing me with financial support and by vigorously discussing scientific and technical questions with me. He has been supportive and available throughout these years whenever I needed his help. His trust in my abilities gave me encouragement to surmount technical and scientific problems that I encountered along the way. I am tremendously thankful for not only his support, but also for his patience when things did not work out right away. I acknowledge my wife Tingting, my parents Brigitte and Hubert, my brother Michael, as well as my grandparents Martha, Oskar, Karoline and Hubert, without whose support and love this accomplishment would not have been possible. I furthermore thank Professors Tobias Baumgart and Jeffery G. Saven for serving on my dissertation committee. Over the years, I have been working together with Michelle Bunagan, Pramit Chowdhury, Robert Culik, Lin Guo, Thomas Measey, Smita Mukherjee, Ileana Pazos, Julie Rogers, Arnaldo Serrano, Kathryn Smith, Jia Tang and Matthew Tucker. Their kind nature has not only made it enjoyable to share the laboratory with them, but also to discuss scientific question with and to learn from them, which helped me greatly in achieving my goals. Finally, my friends

Christopher MacDermaid and Serenity Wang deserve credit for vigorously discussing scientific questions during our first year of study and beyond.

ABSTRACT

ON THE FOLDING AND CONFORMATION OF PEPTIDES AND THE DEVELOPMENT OF NOVEL METHODS FOR THEIR STUDY

Matthias M. Waegele

Feng Gai

Proteins are polymers, whose complex chemistry and physics, in the course of evolution, permitted their recruitment for immensely diverse functions in the biological machinery. Therefore, the study of protein conformational behavior and interactions of proteins with other entities of the intra- and extra-cellular milieu is of great interest. In this Thesis, we do not only quantitatively study the folding of peptide and protein systems using well-established spectroscopic techniques, but also, in recognizing the limits of existing experimental techniques, develop novel methods for studying protein folding, protein-protein and protein-membrane interactions.

In the first part of this Thesis, we are concerned with studying the folding process under various conditions using small peptides or proteins as model systems that allow a tractable, quantitative description of the order-disorder transition. Employing a combined site-directed mutagenesis/temperature-jump kinetics approach, we elucidated the folding mechanism of a helical hairpin, which is an important intermediate in protein folding reactions. We further considered how the folding kinetics of such model peptides and proteins are altered in the presence of high concentrations of

macromolecular cosolutes, which are thought to provide a simplified model system for the interior of the cellular environment. In another study, we addressed current questions concerning the structural properties of the unfolded state ensemble using a polypeptide that is unstructured by design as a model system.

The second part of this Thesis is focused on the development of novel methods to study both protein folding and interactions of proteins with other molecular species. In recent years, the nitrile stretching bands of nitrile-derivatized amino acid side-chains have emerged as attractive probes for monitoring biomolecular processes. In this Thesis, we theoretically and experimentally studied the molecular determinants of the bandwidth of the nitrile stretching band of 5-cyanotryptophan. We further evaluated the feasibility of using light-induced reductive cleavage of disulfide bonds as a means to phototrigger protein folding reactions.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

Proteins, being typically composed of many different monomeric units (i.e. the 20 natural amino acids and their derivatives), are natural heteropolymers whose complex conformational behavior and interactions with their environment have been attracting the interest of researchers for many decades. Of particular interest has been a protein's property to adoption of narrow conformational ensemble of well defined structures, i.e. the "folded" or "native" state. Due to their biological relevance, major efforts have been dedicated towards resolving the native structures of a large number of proteins since the elucidation of the first structure by x-ray crystallography by Kendrew [1] and Perutz [2], whose research groups resolved the structures of the proteins myoglobin and haemoglobin in the late 1950s and early 1960s. Around the same time, seminal experiments on the 124 amino acid protein ribonuclease where carried out by Anfinsen and coworkers [3, 4] that contributed to the understanding of how a structurally diverse protein ensemble is converted into the narrow ensemble that represents the native state and the biologically functional form of the protein. Specifically, they showed that this structural conversion is reversible, implying that, under appropriate solvent conditions, the amino acid sequence conveys the information on the structural characteristics of the native state. These results initiated an increasing interest in the kinetic and thermodynamic aspects of this structural transition, which is referred to as the protein folding transition. Early experiments on the kinetics of renaturation

of various proteins found evidence for the existence of intermediates [5, 6, 7, 8, 9], suggesting that the process of folding is associated with individual kinetic steps. The assignment of the multiple phases observed in the folding process in a number of proteins to specific structural events remained difficult and cumbersome as the employed spectroscopic tools, such as fluorescence or circular dichroism, did either not clearly distinguish between subpopulations of structures or only provided global structural insights, or both. The development of advanced nuclear magnetic resonance techniques in the 1980s provided a more detailed structural interrogation of the folding process.

In the course of the 1970s and 1980s, based on experimental findings and theoretical considerations, specific "nucleation sites", which were thought to facilitate the folding of proteins, were postulated [10, 11, 12, 13, 14, 15]. In this context, "nucleation" does not necessarily refer to a rate-limiting step, but rather describes the initial transient formation of small assemblies of secondary structure elements, such a helical segments, β-hairpins or bends, which allow the remainder of the protein to quickly undergo the structural transition to the native state. Consequently, the elucidation of the mechanisms for α-helix and β-hairpin formation became interesting. However, the most common experimental techniques employed at the time were limited to observation times of > 1 ms, preventing the detailed kinetic study of such small structural elements and only relatively indirect measurements on such

structures were reported [16]. In the mid-1990s, application of the temperature-jump technique [17, 18, 19, 20] allowed the study of folding mechanisms of these structural elements, which has greatly increased our knowledge of these early events in protein folding. Despite the large number of studies, which have been carried out since then, some potential "nucleation site" motifs have not been studied in detail. For example, due to the ubiquitous nature of helices and turns in the native structures of proteins, the helix-turn-helix motif (helical hairpin) might be considered a good candidate for a "nucleation site", which guides the rapid formation of helical proteins. Moreover, the helical hairpin motif plays a key role as a receptor site in DNA binding and protein-protein interactions [21, 22]. Thus, various helical hairpins have recently been developed to assess the factors that control the DNA and/or protein binding affinities of this structural motif and to form synthetic templates for protein and drug design [23, 24, 25]. Despite its importance as a crucial structural element in protein folding and binding, the folding mechanism of the helical hairpin motif has not been thoroughly studied as most helical hairpins are not thermodynamically stable in isolation, i.e. when removed from their native protein environment.

In Chapter 4, we investigate the structural determinants of the folding kinetics of a naturally occurring helical hairpin (porcine PYY) that is free of disulfide bonds and metal ion-induced cross-links using an infrared temperature-jump technique. This study represents the first detailed report on the folding mechanism of a helical hairpin

that is free of cross-links. It was found that mutations in the turn region predominantly increase the barrier of folding irrespective of the temperature, whereas the effect of mutations that perturb the hydrophobic interactions between the two helices is temperature-dependent. At low temperatures, deletion of hydrophobic side chains was found to predominantly affect the unfolding rate, while the opposite was observed at high temperatures. These results were interpreted in terms of a folding mechanism in which the turn is formed in the transition state and also based on the assumption that cross-strand hydrophobic contacts exist in the thermally unfolded state of PYY.

The study summarized in the preceding paragraph was carried out in an aqueous solution of the peptide. However, one might ask to what extent the results of such studies are transferable to the folding scenarios of peptides or proteins in their native environment, i.e. the cytoplasm of the cell. Specifically, the cellular environment exhibits high concentrations of macromolecules on the order of hundreds of g/L, which was shown to modulate biological processes, such as protein folding and protein-protein interactions [26, 27, 28]. Thus, this property of the cellular environment, termed "macromolecular crowding", is therefore intimately coupled to the process of protein folding *in vivo*. Theoretical treatments of the effect of macromolecular crowding on protein folding thermodynamics and kinetics have so far emphasized the role of excluded volume [26], i.e. space which is occupied by the macromolecular cosolute(s) is not available to the protein, which is therefore restricted to the space

between the macromolecular cosolutes. Thus, extended conformations of the protein are removed from the (unfolded) ensemble, giving rise to an entropic stabilization of the (compact) native state with respect to the unfolded state ensemble. While previous studies have provided invaluable insight into the effect of crowding on the stability and folding rate of protein tertiary structures, very little is known about how crowding affects protein folding dynamics at the secondary structure level.

In Chapter 5, we present the first study that addresses the effects of macromolecular crowding on the secondary structure level. Specifically, we examined the thermal stability and folding-unfolding kinetics of three small folding motifs (i.e., a 34-residue α-helix, a 34-residue cross-linked helix-turn-helix, and a 16-residue β-hairpin) in the presence of two commonly used crowding agents, dextran 70 (200 g/L) and ficoll 70 (200 g/L). We found that the presence of these polymers as cosolutes does not induce any appreciable changes in the folding kinetics of the two helical peptides, which is somewhat surprising as the helix-coil transition kinetics have been shown to depend on viscosity. Also to our surprise and in contrast to what has been observed for larger proteins, we found that crowding leads to an appreciable decrease in the folding rate of the shortest β-hairpin peptide, indicating that besides the excluded volume effect, other factors also need to be considered when evaluating the net effect of crowding on protein folding kinetics. A model considering both the static and the dynamic effects arising from the presence of the crowding agent was proposed to rationalize

these results.

The study summarized in the previous paragraph revealed some of the shortcomings of current crowding theories, which almost exclusively postulate the dominance of the excluded volume effect. A shortcoming of many experimental studies is the lack of quantitative comparison to theoretical models, which may help improve current theories. This problem is addressed in Chapter 6, where we performed the first quantitative experimental test of a recently proposed crowding model [29]. Specifically, the dependence of the melting temperature increase (ΔT_m) of the protein ubiquitin on the volume fraction (ϕ) of several commonly used macromolecular crowding agents (dextran 6, 40, and 70 and ficoll 70) was quantitatively examined and compared to a recently developed theoretical crowding model, i.e., $\Delta T_m \sim (R_g/R_c)^{\alpha}\phi^{\alpha/3}$. We found that in the current case this model correctly predicts the power-law dependence of ΔT_m on ϕ but significantly overestimates the role of the size (i.e., R_c) of the crowding agent. In addition, we found that for ubiquitin the exponent α is in the range of 4.1-6.5, suggesting that the relation of $\alpha = 3/(3\nu - 1)$ is a better choice for estimating α based on the Flory coefficient (ν) of the polypeptide chain. Taken together these findings highlight the importance of improving our knowledge and theoretical treatment of the microcompartmentalization of the commonly used model crowding agents.

The summaries of the preceding three Chapters highlight the importance of the

unfolded state ensemble for protein folding thermodynamics and kinetics. In Chapter 4, we saw evidence of pre-exisiting structure in the unfolded state, facilitating fast folding. In Chapter 5 and Chapter 6, the unfolded state was at the center of attention as its properties (specifically, its dimensions) play an important role in the observed modulation of the thermodynamics and kinetics brought about by macromolecular crowding. Additionally, a group of proteins has been discovered, which, unlike most proteins, are mostly void of any structure, yet are biologically relevant [30, 31, 32]. The conformational properties of these so called intrinsically disordered proteins (IDP) are of great interest as the elucidation of those might provide insights into their biological function. However, due to the absence of structure and the resulting heterogeneity, such polypeptides are difficult to study experimentally. In Chapter 7, we attempted to characterize the conformational distribution of a 29-residue polypeptide, which is unstructured by design. Specifically, we measured energy transfer efficiencies between seven different residue pairs. Relating the measured efficiencies to simple models of polymer theory, we found that the polypeptide cannot be described by a freely-jointed chain model [33] under the current solvent conditions, due to correlations between the propagation direction of peptide bond vectors. On the other hand, the Kratky-Porod (worm-like chain) model [34] seemingly provides a physically reasonable description, yet the recovered persistence length of 4.2 Å is surprisingly short. Fitting a model, which takes into account volume-exclusion of

the polymer showed that the peptide is more collapsed than what would be expected from the model, indicating that long-range attractive interactions between residues distant in sequence space might give rise to a more compact conformation, providing a possible explanation for the short apparent persistence length obtained from fitting the Kratky-Porod model, highlighting the difficulties and pitfalls in the application of simple polymer models to unstructured polypeptides.

The last three Chapters of this Thesis are concerned with the development of novel means for studying current problems concerning protein folding, protein-protein interactions, protein-membrane interactions and related problems. Whereas in Chapters 4-6 we employed the amide I band as a global spectroscopic reporter on the conformational transitions discussed therein, the results in Chapter 7 are based on a site-specific non-natural spectroscopic probe, i.e. *p*-cyanophenylalanine. As discussed in great detail in Chapter 3, the spectroscopic transitions of non-natural amino acid based probes sense the *local* environment around a *specific* side chain (as compared to the amide I vibration, which is *delocalized* over the entire backbone). Thus, these probes provide complementary information and are ideally suited for the study of a large array of biophysical problems. Specifically, the nitrile (C≡N) stretching vibration is sensitive to environment, making nitrile-derivatized amino acids an increasingly utilized tool to study various biological processes.

In Chapter 8, we showed that the bandwidth of the C≡N stretching vibration of 5-

cyanotryptophan is particularly sensitive to the degree of hydration of the side chain, rendering it an attractive infrared probe of local hydration status. We confirmed the utility of this probe in biological applications by using it to examine how the hydration status of individual tryptophan side chains of an anti-microbial peptide, indolicidin, changes upon peptide binding to model membranes. Furthermore, we showed that *p*-cyanophenylalanine and 5-cyanotryptophan constitute a useful fluorescence energy transfer pair.

In Chapter 9, we theoretically examined the experimental observation presented in the preceding Chapter that the nitrile stretching band of 5-cyanoindole is significantly broader than the corresponding vibrational transition of comparable nitrile compounds. Specifically, we investigated how interactions of water with various sites of 5-cyanoindole, the side chain of 5-cyanotryptophan, affect its C≡N stretching vibration via a combined electronic structure/molecular dynamics approach. It was found that, besides those interactions with the nitrile group, interactions of water with the indole ring also play a significant role in mediating the C≡N stretching frequency. Thus, this study provides a molecular basis for understanding how hydration affects the C≡N stretching band of 5-cyanotryptophan. In addition, an empirical model, which includes interactions of water with both the nitrile and indole groups, is developed for predicting the C≡N stretching vibrational band via molecular dynamics simulations.

In Chapter 10, we propose a new method for triggering protein folding reactions. As described earlier, in Chapters 4 and 5 laser-induced temperature-jump spectroscopy was used for quantifying the kinetics of conformational changes of peptides. This well-established technique has allowed valuable insights into the folding kinetics of the studied peptides. However, this technique has certain limitations on the conditions under which protein folding can be studied. For example, this technique provides no or limited control over the structures of the unfolded state ensemble. In Chapter 10, we pursued the development of a phototrigger, where a disulfide bond is used to constrain a peptide in a non-native state. In principle, this chemical bond can be introduced between any two residues in the peptide, providing control over the ensemble of unfolded structures. Specifically, we investigated the feasibility of lifting the covalent constraint by reductive cleavage of the disulfide bond via light-induced electron-transfer from an adjacent tryptophan residue. In equilibrium experiments, we found that this method of photoexcitation leads to a sufficiently high yield of bond cleavage such that an increase in helicity of a model helical peptide (sequence: $CW(AAAAK)_3WC$) after irradiation with UV light could be readily observed. In time-resolved experiments involving this model peptide, we observed transients on the μs timescale, which we tentatively attributed to formation of helical structure upon photoexcitation. This timescale suggests that the method is mostly suited for the study of proteins, which fold on slower timescales. We also found evidence of

aggregation upon photoexcitation, which could be due to polymerization of the generated peptide radical species.

Chapter 2

Theory

2.1 Protein Folding Thermodynamics

The thermodynamics of protein folding and unfolding have been comprehensively discussed in a number of reviews [35, 36, 37, 38, 39, 40, 41] and shall not be repeated here in great detail. Rather, a few essential concepts will be discussed that are relevant to the scientific discussion of the data presented in subsequent Chapters of this Thesis. For the sake of simplicity, a two-state scenario shall be considered, i.e. it is assumed that there are only two states, a folded (F) and an unfolded (U) state.

$$F \underset{k_f}{\overset{k_u}{\rightleftharpoons}} U \tag{2.1}$$

The Gibbs free energy of this reaction is [42]

$$\Delta_r G^\circ(T) = \Delta_r H^\circ(T) - T\Delta_r S^\circ(T) \tag{2.2}$$

where $\Delta_r G^\circ(T) = G^\circ_U(T) - G^\circ_F(T)$. The subscripts U and F refer to the unfolded and folded states, respectively. $\Delta_r H^\circ(T)$ and $\Delta_r S^\circ(T)$ are defined analogously. The temperature dependence of the enthalpy of the reaction ($\Delta_r H^\circ$) is given by Kirchhoff's law [42].

$$\Delta_r H^\circ(T) = \Delta_r H^\circ(T_m) + \int_{T_1}^{T_2} \Delta_r C_p \mathrm{d}T \tag{2.3}$$

Similarly, the temperature dependence of the reaction entropy ($\Delta_r S^\circ$) is

$$\Delta_r S^\circ(T) = \Delta_r S^\circ(T_m) + \int_{T_1}^{T_2} \frac{\Delta_r C_p}{T} \mathrm{d}T \tag{2.4}$$

Here, the reference temperature is the melting temperature of the protein, which is the temperature for which $\Delta_r G^\circ(T)$ in Eq. 2.2 is equal to zero, i.e.

$$T_m = \frac{\Delta_r H^\circ(T_m)}{\Delta_r S^\circ(T_m)} \tag{2.5}$$

The following discussion will focus mainly on the thermostability of proteins as in the work presented in this Thesis, heat was used as the primary agent for protein denaturation. First, some general aspects of Eq. 2.2 will be described, then a brief account on the molecular origins of the terms in Eq. 2.2 will be given.

The stability curve of a generic protein, i.e. a plot of $\Delta_r G^\circ(T)$ against temperature, is of concave down shape [43]. This curve crosses the T-axis at two points, T_c and T_m, describing the cold denaturation and heat denaturation points, respectively. The point of maximum stability, T^*, coincides with the maximum of the curve. When comparing two different proteins, the protein with a higher melting temperature is typically described as being "thermodynamically more stable". However, a more precise statement would be to say that the protein exhibits a larger thermal stability as the maximum thermal stability does not necessarily correlate with a higher melting temperature. In particular, one can envision three distinct mechanism that give rise to a higher thermal stability: A shift of the peak maximum of the stability curve to higher temperature, a broadening of the stability curve, and a shift of the stability curve along the free energy axis [43]. Empirically, a correlation between the T^* and T_m was found [41, 43], implying that the latter mechanism typically confers higher

thermal stability to a protein. Thus, if protein X has a higher melting temperature than protein Y, protein X is likely to be more stable than protein Y at any temperature, though this is not a thermodynamic necessity. For a more detailed discussions of protein thermostability, the reader is referred to Refs. [41, 43]. As evident from Eqs. 2.2, 2.3, and 2.4, the Gibbs free energy is composed of three thermodynamic quantities. Based on thermodynamic data of small, globular proteins, Robertson and Murphy found that each of the quantities correlate well with the number of residues N of the protein [40]. Specifically, they found

$$\begin{aligned} \Delta_r H^\circ(T = 333\,\mathrm{K}) &\simeq 2.92N\ \mathrm{kJ\,mol^{-1}\,residue^{-1}} \\ \Delta_r S^\circ(T = 333\,\mathrm{K}) &\simeq 0.088N\ \mathrm{kJ\,mol^{-1}\,residue^{-1}\,K^{-1}} \\ \Delta_r C_p &\simeq 0.058N\ \mathrm{kJ\,mol^{-1}\,residue^{-1}\,K^{-1}} \end{aligned} \tag{2.6}$$

As shown in Eq. 2.6, the thermodynamic quantities for protein unfolding are all positive. The positive value for $\Delta_r H^\circ$ is intuitively expected for the unfolding process as it involves the breaking of non-covalent bonds. One might also expect that the unfolding process is associated with an increase in $\Delta_r S^\circ(T)$, as the narrow set of native conformations is transferred into an ensemble of heterogeneous and nearly isoenergetic structures with large conformational freedom. The origin of the large positive value for $\Delta_r C_p$, however, is less obvious. It has been known for a long time, that the heat capacities of hydrophobic solutes in water are raised with respect to their heat capacities in the pure state [44, 45]. Thus, the increase of the heat capacity

is mainly attributed to the hydrophobic effect as hydrophobic protein residues are exposed to water in the course of the unfolding process [45, 46, 47]. Due to the large heat capacity and entropy changes, the Gibbs free energy of unfolding is strongly temperature dependent. Though the heat capacity is also temperature dependent, the difference in heat capacities between the folded and unfolded states ($\Delta_r C_p$) can typically be assumed to be independent of temperature [36] and therefore be taken outside the integrals in Eqs. 2.3 and 2.4.

2.2 Protein Folding Rates

In this Section, the rate equations for the common two-state folding case are considered. In particular, the kinetic model to which the data are fit is derived and some expressions for the rate constants are discussed. The Section is concluded by making the connection to the thermodynamic expressions described in the previous Section.

2.2.1 Perturbation of Equilibria

In this Section, the following folding/unfolding equilibrium of a two-state folder shall be considered.

$$F \underset{k_f}{\overset{k_u}{\rightleftharpoons}} U \tag{2.7}$$

At equilibrium, the free energy of the reaction is equal to zero ($\Delta_r G = 0$), i.e. there is no *net* change in the concentrations ($\frac{\mathrm{d}[F]}{\mathrm{d}t} = -\frac{\mathrm{d}[U]}{\mathrm{d}t} = 0$) [48]. Consequently, there

is only a static spectroscopic signal, which does not carry any information of the kinetics described in Eq. 2.7. Thus, at the ensemble level, extracting kinetics requires starting the reaction from a non-equilibrium state. The preparation of such a state requires the sudden perturbation of the equilibrium (Eq. 2.7), allowing the subsequent observation of the relaxation of the system from the prepared non-equilibrium state back to an equilibrium state. The choice of the perturbation depends on the specific system. In particular, the perturbation must be fast compared to the rates defined in Eq. 2.7. Common techniques include sudden changes in pressure [49], temperature [17, 18, 19, 20], or solvent composition [50], but also removal of a covalent restraint [51, 52]. The corresponding rate laws for Eq. 2.7 are [48]

$$\frac{\mathrm{d}[F](t)}{\mathrm{d}t} = k_f[U](t) - k_u[F](t) \tag{2.8}$$

$$\frac{\mathrm{d}[U](t)}{\mathrm{d}t} = k_u[F](t) - k_f[U](t) \tag{2.9}$$

If the system is out of equilibrium at a time t, then the concentrations of F and U are [48]

$$[F](t) = [F]_{eq} + \Delta[F](t) \tag{2.10}$$

$$[U](t) = [U]_{eq} + \Delta[U](t) \tag{2.11}$$

The free energy of the system is therefore $\Delta_r G(t) = \Delta_r G^\circ + RT\ln Q(t) > 0$ [48, 42], where $Q = [U]/[F]$. Combining Eqs. 2.10, 2.11, and 2.8, it is

$$\frac{\mathrm{d}\Delta[F](t)}{\mathrm{d}t} = k_f\left([U]_{eq} + \Delta[U](t)\right) - k_u\left([F]_{eq} + \Delta[F](t)\right) \tag{2.12}$$

As $k_f[U]_{eq} = k_u[F]_{eq}$ and $\Delta[F] = -\Delta[U]$ at all times t, it is

$$\frac{\mathrm{d}\Delta[F](t)}{\mathrm{d}t} = -\left(k_f + k_u\right)\Delta[F](t) \tag{2.13}$$

Integration of equation Eq. 2.13 provides the kinetic model to which the spectroscopic data (i.e. the transients observed in T-jump experiments) are to be fitted:

$$\Delta[F](t) = \Delta[F]_0 \exp[-(k_f + k_u)t] \tag{2.14}$$

Thus, the observed signal decays at a rate that is equal to the sum of the folding and unfolding rates, i.e. $k_{obs} = k_f + k_u$.

2.2.2 Expressions for the Rates of Protein Folding Reactions

In the previous Section, it was described how the relaxation rate of a protein folding reaction is measured experimentally. Though the relaxation rate provides information about the timescale of structural rearrangements in the protein, it is more informative to decompose the rate into its folding and unfolding components, which - combined with site-directed mutagenesis [53] - give mechanistic insights into the folding and unfolding processes. This decomposition requires the assumption of thermodynamic and kinetic models for the folding/unfolding processes. In this Section, the choice of the functional dependence of the rate constants on thermodynamic parameters will be discussed. The first functional expression of a rate constant on experimentally

derived parameters was given by Arrhenius [42].

$$k = A\exp\left[-\frac{E_a}{RT}\right] \tag{2.15}$$

In its original meaning, the activation energy of the reaction and the pre-exponential factor of this purely phenomenological expression are assumed to be temperature independent. Based on theoretical considerations, a related expression was later derived by Eyring, Evans and Polanyi [54, 55, 56]. The transition state theory result can be written as,

$$k = \frac{k_B T}{h}\exp\left[-\frac{\Delta G^{\ddagger}}{RT}\right] \tag{2.16}$$

where the pre-exponential factor is on the order of 10^{12} s^{-1}. Though explicit reference to Eyring's work is found in many publications on protein folding reactions, the assumptions underlying the derivation of this equation (bimolecular reaction in the gas-phase) do not apply to the protein folding problem. As a result, the pre-exponential factor is orders of magnitude to large for protein folding reactions, but comparisons of relative free energy barriers extracted from experimental data by fitting the Eyring equation to the data are valid, as long as the pre-exponential factor is a constant for the compared data sets. A better suited expression for a reaction coordinate coupled to the environment was derived by Kramers [57, 58], which takes into account that the crossing of the barrier is assisted by thermal noise and that there is energy transfer from the reaction coordinate to other degrees of freedom [58]. The

hydrodynamic approximation to Kramers' equation in the high friction limit reads

$$k = \frac{A}{\eta(T)} \exp\left[-\frac{\Delta G^{\ddagger}}{RT}\right] \tag{2.17}$$

where A is a constant. The equation has been useful for the investigation of the viscosity dependence of conformational changes in peptides and proteins [59]. However, as the viscosity (η) of the solvent is not necessarily equivalent to the microscopic friction experienced by the polypeptide chain [59], the equation is of limited use. From the above discussion, it is clear that there exists no rigorous expression for the rate constant that could be universally applied to protein folding experiments. In lieu of a theoretical formula for the rate constant of protein folding/unfolding, one can write - in the spirit of the Arrhenius equation - a phenomenological equation.

$$k = D \exp\left[-\frac{\Delta G^{\ddagger}}{RT}\right] \tag{2.18}$$

Eq. 2.18 contains both dynamical (D) and energetic terms (the exponential). The constant D describes the intrachain diffusion on the free energy potential. This diffusion term may be extracted from experiments by the appropriate sudden perturbation of the equilibrium (see Section 2.2.1), which puts a significant population on top of the energetic barrier [60, 61, 62, 63]. In this case, the barrierless diffusion of the chain may be observed. This diffusion process depends on the gradient of the free energy surface, i.e. $\Delta_r G = \left(\frac{\partial G}{\partial \xi}\right)$, along the reaction coordinate (ξ). Thus, the diffusion process, which carries dynamical information and is dependent on local friction, is

not independent of the energetics of the folding reaction. In the analysis of most experiments, the constant D is typically arbitrarily set to a constant for extraction of barrier heights and/or for decomposition of relaxation rates into folding and unfolding rates. A value on the order of 10^6 s^{-1} was proposed to be a reasonable value [64] for D. The significance of Eq. 2.18 in terms of the analysis of experimental relaxation rates as described by Eq. 2.14 is, that it allows the decomposition of the relaxation rate into folding and unfolding rates. Specifically, it is

$$\Delta_r G^\circ = -RT\ln K = RT\ln\frac{k_u}{k_f} \tag{2.19}$$

The free energy of unfolding (obtained from equilibrium experiments) is related to the free energies of activation, i.e. ($\Delta_r G^\circ = \Delta G_F^\ddagger - \Delta G_U^\ddagger$), which are - in analogy to Eq. 2.2 - related to the activation enthalpies, entropies and heat capacities. These quantities are used as fitting parameters in the analysis of relaxation rates.

2.3 Excitonic States and Energy Transfer

In subsequent Chapters, two seemingly unrelated approaches are being used for the characterization of protein conformation. While the first approach, relying on the amide I vibrational band of proteins, provides a global assessment of protein conformation, the second utilizes resonant energy transfer from an electronically excited chromophore (i.e. *p*-cyanophenylalanine) to another chromophore (i.e. tryptophan), providing a local measure of conformation. Specifically, information on the through-

space distance between the two chromophores on the polypeptide chain is obtained. Though the two techniques differ widely in their nature, i.e. whereas the former measures the absorption of vibrational chromophores, the latter measures the emission arising from the radiative decay of electronically excited states, the useful information arises in both methods from the coupling between transition moments. In particular, the former case falls into the strong coupling regime, thus creating an exciton state, whose wavefunction is delocalized along the entire backbone of the polypeptide, rendering the amide I vibrational band a global reporter on protein conformation. The latter case falls into the very weak coupling regime. In this regime, the wavefunctions remain localized on the chromophores and a rate of energy transfer can be defined. This rate is dependent on the distance and orientation between the transition moments of the two chromophores. In Sections 2.3.2 and 2.3.3, an expression for the rate of energy transfer and a coupling model for the so-called amide I subspace are discussed, respectively.

For the sake of simplicity, but without the loss of generality, a system consisting of two molecules (A and B) is considered, each of which possesses only two energy levels, i.e. the ground and excited states. The Hamiltonian of such a system is [65, 66, 67]

$$\hat{H} = \hat{H}_a + \hat{H}_b + \hat{V}_{ab} \tag{2.20}$$

$\hat{V}_{ab}$ has two types of matrix elements, $V'_{a,b} = \left\langle \phi'_a \phi_b \left| \hat{V}_{ab} \right| \phi'_a \phi_b \right\rangle$, which represents the Coulombic attraction and repulsion between the excited molecule A and the unexcited

molecule B, and $U = \left\langle \phi'_a \phi_b \left| \hat{V}_{ab} \right| \phi_a \phi'_b \right\rangle$, which may be interpreted as an interaction energy between the transition charge densities $\phi'_a \phi_b$ and $\phi_a \phi'_b$. The normalized and mutually orthogonal eigenvectors for Eq. 2.20 are

$$|\Psi_+\rangle = \cos\alpha \, |\psi'_a \psi_b\rangle + \sin\alpha \, |\psi_a \psi'_b\rangle \tag{2.21}$$

$$|\Psi_-\rangle = \sin\alpha \, |\psi'_a \psi_b\rangle - \cos\alpha \, |\psi_a \psi'_b\rangle \tag{2.22}$$

where ψ and ψ' correspond to the wavefunctions of the ground and excited states, respectively. The parameter α gives the extent of mixing between the two states defined Eqs. 2.21 and 2.22.

$$\tan 2\alpha = \frac{2U}{E_{a'b} - E_{ab'}} \quad \text{with} \quad 0 \leq \alpha \geq \pi/2 \tag{2.23}$$

where $E_{a'b} = E'_a + E^0_a + V'_{ab}$, and E'_a and w^0_a are the energies of the excited and ground states of molecule A, respectively. $\frac{2U}{E_{a'b} - E_{ab'}}$ is ratio of the magnitude of the coupling element U to the difference between the excitation energies of the two molecules, i.e. this ratio measures the strength of the coupling. The eigenvalues of the wavevectors of $|\Psi_+\rangle$ and $|\Psi_-\rangle$ are

$$W_\pm = \frac{1}{2} \left(E_{a',b} + E_{ab'} \right) \pm \frac{U}{\sin 2\alpha} \tag{2.24}$$

For various degrees of mixing α, there are two limiting cases. In the first case, where $2|U| \gg |E_{a'b} - E_{ab'}|$, i.e. $\alpha \sim \pi/4$, the wavefunctions are linear combinations of the locally excited states and the excitation is delocalized over the dimer system.

$$|\Psi_+\rangle = \frac{1}{\sqrt{2}} \left(|\psi'_a \psi_b\rangle + |\psi_a \psi'_b\rangle \right) \tag{2.25}$$

$$|\Psi_-\rangle = \frac{1}{\sqrt{2}}\left(|\psi'_a\psi_b\rangle - |\psi_a\psi'_b\rangle\right) \tag{2.26}$$

This is the strong coupling case. Experimentally, the strong coupling may be identified by major alterations in the absorption spectra of the dimer system compared to the spectra of the individual components (i.e. in this case, molecules A and B) [65]. This is found for the amide I vibrational absorption bands of proteins, whose band envelopes are typically very different from the amide I vibrational band of the monomer, i.e. N-methylacetamide (NMA).

In the second case, $2|U| \ll |E_{a'b} - E_{ab'}|$ i.e. $\alpha \sim 0$ or $\alpha \sim \pi/2$, the eigenstates are

$$|\Psi_+\rangle = |\psi'_a\psi_b\rangle \tag{2.27}$$

$$|\Psi_-\rangle = -|\psi_a\psi'_b\rangle \tag{2.28}$$

It is for the second, nonresonant (or very weakly resonant) case in which the excitation is essentially located on one molecule and a rate can be defined for the transition from one molecule to the other molecule. It is only in this very weak couling regime, i.e. $2|U| \ll |E_{a'b} - E_{ab'}|$, for which the theory of Förster [68, 65] applies.

2.3.1 Rate Equations of Energy Transfer between Fluorescent Molecules

For a single molecule (A), which was excited at a time $t = 0$ and which is surrounded by a solvent that is transparent in the spectral range of excitation and molecular emission, the change in probability (ρ) of finding the molecule in its excited state at

a time $t > 0$, is given by [68]

$$\frac{\mathrm{d}\rho_{A^*}(t)}{\mathrm{d}t} = -\frac{1}{\tau}\rho_{A^*}(t) = -\frac{1}{\tau}\rho_{A^*}(0)\exp\left(-\frac{t}{\tau}\right), \tag{2.29}$$

where $1/\tau$ is the average rate of the exponentially decaying excited state ensemble and A^* represents the excited state species of molecule A. This average rate is the sum of radiative ($1/\tau_f$) and non-radiative ($1/\tau_n$) rates of the corresponding deexcitation channels.

$$\frac{1}{\tau} = \frac{1}{\tau_f} + \frac{1}{\tau_n} \tag{2.30}$$

The ratio τ/τ_f determines the quantum yield (η) of the fluorescence, which may lie between 0 and 1.

$$\eta = \frac{1}{\tau_f}\int_0^\infty \rho_{A^*}(t)\mathrm{d}t = \frac{\tau}{\tau_f} \tag{2.31}$$

In addition of decaying according to Eq. 2.29, i.e

$$A^* \xrightarrow{\frac{1}{\tau}} A, \tag{2.32}$$

molecule A may also donate energy to a nearby molecule B with a rate k_T.

$$A^* + B \xrightarrow{k_T} A + B^* \tag{2.33}$$

The rate processes in Eqs. 2.32 and 2.33 may be written as

$$\frac{\mathrm{d}\rho_{A^*}(t)}{\mathrm{d}t} = -\frac{1}{\tau}\rho_{A^*}(t) - k_T\rho_{A^*}(t)\rho_B(t) \tag{2.34}$$

From Eq. 2.34 it is apparent, that non-radiative energy transfer between the molecules (the first term on the rhs) occurs simultaneously with radiative decay processes.

This mechanism of radiationless energy transfer between molecules is fundamentally different from the mechanism of reabsorption of fluorescence, which is a sequential process. Using Fermi's Golden Rule [69] as a starting point, Förster expressed the rate of radiationless energy transfer (k_T) in terms of experimental observables [68, 65], namely the absorption and emission spectra of the molecules B and A, respectively. Furthermore, provided that the excitation intensity is sufficiently low (as in most cases), the probability of finding molecule B at $t = 0$ in its ground state is equal to one. $\rho_B(t)$ is also assumed to be unity as back-transfer from B to A is not considered in the equation. Integration of Eq. 2.34 and insertion of the result into Eq. 2.31, where $1/\tau_f$ is substituted with $1/\tau$, gives for the probability that the energy is dissipated by mechanisms other than by transfer to molecule B:

$$\eta = \frac{1/\tau}{(1/\tau + k_T)} \tag{2.35}$$

As this probability (η) and the quantum yield of energy transfer (ϕ_T) need to sum up to unity, i.e. $\eta + \phi_T = 1$, the quantum yield of energy transfer is given by

$$\phi_T = \frac{k_T}{(1/\tau + k_T)} \tag{2.36}$$

In the following, a few key points in the derivation of the Förster equation (i.e., the expression for the rate k_T) and the assumptions implicit in the equation will be highlighted.

2.3.2 The Förster Equation of Resonance Energy Transfer

Let w_a (w_α) and w_b (w_β) represent the vibronic energies of the ground (excited) states of the donor (A) and acceptor (B) molecules, respectively. Each electronic state is associated with a spectrum of vibrational states. Initially, the donor is in a vibronically excited state $\psi_\alpha(q_A, Q_A)$, and the acceptor in one of its vibronic ground states $\psi_b(q_B, Q_B)$. Here, q and Q represent the coordinates of the electrons and nuclei, respectively. Via energy transfer from the donor to the acceptor, the donor is allowed to relax to one of the vibrational states of its electronic ground state and the acceptor is put in a vibrational level of an electronically excited state $\psi_\beta(q_B, Q_B)$. Thus, the process is described by [65].

$$\psi_\alpha\left(q_A, Q_A\right)\psi_b\left(q_B, Q_B\right) \rightarrow \psi_a\left(q_A, Q_A\right)\psi_\beta\left(q_B, Q_B\right) \tag{2.37}$$

For brevity, the vibronic wavefunctions for the initial and final states of the dimer system (donor + acceptor) can be defined as

$$\begin{gathered}\Psi_{\alpha,b}\left(q, Q\right) = \psi_\alpha\left(q_A, Q_A\right)\psi_b\left(q_B, Q_B\right) \\ \text{and} \\ \Psi_{a,\beta}\left(q, Q\right) = \psi_a\left(q_A, Q_A\right)\psi_\beta\left(q_B, Q_B\right),\end{gathered} \tag{2.38}$$

which are eigenfunctions of the vibronic Hamiltonian of the uncoupled system,

$$\hat{H}_0\Psi_{\alpha,b}\left(q_A, q_B\right) = (w_\alpha + w_b)\Psi_{\alpha,b}\left(q_A, q_B\right) \tag{2.39}$$

where the parametric dependence of the vibronic wavefunctions on the nuclear coordinates is suppressed (Born-Oppenheimer approximation). The wavefunctions of the initial states are normalized as usual,

$$\int \mathrm{d}q_A \, \psi_\alpha \left(q_A\right) \psi_{\alpha'} \left(q_A\right) = \delta_{\alpha\alpha'}, \tag{2.40}$$

where $\delta_{\alpha\alpha'}$ is the Kronecker delta. The normalization conditions for the wavefunctions of molecule B are analogously defined. The wavefunctions of the final states are normalized in the Dirac sense on an energy scale [70],

$$\begin{aligned} \int_{w_a} \int_{q_A} \mathrm{d}q_A \mathrm{d}w_a \, \psi \left(q_{A'}, w_{a'}\right) \psi \left(q_A, w_a\right) &= 1 \\ &= \int_{w_a} \mathrm{d}w_a \, \delta \left(w_a - w_a'\right) \\ &= \int_{w_a} \mathrm{d}a \, \frac{1}{n(a)} \delta(a - a') \frac{\mathrm{d}w_a}{\mathrm{d}a} \end{aligned} \tag{2.41}$$

The normalization constant $n(a)$ must obviously be equal to $\mathrm{d}a/\mathrm{d}w_a$ for Eq. 2.41 to hold. The different normalization procedures for the initial and final states arise as the system is initially in a *specific* vibronic state, but can undergo a transition into *any* of the vibrational states of the final electronic states, each being associated with a *continuum* of vibrational states. Thus, for the final states, the normalization procedure for wavefunctions associated with a continuum of states is chosen [70]. The total wavefunction of the system is

$$\begin{aligned} \Xi(q,t) &= a_{\alpha,b}(t) \Psi_{\alpha,b}(q,Q) \exp\left[-\frac{\mathrm{i}}{\hbar} \left(w_\alpha + w_b\right) t\right] \\ &\quad + \int_a \int_\beta a_{a,\beta}(t) \Psi_{a,\beta}(q,Q) \exp\left[-\frac{\mathrm{i}}{\hbar} \left(w_a + w_\beta\right) t\right] \mathrm{d}a \mathrm{d}\beta. \end{aligned} \tag{2.42}$$

Insertion of Eq. 2.42 into the time-dependent Schrödinger equation,

$\mathrm{i}\hbar\frac{\partial\Xi(q,t)}{\partial t} = \left(\hat{H}_0(q) + \hat{V}(q)\right)\Xi(q,t)$, taking into account the initial conditions $a_{\alpha,b}(0) = 1$ and $a_{a,\beta}(0) = 0$, multiplication of the resulting equation with

$\Psi_{a',\beta'}\exp\left[\frac{\mathrm{i}}{\hbar}\left(w_{a'} + w_{\beta'}\right)t\right]$, and integration over space under consideration of the normalization relations Eq. 2.40 and Eq. 2.41, results in

$$\mathrm{i}\hbar\frac{\partial a_{a',\beta'}(t)}{\partial t}\frac{1}{n(a')}\frac{1}{n(\beta')} = \left\langle a',\beta'\left|\hat{V}\right|\alpha,b\right\rangle \mathrm{e}^{-\frac{\mathrm{i}}{\hbar}\Delta W t} \tag{2.43}$$

Where $\Delta W = w_\alpha - w_{a'} + w_b - w_{\beta'}$ is the energy difference between initial and final states. Integration over time and taking the square of the resulting expression gives,

$$\begin{aligned} \left|a_{a',\beta'}(t)\right|^2 &= \left|\frac{n(a')n(\beta')}{\mathrm{i}\hbar}\right|^2 \left|\left\langle a',\beta'\left|\hat{V}\right|\alpha,b\right\rangle\right|^2 \left|\int_0^t \mathrm{d}t\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}\Delta W t}\right|^2 \\ &= \left|n(a')n(\beta')\right|^2 \left|\left\langle a',\beta'\left|\hat{V}\right|\alpha,b\right\rangle\right|^2 \sin^2\left[\frac{\Delta W t}{2\hbar}\right]\frac{4}{\Delta W^2} \end{aligned} \tag{2.44}$$

It is also

$$\int \mathrm{d}q\,\Psi^*\Psi = \sum_{\alpha,b}\left|a_{\alpha,b}\right|^2 + \int_a\int_\beta \left|a_{a,\beta}\right|^2 \frac{1}{n(a)}\frac{1}{n(\beta)}\mathrm{d}a\,\mathrm{d}\beta = 1 \tag{2.45}$$

Thus, $\left|a_{a,\beta}\right|^2\frac{1}{n(a)}\frac{1}{n(\beta)}\mathrm{d}a\,\mathrm{d}\beta$ gives the probability of finding the system in the range of quantum numbers $\mathrm{d}a\,\mathrm{d}\beta$. Integrating over the entire range of vibronic quantum states gives the probability that the system, which is initially in the vibronic states α and b, makes a transition into any of the final states a and β:

$$\begin{aligned} P_{\alpha,b\to a,\beta} &= \int_a\int_\beta \left|a_{a,\beta}\right|^2 \frac{1}{n(a)}\frac{1}{n(\beta)}\mathrm{d}a\,\mathrm{d}\beta \\ &= \int_{w_a}\int_{w_\beta} \left|a_{a,\beta}\right|^2 \frac{1}{n(a)}\frac{1}{n(\beta)}\frac{\partial(a,\beta)}{\partial(w_a,w_\beta)}\mathrm{d}w_a\,\mathrm{d}w_\beta \end{aligned} \tag{2.46}$$

Where $\frac{\partial(a,\beta)}{\partial(w_a,w_\beta)}$ is the Jacobian determinant of the coordinate transformation from quantum numbers to energies. Eq. 2.46 is the form of Fermi's Golden Rule as it was originally derived in Dirac's seminal article [69]. Here, $\frac{\partial(a,\beta)}{\partial(w_a,w_\beta)} = \frac{\mathrm{d}a}{\mathrm{d}w_a}\frac{\mathrm{d}\beta}{\mathrm{d}w_\beta}$ and the differentials of the Jacobian determinant in Eq. 2.46 disappear by virtue of the normalization (Eq. 2.41).

$$P_{\alpha,b\to a,\beta} = \int_{w_a}\int_{w_\beta} \left|\left\langle w_a, w_\beta \left|\hat{V}\right| w_\alpha, w_b\right\rangle\right|^2 \sin^2\left(\frac{\Delta W t}{2\hbar}\right)\frac{4}{\Delta W^2}\mathrm{d}w_a\mathrm{d}w_\beta \tag{2.47}$$

Upon energy transfer from molecule A to molecule B, the energy of molecule A is lowered by

$$W_a = w_\alpha - w_a \tag{2.48}$$

whereas the one of molecule B is increased by

$$W_b = w_\beta - w_b \tag{2.49}$$

The arithmetic mean of Eq. 2.48 and Eq. 2.49 define the total energy of the dimer system, W [68].

$$W = \frac{1}{2}\left(W_a + W_b\right) = \frac{1}{2}\left(w_\alpha - w_a + w_\beta - w_b\right) \tag{2.50}$$

The change in total energy upon transfer is

$$\Delta W = W_b - W_a = w_a + w_\beta - w_\alpha - w_b \tag{2.51}$$

Using the definition in Eq. 2.50 and Eq. 2.51, the vibronic energies of the final state may be expressed as

$$w_a = w_\alpha - W + \tfrac{1}{2}\Delta W$$

and (2.52)

$$w_\beta = w_b + W + \tfrac{1}{2}\Delta W$$

Following Förster's notation [68, 65], the interaction element $\left|\left\langle w_a, w_\beta \left| \hat{V} \right| w_\alpha, w_b \right\rangle\right|$ in Eq. 2.47 is set equal to $u\left(w_a, w_\beta; w_\alpha, w_b\right)$,

$$\begin{aligned} P_{\alpha,b\to a,\beta} &= \lim_{t\to\infty} \int_{w_a}\int_{w_\beta} \left|u\left(w_a, w_\beta; w_\alpha, w_b\right)\right|^2 \sin^2\left(\frac{\Delta W t}{2\hbar}\right) \frac{4}{\Delta W^2} \mathrm{d}w_a \mathrm{d}w_\beta \\ &= \frac{2\pi t}{\hbar} \int_{w_a}\int_{w_\beta} \left|u\left(w_a, w_\beta; w_\alpha, w_b\right)\right|^2 \delta\left(\Delta W\right) \mathrm{d}w_a \mathrm{d}w_\beta \end{aligned} \tag{2.53}$$

We now transform to new energy variables (Eq. 2.52) and integration over ΔW gives

$$P_{\alpha,b\to a,\beta} = \frac{2\pi t}{\hbar} \int_W \left|u\left(w_\alpha - W, w_b + W; w_\alpha, w_b\right)\right|^2 \mathrm{d}W \tag{2.54}$$

The vibronic wavefunctions of the initial and final states (Eq. 2.38) can be written as a product of electronic and nuclear wavefunctions (Born-Oppenheimer approximation).

$$\Psi\left(q_{A,B}, Q_{A,B}; w_\alpha, w_b\right) = \phi_A\left(q_A\right)\phi_B\left(q_B\right)\chi_A\left(Q_A, w_\alpha\right)\chi_B\left(Q_B, w_b\right) \tag{2.55}$$

$$\begin{aligned} &\Psi\left(q_{A,B}, Q_{A,B}; w_\alpha - W, w_b + W\right) \\ &= \phi_A\left(q_A\right)\phi_B\left(q_B\right)\chi_A\left(Q_A, w_\alpha - W\right)\chi_B\left(Q_B, w_b + W\right) \end{aligned} \tag{2.56}$$

Thus, the square of the interaction matrix element is

$$|u|^2 = \left|\int_{q_{A,B}}\int_{Q_{A,B}} \Psi_{a,\beta}\left(q_{A,B}, Q_{A,B}\right)\hat{V}\left(q_{A,B}\right)\Psi_{\alpha,b}\left(q_{A,B}, Q_{A,B}\right)\mathrm{d}q_{A,B}\mathrm{d}Q_{A,B}\right|^2 \tag{2.57}$$

$$\begin{aligned} |u|^2 &= \left|\int_{q_{A,B}} \psi_a(q_A)\psi_\beta(q_B)\hat{V}(q_{A,B})\psi_\alpha(q_A)\psi_b(q_B)\mathrm{d}q_{A,B}\right|^2 \times \\ &\qquad S_A^2(w_\alpha - W; w_\alpha) S_B^2(w_b + W; w_b) \\ &= U^2 S_A^2(w_\alpha - W; w_\alpha) S_B^2(w_b + W; w_b) \end{aligned} \tag{2.58}$$

Where $S_A^2(w_\alpha - W; w_\alpha)$ and $S_B^2(w_b + W; w_b)$ are the Franck-Condon factors of the respective electronic transitions. i.e. the overlap integrals between the ground and excited state vibrational wavefunctions. By truncating the multipole expansion of the interaction operator ($\hat{V} = \frac{1}{2}\sum_{i,j} \frac{e^2}{\left|\vec{R}_{AB} - (\vec{q}_{Ai} - \vec{q}_{Bj})\right|}$), the U^2 term is simply the interaction between the electronic transition dipoles of molecule A and B.

$$U^2 = \frac{\kappa^2 |\vec{m}_A|^2 |\vec{m}_B|^2}{n^4 R_{A,B}^6} \tag{2.59}$$

where κ is a factor that represents the average relative orientation between the donor and acceptor. The average is carried out over all configurations within the excited state lifetime of the donor. $R_{A,B}$ is the separation between donor and acceptor, n is the refractive index of the solvent, and $\vec{m}_A$ and $\vec{m}_B$ are the electronic transition dipole moments of donor and acceptor, respectively. Eq. 2.54 is therefore:

$$P_{\alpha,b\to a,\beta} = \frac{2\pi t}{\hbar} \frac{\kappa^2 |\vec{m}_A|^2 |\vec{m}_B|^2}{n^4 R_{A,B}^6} \int_\nu \left[\int_{w_\alpha} g(w_\alpha) S_A^2(w_\alpha - W; w_\alpha) \mathrm{d}w_\alpha\right] \times \left[\int_{w_b} g(w_b) S_B^2(w_b + W; w_b) \mathrm{d}w_b\right] \mathrm{d}\nu \tag{2.60}$$

Here, $g(w)$ is the (Boltzmann) weight of the energy level w. Using Einstein's coefficients for absorption and emission, the integrand in Eq. 2.60 can be connected to the normalized absorption and emission spectra of donor and acceptor [68, 65].

$$A_A(W) = \frac{4nW^3 |\vec{m}_A|^2}{3\hbar^4 c^3} \int_{w_\alpha} g(w_\alpha) S_A^2(w_\alpha - W; w_\alpha) \mathrm{d}w_\alpha = \frac{1}{\tau_0} f_A(W) \tag{2.61}$$

$$B_B(W) = \frac{2\pi |\vec{m}_B|^2}{3n^2\hbar^2} \int_{w_b} g(w_b) S_B^2(w_b + W; w_b) \mathrm{d}w_b = \frac{c\ln(10)}{nN_A W} \epsilon_B(W) \tag{2.62}$$

Combining Eqs. 2.60, 2.61 and 2.62 and expressing the energy in terms of frequency, i.e. $W = h\nu$, Förster's central equation for the rate of energy transfer is obtained:

$$k_T = \frac{9\ln(10)\kappa^2 c^4}{128\pi^5 n^4 N_A \tau_f R_{A,B}^6} \int_0^\infty f_A(\nu)\epsilon_B(\nu)\frac{1}{\nu^4}\mathrm{d}\nu \tag{2.63}$$

As evident from the derivation, the underlying assumption of this equation is that the coupling can be described by a dipole-dipole (convergent multipole) interaction between the electronic transition dipoles of the donor and acceptor (Eq. 2.59). This assumption may be made as long as the molecular dimensions of donor and acceptor are small compared to their separation distance. It is further assumed that there is very weak coupling between the donor and acceptor (Fermi's Golden Rule [69, 71]), such that the donor fluorescence lifetime, emission lineshape, acceptor absorption lineshape, and oscillator strengths are not perturbed by the coupling. The breakdown of Eq. 2.63 in the context of more complex scenarios has been discussed elsewhere [72]. However, the equation is considered to provide a good description of k_T in the context of the measurements described in this Thesis.

2.3.3 Transition Dipole Coupling in the Amide I Subspace

The amide I band (1600-1700 cm^{-1}) is the by far most frequently utilized vibrational band for the interrogation of protein secondary structure and solvation [73, 74, 75, 76]. The origin of its popularity for these purposes is manifold. First, the band arises from vibrations of the amide groups of the protein backbone and is thus a

spectral observable in all proteins. Second, the band arises mainly from the C=O stretching vibration of the amide group and it is spectrally separated from other features in the infra-red spectrum of proteins. Third, the band is readily observed due to its large extinction coefficient of ~ 720 M^{-1} cm^{-1} [77]. Fourth, through-bond and through-space couplings between the amide oscillators along the protein backbone, which delocalize the amide I vibration along the backbone, depend on the relative distances and orientations of the oscillators with respect to each other, rendering the amide I band particularly sensitive to the secondary structure conformation. Early on, the popularity of the vibrational band as an experimental observable for the characterization of protein conformation combined with the relatively simple nature of the transition (mainly C=O stretch) and the fact that the amide I vibration is largely decoupled from other transitions in proteins have lead to efforts to theoretical model the amide I band. In recent years, the quality of the modeling has seen significant improvements, largely because solvation effects are now taken into account more accurately [78, 79, 80, 81] and the coupling between adjacent oscillators is calculated quantum-mechanically [82]. Over the years, various formalisms have been proposed for the calculation of the amide I band [83, 84, 85, 86, 87, 74]. The model based on the theory of Torii and Tasumi [85] has been widely accepted as a standard model for this purpose. As described above, the amide I mode is not strongly influenced by any other modes in the system. The vibration is considered to be mainly C=O stretch

with only minor contribution from motions of the α-carbons, i.e. in the absence of coupling the amide I vibration is localized on a particular bond, largely unaffected by adjacent amide I vibrations. Thus, a simple model in which one oscillator is assigned to each amide group of the protein can be used as a starting point to calculate the amide I lineshape. Therefore, for a protein with a number of n amide groups, there will be n vibrational frequencies. Torii and Tasumi proceed by employing Wilson's **GF**-matrix formalism [88] for solving for these n frequencies.

$$|\mathbf{GF} - \mathbf{I}\lambda| = 0 \tag{2.64}$$

The **F**-matrix is an n by n matrix containing the force constants and the **G**-matrix arises from the kinetic energy part of the Hamiltonian. A related approach was taken by Hamm and Hochstrasser [87], who recast the problem in terms of a single matrix, representing both the kinetic and potential energies in units of cm^{-1}.

$$H = \sum_{i=1}^{n} H_{ii} + \sum_{i<j}^{n} V_{ij} \tag{2.65}$$

Here, the diagonal matrix elements, H_{ii} , of the Hamiltonian are the vibrational frequencies of the oscillators in the absence of coupling and the off-diagonal matrix elements, V_{ij}, describe the coupling between the oscillators. The index n is equal to the number of amide groups in the protein. The appropriate choice of the diagonal elements of the Hamiltonian has been a long-standing problem as solvation effects influence these elements. As hydrogen bonding with the C=O oscillators is thought

to be the predominate solvation effect, shifts in the diagonal frequency were formerly calculated as [89]

$$\delta\nu_H = \alpha_H(r_{\mathrm{OH}} - 2.6) \tag{2.66}$$

where $\alpha_H$is chosen to by 30 $\mathrm{cm}^{-1}/\text{Å}$ and r_{OH} is the C=O—HX distance in Å. Typically, the C=O—HX angle was required to fall within a predefined range of angles for this interaction to be counted as a hydrogen bond. As noted by Kwac and Cho [90], this simple relationship does not quantitatively account for solvation effects. They proposed to calculate the vibrational shift by calculating the electrostatic potential on selected sites of the *N*-methylacetamide. This electrostatic potential arises from the partial charges on nearby solvent molecules. Within this framework, the diagonal elements of Eq. 2.65 may be calculated as

$$\omega = \omega_{\mathrm{gas}} + \sum_{i=1} \alpha_i \phi_i \tag{2.67}$$

where ϕ_i is the electrostatic potential at site i on *N*-methylacetamide and ω_{gas} is the frequency of the amide I band in vacuum. For the off-diagonal elements, i.e. the coupling terms , V_{ij}, in Eq. 2.65 a transition dipole coupling mechanism (TDM) is typically assumed for non-adjacent oscillators.

$$V_{ij} = A\frac{\hat{\mu}_i \cdot \hat{\mu}_j - 3\left(\hat{\mu}_i \cdot \hat{n}_{ij}\right)\left(\hat{\mu}_j \cdot \hat{n}_{ij}\right)}{r_{ij}^3} \tag{2.68}$$

where the magnitude of the transition dipole moment, A, is equal to 580 $\mathrm{cm}^{-1}\text{Å}^3$ [91], $\hat{\mu}_i$ is the unit vector of the transition dipole moment of the ith amide I mode, $\hat{n}_{ij}$

is the unit vector pointing from dipole i to dipole j, and r_{ij} is the distance between the dipoles. The magnitudes of transition dipoles are assumed to be the same for all amide units. This assumption is justified by the fact that, experimentally, the relative intensities of the components of the amide I band are thought to be proportional to the population of the associated conformation [85]. The location of the transition dipole was measured to be tilted by 20° away from the C=O bond axis [92]. Though the above expression was used to calculate coupling terms for adjacent amide groups [89] it is more appropriate to use quantum-mechanically derived look-up tables for the calculation of these terms due to the possible break down of the point-dipole approximation and through-bond coupling effects [82]. An appropriate ensemble of protein structures is typically generated by evolving the nuclear magnetic resonance structure of the protein of interest using molecular dynamics simulations [93, 94, 95, 96, 97, 91].

Upon diagonalization of the Hamiltonian, the frequencies (eigenvalues) and the different vibrational modes (eigenvectors) of the entire exciton system are obtained. The dot product of an eigenvector $\vec{T}_n$ with the $n \times 3$ transition dipole matrix $\mathbf{d}$ representing the Cartesian components of the transition dipoles for the n peptide groups gives the transition dipole intensities for the corresponding frequency [97]. Thus, the projection of the Cartesian transition dipole moment components onto the vibrational mode vector (of the n-dimensional subspace) gives the transition dipole

moment for that particular mode.

The procedure described above results in a stick spectrum. For comparison of the theoretical result with an experimental spectrum, each spectral line in the stick spectrum is convoluted by a Lorentzian with FWHM of 5 cm^{-1} to account for the vibrational lifetime of the vibrationally excited state (1 ps). Inhomogeneous broadening is taken into account by averaging over multiple structures.

It is noteworthy to point out that in some older formalisms, the phases of the oscillators explicitly appear in the arguments of cosine terms [74]. The relative phase of the oscillators is important and may give rise to complicated band structures. For example, β-sheets exhibit a lower frequency peak ($\sim$ 1630 cm^{-1}) and a higher frequency peak ($\sim$ 1680 cm^{-1}), which are attributed to different spatial patterns of the relative phase relationships between amide I oscillators [74, 76]. Though it is not apparent that the formalism described above contains this phase information, additional mathematical procedures may be used to extract this information [97].

2.4 Theory of Vibrational Lineshapes

The absorption line shape function may be written in terms of the time autocorrelation function of the transition dipoles [98, 99, 66].

$$I(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega t} \left\langle \vec{\mu}(t) \cdot \vec{\mu}(0) \right\rangle \tag{2.69}$$

Eq. 2.69 can be derived from Fermi's Golden Rule [69, 71] and the interested reader is referred to Refs. [98, 99] where detailed derivations are presented. Classically, fluctuations in the frequency ω and transition dipole moment (μ) may be considered coupled according to [100, 66]

$$\frac{\mathrm{d}\mu(t)}{\mathrm{d}t} = -\mathrm{i}\omega(t)\mu(t) \tag{2.70}$$

Integration of this equation with respect to time gives

$$\mu(t) = \mu(0)\exp\left[-\mathrm{i}\int_0^t \omega(\tau)\mathrm{d}\tau\right] \tag{2.71}$$

which, combined with the definition of the correlation function defined in Eq. 2.69, provides the following classical expression for the dipole-dipole correlation function:

$$C(t) = \langle\vec{\mu}(t)\cdot\vec{\mu}(0)\rangle = |\vec{\mu}(0)|^2\exp\left[-\mathrm{i}\left\langle\omega\right\rangle t\right]\left\langle\exp\left[-\mathrm{i}\int_0^t \delta\omega(\tau)\mathrm{d}\tau\right]\right\rangle \tag{2.72}$$

where $\omega = \langle\omega\rangle + \delta\omega(t)$ was used. Using the cumulant expansion method [101], the ensemble average may be conveniently moved to the exponential:

$$\left\langle\exp\left[-\frac{\mathrm{i}}{\hbar}\int_0^t \delta\omega(\tau)\mathrm{d}\tau\right]\right\rangle$$
$$= \exp\left[-\frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}\tau\,\langle\delta\omega(\tau)\rangle + \left(\frac{-\mathrm{i}}{\hbar}\right)^2\int_0^t \mathrm{d}\tau_2\int_0^{\tau_2}\mathrm{d}\tau_1\,\langle\delta\omega(\tau_2)\delta\omega(\tau_1)\rangle + ...\right] \tag{2.73}$$

For a system with a Gaussian frequency distribution, all terms of the cumulant expansion, but the second one, vanish. The lineshape is therefore given by

$$I(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}t\,e^{\mathrm{i}\omega t}\exp\left[-g(t) - |t|/2T_1\right] \tag{2.74}$$

The $\exp[-|t|/2T_1]$ term represents the decay of the population in the excited state, which has a lifetime of T_1. For vibrational excitation, T_1 is typically on the order of a few picoseconds. The $\exp[-g(t)]$ term contains the structural and dynamical properties of the bath (i.e. the part of the system that is not the oscillator of interest), whose interaction with the oscillator modulate its vibrational frequency. $g(t)$ is the integral over the frequency-frequency time correlation function, i.e.

$$g(t) = \int_0^t \mathrm{d}\tau_2 \int_0^{\tau_2} \mathrm{d}\tau_1 R(\tau_1) \tag{2.75}$$

where $R(t)$ is

$$R(t) = \langle \delta\omega(t)\delta\omega(0) \rangle \tag{2.76}$$

The fluctuations described by Eq. 2.76 are characterized by a variance ($\Delta = \langle \delta\omega^2 \rangle^{1/2}$) and a correlation time ($\tau_c = \frac{1}{\Delta^2} \int_0^\infty \langle \delta\omega\delta\omega \rangle \, \mathrm{d}t$). It is instructive to assume an exponential form for the frequency-frequency time correlation function (Eq. 2.76), i.e. $R(t) = \Delta^2 \exp[-t/\tau_c]$ [66], and to consider the following limiting cases: If the correlation time is very long ($t \ll \tau_c$), the correlation function is a constant equal to the variance of the sampled frequencies, i.e. $R(t) = \Delta^2$, and the lineshape is a Gaussian function with a full width at half maximum of 2Δ. In this regime, the lineshape is inhomogeneously broadened and is a direct measure of the structural heterogeneity of the system. On the other hand, if the correlation time is short ($t \gg \tau_c$), then the oscillator samples all the frequencies on a very fast timescale compared to $1/\Delta$, which results in a motionally narrowed Lorentzian lineshape.

More generally, Eq. 2.72 may be recast in quantum-mechanical terms, where $\delta\omega$ is represented by the time-dependent operator $\delta\Omega(t)$ and $\langle ... \rangle$ represents a quantum-mechanical trace over bath degrees of freedom rather than a classical ensemble average: $\left\langle \exp_+ \left[-\mathrm{i} \int_0^t \delta\Omega(\tau) \mathrm{d}\tau \right] \right\rangle$ [90]. Then, $g(t)$ is the integral over a quantum-mechanical correlation function $\xi(t)$ of the fluctuation operator $\delta\Omega$, i.e.

$$\xi(t) = \langle \delta\Omega(t)\delta\Omega(0) \rangle = \xi_R(t) + \xi_I(t) \tag{2.77}$$

whose real and imaginary parts are

$$\xi_R(t) = \frac{1}{2} \left[\langle \delta\Omega(t)\delta\Omega(0) \rangle + \langle \delta\Omega(0)\delta\Omega(t) \rangle \right] \tag{2.78}$$

$$\xi_I(t) = \frac{\mathrm{i}}{2} \left[\langle \delta\Omega(t)\delta\Omega(0) \rangle - \langle \delta\Omega(0)\delta\Omega(t) \rangle \right] \tag{2.79}$$

$\xi_R(t)$ is an even function of time and $\xi_I(t)$ is an odd function of time [98]. The spectral density if defined as the Fourier transform of the correlation function (Eq. 2.77)

$$\tilde{\xi}(\omega) \equiv \int_{-\infty}^{\infty} \mathrm{d}t \, \mathrm{e}^{\mathrm{i}\omega t} \xi(t) = 2\mathrm{Re} \int_0^{\infty} \mathrm{d}t \, \mathrm{e}^{\mathrm{i}\omega t} \xi(t) \tag{2.80}$$

The second equation in Eq. 2.80 holds due to the symmetry properties of QM correlation functions [98]. Thus, it is

$$\tilde{\xi}(\omega) = \tilde{\xi}_R(\omega) + \tilde{\xi}_I(\omega) \tag{2.81}$$

Both $\tilde{\xi}_R(\omega)$ and $\tilde{\xi}_I(\omega)$ are real quantities. $\tilde{\xi}_R(\omega) = \tilde{\xi}_R(-\omega)$ is an even function, and $\tilde{\xi}_I(\omega) = -\tilde{\xi}_I(-\omega)$ is an odd function. Moreover, $\tilde{\xi}(\omega)$ fulfills the detailed balance condition

$$\tilde{\xi}(-\omega) = \exp\left[-\beta\hbar\omega\right] \tilde{\xi}(\omega) \tag{2.82}$$

Eq. 2.82 follows from the property of the quantum-mechanical frequency correlation function, whose Fourier transform is: $\tilde{\xi}(\omega) = \sum_{m,n} |\mu_{mn}|^2 \delta(\omega - \omega_{mn})$. Considering up/down transitions between specific states i and j, it is $\tilde{\xi}(\omega_{ij})/\tilde{\xi(\omega}_{ji}) = p_i/p_j = \exp(-\beta\omega_{ij})$, which is equivalent to Eq. 2.82 [66]. Solving for $\tilde{\xi}_R(\omega)$ or $\tilde{\xi}_I(\omega)$, it is

$$\tilde{\xi}_R(\omega) = \frac{\exp\left[-\beta\hbar\omega\right] + 1}{\exp\left[-\beta\hbar\omega\right] - 1}\tilde{\xi}_I(\omega) = \coth\left(\frac{\beta\hbar\omega}{2}\right)\tilde{\xi}_I(\omega) \tag{2.83}$$

$$\tilde{\xi}_I(\omega) = \frac{\exp\left[-\beta\hbar\omega\right] - 1}{\exp\left[-\beta\hbar\omega\right] + 1}\tilde{\xi}_R(\omega) = \tanh\left(\frac{\beta\hbar\omega}{2}\right)\tilde{\xi}_R(\omega) \tag{2.84}$$

The real part of the quantum-mechanical frequency time autocorrelation function may be approximated by a classical correlation function, such as the one obtained from a molecular dynamics simulation ($\xi_R(\omega) \cong M(t)$). From Eq. 2.80, it is $\tilde{\xi}_R(\omega) = 2\int_0^\infty \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t} M(t)$ and the imaginary part is obtained by the fluctuation-dissipation relationship given in Eq. 2.84. The lineshape broadening function is then given by

$$g(t) = \frac{1}{\pi}\int_0^\infty \mathrm{d}\omega \coth\left(\frac{\hbar\omega}{2k_BT}\right)\tilde{\xi}_I(\omega)\frac{1-\cos(\omega t)}{\omega^2} + \frac{\mathrm{i}}{\pi}\int_0^\infty \mathrm{d}\omega\, \tilde{\xi}_I(\omega)\frac{\sin(\omega t) - \omega t}{\omega^2} \tag{2.85}$$

Chapter 3

Methods

3.1 Introduction

3.2 Laser-induced Temperature-jump Method

The temperature-jump (T-jump) methods owes its utility to the temperature sensitivity of most chemical equilibria, i.e.

$$\left(\frac{\partial K_{eq}}{\partial T}\right)_p = \frac{\Delta H_r^\circ}{RT^2} \tag{3.1}$$

As discussed in Chapter 2, for the observation of chemical kinetics, a non-equilibrium state needs to be created, i.e. the rise in temperature must be much faster than the relaxation time of the system. A T-jump apparatus based on the discharge of a capacitor through a conducting sample solution was introduced by Eigen and coworkers in the late 1950s [102], which allowed them to observe relaxation processes of chemical reactions that occur on timescales that were previously thought to be inaccessible to experimental scrutiny. However, to achieve sufficiently high temperature jumps of a few Kelvins ($\Delta T \sim CU^2$), a capacitor with a high capacity, C, needs to be employed. As the charge is driven through the solution by high voltages in the kilovolt range, undesired sample polarization may occur in the case of biological molecules [103], which often exhibit an anisotropic charge distribution. In addition, as the charge of a capacitor decays with $\mathrm{e}^{-t/RC}$, achievement of sufficiently short heating times (typically on the order of μs) requires samples of low resistivity, R, i.e. the sample must be necessarily conductive, requiring samples with high ionic strength [104, 103]. With the

advent of pulsed lasers, the limitations of the capacitor based T-jump method could be circumvented by employing absorptive heating of the sample through appropriate laser pulses [104, 105, 17, 18, 19, 20]. While some laser-induced T-jump experiments rely on the absorption of light energy by dye cosolutes of the sample of interest [106], an alternative approach, which was first demonstrated by Beitz *et al.* [107], utilizes mid-infrared laser pulses, which heat the sample through direct absorption of light energy by the solvent. The time resolution of the laser induced T-jump method is limited only by the thermalization of solvent and solute, which occurs within a few tens of picoseconds [106]. The laser-induced T-jump method was used for some of the work presented in this Thesis and is discussed in more detail in the following.

The setup utilizes pulses (ω = 1064 nm, 3 ns, 300 mJ/pulse, 10 Hz) originating from a Q-switched neodymium-doped yttrium aluminum garnet (Nd:YAG) based laser [108]. Our setup (Fig. 3.1) includes a commercial Nd:YAG laser (Infinity, Coherent, Inc., Santa Clara, CA), whose master oscillator is composed of two concave mirrors, a Nd:YAG prism and an acousto-optical modulator, which together form a ring oscillator [109]. The oscillator is pumped by a GaAs laser diode, emitting light at $\sim$ 800 nm, which matches very well with the major Nd:YAG absorption band ($^4F^{5/2}$ band) [109]. The acousto-optical modulator serves as a Q-switch: When the acousto-optical modulator is on, light is diffracted out of the laser cavity. When the modulator is switched off, the Q-factor of the cavity is increased, resulting in stim-

ulated emission, which produces a nanosecond pulse. This pulse is further amplified by passing it through Nd:YAG rods, which are pumped by a flash lamp [109]. Part of the pulse energy ($\sim$ 20%) is focused into a H_2 pressurized cell ($\sim$ 700 psi) with a length of 1 m. Interaction between the light pulses and the hydrogen gas produces pulses at a wavelength of 1.9 μm, which are separated from other wavelengths by a dispersive prism. The 1.9 μm pulses (3 ns, $\sim$ 10 mJ/pulse) are weakly focused (spot size $\sim$ 1 mm^2) onto the sample cell. Changes in sample absorption in the infrared region (typically $\sim$ 1630 cm^{-1}) are probed by a quantum-cascade continuous wave laser (DayLight Solutions, San Diego, CA), whose light is focused onto the sample (spot size $\sim$ 200 μm^2).

The generated temperature difference in the sample varies with both interaction length along the beam path and across the excitation beam profile [20], i.e.

$$\Delta T(r,z) = \frac{k}{\rho C_v} I(r,z) \tag{3.2}$$

where, ρ, C_v, k and $I(r,z)$ are the density and heat capacity of the sample, the absorption coefficient of the sample, the intensity of the heating pulse as a function of the radial distance, r, from the center of the excitation beam axis and the interaction length along the propagation direction, z, of the excitation beam, respectively. It is important to probe only a population of molecules, which experience the same temperature jump, $\Delta T(r,z)$. The radial temperature gradient is minimized in our setup by using heating pulses with a flat-topped beam profile. In addition, the comparatively

smaller spot size of the probe beam ensures that only a small, homogeneously heated cross-section in the center of heating pulse is probed. The gradient of the temperature difference generated by the heating pulse must further be nearly uniform along the interaction pass through the sample. This temperature difference gradient along the beam path is determined by the absorbance of the heating pulse by the sample [20]:

$$I(r,z) = I_0(r,z)\exp(-kz) \tag{3.3}$$

The utilized excitation pulses of a wavelength of 1.9 μm are only weakly absorbed by the DO overtone stretch of the D_2O solvent, thus minimizing attenuation of the excitation beam along the short pathlength ($\sim 50 - 100$ μm) through the sample.

Technical issues related to the T-jump method involve cavitation, thermal lensing and photo-accousitc effects [20]. Cavitation [110, 111] refers to the laser-induced formation of solvent vapor bubbles in the sample. This problem is minimized by degassing the samples before measurements, by measuring at sufficiently low temperatures and by limiting the magnitude of the temperature jumps to < 10 K. Thermal lensing, i.e. the transient pump-induced change of the sample refractive index, leading to a large apparent change in optical density due to deflection of the beam away from the active area of the detector, is also minimized by weakly focusing the pump beam onto the sample.

3.3 Flash-photolysis Method

Our flash photolysis setup is based on the Nd:YAG laser system described in the previous section. The required UV-pump ($\omega = 266$ nm) pulses for electronic excitation of tryptophan are generated by passing the fundamental beam (~ 100 mJ/pulse) through two appropriately aligned birefringent β-barium borate (BBO) crystals, producing the second- (532 nm) and fourth- (266 nm) harmonics of the Nd:YAG fundamental, respectively. The fundamental and generated beams are separated by dispersive optics: The beams are positioned onto the input face of a Pellin-Broca prism. After passing the beams through a triangular prism, the fundamental and second harmonic beams are directed onto beam dumps. The UV beam (a few μJ) is collected by another triangular prism and mirrors and subsequently weakly focused onto the sample (spot size ~ 1 mm^2).

3.4 Circular Dichroism Spectroscopy of Peptides

Although the molecular determinants of circular dichroism (CD) spectra are very difficult to understand theoretically [112], well established empirical relationships between the circular dichroism spectra and the corresponding peptide structures have been established [113]. Due to the availability of reliable commercial CD spectrometers and the low sample concentrations (20-50 μM) and volumes (~ 300 μL), CD spectroscopy is a fast and reliable way for determining the secondary structure of pep-

tides. In this Thesis, CD spectroscopy is mainly used to extract the thermodynamics of unfolding from the analysis of the thermal denaturation curves, i.e. the circular dichroism signal at a specific wavelength as a function of temperature. Employing a two-state model, the signal, $\theta(T)$, is proportional to the folded state population, α_F, i.e. $\theta(T) = [\theta_F(T) + \theta_U(T)]\,\alpha_F + \theta_U(T)$. The equilibrium constant of folding, $K_{eq} = \frac{\alpha_F}{1-\alpha_F}$, is therefore related to the circular dichroism signal as

$$\theta(T) = \frac{\theta_F(T)K_{eq} + \theta_U(T)}{1 + K_{eq}} \tag{3.4}$$

The equilibrium constant is related to the thermodynamic quantities of unfolding as described in Chapter 2, which are used as fitting parameters. The spectroscopic signals of the folded and unfolded states are further assumed to be linearly dependent on temperature, i.e. $\theta_F(T) = a + bT$ and $\theta_U(T) = c + dT$. This linear dependence of the signals manifests itself in the small, steady changes in the spectroscopic signal before and after the major cooperative unfolding transition. These small changes are typically linear and are attributed to small structural changes, e.g fraying of helical ends, that do not coincide with the global order-disorder transition.

3.5 Fourier-transform Infrared Spectroscopy of Peptides

Our Fourier-transform infrared (FTIR) spectrometer setup is based on a commercial Nicolet Magna-860 spectrometer. In the spectrometer, light originating from an infrared source is collimated and directed onto a beamsplitter, which reflects one half

of the beam onto a fixed mirror, and the other half onto a mirror that is movable. The light beams are reflected at the mirrors and recombined at the beam splitter. The recombined beam is passed through a sample cell and focused on a nitrogen-cooled mercury cadmium telluride (MCT) detector. As the two components in the recombined beam interfere constructively and deconstructively as a function of mirror position, an interferogram can be recorded [114].

$$I(x) = \int_0^\infty B(\omega)\left[1 + \cos(2\pi\omega x)\right] \mathrm{d}\omega \tag{3.5}$$

Here, $B(\omega)$ is the intensity of the light of the wavenumber ω. Taking the inverse Fourier-transform of Eq. 3.5, gives the intensity as a function of wavenumber.

$$B(\omega) = \int_0^\infty I(x)\cos(2\pi\omega x)\mathrm{d}x \tag{3.6}$$

$B(\omega)$ is measured for both the sample (e.g. peptide+solvent) and a reference (e.g. solvent without peptide). The optical density as a function of wavenumber is then calculated according to Beer's law [42], i.e.

$$OD(\omega) = -\log\frac{B_s(\omega)}{B_{ref}(\omega)} \tag{3.7}$$

In our setup, the cell consists a teflon spacer (thickness $\sim 50 - 100$ μm), which is sandwiched between two calcium fluoride windows, The spacer divides the cell into a sample and reference side, which are alternatingly moved in and out of the beam path to correct of experimental drift. The cell is mounted onto a brass holder, which is temperature controlled to ± 0.2 K by a water bath.

3.6 Non-natural Spectroscopic Reporters

3.6.1 Vibrational Spectroscopy

The application of non-natural vibrational and/or fluorescence reporters based on nitriles, thiocyanates and azides represents a novel methodology to study processes involving peptides and proteins. In the following, a few theoretical concepts and experimental observations related to the vibrational bands of nitrile, thiocyanate and azide compounds are discussed that are relevant for their interpretation in a biomolecular context. The current state of fluorescent labels based on these molecules is furthermore summarized. The Chapter is concluded with a critical assessment of the structural and energetic perturbations caused by these probes.

3.6.2 Properties of the Nitrile Stretching Mode

The nitrile (C≡N) stretching band of a variety of nitrile compounds have been extensively studied in the gas [115, 116, 117, 118] and condensed phases [119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148]. In particular, the solvent [127, 128, 129, 130, 132, 133, 134, 135, 136, 143, 144, 146, 147], temperature [129, 130] pressure [119, 135] electrostatic field [139, 140, 141], or complexation [121, 122, 123, 124, 125, 126, 131, 148] induced vibrational frequency shifts and intensity changes of this normal mode have been the focus of many investigations over the years. The

nitrile stretching band of acetonitrile (CH_3CN), which is partly affected by a Fermi resonance [134, 138] between this mode and the combination band of the CC stretch and the symmetric CH_3 bend and a hot band arising from the $\nu_8 = 369$ cm^{-1} mode [115, 116, 117, 118, 130, 138], has been studied in the neat liquid [129, 130, 135, 138], under dilute conditions in a large range of solvents [132, 133, 134, 136, 138], in mixtures of solvents [132, 133], at air/liquid interfaces [137, 142], as well as in solid matrices [139, 140, 141]. Though no model has been found that quantitatively predicts the frequency shifts of the C≡N stretch in all the solvents studied [138], the shifts are well understood on a semi-quantitative level. When dissolved in an aprotic or protic solvent, the peak position of the C≡N stretching band of acetonitrile red-shifts compared to its gas phase value of 2266.5 cm^{-1} [115, 116, 117, 118], except when dissolved in fluorinated alcohols, for which small blue-shifts are detected [133, 136, 138]. This observation was interpreted as a complex interplay between non-specific and specific interactions between acetonitrile and the solvent [138]. Generally, the solvatochromatic shift of the C≡N stretching band may be expressed as [138]

$$\Delta\nu = \Delta\nu_{NS} + \Delta\nu_S = \Delta\nu_C + \Delta\nu_R + \Delta\nu_A + \Delta\nu_S \tag{3.8}$$

Here, $\Delta\nu_{NS}$ and $\Delta\nu_S$ refer to the solvent-induced shifts by non-specific and specific interactions, respectively. The nonspecific contribution can be decomposed into three independent terms: Upon solvation, the centrifugal distortion of the acetonitrile molecule due to free rotation in the gas phase is reduced, resulting in a red-

shift of $\Delta\nu_C = -0.5$ cm^{-1} [135, 138]. Short-range repulsive forces, which arise from collisions of acetonitrile with solvent molecules were estimated to give rise to a $\Delta\nu_R = +4.2$ cm^{-1} blue-shift [135]. Thus, the observed red-shifts arise mainly from attractive forces, i.e. dispersive and dipole-dipole interactions between solvent and solute [135, 138], i.e.

$$\Delta\nu_A = DA_\epsilon + PA_\alpha \tag{3.9}$$

where D and P are the dielectric strength and the polarizability of the solvent, respectively. No expression has been found for the solvent specific term, $\Delta\nu_S$. However, Reimers and Hall [138] estimated the relative contributions of the terms in Eq. 3.8 by setting $\Delta\nu_C$ and $\Delta\nu_R$ to the values mentioned above and by fitting the non-specific component of Eq. 3.8 (i.e., $\Delta\nu_{NS} = \Delta\nu_C + \Delta\nu_R + \Delta\nu_R$) to the frequency shifts of C≡N stretching band in 33 different solvents using A_ϵ and A_α as fitting parameters. The difference between this fit and the experimentally observed shift in a given solvent was then taken to be the solvent specific contribution, $\Delta\nu_S$. They found that whereas PA_α, falling into the range of -8.2 to -15.1 cm^{-1}, makes the largest contribution to $\Delta\nu_A$, PA_ϵ, which was found to be between -1.5 to -6.5 cm^{-1} also contributes significantly to the observed red-shifts. The solvent specific contribution varies greatly. For example, whereas $\Delta\nu_S = -0.8$ cm^{-1} for benzene, it is $+6.7$ cm^{-1} in water and $+15.4$ cm^{-1} in trifluoroethanol. The modulation of the vibrational potential surface induced by specific interactions between acetonitrile and Lewis acid, such

as water, trifluoroethanol, or metal cations has been of great interest. For example, compared to the C≡N stretching band in an aprotic solvent, such as tetrahydrofuran, the band shifts increasingly to the blue in the series of solvents: methanol, trifluoroethanol, and hexafluoroisopropanol [133]. This observation is attributed to an increasing strength of the hydrogen bond formed between the nitrogen of the nitrile group and the hydroxyl moiety along this series of alcohols. It is well established that in compounds, such as $Ag(CH_3CN)_4^+$, whose nitrile stretching frequency is blue-shifted with respect to neat acetonitrile, the coordination of the metal ion occurs with the nitrogen lone-pair [133]. Molecular orbital calculations based on Hückel theory by Purcell and Drago [125] indicate that for the $CH_3CN{\cdot}BF_3$ adduct the N_{2s} orbital overlaps more strongly with the C_s and $C_{p\sigma}$ orbitals in the adduct compared to free acetonitrile, thus the stronger (and therefore shorter) C≡N in the adduct. The oscillator strength of the nitrile group was found to be strongly dependent on solvent environment [133, 132]. Specifically, Eaton and Symons [133] found that the C≡N band intensity increases with increasing Gutmann acceptor number of the solvent. In particular, the transition dipole strength in aprotic solvents is significantly smaller than in protic solvents. Similarly, gains in C≡N band intensity were found for nitriles when complexed with Lewis acids [120, 121]. These observations can be understood in terms of an increased polarization of the nitrile group upon interaction of the nitrogen-lone pair with a Lewis acid. If the dipole moment of the C≡N group

is thought of as arising from two partial atomic charges $\pm\delta q$, then, upon displacement δr from the equilibrium C≡N bond length r_c, the dipole is $\mu = r_e\delta q + \delta r\delta q$. Thus, an increase in the magnitude of the partial charges $\pm\delta q$ results in a larger vibrational transition dipole moment [42], i.e.

$$\mu_{01} = \left(\frac{\mathrm{d}\mu}{\mathrm{d}r_e}\right) \langle\nu_0|\, r\, |\nu_1\rangle = \delta q\, \langle\nu_0|\, r\, |\nu_1\rangle \tag{3.10}$$

where ν_0 and ν_1 are the wavefunctions of the vibrational levels involved in the transition. It is worth noting that kinematic coupling between the C≡N stretching vibration and the N-Lewis acid bond vibrations has also been evoked as an explanation for both frequency shifts and intensity gains in the case of complexation of the nitrile group with various Lewis acids [120], but theoretical considerations by Beattie and Gilson [148], as well as those by Purcell and Drago [125] are more consistent with the explanation given above. For other nitriles, such as alkyl or aryl nitriles or cyanogen halides, similar trends for the frequency shifts and changes in transition intensities are observed, though their magnitudes depend on the specific molecule. Of particular interest for probe development is the dependence of the transition dipole strength of substituted benzonitriles [120, 121, 139, 147]. For example, whereas the transition dipole of *p*-tolunitrile was found to be approximately a factor of 1.6 times larger than the one of benzonitrile, the oscillator strengths of *o*-nitrobenzonitrile and *m*-nitrobenzonitrile were found to be about 6 and 3 times smaller, respectively [121]. Similarly, Andrews and Boxer [139] found a correlation between the Hammett num-

ber and the transition dipole strength of substituted benzonitriles. Cho and co-workers [147] found that the extinction coefficient of 4-cyanophenol is smaller than the one of 4-cyanophenoxide by a factor of two. These observations are understood in terms Eq. 3.8 as greater electron donation to the aromatic ring leads to larger partial charges on the nitrile atoms. For experiments in which the chemical dynamics of a system are of interest, the lifetime of the vibrationally excited state is of great importance as it determines the time window in which dynamics can be observed. The vibrational lifetimes of a few nitriles have been measured and are on the order of a few picoseconds [149, 147, 146].

3.6.3 Properties of the Nitrile Stretching Mode of Thiocyanates (R-SCN) and Cyanates (R-OCN)

Compared to nitriles, the C≡N stretching band of alkyl or aryl thiocyanates (R-S-C≡N) have received considerably less attention. However, the infrared spectrum of methyl thiocyanate was studied in the gas phase [150]. More recently, Maienschein-Cline and Londergan [151] studied the nitrile stretching band shape, peak position and extinction coefficient of methyl thiocyanate in nine different solvents. The line shape of the C≡N stretching band was found to be symmetric in all solvents studied, except in fluorinated alcohols, which was attributed to the existence of two or more distinct populations of hydrogen-bonded complexes of the alcohols and methyl thio-

cyanate. Contrary to acetonitrile, no evidence of a Fermi resonance between the C≡N stretching mode and other modes or linear combinations thereof were found, making the interpretation of solvent-induced shifts in the C≡N stretch in methyl thiocyanate easier than the one in acetonitrile. As for nitriles, the extinction coefficient of the C≡N stretch was found to depend on solvent environment, being smaller by about a factor of two in CCl_4 compared to protic solvents, such as water of trifluoroethanol. Organic cyanates (R-OCN) are also interesting candidates to serve as vibrational probes due to the comparatively large extinction coefficient of the OCN asymmetric stretching mode [152, 153]. A recent infrared study on phenyl cyanate found that the infrared spectrum of this compound exhibits a complex band shape in the OCN stretching region due to the involvement of the mode in at least two Fermi resonances with overtones [153], rendering derivatives of phenyl cyanate less likely to be useful as vibrational probes.

3.6.4 Properties of the Asymmetric Stretching Mode of Organic Azides (R-N_3)

The asymmetric stretching vibration of organic azides (R-N_3) is an attractive candidate as an infrared probe as its vibrational transition dipole moment is typically stronger than that of the C≡N stretching mode. Unfortunately, contrary to the infrared spectra of most nitrile compounds, which typically show a single, sharp

band (fwhm $\sim 6 - 15$ cm^{-1}) in the C≡N stretching region, the band of the N_3 asymmetric stretch of organic azides is much broader (fwhm $\sim 25 - 40$ cm^{-1}) and typically convolved with one or more bands in the same region. The origin of the complicated band structure of organic azides has been the topic of various studies [150, 154, 155, 156, 157]. Early studies tentatively attributed the band structure to mixing between a combination tone of low frequency vibrations and the N_3 asymmetric stretch due to anharmonic coupling [154, 156]. For example, in a systematic study of the vibrational spectra of various aromatic acid azides (R-CON_3), Lieber and coworkers [154] found that the infrared spectra of all aryl acid azides have two peaks in the N_3 asymmetric stretching region, except those whose CON_3 group is separated from the aromatic phenyl ring by one or more methylene groups. For those compounds, where the CON_3 group is directly connected to the benzyl ring, the degree of the splitting (as judged by the relative intensity of the two peaks) was found to depend on the substituent on the benzyl ring. Lieber *et al*. interpreted the observed the band structure in the region ~ 2200 cm^{-1} to a Fermi resonance between a combination band and the N_3 asymmetric stretching band. The combination band was thought to arise from the combination of the N_3 symmetric stretch (~ 1200 cm^{-1}) and a band at ~ 1000 cm^{-1}, which was tentatively attributed to vibrations of the aromatic ring. Similarly, investigating the origin of the band structure in the N_3 stretch region in a series of aryl azides, Dyall and Kemp [156] found that the observed splittings are more

likely to arise from a Fermi resonance than from hot bands. In a recent experiment, Cheatum and coworkers [157] found direct evidence of a Fermi resonance interaction between the N_3 asymmetric stretch and a combination or overtone state in the form of cross-peaks in the 2D IR spectrum of 3-azidopyridine at early delay times. The presence of a Fermi resonance greatly complicates the interpretation of frequency shifts. As the vibrations involved in the resonance (combination or overtone band and N_3 asymmetric stretch) are differently affected by environmental changes, the coupling between them is also modulated, thus changing peak position and intensity of the vibrational band. It is interesting to note that the transition dipole strength of the N_3 asymmetric stretch of aromatic azides was found to be strongly dependent on the nature of the aromatic ring. For example, though the extinction coefficients of various aryl azides studied by Dyall and Kemp varied significantly with substitutions on the benzyl ring [156], their reported coefficients are significantly higher than that of 3-azidopyridine ($\sim$ 250 M^{-1} cm^{-1}) [157]. Despite these difficulties, some useful azide probes have been identified, which will be discussed in a later section.

3.6.5 Theoretical Modeling of Vibrational Lineshapes of Nitriles, Thiocyanates, and Azides

Two distinct approaches for the calculation of vibrational lineshapes are currently used. The first approach originated in the groups of Cho [158, 159, 160, 161, 162] and

Skinner [163], the second was recently developed by Corcelli and coworkers [164, 165, 166]. Both methods rely on a robust classical molecular mechanics (MM) or quantum mechanical/molecular mechanical (QM/MM) simulation of the system of interest. In the first approach [158, 159, 160, 161, 162, 163], QM harmonic or anharmonic frequency analyses are performed on a large number (tens to thousands) of probe-solvent clusters, which are randomly chosen from the trajectory. For N different clusters, there will be N unique frequencies as the vibrational frequency of the probe in each cluster depends on the exact orientation of all the molecules with respect to each other. The frequencies are assumed to be related to the electrostatic field or potential in the following manner:

$$\omega = \omega_{\text{gas}} + \sum_{i=1} \alpha_i \phi_i + \sum_{i,j} \beta_{i,j} E_{i,j} \tag{3.11}$$

where ϕ_i and $E_{i,j}$ are the electrostatic potential and Cartesian electric field component j at position i, arising from the solvent environment, and the constants of proportionality α and β are determined by fitting the model to the *ab initio* determined vibrational frequencies of the probe-solvent clusters, assuming the appropriate gas phase frequency ω_{gas} of the probe. Once the parameters α and β are known, the fluctuation of the frequency along the molecular dynamics (MM or QM/MM) trajectory may be easily calculated via Eq. 3.11 by summing over all solvent partial charges.

The second approach [164, 165, 166], rather than employing Eq. 3.11 to obtain the vibrational frequencies along the trajectory, uses a QM/MM method to calcu-

late the vibrational frequency in each snapshot along the trajectory. The QM/MM method of Corcelli and coworkers is based on the PM3 semiempirical method, which was optimized to reproduce the C≡N vibrational frequencies of acetonitrile-water clusters. Using the single set of parameters obtained via this optimization procedure, they found quantitative agreement between the experimental and the calculated C≡N vibrational bandwidths of acetonitrile in both water and tetrahydrofuran (THF). Furthermore, they correctly predicted the solvent-induced shift when going from THF to water [165]. Corcelli *et al.*'s results demonstrate that this method, once parameterized using one type of solvent (e.g. water), can be used for lineshape calculations of the C≡N stretch of acetonitrile in different solvent environments, i.e. the method is "solvent transferable". As the purpose of spectroscopic reporters is to probe the heterogeneous environments found in biomolecular systems, solvent transferability of a method for lineshape calculations is of paramount importance. Though the former method based on Eq. 3.11 typically provides good agreement between experimental and calculated lineshapes [158, 159, 160, 161, 162, 163, 167], it is not clear to what extent such a map is solvent transferable.

Force field parameters are currently available for a range of model compounds of the probes [159, 165, 167, 168, 169, 170], but not all of those might be entirely compatible with commonly used force fields as they were derived by procedures that are different from those typically employed for a particular force field. To ensure the

integrity of a molecular dynamics simulation of a more complex system containing a vibrational probe, it would be beneficial to have parameters available that were derived using the standard procedures for a given force field.

3.6.6 Infrared Probes Incorporated into Biologically Relevant Molecules

3.6.6.1 Nitriles Probes

p-Cyanophenylalanine (p-Phe$_{CN}$) and m-Cyanophenylalanine (m-Phe$_{CN}$). Since the first demonstration of the synthetic incorporation of *p*-cyanophenylalanine (*p*-Phe$_{CN}$) into a peptide system [171], this amino acid has become one of the most popular site-specific vibrational probes in the context of proteins and peptides. Getahun *et al.* [171] prepared two derivatives of a peptide, which strongly binds to the hydrophobic pocket of the C-terminal lobe within the protein calmodulin (CaM). In one derivative (MLCK$_{5CN}$), *p*-Phe$_{CN}$ was incorporated at a position that would remain exposed to the solvent, whereas in the other derivative (MLCK$_{3CN}$) *p*-Phe$_{CN}$ was placed in a position of the peptide sequence that would be sequestered from the aqueous environment. Indeed, whereas the infrared spectrum of the MLCK$_{5CN}$-CaM complex in the C≡N stretch region is virtually identical to the one of MLCK$_{5CN}$ in aqueous solution, the peak position of the C≡N stretch of MLCK$_{3CN}$-CaM complex is shifted to the red by ~ 6.8 cm^{-1}. These experiments established that *p*-Phe$_{CN}$ is

not only highly sensitive to environmental changes, but also provides site-specificity in a complex system. Schultz *et al.* [172] genetically incorporated *p*-Phe$_{\mathrm{CN}}$ into the active site of the protein myoglobin, demonstrating that this probe may also be used in the context of longer proteins, which cannot be chemically synthesized. Huang *et al.* [173] investigated the temperature dependence of the C≡N stretch. Recently, Boxer an coworkers found that the spectral properties of *m*-Phe$_{\mathrm{CN}}$ are similar to those of *p*-Phe$_{\mathrm{CN}}$ [174].

p-Cyanobenzylcysteine (p-Cys$_{\mathrm{BCN}}$) and p-Cyanophenylcysteine (p-Cys$_{\mathrm{PCN}}$). An alternative approach was taken by Jo *et al.* [175], who site-specifically incorporated the *p*-tolunitrile moiety into proteins via post-translational alkylation and arylation of cysteines. Following the incorporation of a *p*-tolunitrile moiety into the peptide mastoparan-X, via reaction with *p*-cyanobenzyl bromide, the infrared spectra of the peptide were taken in aqueous solution and when bound to calmodulin. In analogy to the model system involving MLCK$_{\mathrm{3CN}}$-CaM, there is a frequency shift of about 5.3 cm^{-1} upon binding to CaM compared to the mutant in free solution.

5-Cyanotryptophan (5-Trp$_{\mathrm{CN}}$). Ideally, a native residue in a protein should be replaced with its nitrile analogue in order to minimize energetic and structural perturbations. An interesting alternative to *p*-Phe$_{\mathrm{CN}}$ is 5-cyanotryptophan (5-Trp$_{\mathrm{CN}}$), whose extinction coefficient is about 1.5 times larger than the one of *p*-Phe$_{\mathrm{CN}}$. As tryptophan residues are frequently found at functionally relevant sites of proteins, 5-

Trp_{CN} provides an important addition to the toolbox of the spectroscopist. Waegele *et al.* [176] showed that the C≡N stretching band of a mutant of the membrane-binding peptide indolicidin containing 5-Trp_{CN} is sensitive environmental changes. Unlike in the case of *p*-Phe_{CN}, the changes in solvent environment do not significantly change the peak position of the band, but manifest themselves in readily detectable changes in the width of the stretching band.

β-Cyanoalanine (β-Ala_{CN}). The spectroscopic properties of β-Ala_{CN} were characterized [171], but the amino acid has not been used in the context of a biologically relevant systems due to its comparatively small extinction coefficient of ~ 30 M^{-1} cm^{-1} in water.

5-Cyano-2'deoxyuridine (dU_{CN}). The infrared spectrum in the ν_{CN} region of dU_{CN} in 2-methyltetrahydrofuran shows a single, sharp peak (fwhm ~ 8 cm^{-1}) at 2231 cm^{-1} [177]. The large extinction coefficient of 332 M^{-1} cm^{-1} and a Stark tuning rate of 1.1 cm^{-1} MV/cm render this band an attractive sensor in the major groove of DNA molecules.

N2-nitrile-2-deoxyguanosine (dG_{CN}). The band of ν_{CN} peaks at 2170 cm^{-1} with a fwhm of $\sim$ 29 cm^{-1} in 2-methyltetrahydrofuran is unusually broad for nitriles [177]. However, the large extinction coefficient of 412 M^{-1} cm^{-1} and a Stark tuning rate of 1.1 cm^{-1} MV/cm make this band an attractive sensor in the minor groove of DNA molecules.

3.6.6.2 Thiocyanate Probes

β-Thiocyanatoalanine (β-Ala$_{SCN}$). Fafarman *et al.* [178] post-translationally incorporated β-Ala$_{SCN}$ into proteins via cysteine cyanylation. The infrared spectrum of β-Ala$_{SCN}$ in a variant of S-peptide shows a peak at 2161.2 cm^{-1} in aqueous solution, shifting to the red by 5.8 cm^{-1} upon complex formation with the ribonuclease S-protein. They further demonstrated that the probe may be selectively incorporated into larger proteins and used β-Ala$_{SCN}$ to quantify electric fields using vibrational Stark spectroscopy [179].

Recently, Londergan and coworkers have assessed the utility of the C≡N stretch of β-Ala$_{SCN}$ to report the degree of burial of side-chains in membrane binding peptides and upon structural transitions in peptides and proteins [180, 181, 182]. They prepared two singly-labeled cyanylated cysteine variants of the membrane binding peptide MBP [180]. The peptide is unstructured in water, but undergoes a coil to helix transition upon binding to DPC micelles. One version of the peptide contains the label in a region that is known to remain solvent exposed upon binding, the other contains the label in a region that is buried in the hydrophobic interior of the micelle upon binding. Whereas the infrared spectra in the C≡N stretch region of the former variant in aqueous solution and in the presence of the membrane was virtually identical, a significant red-shift of ~ 4 cm^{-1}, accompanied by a narrowing of the vibrational band upon binding was detected in the spectra of the later variant. Similar

results were obtained for another model peptide. In another study [181], using a series of singly-labeled helical peptides with the sequence $(AAAAC)_n$ as model systems, Edelstein *et al.* [181] found that width of the C≡N stretch band is broader for the helical conformation than for the coil conformations. The broadening of the lineshape upon gain in structure was thought to originate either from reduced motional narrowing due to slowed solvent dynamics in the vicinity of the structured helix or from a conformational dependent modulation of the vibrational potential due to the helix dipole. The small, but reproducible changes in bandwidth upon formation of helical structure was recently used by Bischak *et al.* [182] to reexamine the order-to-disorder transition of an intrinsically unstructured protein.

3.6.6.3 Azide Probes

p-Azidophenylalanine (p-Phe$_{N_3}$). Ye *et al.* [183] showed that the peak position of the N_3 asymmetric stretching band of *p*-Phe$_{N_3}$ significantly red-shifts by $\sim$ 13.1 cm^{-1} upon changing the solvent from water to isopropanol, suggesting a high sensitivity of this probe to electrostatic changes in the environment. However, as in the case of simpler aryl azides, the quantitative interpretation of the N_3 asymmetric stretching band of *p*-Phe$_{N_3}$ is complicated by its rather complex structure due to Fermi resonances. Via genetic incorporation of this vibrational probe at various sites of the protein rhodopsin, Ye *at al.* obtained site-specific information on conformational

changes associated with the photoisomerization of rhodopsin [183, 184].

β-Azidoalanine (β-Ala$_{N_3}$). Oh *et al.* [77] found that the infrared band of the N_3 asymmetric stretch of β-Ala$_{N_3}$ shifts by $\sim$ 13.8 cm^{-1} to the red upon changing the solvent from water to DMSO. The extinction coefficient of band in water is 414 M^{-1} cm^{-1}, which is roughly a factor of two larger than that of *p*-Phe$_{CN}$, rendering this probe attractive for the use in conjunction with photon echo experiments. They incorporated β-Ala$_{N_3}$ into a short peptide, Aβ(16-22), which is known to form β-sheet aggregates in water, but not in DMSO. The N_3 asymmetric stretching bands of this peptide in DMSO and water were found to be virtually identical, suggesting that the probe in the aggregate samples an electrostatic environment similar to that in DMSO. Unfortunately, the azido group of β-Ala$_{N_3}$ was recently found to disrupt the natural backbone conformation of short model peptides [185].

Azidohomoalanine (homo-Ala$_{N_3}$). This amino acid has been incorporated into proteins both genetically [186, 187] and synthetically [187]. The N_3 asymmetric stretching band of homo-Ala$_{N_3}$ (fwhm $\sim$ 35 cm^{-1}) shifts from 2098 cm^{-1} to 2115 cm^{-1} when changing the solvent from DMF to D_2O [187]. Taskent-Sezgin *et al.* further demonstrated that via incorporation of this probe into the protein NTL9, they could obtain site-specific information on the unfolding process of this protein. They further demonstrated that the N_3 asymmetric stretching band of a mutant which contains homo-Ala$_{N_3}$ at the N-terminus reports on the protonation state of the N-terminus.

2'-Azido-2'-deoxyuridine (dU_{N_3}). The N_3 asymmetric stretching band of dU_{N_3} (fwhm $\sim$ 21 cm^{-1} in THF) was found to be centered at 2110.7 cm^{-1} in THF, shifting by $\sim$ 13 cm^{-1} to the blue in water with a peak extinction of $\sim$ 1160 M^{-1} cm^{-1} [188, 189]. Interestingly, due to the position of the N_3 group on the sugar ring, the N_3 band was found to be insensitive to heterodimer formation between dU_{N_3} and another model base. This is in contrast with the previously mentioned major groove probe dU_{CN}, whose C≡N stretch was found to be sensitive to base pairing [188]. This result suggests that dU_{N_3} is a complementary reporter to dU_{CN}. Whereas dU_{CN} is a site-specific reporter of environmental changes in the major groove, dU_{N_3} may be used to report on events in the phosphate sugar region. Tucker *et al.* [189] showed that the 2D IR spectra and frequency-frequency correlation functions of the N_3 stretch of dU_{N_3} can be readily measured in photon echo experiments due to its large extinction coefficient. Though the short vibrational lifetime of $\sim$ 1 ps of this mode limits the accessible timescales, the experiments suggest that dU_{N_3} may serve as a reporter on local water dynamics. Contrary to aryl azides, no evidence for a Fermi resonance was found. The origin of the slight asymmetry of the N_3 band attributed to the possible conformational diversity of the molecule.

Azido-NAD^+. Being a cofactor for a large range of enzymes, this azido analogue of nicotinamide adenine dinucleotide (NAD^+) has been developed by Dutta *et al.* [190] as a sensor of enzyme active-site dynamics on the femto and picosecond timescales.

The N_3 band of azido-NAD^+ is centered at 2140 cm^{-1} (250 M^{-1} cm^{-1}) with a fwhm $\sim$ 33 cm^{-1}. Though the peak position of the N_3 band shifts by $\sim$ 2 cm^{-1} to the red upon binding of azido-NAD^+ to the enzyme formate dehydrogenase, a substantial narrowing of the band to $\sim$ 26 cm^{-1} is observed. The binding constants of azido-NAD^+ to three enzymes are comparable to those of NAD^+, suggesting that azido-NAD^+ is a viable probe.

3.6.7 *p*-Cyanophenylalanine as a Fluorescence Probe

p-Phe$_{CN}$ fluorescence. Besides its application as a vibrational probe, *p*-Phe$_{CN}$ can also be used as a fluorescent probe due to its appreciable quantum yield of about 0.11 in water [191, 192]. Tucker *et al.* [191] found that the fluorescence of *p*-Phe$_{CN}$ is very sensitive to solvent environment and can be selectively excited in the presence of other aromatic amino acids. For example, they found that the fluorescence intensity of *p*-Phe$_{CN}$ to be about 10 times lower in acetonitrile compared to water. Tucker *et al.* further demonstrated the utility of this probe in the context of a biological problem by determined the binding affinity of the MLCK$_{3CN}$ peptide to CaM (see first section on *p*-Phe$_{CN}$ above) while monitoring the degree of quenching of the *p*-Phe$_{CN}$ fluorescence in a series of stoichiometric titrations. Recently, the photophysics of *p*-Phe$_{CN}$ was investigated in more detail by Serrano *et al.* [192] as well as Raleigh and coworkers [193, 187].

p-Phe$_{CN}$/Trp Förster resonance energy transfer (FRET). As the fluorescence of *p*-Phe$_{CN}$ overlaps with the UV/vis absorption profile of tryptophan, these two amino acids form a FRET pair [194]. With a Förster radius of 16.0 Å, this FRET pair is ideally suited to probe distances within small proteins and peptides and has been used to probe the conformational distribution of unstructured peptides as well as protein unfolding [195, 196]. Recently, Rogers *et al*. showed that 5-L-hydroxytryptophan (5-Trp$_{HO}$) is an interesting substitute for Trp as its absorption spectrum overlaps more strongly with the fluorescence emission of *p*-Phe$_{CN}$, giving rise to a Förster radius of 18.5 Å [196]. In particular, the fluorescence quantum yield of 5-Trp$_{HO}$ is higher than that of Trp and is less sensitive to solvent polarity, opening up the possibility of analyzing acceptor fluorescence in case where the analysis of donor fluorescence is complicated by changes in solvent environment. Alternatively, 7-azatryptophan (7-Trp$_{N}$) may also be used in conjunction with *p*-Phe$_{CN}$ [196]. As 7-Trp$_{N}$ has a very low quantum yield and its fluorescence is well separated form *p*-Phe$_{CN}$ fluorescence, this FRET pair is particularly suitable in the context of time-resolved fluorescence experiments where leakage of acceptor fluorescence into the donor channel can be problematic. Unfortunately, this amino acid is currently only available as a racemic mixture of the D and L forms, which may complicate the extraction of quantitative information involving the *p*-Phe$_{CN}$/ 7-Trp$_{N}$ FRET pair. An alternative acceptor in this context is 5-Trp$_{CN}$, which also has a low quantum yield, but as *p*-Phe$_{CN}$ also

has an absorption maximum close to $\sim$ 240 nm [176]. Taskent-Sezgin *et al.* [193] found that tyrosine (Tyr) also constitutes an acceptor of *p*-Phe_{CN} fluorescence with a Förster radius of 12.0 Å.

3.6.8 Structural and Energetic Perturbation

Frequently, the comparatively small sizes of nitrile, thiocyanate, or azide probes, supposedly giving rise to a minimal perturbation of the system of interest, are cited as their primary advantage over other ubiquitously used spin- or fluorescence-labels. However, any extrinsic probe will always lead to some degree of perturbation. An important molecular parameter that may be used to estimate the expected structural or energetic perturbations is the permanent dipole moment of the probe compared to the unmodified molecule. For example, whereas the dipole moments of toluene ("phenylalanine") phenol ("tyrosine") are 0.4 D and 1.2 D, respectively [197], the dipole moment of benzonitrile is 4.2 D. The dipole moments of acetonitrile, methyl cyanate, and methyl azide are 3.9 D, 3.6 D, and 2.2 D, respectively [197]. Given these values, it is not surprising that substitution of Phe58 in the hydrophobic cluster of 35-residue villin headpiece subdomain (HP35) with *p*-Phe_{CN} results in a strong shift of the equilibrium towards the unfolded state as evidence by the $\sim$ 20 fold decrease in the unfolding time [198]. This significant decrease in the unfolding time, corresponding to a $\sim$ 60% decrease in the free energy of unfolding is expected in

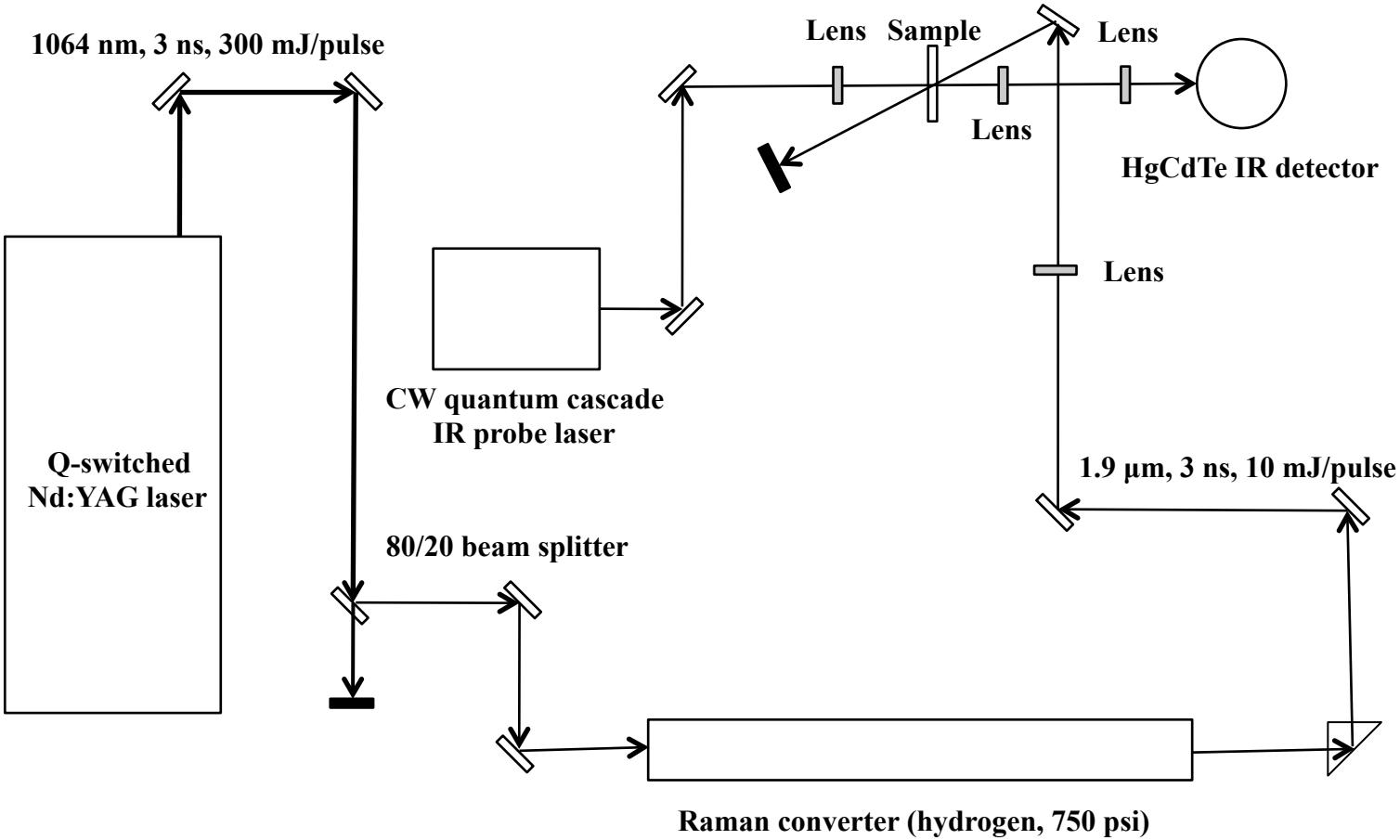
this case as only a few hydrophobic residues stabilized the fold of this small protein. Comparable perturbations were found in the case of NTL9*-Ile4/homo-Ala_{N_3} in which a hydrophobic side-chain (Ile) is replaced by a strongly hydrophilic one (homo-Ala_{N_3}) [187]. A much smaller destabilization ($\sim$ 20%) was observed though, when replacing a less hydrophobic amino acid (Met) at the beginning of the sequence with homo-Ala_{N_3}. Despite these destabilizations, the HP35-Phe58/*p*-Phe_{CN} mutant remains in its folded state (97%) at 300 K as judged by circular dichroism (CD) spectroscopy, and the NTL9-Met1/homo-Ala_{N_3} mutant exhibits NMR and CD spectra comparable to the wild-type. Substantial reductions in the thermal stabilities were found in β-Ala_{SCN} derivatized versions of the α-helix forming peptide of the sequence Ac-$(AAAAK)_5$-NH_2 [181]. However, all versions showed spectral signatures characteristic of helical structure and helix-coil transitions were observed qualitatively similar to the wild-type. Significant structural differences between Ac-Ala-NHMe, *i.e.* acetylated and methylated alanine, and its β-Ala_{N_3} analogue were observed [185]. Whereas Ac-Ala-NHMe forms a polyproline II helical structure, Ac-β-Ala_{N_3}-NHMe adopts backbone dihedral angles that give rise to a cyclic structure, which is stabilized by intramolecular hydrogen bonds between the azide group and the peptide backbone, whose formation is facilitated by the azido-gauche effect.

Dutta *et al.* [190] measured the activities of the enzymes formate dehydrogenase (FDH) and glucose dehydrogenase (GDH) with azido-NAD^+ as a substrate. Com-

pared to the complexes with native NAD^+, the Michaelis-Menten constants (K_m) of the azido derivatives were found to be ~ 4 times smaller and ~ 5 times larger for azido-NAD^+-FDH and azido-NAD^+-GDH, respectively. Compared to the wild-type, the turnover rate (k_{cat}) of azido-NAD^+-FDH was found to be smaller by a factor of two. Similar effects on the values of K_m and k_{cat} were found by Fafarman and Boxer in the case of nitrile derivatives of RNase [174]. These changes are on the order of the uncertainty of these values commonly reported in enzymology. Moreover, Ala_{SCN}-RNase shows quite small decreases in K_m and k_{cat} and both the global and local structural perturbations in the vicinity of the probe of all nitrile derivatives were shown to be small compared to the accuracy of the crystallographic data [174].

In conclusion, though energetic perturbations due to incorporation of probes into small proteins are often substantial as determined by the folding equilibrium constant, in most cases the population of the folded state is still significant under ambient conditions and its structural characteristics are similar to the one of the native protein. It was further shown that enzymes typically retain their catalytic activity upon probe incorporation. In the case of RNase, it was shown that the structure in the vicinity of the probes is not significantly altered by the presence of the nitrile [174], corroborating the notion that these probes can be considered minimally perturbative.

Figure 3.1: Temperature-jump setup.


1064 nm, 3 ns, 300 mJ/pulse
Lens
Sample
Lens
Lens
HgCdTe IR detector
Lens
Q-switched
Nd:YAG laser
CW quantum cascade
IR probe laser
1.9 µm, 3 ns, 10 mJ/pulse
80/20 beam splitter
Raman converter (hydrogen, 750 psi)

Chapter 4

Infra-red Study of the Folding Mechanism of a Helical Hairpin: Porcine PYY

4.1 Introduction

The α-helical hairpin, or helix-turn-helix (HTH) motif, can be regarded as a super-secondary structural element of helical proteins. In many cases, the HTH motif is found to play an important functional role. For example, it is the primary recognition element of a large number of DNA-binding proteins [21, 22] and is also involved in protein-protein interactions [22] . Therefore, it is not surprising that there has been great interest in designing stable helical hairpins that exhibit high affinity and specificity towards DNA or proteins [23, 24, 25]. In addition, the HTH motif has been suggested to play an important role in protein folding [11, 199, 200]. Because it encompasses both secondary structural elements and long-range tertiary contacts, a rapidly formed local HTH conformation could serve as a folding nucleus or kernel [11, 199], allowing rapid assembly of the entire tertiary structure. For example, Religa *et al*. have shown that the fast phase in the folding kinetics of Engrailed homeodomain (EnHD) arises from the formation of a folding kernel consisting of helices 2 and 3 of the protein [201].

Because of the structural and functional roles of the HTH motif, the structural determinants and folding/unfolding thermodynamics of various helical hairpins have been extensively studied in the past few years [25, 202, 203, 204]. Though these equilibrium studies provided valuable insights into the factors that control the thermal stability of the HTH motif, they revealed little information on how folding and

unfolding actually take place. As a result, our current understanding of the folding mechanism of the HTH motif is mainly based on the kinetic study of a cross-linked α-helical hairpin, Z34C [205], whose termini are covalently connected via a disulfide bond that is introduced to stabilize the folded structure [23]. While Z34C is a useful template for protein- and drug-designs [23], it may not be a good model system for understanding the mechanism of α-helical hairpin formation as the disulfide constraint artificially limits the accessible conformational space of the unfolded state ensemble.

To provide further insight into the folding mechanism of the HTH motif, herein we study the folding dynamics of a monomeric, naturally occurring helical hairpin (porcine PYY; sequence: YPAKP - EAPGE - DASPE - ELSRY - YASLR - HYLNL - VTRQR - Y - NH_2) by means of temperature-jump infrared (T-jump IR) spectroscopy [17] in conjunction with site-directed mutagenesis [53]. As shown (Fig. 4.1), porcine PYY (referred to hereafter as PYY) is free of disulfide bonds and assumes a structure known as the PP-fold in which an N-terminal type II polyproline (PPII) helix folds onto a C-terminal α-helix to form a stable, well defined hydrophobic cluster between the interfaces of the two helices [203, 206, 207]. To dissect the kinetic roles of the two most important structural elements in the PP-fold, i.e., the hydrophobic cluster and the turn, we studied the folding-unfolding kinetics of two sets of PYY mutants. The first set of mutants, A7Y, Y21A, and Y27A, involves residues that contribute to the hydrophobic interactions between the two helices. The second set of mutants, S13A

and P14A, targets residues in the turn region of the peptide. Our results support the notion that the turn sequence has a strong influence on the folding rate of hairpin structures [205, 208, 209] and that the turn sequence in wild-type PYY is highly optimized for folding, whereas the kinetic effect of mutations in the hydrophobic cluster is more complicated.

4.2 Results and Discussion

The HTH motif has recently attracted much attention due to its DNA- and protein-binding properties [23, 24, 25, 210] and because of its potential role as a productive (on pathway) intermediate in protein folding [11, 199, 200], such as in the case of EnHD [201]. Previously, we have shown that the folding mechanism of a cross-linked α-helical hairpin, Z34C, is similar to that of β-hairpins [205], with turn formation playing a dominant role in the folding kinetics of the hairpin. However, because of the disulfide cross-linker it is unclear whether the folding mechanism of Z34C is representative of that of unconstrained α-helical hairpins. Therefore, we extend our study on how α-helical hairpins fold to porcine PYY [203, 206, 207], a naturally occurring HTH motif free of disulfide bonds.

Folding thermodynamics and kinetics of PYY. As shown (Fig. 4.2), the far-UV CD spectrum of PYY in phosphate buffer (20 mM, pH* 5.2) at 25 °C exhibits the characteristic double minima of α-helices, indicating that it is mostly folded at this

temperature. To better quantify the folding/unfolding thermodynamics of PYY and the mutants studied here (see below), we globally analyzed their CD melting curves measured at 222 nm using a global fitting method (see Methods Section for detail). As shown (Fig. 4.3 and Table 4.1), the thermal melting temperature (T_m) of wild-type PYY thus obtained is 44.0 ± 1.6 °C, which is almost identical to that determined by Zerbe and coworkers using NMR spectroscopy [203].

The conformational relaxation kinetics of PYY in response to a T-jump were measured using a time-resolved infrared technique [211] with a probing frequency of 1631 cm^{-1}, where solvated α-helical amides show strong absorbance [212, 213, 214]. As shown (Fig. 4.4), the T-jump induced conformational relaxation of PYY obtained at 32.2 °C, a temperature that is below T_m, occurs on the microsecond timescale, indicating that it folds very rapidly. Indeed, decomposition of the relaxation rate constants into folding and unfolding rate constant based on a two-state analysis (see Methods section for detail) shows that the folding time constant of PYY is around 3 μs in the temperature range studied (Figure 5). For example, at 30 °C the folding and unfolding rate constants are $(2.8 \pm 0.5\ \mu s)^{-1}$ and $(7.1 \pm 1.3\ \mu s)^{-1}$, respectively (Table 4.2). Interestingly, the α-helical hairpin Z34C, whose folded structure is stabilized by a disulfide bond that links the N- and C-termini of the peptide, folds with a similar rate [205]. Taken together, these results thus corroborate our previous notion that such disulfide bonds stabilize the folded state of the HTH motif by primarily

decreasing its unfolding rate [205].

Effect of the turn sequence on the folding thermodynamics and kinetics of PYY. For β-hairpins, it has been shown that the sequence of the reverse turn plays a critical role in both the stability and folding kinetics of the native fold [209, 208]. To explore the role of the turn in PYY folding, we further studied the folding thermodynamics and kinetics of two PYY mutants, S13A and P14A. The choice of these mutations is based on the study by Zerbe and co-workers which showed that both Ser13 and Pro14 are located in the turn region that links the C-terminal α-helix to the N-terminal polyproline helix and that the Ser13 to Ala mutation is especially disruptive to the native fold of PYY [203]. Consistent with their NMR study, our CD results show that both mutations result in a decrease in the thermal stability of the PYY fold and that replacement of Ser13 with Ala is strongly destabilizing (Fig. 4.3 and Table 4.1). This observation suggests that the native turn sequence in PYY is optimized for folding and is also in agreement with the finding of Hodges and Schepartz [25] that mutations in the turn region destabilize the folded state of a designed derivative of avian polypeptide (aPP), whose sequence and fold are similar to those of PYY.

The thermodynamic consequence of a mutation is a direct outcome of its effects on the folding and unfolding rates. For example, if a mutation preferentially destabilizes the folding transition state and the folded state, the folding rate of the mutant is expected to be slower than that of the wild-type, whereas the unfolding rate is

largely unaffected. Thus, measurements of the folding kinetics of the mutant yield additional information that can be used to assess the characteristic of the transitions state [53]. As shown (Fig. 4.5), both mutations in the turn region perturb the relaxation rate of PYY, consistent with the equilibrium CD results. A mutation of the turn sequence may increase the chain entropy of the peptide and/or remove a native sidechain-sidechain or sidechain-backbone interaction. Thus, if the turn structure is formed (or becomes native-like) in the folding transition state, such a mutation would destabilize not only the native state but also the transition state, resulting in a decrease in the folding rate. As indicated (Table 4.2), at 30 °C the folding time of P14A is increased by about 146 % from that of the wild-type, whereas the unfolding time is only increased by about 34 %, suggesting that the transition state ensemble contains a native-like turn, which in the case of PYY, is stabilized by the conformational constraint on the backbone ϕ dihedral angle of Pro14. Further evidence supporting this notion comes from the folding time of S13A (Table 4.2), which also shows an appreciable increase from that of the wild-type (i.e., from 2.8 to 5.6 μs). In addition, the results obtained with S13A and P14A are consistent with the fact that the sequence of SPE is frequently found at the N-terminal end of α-helices in proteins as it is an efficient end-capping group and can increase the rate of α-helix formation [215]. Furthermore, because the sidechain of Ser13 can form two hydrogen bonds with the main chain nitrogen and the sidechain of Glu16 [203], respectively, it

is expected that elimination of these hydrogen bonds via mutation would result in not only a decrease in the folding rate but also an increase in the unfolding rate if only one of these hydrogen bonds is formed in the transition state, as observed (Table 4.2).

Effect of the hydrophobic cluster on the folding thermodynamics and kinetics of PYY. It is well known that the formation of a tightly packed hydrophobic cluster involving residues from both helices is essential to stabilize the folded hairpin structure [23, 24, 25, 210]. However, kinetically this can be achieved by either increasing the folding rate or decreasing the unfolding rate. To dissect these possibilities, we studied the thermodynamic and kinetic effects of two mutations (i.e., Y21A and Y27A) that weaken the hydrophobic interactions between the two helices [203]. As expected, both mutations destabilize the PP-fold (Fig. 4.3 and Table 4.1), although the Y27A mutation has a much larger effect than the Y21A mutation, which is in agreement with the NMR measurements of Zerbe and co-workers [203].

As shown (Fig. 4.5), mutation of either Y21 or Y27 to Ala results in a change in the T-jump induced relaxation time of the peptide. It is clear that at 30 °C this change mostly arises from a change in the unfolding time (Table 4.2). For example, the unfolding time of Y27A is 2.1 μs, which is significantly shorter than that (7.1 μs) of the wild-type, whereas its folding time only shows a marginal increase in comparison to that of the wild-type (i.e., from 2.8 to 4.3 μs). Thus, taken together, the results obtained with Y21A and Y27A suggest that the native hydrophobic cluster in PYY

is formed at the downhill side of the folding free energy barrier.

Effect of temperature. The kinetic data obtained at 30 °C support a folding mechanism wherein the turn, but not the hydrophobic cluster, is formed in the transition state ensemble of PYY. To check whether such a conclusion remains true at higher temperatures, we further analyzed the folding and unfolding times of these PYY peptides at 50 °C. As shown (Table 4.3), the results of S13A and P14A corroborate the folding mechanism reached above. However, those of Y21A and Y27A seem to indicate that the hydrophobic cluster is partially formed in the transitions state, as their folding times are significantly longer than that of the wild-type. In other words, the kinetic data obtained at 50 °C suggest a folding mechanism of PYY that is different from that deduced from the 30 °C data. While it is possible that the characteristics of the transition state ensemble of PYY shows such a dependence on temperature, another possibility is that this apparent discrepancy arises from mutation-induced changes in the unfolded state. Many studies have shown that native and/or non-native hydrophobic contacts can exist in the denatured state of proteins [216, 217, 218, 219, 194, 220, 221, 222]. If such contacts exist in the thermally denatured state of wild-type PYY, they are expected to have a strong effect on the folding rate of the peptide. The NMR structure of PYY shows that Y21 and Y27 lie on the α-helix side, while their sidechains extend towards the PPII helix (Fig. 4.1). Thus, it is possible that even in the unfolded state of PYY, interactions involving Y21

and Y27 and/or multiple residues on the PPII helix side persist. It is expected that such cross-strand interactions will reduce the entropic cost associated with the turn formation by effectively restricting the conformational space of the unfolded polypeptide chain, thus speeding up folding. In other words, mutations that eliminate or decrease such cross-strand interactions are expected to decrease the folding rate, as observed. In addition, we investigated the folding kinetics of the A7Y mutant, which was previously shown to be more stable than the wild-type peptide [203]. Our CD results confirm that the A7Y mutation increases the thermal stability of the PP fold (Fig. 4.3 and Table 4.1). However, the changes in the folding and unfolding rates induced by this mutation are close to the uncertainty of the experiments (Table 4.2 and Table 4.3), thus preventing us from making further analysis.

4.3 Conclusions

In summary, we present a detailed mechanistic study of the folding mechanism of a naturally occurring helical hairpin (porcine PYY) that is free of disulfide bonds and other types of cross-links. At relatively low temperature (e.g., 30 °C) the effect of mutations in both the turn and hydrophobic cluster on the folding and unfolding rates of PYY supports a folding mechanism in which the turn is formed in the transition state, whereas the hydrophobic cluster is consolidated on the downhill side of the folding free energy barrier. However, a more complex picture emerges at relatively higher

temperatures (e.g., 50 °C) at which the folding rate of PYY is more significantly affected, even by mutations that weaken the cross-helix interactions. While this result could manifest a temperature dependent transition state ensemble, a more likely scenario is that it reflects changes in the unfolded conformational ensemble induced by such mutations. As the strength of hydrophobic interaction increases with increasing temperature, it is conceivable that the unfolded state ensemble of PYY is more compact at 50 °C than at 30 °C due to residual hydrophobic interactions between the two strands. Therefore, the comparatively slower folding rates of the hydrophobic deletion mutations at higher temperatures are reflections of their more flexible unfolded state ensembles, which allow the turn sequence to explore a larger region in the allowed conformational space, thus increasing the entropic penalty associated with turn formation and hence the folding time of the helical hairpin.

4.4 Materials and Methods

Materials. All materials were used as received. D_2O (D, 99.96 %) was purchased from Cambridge Isotope Laboratories (Andover, MA). Amino acids for peptide synthesis were obtained from Advanced ChemTech (Louisville, KY).

Peptide sample preparation. The peptides were synthesized using standard flouren-9-ylmethoxycarbonyl (Fmoc)-chemistry on a PS3 peptide synthesizer from Protein Technologies, (Tucson, AZ) and were purified by reversed-phase HPLC (Agilent Tech-

nologies, Santa Clara, CA). The mass of each peptide was verified by matrix-assisted laser desorption ionization mass spectrometry (Voyager-DE RP, Applied Biosystems, Foster City, CA). Following purification, the peptide solutions were lyophilized, dissolved in D_2O and titrated to pH* 5.2 $\pm$ 0.1 (pH meter reading) with a solution of sodium hydroxide in D_2O. Another round of lyophilization in D_2O ensured complete removal of H_2O. Peptide sample used in spectroscopic measurements were prepared by dissolving an appropriate amount of lyophilized peptide solid in 20 mM D_2O 2-(N-morpholino)ethanesulfonic acid (MES) buffer (pH* 5.2). For circular dichroism (CD) measurements, the peptide concentration was about 40 μM, determined optically using the absorbance of tyrosine at 280 nm and an extinction coefficient of 1492 cm^{-1} M^{-1}. For IR measurements, the peptide concentration was in the range of 1-4 mM.

CD measurement and analysis. CD data were collected on an Aviv Model 410 instrument (Aviv Biomedical, Lakewood, NJ) using a 1 mm cuvette. For thermal unfolding measurements, the CD signal at 222 nm was averaged for 30 s at each temperature. After the highest temperature was reached, the sample was cooled to 25 °C and another full CD spectrum was measured to ensure that unfolding was reversible. The folding-unfolding thermodynamics of PYY and the mutants were determined by globally fitting their CD thermal melting transitions to a two-state model, the detail of the global fitting method has been described previously [223].

Briefly, the folded and unfolded CD baselines $\theta_F(T)$ and $\theta_F(T)$ were assumed to be linear functions of temperature, i.e., $\theta_F(T) = m + n \cdot T$, $\theta_F(T) = p + q \cdot T$, and n and q (m and p) were treated as global (local) fitting parameters.

T-jump relaxation measurement. The T-jump induced relaxation kinetics of PYY and mutants were measured by probing the time-dependent absorbance change of the helical amide I′ band of the peptide at 1631 cm^{-1}. A detailed description of the T-jump IR setup can be found elsewhere [211]. The only difference is that in the current study the transient IR signals were probed by a tunable quantum cascade laser (Daylight Solutions, Poway, CA). The T-jump induced relaxation kinetics of the PYY peptides were found to develop in two distinct phases. The fast kinetic phase is unresolvable due to the 20 ns rise time of the IR detector and is attributed to temperature-induced spectral shift and/or imperfect background subtraction. Thus, only the slow phase, which can be described by a single-exponential function, is considered in the following discussion.

4.5 Acknowledgement

We gratefully acknowledge financial support from the National Institutes of Health (GM-065978 and RR-01348).

4.6 Original Publication

This Chapter has been published: Matthias M. Waegele and Feng Gai, *Biochemistry* (2010) 49, 7659-7664. DOI: 10.1021/bi100851c

Table 4.1: Unfolding thermodynamic parameters of PYY and mutants obtained from global fitting of their CD thermal melting curves.

Peptide	ΔH_m° (kcal mol^{-1})	ΔS_m° (cal K^{-1} mol^{-1})	ΔC_p (cal K^{-1} mol^{-1})	T_m ($^\circ$C)
PYY	14.5 ± 1.6	45.8 ± 4.8	221 ± 25	44.0 ± 1.6
A7Y	13.5 ± 1.3	41.9 ± 3.9	333 ± 47	48.7 ± 2.0
S13A	5.5 ± 0.7	19.1 ± 2.5	221 ± 25	13.5 ± 2.0
P14A	10.2 ± 0.7	33.1 ± 2.1	221 ± 25	36.0 ± 1.9
Y21A	15.8 ± 0.2	51.4 ± 0.7	139 ± 41	35.1 ± 0.9
Y27A	3.3 ± 0.5	11.8 ± 1.9	100 ± 29	4.1 ± 2.7

Table 4.2: Folding and unfolding time constants of PYY and mutants at 30 °C.

Peptide	k_f^{-1} (μs)	k_u^{-1} (μs)
PYY	2.8 ± 0.5	7.1 ± 1.3
A7Y	3.6 ± 0.6	9.6 ± 1.7
S13A	5.6 ± 1.0	2.8 ± 0.5
P14A	6.9 ± 1.2	9.5 ± 1.7
Y21A	2.8 ± 0.5	4.2 ± 0.8
Y27A	4.3 ± 0.8	2.1 ± 0.4

Table 4.3: Folding and unfolding time constants of PYY and mutants at 50 °C.

Peptide	k_f^{-1} (μs)	k_u^{-1} (μs)
PYY	2.4 ± 0.4	1.6 ± 0.3
A7Y	2.2 ± 0.4	2.0 ± 0.4
S13A	7.5 ± 1.4	1.2 ± 0.2
P14A	3.6 ± 0.6	1.6 ± 0.3
Y21A	4.6 ± 0.8	1.3 ± 0.3
Y27A	5.5 ± 1.0	1.3 ± 0.2

Figure 4.1: NMR structure of porcine PYY (PDB code: 2RLK). The side chains of those residues targeted for mutation are shown in stick representation.

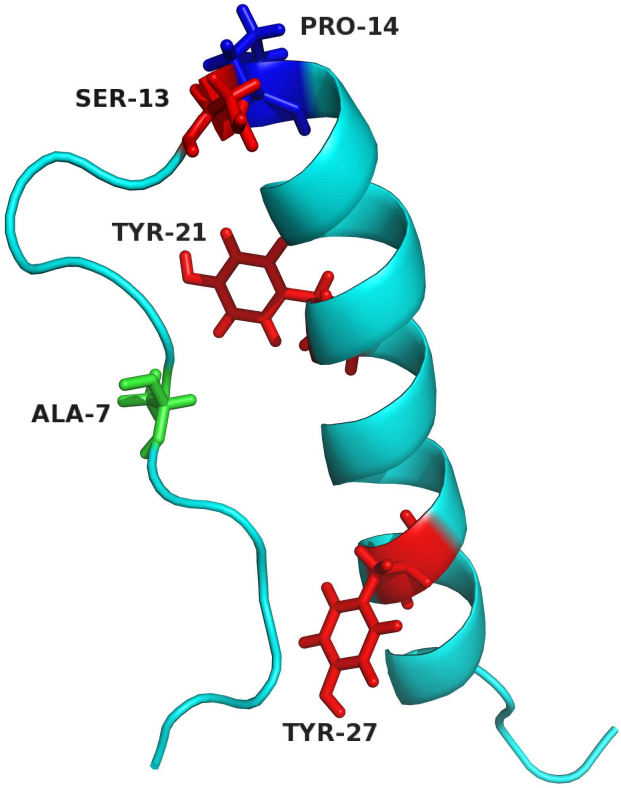

PRO-14
SER-13
TYR-21
ALA-7
TYR-27

Figure 4.2: Far-UV CD spectrum of PYY at 25 °C.

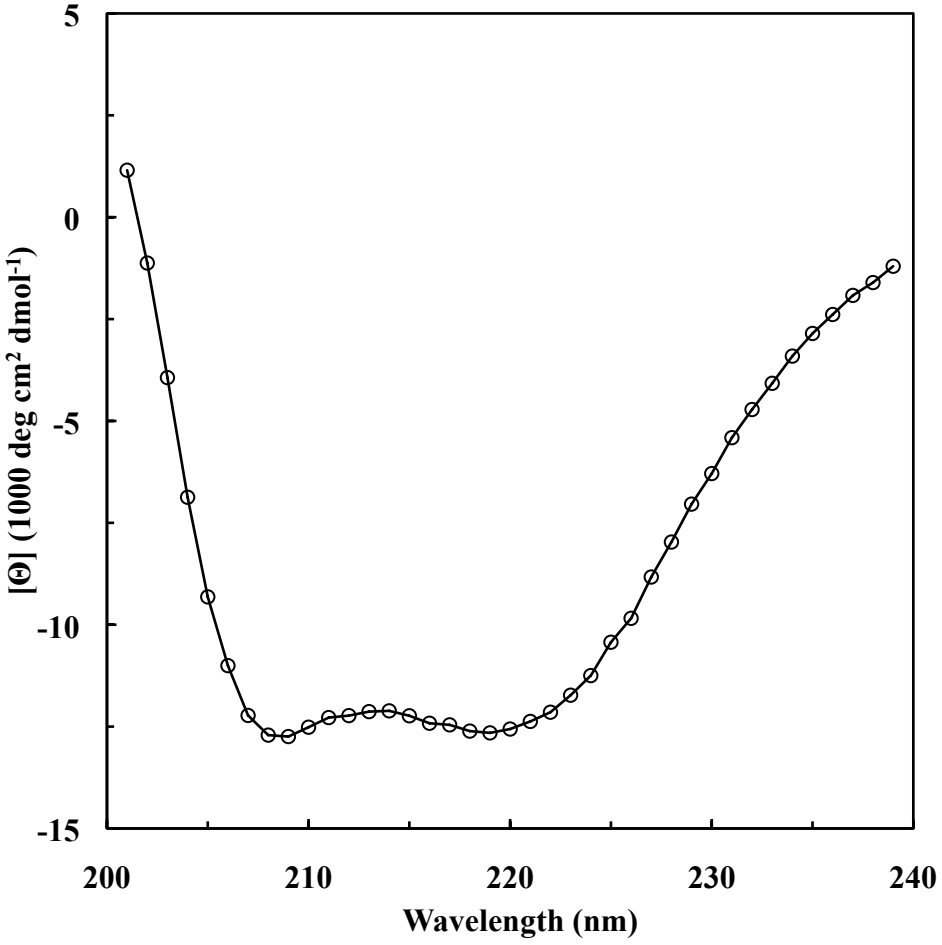

5
0
-5
-10
-15
[Θ] (1000 deg cm2 dmol-1)
200
210
220
230
240
Wavelength (nm)

Figure 4.3: Thermal unfolding CD curves of PYY and its mutants, as indicated. Lines are fits to the two-state model described in the text.

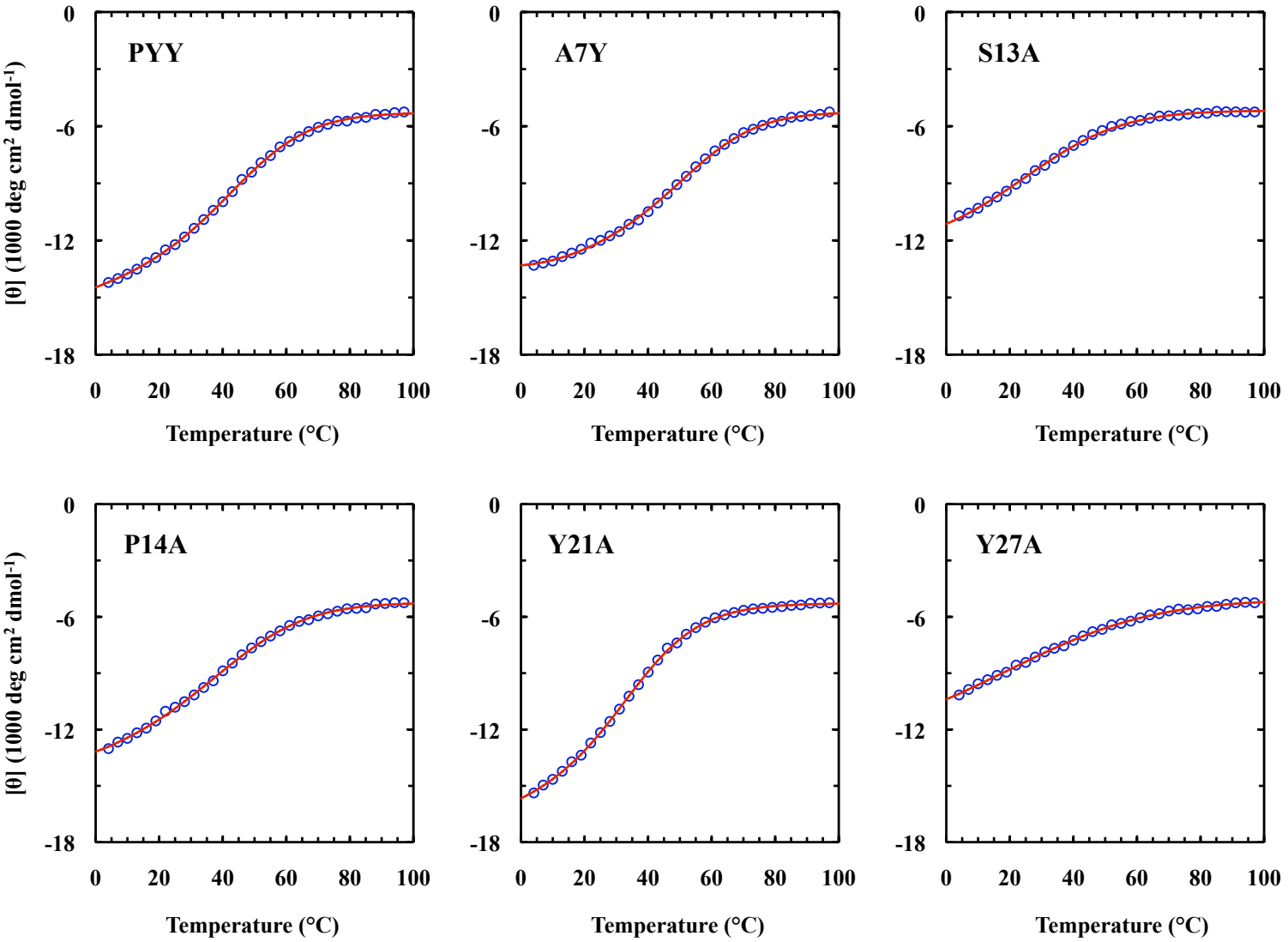

PYY
A7Y
S13A
P14A
Y21A
Y27A
[θ] (1000 deg cm² dmol⁻¹)
Temperature (°C)
0
-6
-12
-18
0
20
40
60
80
100

Figure 4.4: A representative relaxation kinetic trace of PYY in response to a T-jump of 9.2 °C, from 23.0 to 32.2 °C. For clarity, an unresolved kinetic phase has been subtracted from these data. The smooth line represents the best fit of these data to a single-exponential function with a life time of 1.9 μs.

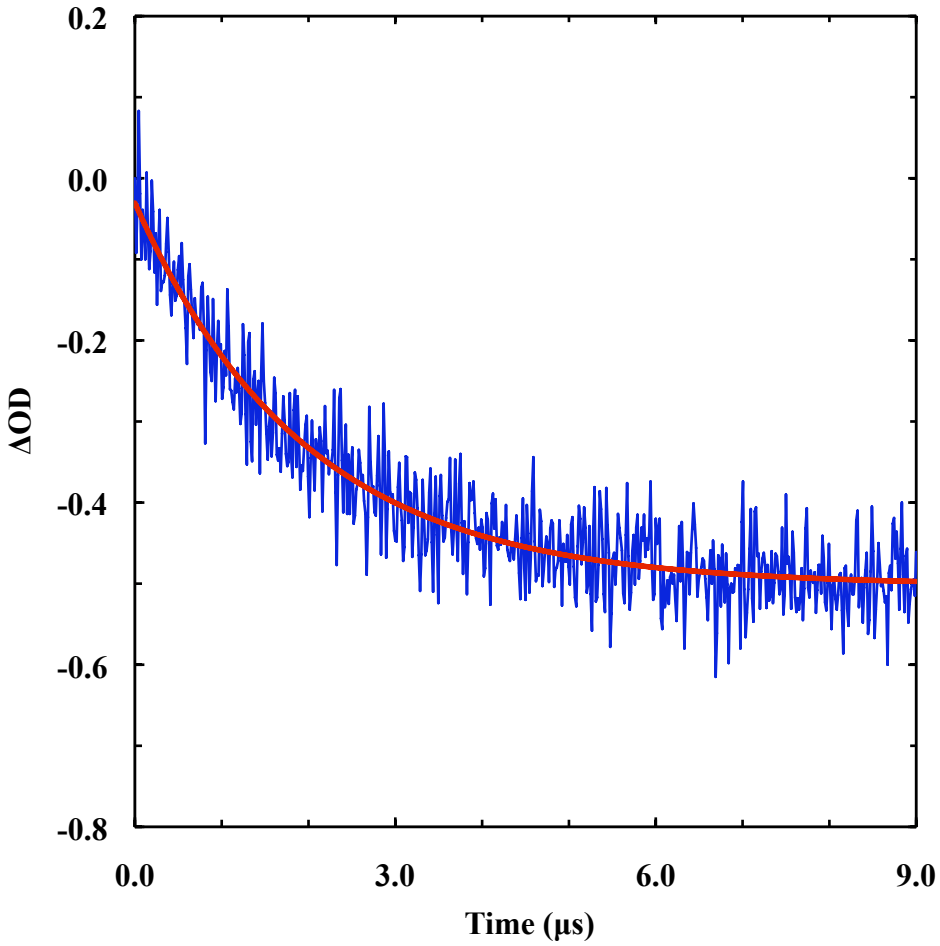
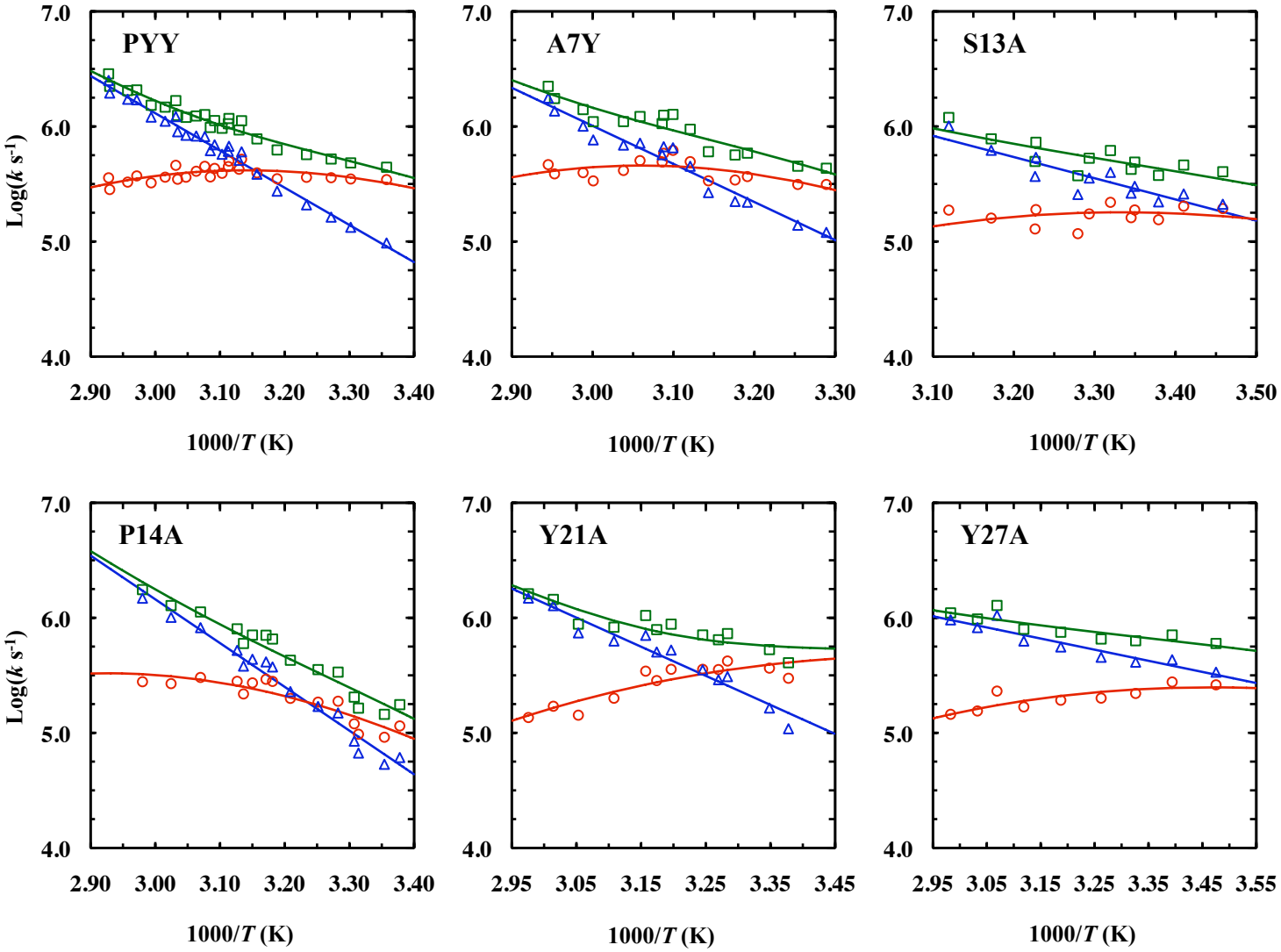

0.2
0.0
-0.2
-0.4
-0.6
-0.8
ΔOD
0.0
3.0
6.0
9.0
Time (μs)

Figure 4.5: Arrhenius plots of the observed relaxation rates (□) as well as folding (◯) and unfolding (△) rate constants for the peptides as indicated in the respective panels. For each peptide, lines represent global fits of the folding and unfolding rate constants to the Eyring equation, $\ln(k) = \ln(D) - \Delta G^{\ddagger}/(RT)$, where D is a constant and $\Delta G^{\ddagger}$ is the free energy of activation. In this work, D was arbitrarily set as 10^{10} s^{-1} and $\Delta G_f^{\ddagger}$ and $\Delta G_u^{\ddagger}$ were constrained so that $\Delta G_f^{\ddagger}(T) - \Delta G_u^{\ddagger}(T) = \Delta G^{\circ}(T)$, where $\Delta G^{\circ}(T)$ is the equilibrium free energy of unfolding at temperature T.


PYY
A7Y
S13A
P14A
Y21A
Y27A
Log(k s-1)
1000/T (K)
7.0
6.0
5.0
4.0
2.90 3.00 3.10 3.20 3.30 3.40
2.90 3.00 3.10 3.20 3.30
3.10 3.20 3.30 3.40 3.50
2.95 3.05 3.15 3.25 3.35 3.45
2.95 3.05 3.15 3.25 3.35 3.45 3.55

Chapter 5

Effect of Macromolecular Crowding on Protein Folding Dynamics at the Secondary Structure Level

5.1 Introduction

Understanding how proteins fold *in vivo* poses a formidable challenge. Thus, a majority of protein folding studies have been carried out *in vitro* and under conditions wherein only dilute aqueous solutions were used. While such studies have provided invaluable insights into our understanding of the protein folding problem, these did not take into account the possible effects arising from macromolecular crowding - an important but often neglected aspect of the intracellular environment [224, 225, 26, 226]. For instance, the presence of macromolecules near a protein could alter its folding energy landscape simply through the excluded volume effect since folding results in a compaction of the polypeptide chain [227]. Indeed, several recent experimental and computational studies have shown that such effects of volume exclusion can significantly alter the stability and folding rate of a protein [228, 229, 230, 231, 232, 29, 233, 234, 235, 236, 237]. Besides such entropic effects, macromolecular crowding could also affect the dynamics of protein folding through other mechanisms. [In the present case, the effect of macromolecular crowding is referred to as the net effect of an inert macromolecule on the folding properties of the protein or peptide molecule in question, which includes both static effects, such as that arising from confinement, and dynamic effects, such as modulation of the frictional drag experienced by conformational motions along the reaction coordinate.] For example, it is well known that a high-mass macromolecular crowding agent (i.e.,

inert polymers) not only changes the dynamic viscosity of the solution (i.e., macroviscosity) but also could modulate the microviscosity of the protein environment in which the folding reaction occurs [238, 239, 240, 241], thus perturbing the underlying chain dynamics, as shown by Neuweiler *et al*. [242]. However, most of the experimental studies on the effects of macromolecular crowding on protein folding carried out thus far have dealt with proteins of fairly large size [26, 226, 228, 229, 232, 233, 234, 235, 236], wherein the excluded volume effect appears to dominate, thus obscuring other subtle but important effects arising from the presence of crowding agents. In this work, we studied the folding thermodynamics and kinetics of three relatively small protein motifs in the presence of two commonly used crowding agents, dextran 70 and ficoll 70, with the aim of providing new insights into the effect of macromolecular crowding on folding events taking place over a relatively short length scale. These peptides form different types of protein secondary and/or supersecondary structures in solution, specifically, a 34-residue monomeric α-helix (L9:41-74) [243], a 34-residue cross-linked helix-turn-helix (HTH) motif (Z34C-m1, which is the D20A mutant of Z34C) [205], and a 16-residue β-hairpin (trpzip4-m1) [208]. Given the critical importance of protein secondary structure formation in several protein folding models, such as the framework model [199], this study is expected to also have considerable implications for the applicability of those models in describing *in vivo* protein folding. Dextran 70 is a flexible and linear ($< 5\%$ branching) polymer of D-glucopyranose

that behaves as a quasirandom coil (R_h of $\sim$ 63 Å) [241, 244, 245, 246, 247] whereas ficoll 70 is a compact and highly cross-linked and branched copolymer of sucrose and epichlorohydrin that can be approximated as a semi-rigid sphere (R_h of $\sim$55 Å) [245, 246, 247]. Thus, comparative studies employing these two polymers allow one to examine how the nature and the geometric shape of the respective crowding agent affect the folding dynamics of the protein system in question. Interestingly, only the thermodynamic stability of the shortest peptide studied here (i.e., trpzip4-m1) shows an appreciable change when its environment is crowded by ficoll 70. Similar to that observed for large proteins, macromolecular crowding leads to an increase in the thermal stability of trpzip4-m1 in the presence of 200 g/L of ficoll 70, a concentration that falls within the concentration range that has been used in previous crowding studies [224, 225, 228, 229, 232, 29, 234]. However, in contrast to the common notion that macromolecular crowding increases the rate of protein folding, our results show that the folding rate of trpzip4-m1 in fact decreases in the presence of either ficoll 70 or dextran 70. Taken together, these results indicate that besides the commonly encountered excluded volume effect, other factors need to be considered when assessing the effect of macromolecular crowding on protein folding.

5.2 Results

All crowding experiments were carried out in 20 mM phosphate buffer in D_2O at pH* 7 in the presence of either 200 g/L of dextran 70 or 200 g/L of ficoll 70, whereas all other experiments were carried out in 20 mM phosphate buffer in D_2O at pH* 7.

L9:41-74. The far-UV circular dichroism (CD) spectra of L9:41-74 in both dextran 70 and ficoll 70 solutions at 4 °C exhibit the characteristic double minima of α-helices at 208 and 222 nm, respectively, and overlap with the CD spectrum of L9:41-74 in D_2O buffer solution (Fig. 5.1). In addition, as shown in Fig. 5.2, the presence of these crowding agents has little effect on the CD thermal denaturation profile of L9:41-74, indicating that the thermodynamics of the underlying helix-coil transition is not sensitive to the environmental changes induced by these crowding agents. The relaxation kinetics of L9:41-74 in these polymer solutions were studied by the laser-induced temperature jump (T-jump) infrared method, the details of which have been described elsewhere [211]. Similar to those observed in dilute solution [243], the T-jump-induced relaxation kinetics of L9:41-74 contain two distinct phases (Fig. 5.3). The fast phase cannot be resolved by our setup and has been attributed to temperature-induced spectral shift of the amide-I' band [214]. On the other hand, the slow phase, which arises from the conformational redistribution process of the peptide in response to the T-jump, is well resolved and can be described by a single-exponential function. As shown in Fig. 5.4, within the temperature range studied, the

T-jump induced conformational relaxation rates of L9:41-74 in both dextran 70 and ficoll 70 solutions are almost identical with the relaxation rate of the peptide in D_2O buffer solution. For example, at 20 °C, the relaxation time constants of L9:41-74 in dextran 70 and ficoll 70 solutions are 1.4 ± 0.2 and 1.5 ± 0.2 μs, respectively, whereas the relaxation time constant in D_2O buffer solution is 1.17 ± 0.15 μs [243]. Taken together, these thermodynamic and kinetic results indicate that the folding-unfolding transition of this α-helical peptide is not sensitive to macromolecular crowding, at least not to that induced by 200 g/L of dextran 70 or ficoll 70.

Z34C-m1. The far-UV CD spectra of Z34C-m1 obtained at 4 °C in both dextran 70 and ficoll 70 solutions are almost completely superimposable with the CD spectrum obtained in D_2O buffer solution (Fig.5.5), indicating that the helical nature of the folded conformation is not affected by the crowding agents. The thermal denaturation of Z34C-m1 in both dextran 70 and ficoll 70 solutions, measured by monitoring the change in its helical CD signal at 222 nm with increasing temperature (Fig. 5.6), also shows that these crowding agents do not change the thermal stability of the peptide to any appreciable extent. Indeed, globally fitting the CD data to a two-state model reveals that the thermal melting temperature (T_m) of Z34C-m1 in these crowded environments is only 2-3 °C higher than that [205] in D_2O buffer (Table 5.1). Consistent with these thermodynamic assessments, the T-jump-induced relaxation rates of Z34C-m1 in these polymer solutions also show only a moderate decrease from

those measured in dilute aqueous solution (Fig. 5.7 and Table 5.1).

trpzip4-m1. Similar to the CD spectrum obtained in D_2O buffer solution [208], the far-UV CD spectra of trpzip4-m1 in both dextran 70 and ficoll 70 solutions exhibit a positive band centered at ∼229 nm (Fig. 5.8), arising from excitonic coupling between the paired tryptophan side chains. As shown in Fig. 5.9, the thermal unfolding transitions of trpzip4-m1 in these solutions, as monitored by the change in the CD signal at 229 nm, show characteristics of a cooperative thermal unfolding process. As indicated in Table 5.1, the thermal melting temperature (T_m) of trpzip4-m1 in dextran 70 solution is increased only slightly compared with that (32 °C) in D_2O buffer, whereas in ficoll 70 solution, the T_m of this peptide shows a substantial increase to ∼ 44 °C (Table 5.1). The relaxation kinetics of trpzip4-m1 in response to a T-jump were also probed at 1631 cm^{-1}, where antiparallel β-sheets are known to absorb [208]. As shown in Figs. 5.10 and 5.11, unlike L9:41-74, the folding and unfolding rates of trpzip4-m1 are distinctly slower in the presence of the crowding agents. For example, at 35 °C, the folding and unfolding time constants of trpzip4-m1 in dextran 70 solution are 96 ± 14 and 92 ± 14 μs, respectively, whereas in dilute aqueous solution [208], this peptide folds in 47.5 ± 2.3 μs and unfolds in 38.1 ± 2.0 μs (Table 5.1).

5.3 Discussion

Since macromolecular crowding is an intrinsic feature of the cellular environment [248, 249], there has been considerable interest in recent years in investigating its effect on protein folding, both experimentally [26, 226, 228, 229, 232, 233, 234, 235, 236] and computationally [26, 226, 230, 231, 29, 234, 237]. However, almost all of the previous experimental efforts in this area have been focused on large proteins [26, 226, 228, 229, 232, 233, 234, 235, 236], thus providing little, if any, information on the effect of macromolecular crowding on protein folding at the secondary structure level. While such a bias in focus is understandable because most protein secondary structural elements (e.g., monomeric α-helix and β-hairpin) are probably too small to be profoundly affected by the excluded volume effect [26, 227], understanding the influence of crowding on the folding dynamics of such small structural moieties could provide new insights into the otherwise complex interplay between different crowding effects. In addition, in view of the important role of secondary structure formation in existing protein folding models, such as the framework model [199], there is a strong need for investigation of folding dynamics of secondary structural elements in crowded environments in order to understand protein folding *in vivo*. In this work, we studied the folding thermodynamics and kinetics of three distinct secondary structural elements [i.e., a monomeric α-helix (L9:41-74), a β-hairpin (trpzip4-m1), and an HTH motif (Z34C-m1)] in the presence of two commonly used crowding agents, namely,

dextran 70 and ficoll 70.

L9:41-74. L9:41-74 is the central α-helix of ribosomal protein L9 from the bacterium Bacillus stearothermophilus. Owing to a series of favorable side chain-side chain interactions, mostly electrostatic in nature, this peptide remains folded even in isolation in aqueous solution [250]. Previously, we have shown that this peptide folds on a timescale that is significantly slower than that of alanine-based peptides and that its folding time may be a more realistic representation of the timescale in which α-helices in proteins fold [243]. Thus, it would be quite interesting to further examine how crowding affects its folding-unfolding dynamics. Our CD measurements (Fig. 5.2) show that the thermal stability of L9:41-74 in either dextran 70 or ficoll 70 solution remains practically unchanged from that in dilute aqueous solution, suggesting that the helix-coil transition experiences little, if any, crowding effect. This observation is however unexpected because the chain length of this α-helix is estimated to be 50 Å (assuming a full helical structure), which is comparable with the hydrodynamic radii of dextran 70 and ficoll 70. Besides the apparent excluded volume effect, addition of a polymer to aqueous solution is also known to affect the dynamic viscosity of the solution. For example, the dynamic viscosities of 200 g/L of dextran 70 and 200 g/L of ficoll 70 are about 18 and 10 times greater than the dynamic viscosity of water at 20 °C [245, 251], respectively. Previously, Jas *et al.* [59] showed that the T-jump-induced relaxation rate of an alanine-based helical peptide is inversely proportional

to $\eta^{0.6}$ (where η represents the viscosity of the solution) when small viscogens, such as glucose and sucrose, were used to increase the viscosity of the solution. Thus, our observation that the T-jump-induced relaxation kinetics of L9:41-74 in both dextran 70 and ficoll 70 solutions are almost the same as those obtained in dilute D_2O solution (Fig. 5.4) is quite interesting and warrants further discussion.

The viscosity dependence of protein folding rates stems from the diffusive nature of the associated barrier-crossing events [252, 253, 254, 255], namely, from the requirement of polypeptide chain motion to form the stabilizing native contacts via diffusion. Previous studies have shown that under the current crowding conditions the polymer molecules can form porous networks [238, 240, 242] wherein the viscosity determined via the diffusion of a probe molecule can be significantly lower than the dynamic viscosity of the bulk solution [238, 242, 256]. To differentiate these viscosities, we referred to the one measured via molecular diffusion as the microviscosity of the solution. Since the sizes of the three peptides used in this study are relatively small, it is reasonable to assume that their folding rates would respond more to the microviscosity rather than to the dynamic viscosity of the polymer solution. Thus, to further probe the microviscosity of the current crowding solutions, we have also measured the characteristic translational diffusion times (τ_D, a quantity that is inversely proportional to the diffusion constant D) of three molecular systems of varying sizes using fluorescence correlation spectroscopy (FCS), namely, (i) a fluorescent

dye [rhodamine 6G (R6G)], (ii) a 37-residue fluorescently labeled pH (low) insertion peptide (pHLIP), and (iii) the human serum albumin (HSA) protein complexed with the dye Nile red (NR). As shown in Table 5.2, the translational diffusion times of all three probes indicate that the microviscosities of ficoll 70 and dextran 70 solutions are about four to six times larger than the microviscosity of water, depending on the size of the probe and on the crowding agent.

These FCS measurements indicate that not only the dynamic (or bulk) viscosity but also the microviscosity measurements fall short of providing a rationale behind the observation that the relaxation rate of L9:41-74 is insensitive to the presence of crowding agents - this being in stark contrast to the well-documented notion that the rate of α-helix formation is susceptible to viscosity [59, 257]. However, this apparent discrepancy may be reconciled by taking into account the notion that the diffusion constant could also be distance dependent [258, 259]. For example, the diffusion constant (D) of a probe molecule in a polymer solution (Fig. 5.12) consisting of porous networks [260] is:

$$D = D_0 \exp\left(-ac^{\frac{1}{2}}\right) \tag{5.1}$$

where c is the concentration of the polymer and D_0 is the diffusion constant of the probe at $c = 0$. The constant a has been shown to characterize the length scale (ξ) of the pores embedded in the porous networks, at least for a simple pore model [259]. Thus, for short diffusion modes or diffusion events during which the diffusing

particles travel a distance smaller than the pore size ξ of the polymer solution, the particles rarely come into contact with the polymer chains and their diffusion constant is hence close to that found in bulk water. On the other hand, for long diffusion modes or diffusion events during which the particles travel a distance much larger than ξ, the resultant diffusion constant becomes smaller than that measured in bulk water, similar to the diffusion constants observed in our FCS measurements. In this context, it is clear that the viscosity effect exerted by a polymeric crowding agent on the folding rate of a peptide depends on the length scale that the polypeptide chain has to traverse from the unfolded state to the transition state along the folding coordinate. The formation of an α-helix can be regarded as a series of local events wherein a hydrogen bond between the amino acids at "i" and "i+4" positions along the polypeptide chain is formed. Thus, the length scale (a few ångströms) over which the key α-helix folding events take place is far less than the pore size of the polymer solutions used in the current study and the corresponding folding rate therefore does not show any significant deviation from that in dilute aqueous solution. In contrast, in studies where small viscogens (e.g., glucose and sucrose) are used to increase the viscosity, the resulting solution viscosity is microscopically homogeneous and the effect of increased solvent friction can therefore be experienced down to a very short length scale. Thus, in such cases, the viscosity dependence of the folding kinetics of protein secondary structural elements might become more pronounced, as has been shown in

a previous study [59].

Z34C-m1. Z34C-m1 is a mutant of Z34C (i.e., D20A) that forms a crossed-linked HTH structure, a common structural motif found in DNA-binding proteins [23]. We have previously shown that Z34C-m1 folds significantly slower than Z34C does since the mutation deletes a hydrogen bond that is critical to the stability of the reverse turn [205]. Z34C-m1 constitutes a good model system to further examine the effect of microviscosity on the folding-unfolding kinetics of α-helices in a protein context as the disulfide cross-linker prevents the protein to become extended upon unfolding and hence makes it less likely to experience the excluded volume effect. As expected, the folding thermodynamics and kinetics of Z34C-m1 are only moderately affected by the crowding agents employed herein (Figs. 5.6 and 5.7 and Table 5.1). Thus, these results are consistent with those for L9:41-74 and further corroborate the notion that macromolecular crowding does not affect to any significant extent the folding dynamics of α-helices that can fold independently. However, in a protein context, the frictional drag along the folding coordinate of an α-helix may be position dependent and thus could exert a more complex effect on the folding dynamics. Apparently, for β-sheets whose folding requires relatively large-scale chain diffusions, the effect of microviscosity is expected to play a more important role, as discussed below.

trpzip4-m1. To further explore the extent to which the nature of the folded topology determines the effect of macromolecular crowding, we have also studied the folding

thermodynamics and kinetics of trpzip4-m1, the D46A mutant of the 16-residue β-hairpin trpzip4 [208] in both dextran 70 and ficoll 70 solutions. While the thermal denaturation properties of trpzip4-m1 are almost unaffected by addition of dextran 70, its thermal melting temperature in ficoll 70 solution shows a substantial increase from that in D_2O buffer [208]. This difference manifests the disparity in terms of crowding efficiency of these crowding agents and can be rationalized based on the overall difference in the molecular structures of these polymers. Previous studies have shown that the persistence length of Dextran is about 0.4 nm [261], which renders it to be a relatively flexible polymer. On the other hand, ficoll 70 is comparatively more rigid due to its cross-linked structure [247, 262]. Thus, based on these considerations, it is conceivable that dextran 70 can accommodate more interstitial spaces in comparison with ficoll 70 at the same concentration, therefore allowing the unfolded polypeptide chain of trpzip4-m1 more space to escape and hence leading to a weaker excluded volume effect. In contrast to that observed in other studies, [228, 231, 29] macromolecular crowding results in a decrease in the folding rate of trpzip4-m1. This result thus clearly indicates that a macromolecular crowding agent can affect not only the protein stability via the excluded volume effect but also the folding dynamics through modulation of the effective viscosity of the protein environment. In the case of dextran 70, the thermal stability of the β-hairpin exhibits only a very modest change (Table 5.1), thus implying that the observed reduction in the folding and unfolding

rates is unlikely to result from the excluded volume effect and rather arises from a viscosity effect. This interpretation is consistent with the notion that the dynamics of peptide loop closure and β-hairpin folding require relatively large-scale and nonlocal motions, in contrast to the folding of an α-helix, wherein the chain primarily undergoes a series of local structural reorganizations. Thus, during such a folding event, the peptide chain has a higher chance of colliding with the polymer network formed by the crowding molecules, resulting in a slower folding-unfolding rate. This picture is consistent with a recent FCS study by Neuweiler *et al.* [242], who showed that the intrachain diffusion rate of a series of fluorescently labeled short poly(GS)-peptides in ficoll 70 depends on the net effect of two opposing forces: the excluded volume effect and the increased viscous drag of the polymer solution. The effect of viscosity on protein folding or unfolding rate constant (k_{obs}) can often be described by the hydrodynamic approximation of the Kramers equation in the high friction limit; that is,

$$k_{obs} = \frac{A}{\eta^{\alpha}(T)} \exp\left(-\frac{\Delta G^{\ddagger}}{RT}\right) \tag{5.2}$$

where $\eta(T)$ is the effective viscosity of the medium at temperature T and has been assumed here to be independent of the spatial location of the peptide, α is equal to unity in many cases, A is a constant, $\Delta G^{\ddagger}$ is the apparent free-energy barrier at temperature T, and R is the gas constant. Thus, a crowding agent can modulate the folding and unfolding rate constants of a protein by changing η, $\Delta G^{\ddagger}$, or both.

Apparently, the effect related to η is a dynamic one, whereas that related to $\Delta G^{\ddagger}$ is static (or the excluded volume) in nature. For a simple two-state scenario, assuming that A is the same for both folding and unfolding and that it is independent of crowding, one can easily show that (assuming the position of the transition state is not affected by crowding)

$$\ln\left(\frac{k_f^c}{k_f^0}\right) + \ln\left(\frac{k_u^c}{k_u^0}\right) = 2\alpha \ln\left(\frac{\eta_0(T)}{\eta_c(T)}\right) + \frac{1}{RT}\left(\Delta G_u^{c-0} + \Delta G_f^{c-0}\right) \tag{5.3}$$

where k_f and k_u are the folding and unfolding rate constants, respectively; $\Delta G_u^{c-0} = G_u^c - G_u^0$ and $\Delta G_f^{c-0} = G_f^c - G_f^0$, where G_u and G_f represent the free energies of the unfolded and folded states, respectively; and, in all cases, the superscripts "0" and "c" represent the dilute and crowded solutions, respectively. In addition, $\eta_0(T)$ and $\eta_c(T)$ are the effective viscosities of the dilute and crowded solutions at temperature T, respectively. For dextran 70, the increased stabilization of the β-hairpin conformation of trpzip4-m1 arising from crowding is considered to be negligible (i.e., $\Delta G_u^{c-0} + \Delta G_f^{c-0} \sim 0$). Thus, with the use of the measured folding and unfolding rate constants of trpzip4-m1 in the presence and in the absence of crowding agents, in conjunction with the assumption that α is approximately unity for β-hairpin folding [59], the value of η_c/η_0 is calculated to be 3.5 ± 0.5 at 20 °C, which is comparable with the ratio of diffusion times $\tau_c/\tau_0 = 3.8$ of the pHLIP peptide in 200 g/L of dextran 70 and in water at room temperature (Table 5.2). Therefore, this agreement further corroborates the idea that dextran 70 affects trpzip4-m1 folding via mostly the dynamic (or viscosity)

effect.

In the case of ficoll 70, the presence of the crowding agent leads to not only a decrease in the folding rate of trpzip4-m1 but also a significant increase in the stability of β-hairpin conformation (Figs. 5.9 and 5.11 and Table 5.1), indicating that the folding and unfolding dynamics of trpzip4-m1 experience both the excluded volume and viscosity effects. As a result, the unfolding rate is decreased to a larger extent than the folding rate, compared with their respective values in D_2O buffer solution (Table 5.1). This is because both the excluded volume and viscosity effects slow the rate of unfolding, whereas they exhibit opposite effects on the folding rate. Furthermore, it is easy to show that

$$\frac{k_f^c}{k_f^0} = \left(\frac{\eta_0}{\eta_c}\right)^{\alpha} \exp\left(\frac{\Delta\Delta G^{\ddagger}}{RT}\right) \tag{5.4}$$

$\Delta\Delta G^{\ddagger} = \Delta G_0^{\ddagger} - \Delta G_c^{\ddagger}$, which is a quantitative measure of the excluded volume effect on the folding dynamics. Thus, it is possible to dissect the static and dynamic contributions to the total crowding effect if η_c/η_0 is known. While η_c/η_0 cannot be determined independently from the current study, the result obtained with dextran 70 nevertheless suggests that it approaches the τ_c/τ_0 ratio measured by the FCS method. Thus, using the ratio $\tau_c/\tau_0 = 4.1$ (Table 5.2), we estimated the $\Delta\Delta G^{\ddagger}$ value to be 0.3 ± 0.1 kcal mol^{-1} at 35 °C for the folding of trpzip4-m1 in 200 g/L of ficoll 70 solution. As expected, the decrease in the folding free-energy barrier of trpzip4-m1 due to the excluded volume effect is very modest, as the additional stabilization (i.e.,

$\Delta\Delta G^0 = |\Delta G_c^0| - |\Delta G_0^0|$) of trpzip4-m1 due to the presence of 200 g/L of ficoll 70 at 35 °C is only about 0.5 ± 0.1 kcal mol^{-1}, owing to the small size of the peptide relative to the size of the crowding agent, which is in agreement with the theoretical prediction by Minton [227]. However, we expect that the approach outlined above will be very useful for dissecting the static and dynamic effects of a crowding agent for other protein systems.

5.4 Conclusions

Employing two commonly used crowding agents, dextran 70 (200 g/L) and ficoll 70 (200 g/L), we have studied how macromolecular crowding affects the folding-unfolding kinetics of three peptides that fold into different conformations in solution, namely, an α-helix (L9:41-74), a cross-linked HTH (Z34C-m1), and a β-hairpin (trpzip4-m1). Interestingly, only the folding thermodynamics and kinetics of the shortest peptide, trpzip4-m1, were found to show significant changes in the ficoll 70 solution, thus indicating that the effect of macromolecular crowding is sensitive to the size and the shape of both the peptide and the crowding agent. In addition, the observation that macromolecular crowding does not significantly alter the folding dynamics of the two helical peptides is consistent with the notion that α-helix formation involves mostly local interactions. Furthermore, and in contrast to that observed for large proteins, macromolecular crowding results in a decrease in the folding rate of trpzip4-

m1, indicating that the motion of the peptide chain experiences higher friction arising from the presence of the crowding agent and that the viscosity effect in this case outplays the excluded volume effect on the folding kinetics of the β-hairpin.

5.5 Materials and Methods

Peptide synthesis and purification. All peptides used in this study were synthesized based on standard Fmoc protocols on a PS3 automated peptide synthesizer (Protein Technologies, MA) and purified by reversed-phase chromatography. The peptide sequences are as follows:

Ac-41PANLKALEAQ-51KQKEQRQAAE-61ELANAKKLKE-71QLEK-NH_2

(L9:41-74),

6FNMQCQRRFY-16EALHAPNLNE-26EQRNAKIKSI-36RDDC-NH_2 (Z34C-m1),

GEWTWADATKTWTWTE-NH_2 (trpzip4-m1).

TMR maleimide (Molecular Probes, CA), a thiol reactive dye, was used to label the cysteine variant of pHLIP peptide [263] (sequence: ACEQNPIYWA-RYADWLFTTP-LLLLDLALLV-DADEGTG) following the protocol available in the Molecular Probes handbook. Oxidation of the Z34C mutant was done as previously described [205]. The identity of the peptide sample was further verified by matrix-assisted laser desorption ionization mass spectroscopy. Multiple rounds of lyophilization against a solution of 0.1 M DCl/D_2O were used to remove the exchangeable hydrogen atoms and the

residual trifluoroacetic acid from the peptide synthesis. Peptide solutions used in both CD and infrared measurements were prepared by directly dissolving lyophilized peptide solids in 20 mM phosphate buffer, pH* 7, with or without the presence of the crowding agent. The final peptide concentration for L9:41-74 was determined by comparing the published CD data [250], and the concentrations for other peptides were determined as previously described [205, 208] and were about 16-134 μM and 0.8-1.5 mM for CD and infrared samples, respectively.

Crowding agents. Dextran 70 and ficoll 70 were purchased from Fisher Scientific (PA) and GE Healthcare (NJ), respectively, and were used as received.

CD spectroscopy. All CD data were collected on an Aviv 62A DS CD spectrometer (Aviv Associate, NJ). The folding-unfolding thermodynamics of each peptide were obtained by fitting its CD thermal melting transition obtained at either 222 nm (for the helical peptides) or 229 nm (for the β-hairpin) to a two-state model described in detail previously [264].

Infrared T-jump setup. The time-resolved T-jump infrared apparatus used in the current study has been described in detail elsewhere [211]. The relaxation kinetics of all peptides in response to a T-jump were probed at 1631 cm^{-1}. The observed relaxation rate constants (k_R) were further separated into folding (k_f) and unfolding (k_u) rate constants based on the thermodynamic results obtained from CD studies

using the following equations:

$$k_R = k_f + k_u \tag{5.5}$$

$$K_{eq} = \frac{k_f}{k_u} \tag{5.6}$$

Fluorescence correlation spectroscopy. Details of the FCS setup have been described elsewhere [265]. For each measurement, either 1 nM R6G or 1 nM TMR-labeled pHLIP peptide solution or a solution of 40 nM NR and 20 μM HSA was loaded on PEG (polyethylene glycol)-silane (Gelest Inc., Morrisville, PA)-modified glass slip. Excitation of the dye was accomplished by the 514 nm line of an Ar^+ ion laser ($\sim$ 100 μW before entering the microscope), and the resultant fluorescence was equally split by a nonpolarizing beamsplitter (Newport Corporation, CA) and detected by two avalanche photodiodes (Perkin Elmer, NJ) using an integration time of 0.1 μs. Correlating the fluorescence signals in the cross-correlation mode was accomplished by a Flex 03-LQ-01 correlator card (Correlator.com, NJ) for a duration of 120 s, and the resulting autocorrelation traces were analyzed using the following equation [266]:

$$G(\tau) = \left(\sum_{i=1}^{n} \frac{1}{N} \left(\frac{f_i}{1 + \frac{\tau}{\tau_D^i}} \right) \left(\frac{1}{1 + \frac{\tau}{\omega^2 \tau_D^i}} \right)^{1/2} \right) \left(1 - T + T \exp\left[-\frac{\tau}{\tau_T} \right] \right) \tag{5.7}$$

where τ_D^i represents the characteristic diffusion time constant of species i, ω refers to the axial-to-lateral dimension ratio of the confocal volume element, N represents the number of fluorescent molecules in the confocal volume, f_i represents the frac-

tion of species i, τ_T is the triplet lifetime of the fluorophore, and T represents the corresponding triplet amplitude.

5.6 Acknowledgement

We gratefully acknowledge financial support from the National Institutes of Health (GM-065978 and RR-01348). M.M.W. further thanks Smita Mukherjee, who collected that data on both L9:41-74 and trpzip4-m1, for fruitful collaboration. Lin Guo is credited for collecting the FCS data.

5.7 Original Publication

This Chapter has been published: Smita Mukherjee*, Matthias M. Waegele*, Pramit Chowdhury, Lin Guo, Feng Gai, *Journal of Molecular Biology* (2009), 393, 227-236. DOI:10.1016/j.jmb.2009.08.016 *equal contribution

Table 5.1: Summary of the thermodynamic and kinetic data for Z34C-m1 and trpzip4-m1 obtained under different conditions.

		Z34C-m1	
	D_2O buffer	Dextran 70	Ficoll 70
ΔH_m (kcal mol^{-1})	16.2 ± 0.7	17.1 ± 0.8	16.8 ± 0.7
ΔS_m (cal K^{-1} mol^{-1})	49.4 ± 3.8	52.0 ± 2.6	50.9 ± 2.0
ΔC_p (cal K^{-1} mol^{-1})	118 ± 34	245 ± 105	120 ± 54
T_m (°C)	54.7 ± 1.0	55.3 ± 1.5	57.7 ± 0.5
τ_f (μs)	5.3 ± 0.8	8.6 ± 1.3	7.3 ± 1.1
τ_u (μs)	5.1 ± 0.8	8.8 ± 1.3	9.0 ± 1.4
		trpzip4-m1	
	D_2O buffer	Dextran 70	Ficoll 70
ΔH_m (kcal mol^{-1})	13.8 ± 0.2	12.5 ± 0.3	15.6 ± 0.2
ΔS_m (cal K^{-1} mol^{-1})	45.1 ± 2.5	40.6 ± 0.9	49.1 ± 0.7
ΔC_p (cal K^{-1} mol^{-1})	343 ± 41	238 ± 19	223 ± 15
T_m (°C)	32.1 ± 0.9	34.3 ± 0.5	44.0 ± 0.2
τ_f (μs)	47.5 ± 2.3	96 ± 14	122 ± 18
τ_u (μs)	38.1 ± 2.0	92 ± 14	240 ± 36

Table 5.2: Translational diffusion times determined by FCS at room temperature for R6G, the pHLIP peptide, and the HSA-NR complex.

	D_2O	Dextran 70 (200 g/L)	Ficoll 70 (200 g/L)
R6G (μs)	53 ± 2	216 ± 5	256 ± 5
pHLIP (μs)	75 ± 3	286 ± 5	305 ± 8
HSA-NR (μs)	419 ± 15	2000 ± 30	2460 ± 60

Figure 5.1: Far-UV CD spectra of L9:41-74 at 4.0 °C in D_2O, dextran 70 (200 g/L), and ficoll 70 (200 g/L) solutions, as indicated. All solutions contain 20 mM phosphate buffer ($pH^* = 7$).

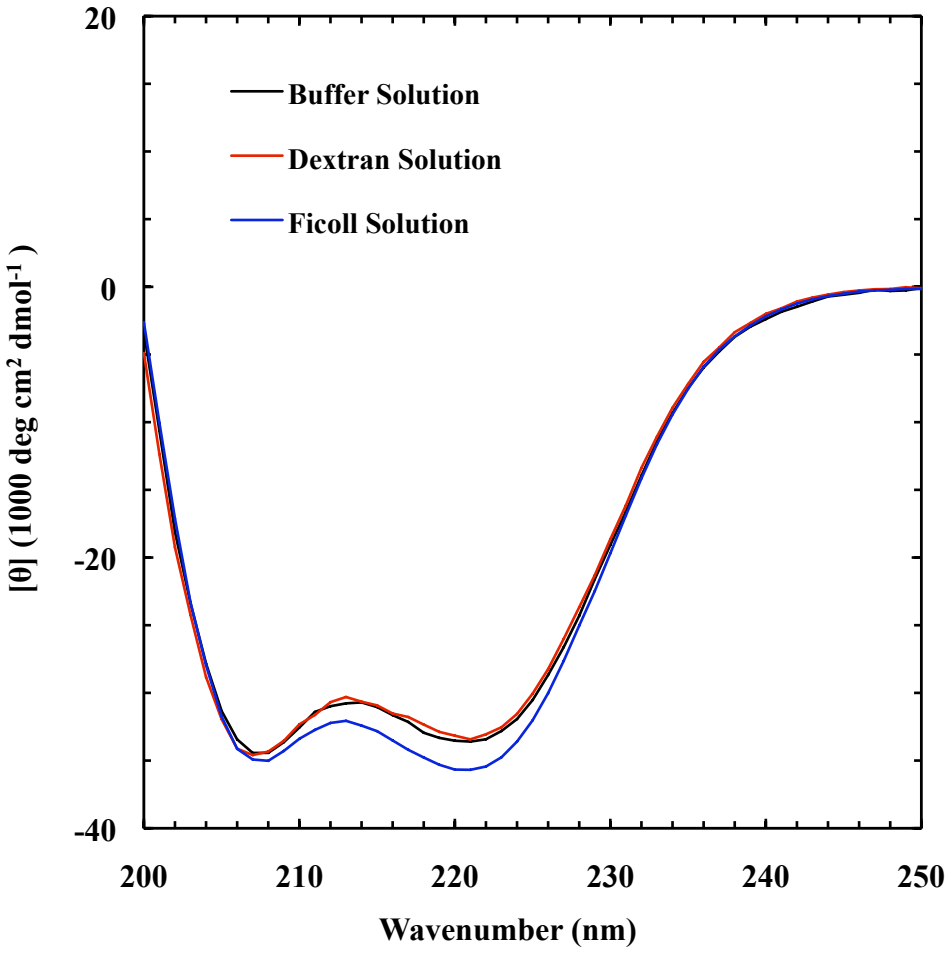

Buffer Solution
Dextran Solution
Ficoll Solution
[θ] (1000 deg cm² dmol⁻¹)
20
0
-20
-40
200
210
220
230
240
250
Wavenumber (nm)

Figure 5.2: CD thermal melting curves of L9:41-74 in 200 g/L of dextran 70 (red) and in 200 g/L of ficoll 70 (blue). Also shown for comparison are the CD thermal melting curve (black crosses) of the same peptide in 20 mM phosphate D_2O buffer (derived from Mukherjee *et al.* [243].)

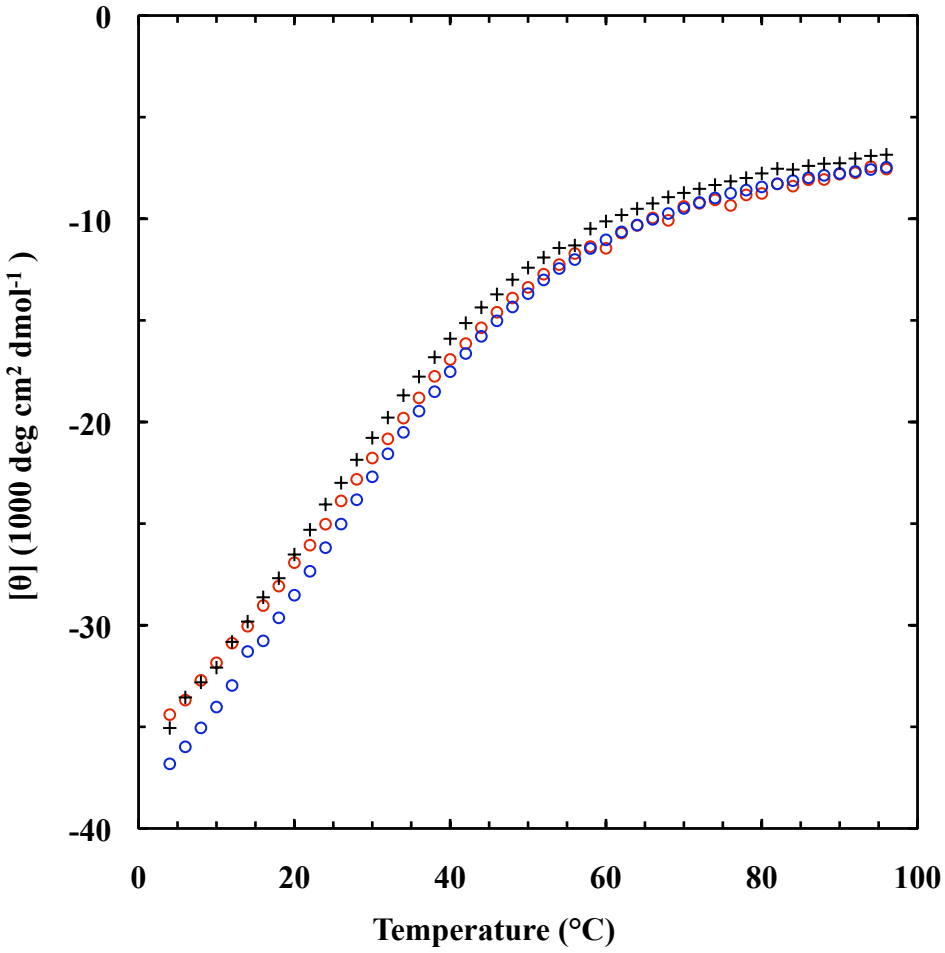

0
-10
-20
-30
-40
0
20
40
60
80
100
[θ] (1000 deg cm^2 dmol^{-1})
Temperature (°C)

Figure 5.3: Representative T-jump relaxation traces (blue and red) of L9:41-74 in 200 g/L dextran 70 (20 mM phosphate buffer, $pH^* = 7$) in response to a T-jump of 5.5 °C (blue), from 22.9 to 28.4 °C, and 5.9 °C (red), from 8.1 to 14.0 °C (scaled 1.5 times). The smooth lines are convolutions of the instrument response function with $\Delta OD(t) = A \cdot [1 - B \cdot \exp(-t/\tau)]$, with $A = -0.0033$, $B = 0.59$, and $\tau = 0.93$ μs for the higher final temperature data and $A = -0.0019$, $B = 0.78$, and $\tau = 1.9$ μs for the lower final temperature data.

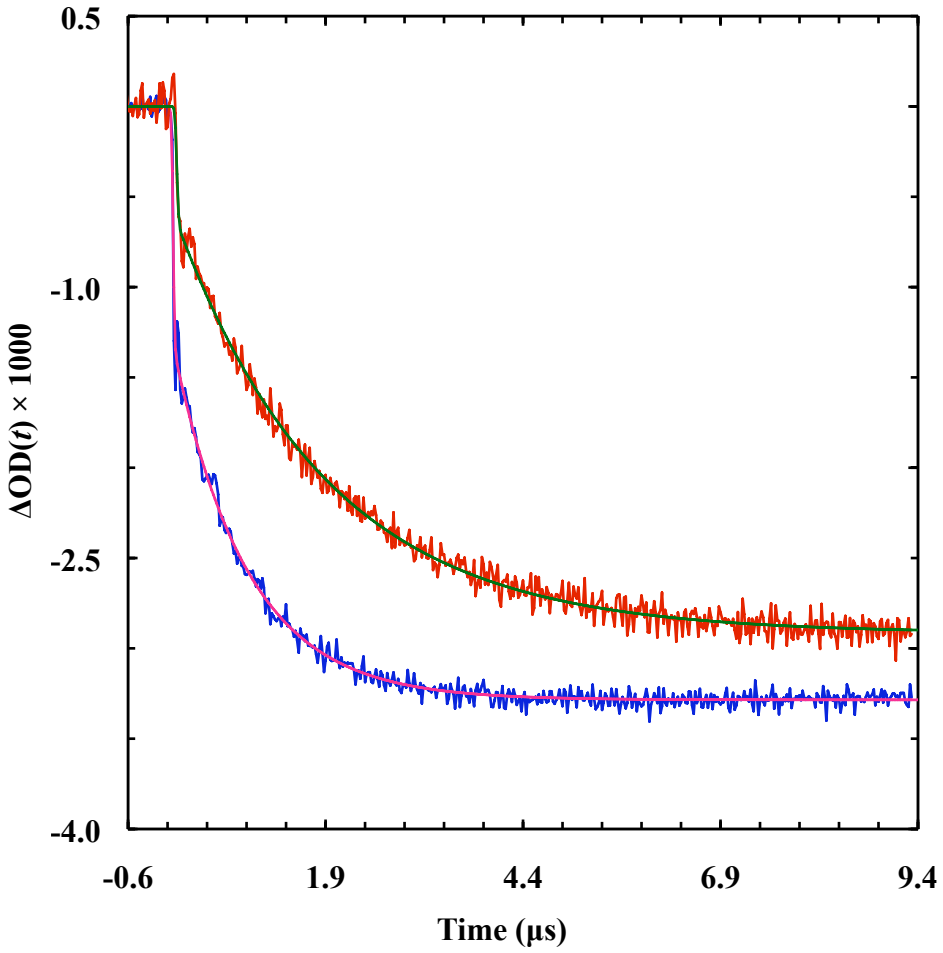

0.5
-1.0
-2.5
-4.0
ΔOD(t) × 1000
-0.6
1.9
4.4
6.9
9.4
Time (μs)

Figure 5.4: Arrhenius plot of the T-jump-induced conformational relaxation rates of L9:41-74 in 200 g/L of dextran 70 (red) and in 200 g/L of ficoll 70 (blue). Also shown for comparison are the relaxation rates (black dashed line) of the same peptide in 20 mM phosphate D_2O buffer (derived from Mukherjee *et al.* [243]).

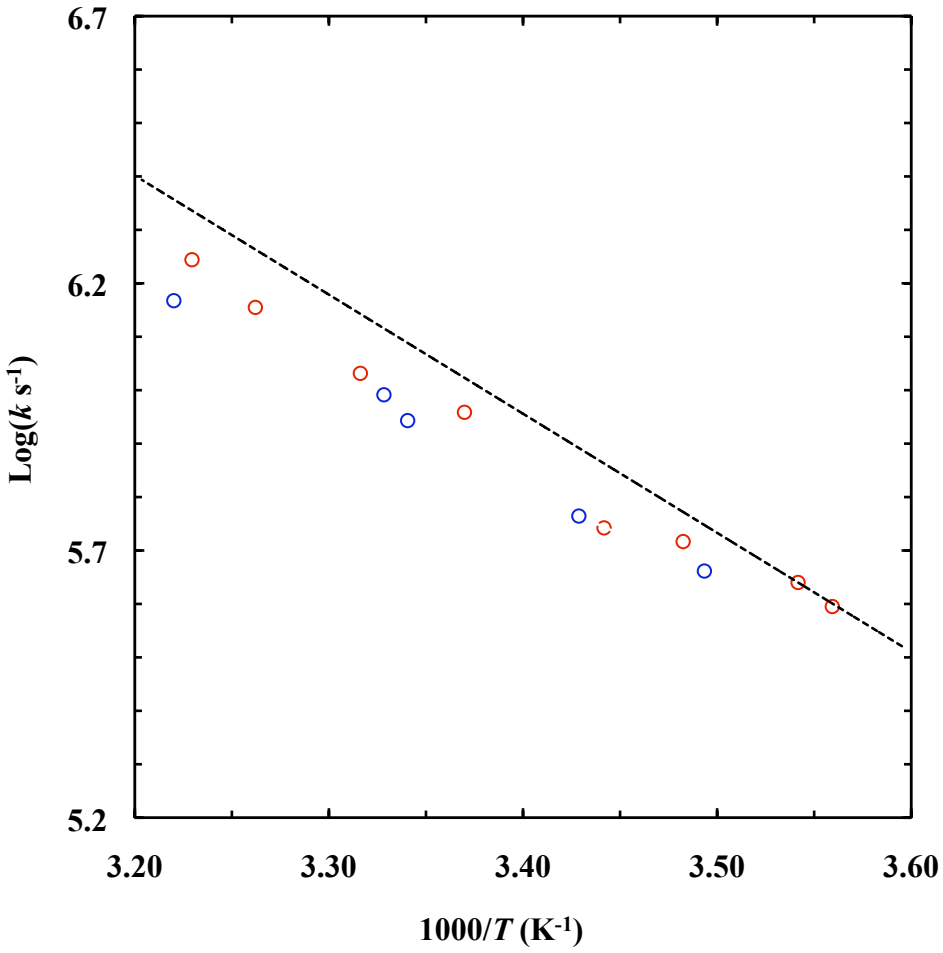

6.7
6.2
5.7
5.2
Log(k s-1)
3.20
3.30
3.40
3.50
3.60
1000/T (K-1)

Figure 5.5: Far-UV CD spectra of Z34C-m1 at 4.0 °C in D_2O, dextran 70 (200 g/L) and ficoll 70 (200 g/L) solutions, as indicated. All solutions contain 20 mM phosphate buffer ($pH^* = 7$).

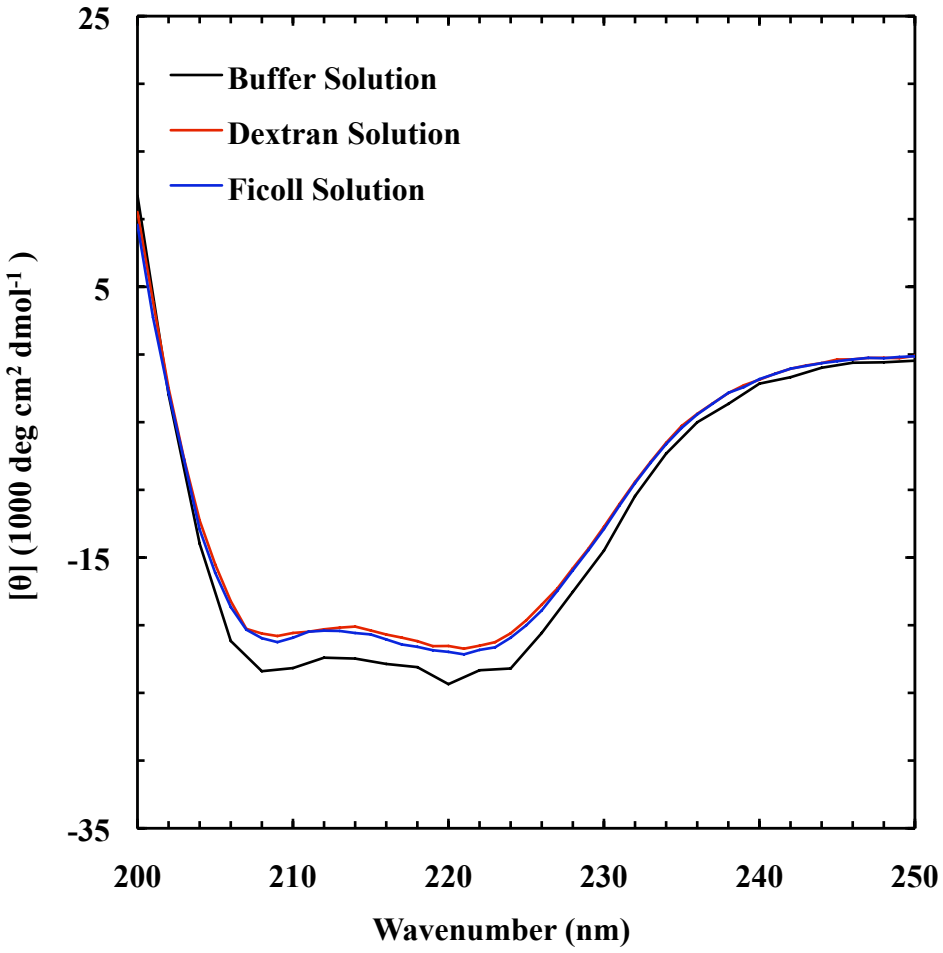

Buffer Solution
Dextran Solution
Ficoll Solution
[θ] (1000 deg cm² dmol⁻¹)
25
5
-15
-35
200
210
220
230
240
250
Wavenumber (nm)

Figure 5.6: CD thermal melting curves of Z34C-m1 in 200 g/L of dextran 70 (red) and in 200 g/L of ficoll 70 (blue). Lines are global fits of these data to the two-state model described in the text. Also shown are the CD thermal melting curve (black crosses) of Z34C-m1 in 20 mM phosphate D_2O buffer (derived from Du and Gai [205]).

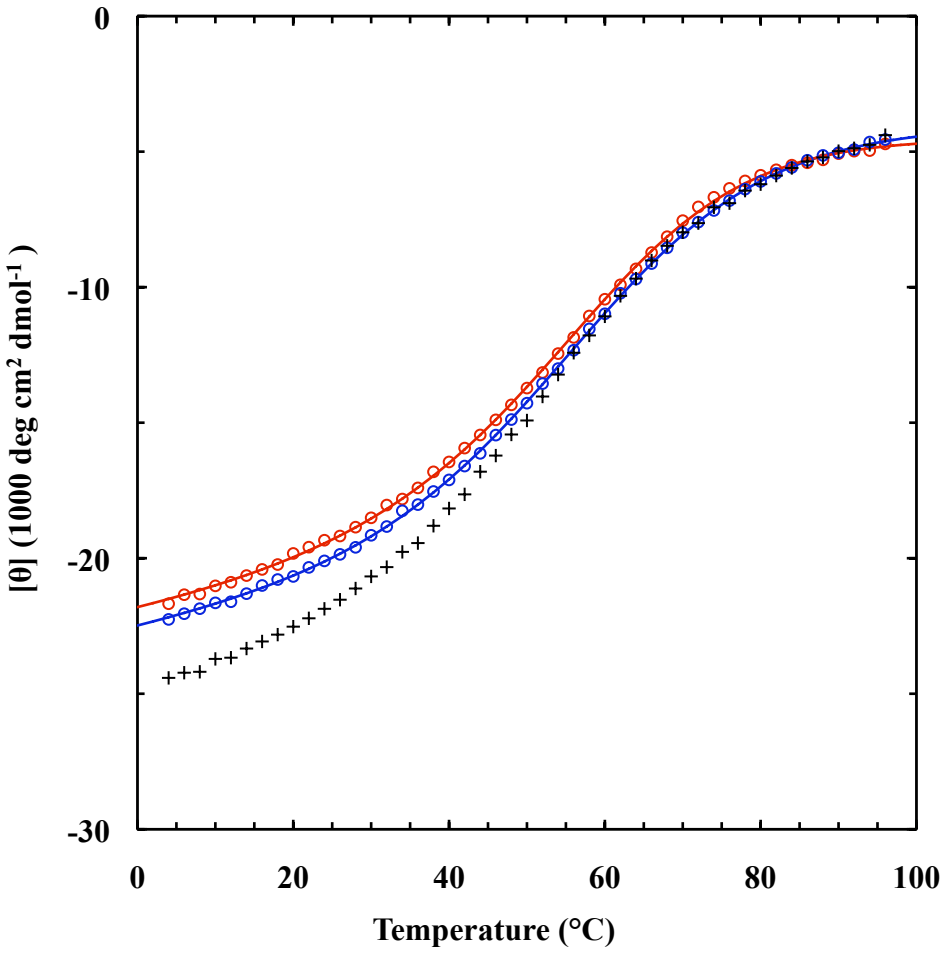

0
-10
-20
-30
[θ] (1000 deg cm^2 dmol^{-1})
0
20
40
60
80
100
Temperature (°C)

Figure 5.7: Arrhenius plot of the T-jump-induced conformational relaxation rates (open circles) of Z34C-m1 in 200 g/L of dextran 70 (red) and in 200 g/L of ficoll 70 (blue). Open triangles and squares correspond to the two-state folding and unfolding rates of Z34C-m1 in 200 g/L of dextran 70, respectively. Also shown are folding (black dashed line) and unfolding (black continuous line) rate constants of Z34C-m1 in 20 mM phosphate D_2O buffer (derived from Du and Gai [205]).

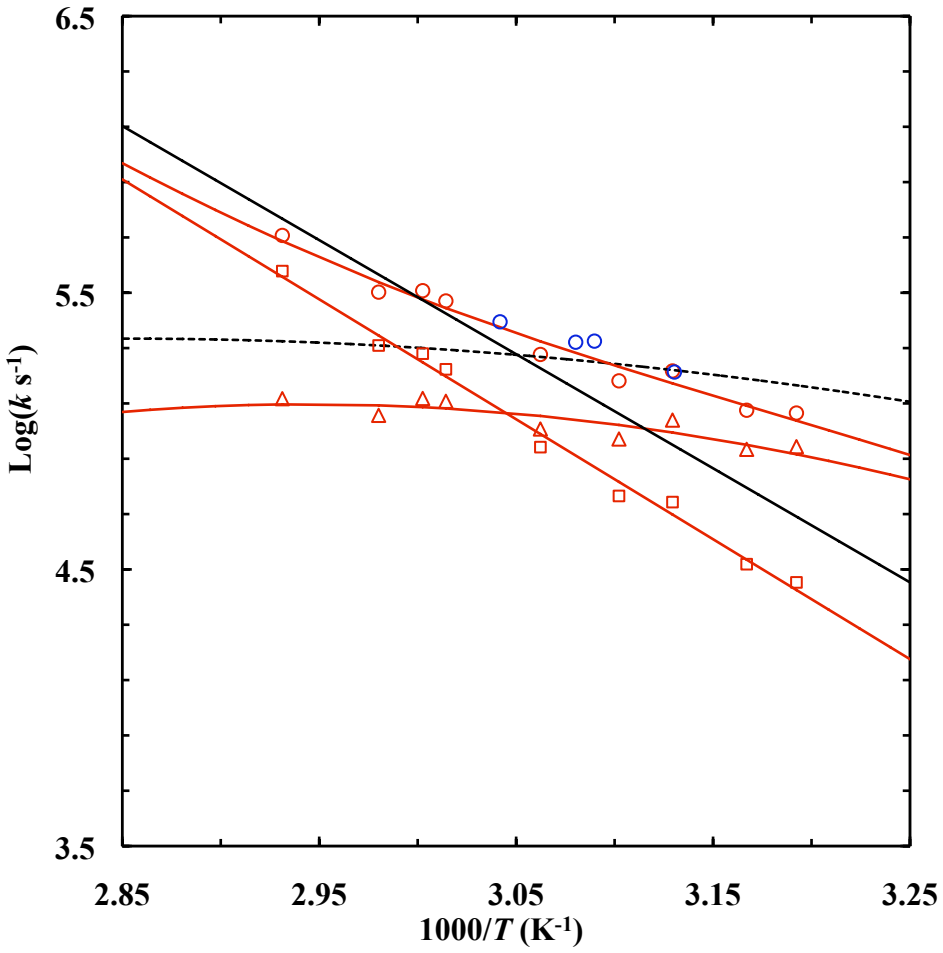

6.5
5.5
4.5
3.5
Log(k s-1)
2.85
2.95
3.05
3.15
3.25
1000/T (K-1)

Figure 5.8: Far-UV CD spectrum of trpzip4-m1 at 4.0 °C in D_2O, dextran 70 (200 g/L), and ficoll 70 (200 g/L) solutions, as indicated. All solutions contain 20 mM phosphate ($pH^* = 7$).

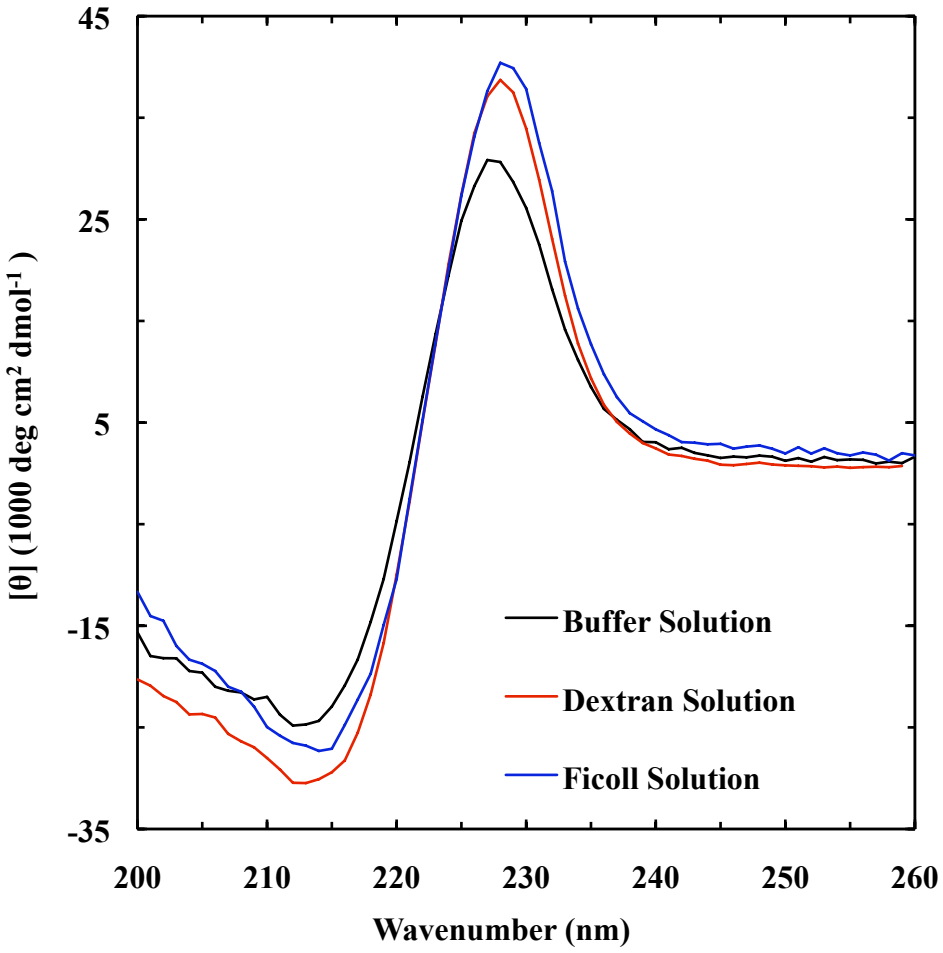

45
25
5
-15
-35
200
210
220
230
240
250
260
[θ] (1000 deg cm² dmol⁻¹)
Wavenumber (nm)
Buffer Solution
Dextran Solution
Ficoll Solution

Figure 5.9: CD thermal melting curves of trpzip4-m1 in 200 g/L of dextran 70 (red) and in 200 g/L of ficoll 70 (blue). Lines are global fits of these data to the two-state model described in the text. Also shown is the CD thermal melting curve (black) of the same peptide in 20 mM phosphate D_2O buffer (derived from Du *et al.* [208]).

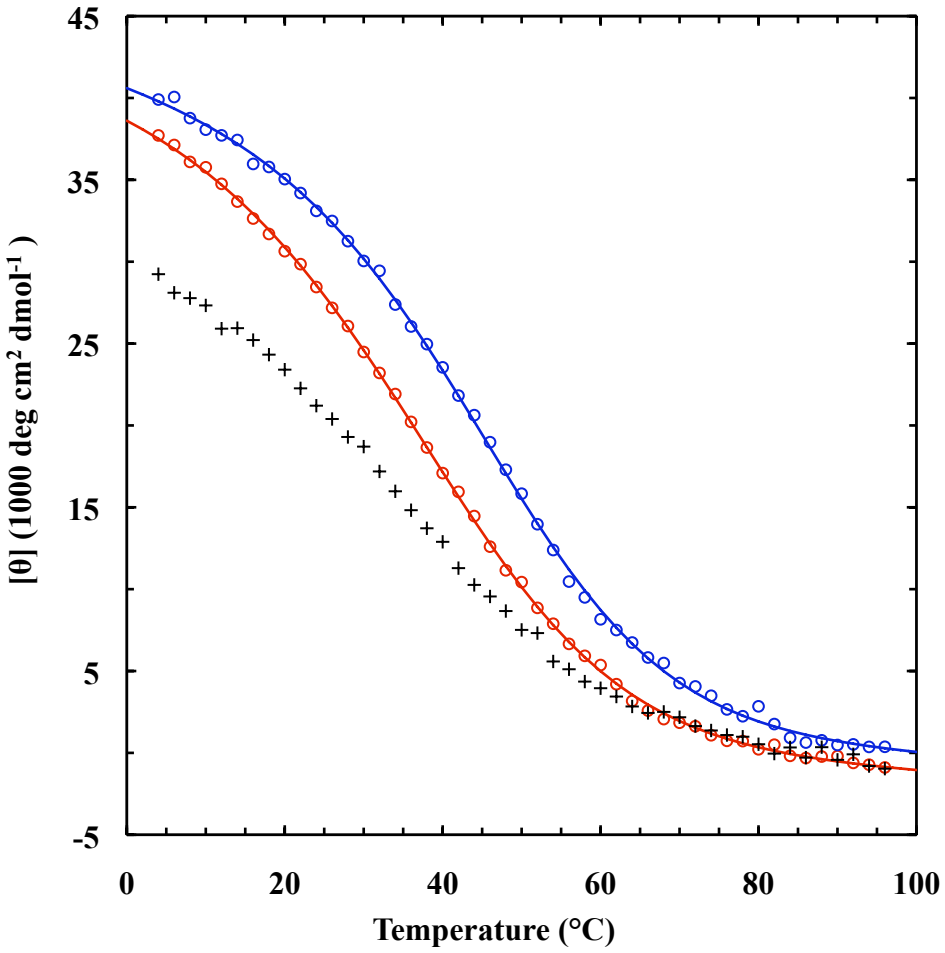

45
35
25
15
5
-5
0
20
40
60
80
100
[θ] (1000 deg cm^2 dmol^{-1})
Temperature (°C)

Figure 5.10: Arrhenius plots of the T-jump-induced conformational relaxation rates (open circles) of trpzip4-m1 in 200 g/L of dextran 70. Open triangles and squares correspond to their respective two-state folding and unfolding rates. Also shown are the folding (blue dashed line) and unfolding (red dashed line) rate constants of trpzip4-m1 in 20 mM phosphate D_2O buffer (derived from Du *et al.* [208]).

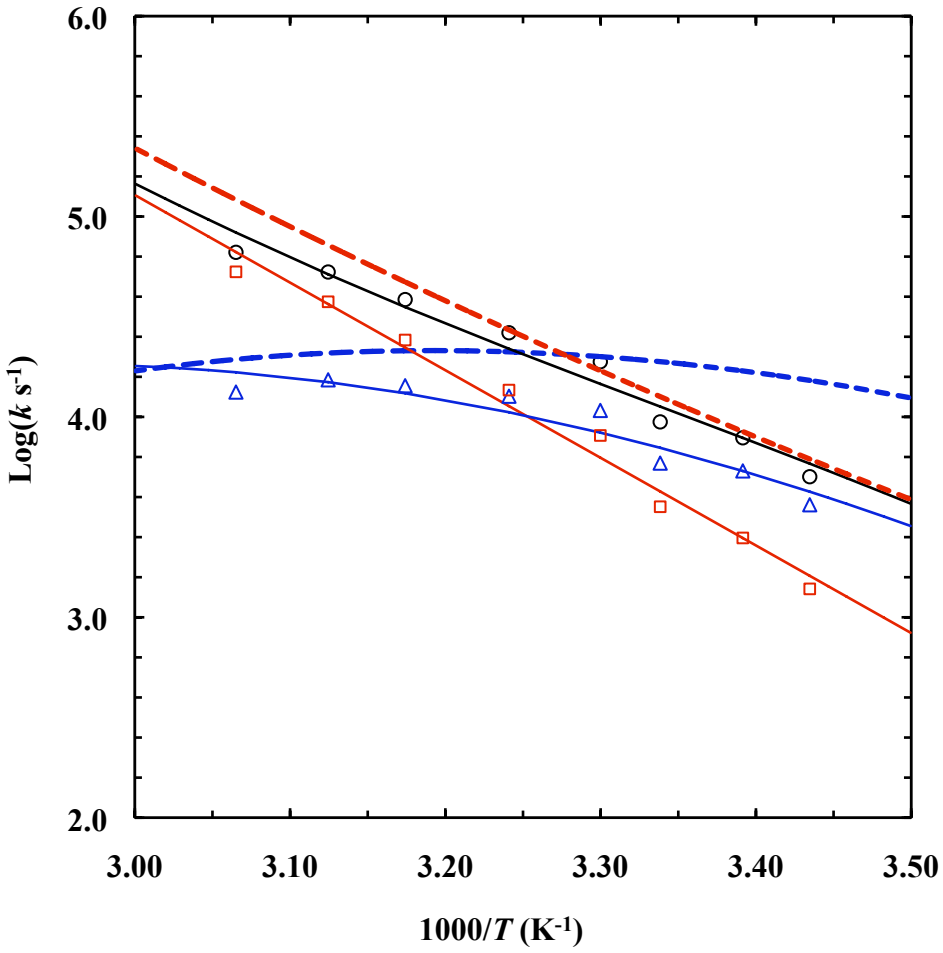

6.0
5.0
4.0
3.0
2.0
Log(k s-1)
3.00
3.10
3.20
3.30
3.40
3.50
1000/T (K-1)

Figure 5.11: Arrhenius plots of the T-jump-induced conformational relaxation rates (open circles) of trpzip4-m1 in 200 g/L of ficoll 70. Open triangles and squares correspond to their respective two-state folding and unfolding rates. Also shown are the folding (blue dashed line) and unfolding (red dashed line) rate constants of trpzip4-m1 in 20 mM phosphate D_2O buffer (derived from Du *et al.* [208]).

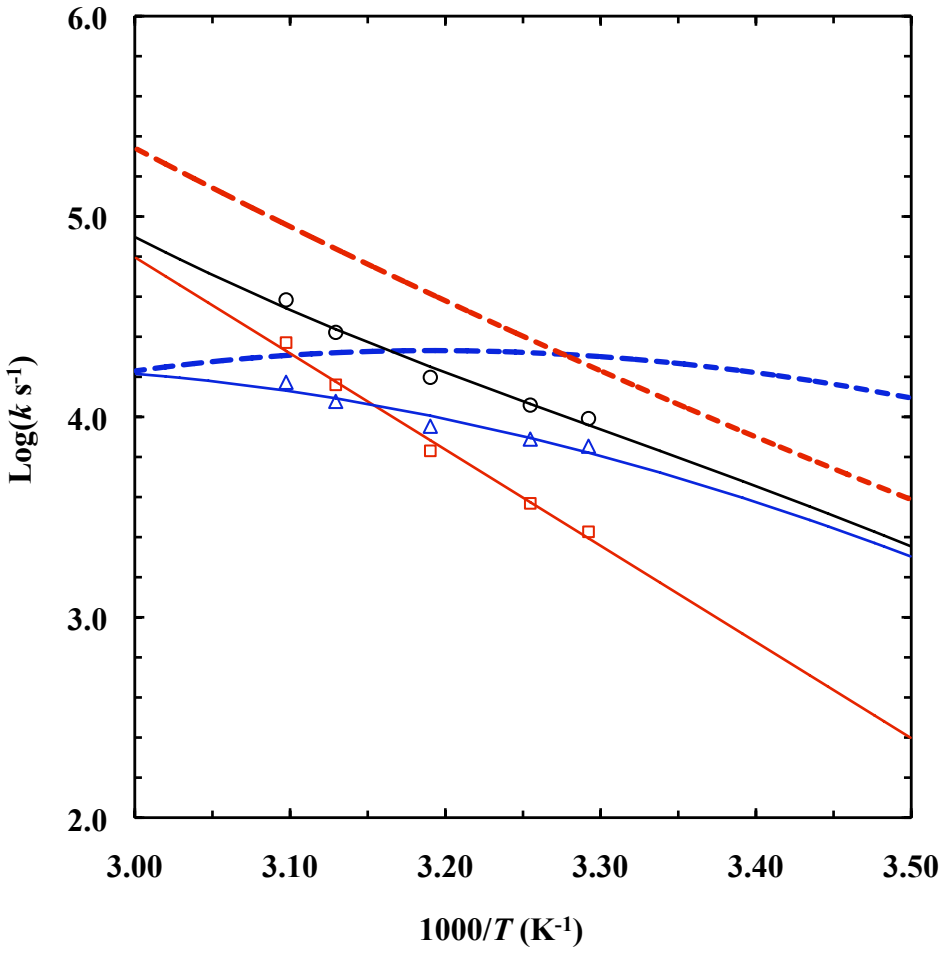
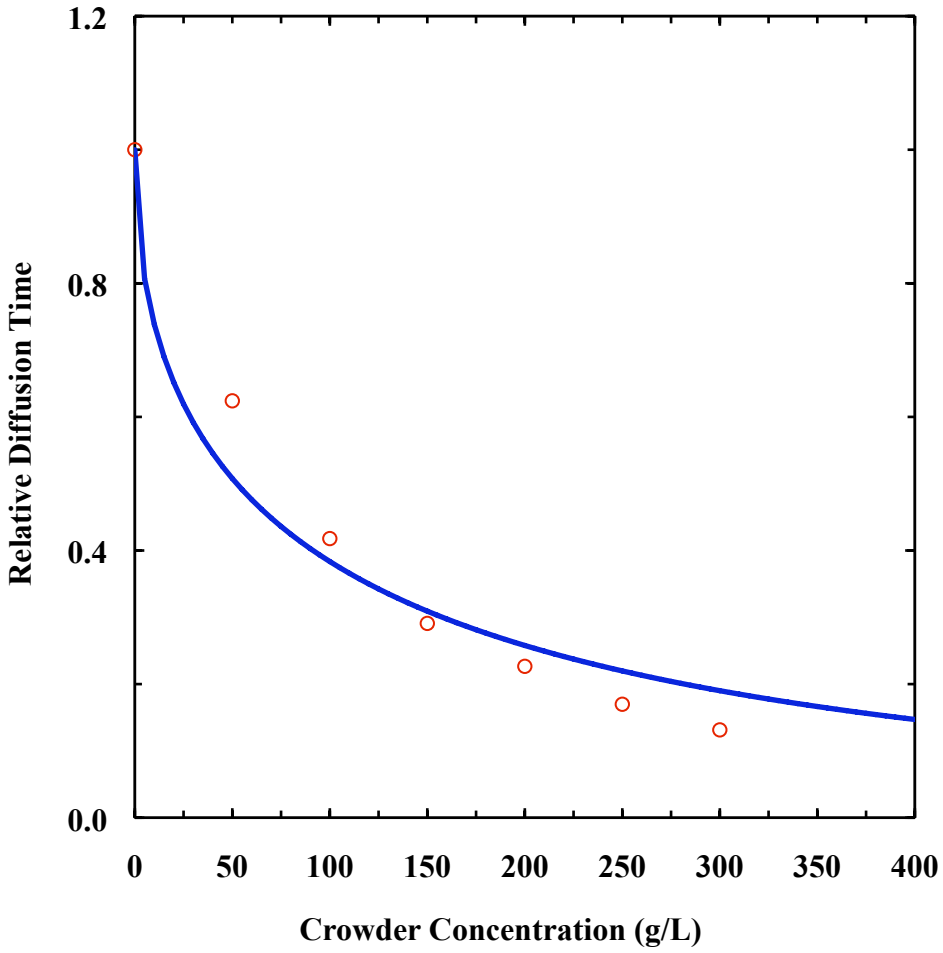

6.0
5.0
4.0
3.0
2.0
Log(k s-1)
3.00
3.10
3.20
3.30
3.40
3.50
1000/T (K-1)

Figure 5.12: Relative diffusion time of R6G versus ficoll 70 concentration, measured via FCS at room temperature. The solid line represents the best fit of these data to Eq. 5.1 with $a = 0.096$. The model given in Ref. [259] suggests that the characteristic length scale (ξ) of the pores of a polymer solution is related to a via the following equation, $\xi = (a/b)c^{-1/2}$, where b is the diameter of the diffusing probe molecules and c is the concentration of the polymer. However, this simple pore model seems to fail in the present case as the length scale of ficoll 70 at 200 g/L is predicted to be smaller than the hydrodynamic diameter of the Rhodamine 6G dye.


1.2
0.8
0.4
0.0
Relative Diffusion Time
0
50
100
150
200
250
300
350
400
Crowder Concentration (g/L)

Chapter 6

Power-law Dependence of the Melting Temperature of Ubiquitin on the Volume Fraction of Macromolecular Crowders

6.1 Introduction

Cellular compartments contain high concentrations of macromolecules (80-400 g/L), which are expected to affect the thermodynamics and kinetics of many biological processes that take place in such crowded environments [26, 28, 27]. Thus, many *in vitro* experiments [267, 229, 268, 232, 233, 269, 270, 271, 234, 235, 272, 273, 251, 274, 236, 275, 276, 277] and computer simulations [270, 272, 273, 277, 278, 29, 279, 280, 281, 282, 283, 284] have been carried out to determine, for example, the influence of macromolecular crowding on the energy landscape of protein folding and protein-protein interactions. Additionally, several models have been developed, aiming to provide a theoretical basis for understanding the effect of macromolecular crowding on various biological events [26, 29, 285, 286, 287, 288, 289, 282]. In the case of protein folding, while the existing experimental results are consistent with the theoretical prediction that an inert macromolecular crowder will destabilize the unfolded state of proteins via a repulsive depletion force arising from the volume occupied by the macromolecular cosolutes [26, 29, 285, 286, 290], quantitative comparisons of the experimentally observed and theoretically predicted crowding effects are hardly made [291]. A rare exception is the recent work of Zhou and coworkers, [272, 273] which showed that in some cases their theoretical model is able to provide a quantitative fit to the experimental results [272]. Thus, there is a need for further experimental validation of the existing crowding models [29, 285, 286, 287, 288, 289, 290]. The scaling laws of

polymer physics [292, 293] predict that the thermal melting temperature (T_m) of a protein that is subject to confinement exhibits a simple power-law dependence on the size of the confining space, i.e., $\Delta T_m \sim (R_g/R)^{\alpha}$; where α is a constant, R_g is the radius of gyration of the unfolded protein in the absence of confinement, and R is a characteristic length of the confining cavity. Indeed, recent computer simulations have shown that such a power-law dependence exists for proteins confined in spherical [294, 295] or cylindrical cavities [295, 296, 297]. Based on this model, Cheung and Thirumalai [29] further proposed that the effect of macromolecular crowding on a protein's T_m could be quantified by evaluating the underlying confinement effect, which was shown to be a function of the radius (R_c) and volume fraction (ϕ) of the crowder. For $0 < \phi < 0.1$, they proposed that the confinement effect is equivalent to that arising from a spherical pore of radius $R = (4\pi/3\phi)^{1/3} R_c$, namely:

$$\Delta T_m \sim \left(\frac{R_g}{R_c}\right)^{\alpha} \phi^{\alpha/3} \tag{6.1}$$

While it has been shown in previous studies that T_m increases nonlinearly with ϕ [268, 273], there is no rigorous experimental test of Eq. 6.1. In addition, there is some ambiguity about how to relate the exponent α in Eq. 6.1 to the Flory coefficient (ν) of a volume-excluded random coil polymer chain ($\nu \approx 0.6$) [292, 293]. For example, in the case of confining a polymer to a spherical cavity, theory predicts $\alpha = 3/(3\nu - 1)$; whereas in the case of confinement to slits or capillaries, it is found $\alpha = 1/\nu$ [293]. Though the latter expression was adopted in a theoretical analysis of simulations

of proteins in crowded environments, [29, 281] it is not clear whether such choice is also applicable to other crowding systems. Furthermore, we note that Eq. 6.1 was tested via simulations only for a special case where $R_c = R_g$; its application to more complex scenarios could be complicated by other factors such as protein-protein association [29, 298] and/or entropically driven phase separation [29, 299]. Herein, we aim to test both the validity of Eq. 6.1 at physiologically relevant volume fractions of various crowding agents and the appropriate choice of the exponent α. This is accomplished by examining how the thermal melting temperature of ubiquitin [300] varies with ϕ using model crowding agents of different chemical structures and molecular weights, i.e., dextran 6 (6 kDa), dextran 40 (40 kDa), dextran 70 (68 kDa), and ficoll 70 (68 kDa). Ubiquitin is a 76-residue, single domain protein consisting of both α and β secondary structures, whose folding/unfolding thermodynamics in dilute solution have been well characterized [301, 302, 303, 304, 60, 63]. We find that the ΔT_m of ubiquitin exhibits a power-law dependence on the volume fraction of these crowders as predicted by Eq. 6.1 and that setting $\alpha = 3/(3\nu - 1)$ results in a reasonable prediction of how ΔT_m changes with ϕ. Interestingly, we find that the increase in the thermal melting temperature of ubiquitin is independent of the molecular weight of the dextrans used in the current study, which is surprising as the hydrodynamic radii of these dextran molecules are significantly different. Thus, taken together these results suggest that the microcompartmentalization of such polymer

solutions is different from that created by hard spheres with a radius equivalent to the hydrodynamic radius of the crowder molecule.

6.2 Results and Discussion

Temperature dependence of the amide I' band of ubiquitin. The thermal unfolding transitions of ubiquitin under dilute and crowded conditions were quantitatively assessed using infrared spectroscopy. In particular, the amide I' band of the protein under each condition, which has been shown to be a sensitive reporter of both α-helical [213, 214] and β-sheet structures, [305] was collected over a wide temperature range. As shown (Fig. 6.1), the amide I' band of ubiquitin in dilute aqueous solution loses/gains intensity in the lower/higher frequency region of the spectrum with increasing temperature, indicative of its usefulness in monitoring protein unfolding. To extract quantitative thermodynamic information from these spectra, we analyzed them via SVD (Ref. [306]) following Tokmakoff and coworkers, who have shown that the second singular value, $C_2(T)$, is a good reporter of the unfolding transition of ubiquitin [60, 63]. As shown (Fig. 6.2), the thermal melting curve thus obtained, indeed, shows a sigmoid shape, consistent with a cooperative unfolding transition [60, 63]. Thus, these data were further fit to the following two-state model,

$$\frac{C_2(T)}{C_2(T_{ref})} = A_2 + \frac{(A_1 - A_2)}{1 + \exp[(T - T_m)/W]} \tag{6.2}$$

where T is temperature, T_m and W represent the midpoint and width of the sigmoidal transition, respectively. In addition, $A_1 = mT + b_1$ and $A_2 = mT + b_2$ represent the folded and unfolded baselines, respectively. As shown (Fig. 6.2), the fit is satisfactory and the resultant $T_m = 66.2 \pm 1.1$ °C, is in good agreement with that obtained in previous studies [304, 63]. We note that fitting Eq. 6.2 to the thermal melting curve determined using the absorbance at 1624 cm^{-1}, which contains predominantly contributions from both α-helical and β-sheet structures, yields essentially identical results, thus further corroborating the notion that the thermal unfolding transition of ubiquitin occurs between two conformational ensembles.

Dependence of T_m on crowder concentration. Following the method discussed above, the thermal melting curves of ubiquitin under different crowding conditions were also determined and fitted to Eq. 6.2. As shown (Fig. 6.2 and Table 6.1), the T_m of ubiquitin increases with increasing crowder concentration, a trend that has been observed in other experimental studies [268, 270, 277] and also computer simulations [277, 29, 283]. While such an increase is consistent with the excluded volume effect of the crowding agents, which preferentially destabilizes the unfolded state of proteins [26, 29, 285, 286, 287, 288], a quantitative assessment of the underlying relation between T_m and the degree of crowding (or volume fraction of the crowder, ϕ) was rarely performed. As shown (Fig. 6.3), the T_m of ubiquitin increases nonlinearly with increasing ϕ (within the experimentally accessible range of crowder concentrations).

In fact, the data can be well described by Eq. 6.1 (Fig. 6.3 and Table 6.2), suggesting that the theoretical model of Cheung and Thirumalai [29], which is based on considerations of the entropy loss associated with confining an unstructured polymer to a spherical pore [292, 293], possibly captures the essential physics underlying the crowding effect. What is more interesting, however, is that the exponents (i.e., α) extracted from the fits of the data obtained with different crowding agents fall within the range of 4.1-6.5 (Table 6.2). As discussed above, polymer theory provides two possible expressions for α, one is based on confinement to a spherical cavity, which provides repulsive boundaries in all three spatial dimensions, and the other is based on confinement to a capillary/slit, which restricts the polymer in two-dimensions/one-dimension [293]. Using the Flory coefficient for an unstructured excluded-volume polymer in a good solvent, which yields $\nu \approx 0.6$, the theoretically predicted values of α are 1.67 and 3.75 for capillary/slit and spherical confining geometries, respectively. Thus, the experimentally determined value of α suggests that the crowding effect of these macromolecular crowders is more similar to the confinement exerted by a spherical cavity than the one of a capillary/slit. This finding is remarkable as intuitively we might think that the confinement effect exerted by a flexible and dynamic crowding agent would bear more similarity to those of a capillary/slit confining geometry rather than the ones of a spherical cavity. The question arises, however, whether the power-law dependence observed herein applies universally to different

proteins, as observed in simulation studies of the confinement effect by Takagi *et al.* [296], or if it is unique to a specific protein as shown in simulations by Rathore *et al.* [294]. While more experiments are needed to answer this question, the few existing experimental data seem to suggest that setting $\alpha = 1/\nu$ provides too small a value for α and ΔT_m. For example, Sasahara and Minton [268] found a quadratic dependence of the thermal melting temperature of hen egg-white lysozyme (129 residues) on crowder concentration, which is in good agreement with our results as their data essentially suggest an α value of 6. Similarly, we find that the experimental data of Stagg *et al.* [270] on flavodoxin (148 residues) can also be fit by Eq. 6.1, yielding a value of about 3.7 for α.

While it is noted that residual structures in the unfolded state could also lead to a larger α due to $\nu < 0.6$, this seems to be unlikely for ubiquitin. Simulation studies have shown that if partially folded structures are populated under a crowded environment the width of the unfolding transition would show a broadening with increasing crowder concentration [29, 283], due to a decrease in the cooperativity of the folding-unfolding transition. As shown (Table 6.1), our results indicate that the width of the thermal unfolding transition of ubiquitin does not show any measurable dependence on the concentration of the crowding agents used, which further corroborates the two-state model used to analyze the temperature melting curves and suggests that the large α value observed is unlikely to be caused by residual structures in the unfolded

state of ubiquitin.

Dependence of T_m on crowder size and structure. According to excluded volume models [26, 29, 285, 286, 287, 288], destabilization of the unfolded state of proteins by crowding is a consequence of reduction of the space available to the polypeptide chain, as confinement effectively eliminates conformations whose radii of gyration are larger than the size of the confining space. The radius of the most probable spherical volume that is void of crowders in a solution at a volume fraction ϕ of crowders is given by $(4\pi/3\phi)^{1/3}R_c$ [29]. Thus, there should be a functional dependence of the destabilization of the unfolded state on the size of the crowder as predicted by Eq. 6.1. For example, using the experimentally determined hydrodynamic radii [307] of dextran 70 ($R_{c,D70} = 6.49$ nm) and dextran 40 ($R_{c,D40} = 4.78$ nm), and an exponent $\alpha = 3.75$, the ratio of $\Delta T_{m,D40}/\Delta T_{m,D70} \sim (R_{c,D70}/R_{c,D40})^{\alpha}$ is estimated to be about 3.1. Similarly, using the hydrodynamic radius of dextran 10 kDa ($R_{c,D10} =$ 1.86 nm) [307], one predicts that $\Delta T_{m,D10}/\Delta T_{m,D70}$ is about 108 for $\alpha = 3.75$ and about 8 for $\alpha = 1.67$. Though the constant of proportionality in Eq. 6.1, which is equal to the inverse of the entropy of unfolding will lower the calculated ratios somewhat, the data clearly do not support such a strong dependence for ubiquitin. As a matter of fact, the ratio is found to be close to unity (Table 6.1). Such a stark difference between theoretical prediction and experiment cannot be attributed to experimental uncertainty, but must arise from other reasons. Similarly, other

experimental studies have also found either no [271] or a weak [272, 277] dependence of the thermal stability of proteins on the crowder size. The origin of the observed discrepancy may be because when mapping a given crowded solution onto a spherical cavity via $(4\pi/3\phi)^{1/3}R_c$, void formation due to density fluctuations in the crowded solution is not taken into account. In other words, protein conformations with $R_g > R$ are strictly not allowed according to such mapping relationship, but they could exist in a real crowded solution which is subject to density fluctuations. Interestingly, the probability of void formation has been shown to be a function of the number density (ρ_c) of the crowder, i.e., $(1-\phi)/\rho_c$ [236]. Alternatively, these findings may not indicate a failure of the theory, but rather suggest that picturing the crowder molecule as a hard sphere of a given radius and the polymer solution as a suspension of hard spheres is an oversimplification. In particular, it seems that the effective "crowding radii" of dextrans of different molecular weights are not as different as one might expect based on their hydrodynamic radii, at least for proteins that have a similar size to ubiquitin. For example, it is well known that dextran and ficoll polymers undergo compression as concentration is increased [308, 245, 309]. At high concentrations, the degree of compression was found to be dependent on molecular weight [308], providing a possible explanation for the observed insensitivity of ΔT_m on dextran molecular weight. Taken together, our results highlight the need for experimental and theoretical efforts to gain a more detailed picture of the microscopic

structure and dynamics of these polymer solutions. Moreover, numerous experimental studies have compared the efficacy of various polymers as a crowding agent [267, 272, 273, 274, 277]. For example, previous studies have found that dextran 70 is more effective than ficoll 70 in stabilizing the folded state of proteins [272, 273, 277]. As shown (Table 6.1 and Fig. 6.3), our results are in agreement with this finding. Though the most probably geometric shapes of either dextran 70 or ficoll 70 are not definitely determined experimentally, it is typically postulated that dextran 70 assumes a rodlike shape whereas ficoll 70 has a spherelike geometry [277]. Thus, under identical solution conditions dextran 70 should exert a larger confinement effect to the unfolded polypeptide chain than ficoll 70, because according to the excluded-volume theory the size of the most probable interstitial space found between rod-shaped objects is smaller than that created by spherical particles of similar molecular weight [285].

6.3 Conclusions

In summary, we demonstrated the power-law dependence of the thermal melting temperature of ubiquitin on the crowder volume fraction, a theoretical prediction that has not been experimentally verified. The exponent α of the power-law was found to be in the range of 4.1-6.5 for both dextran and ficoll polymers. This result suggests, as previously proposed, that the effect of macromolecular crowding on protein folding

stability can be modeled by confinement of the protein of interest to a spherical cavity [29]. However, in contrast to theoretical prediction, but in agreement with other experimental studies [271, 272, 277], we found that the stabilization arising from crowding shows a weak, if any, dependence on the size of the crowding agents used in the current study. Taken together, these findings suggest that the microscopic compartmentalization of these polymer solutions is different from what is expected based on the hydrodynamic radii of the crowders and underscore the need for further improvement of existing crowding theories.

6.4 Materials and Methods

Materials. All materials were used as received. D_2O (D, 99.96%) and deuterium chloride (D, 99.5%) were purchased from Cambridge Isotope Laboratories (Andover, MA), 6-kDa and 40-kDa dextrans and bovine ubiquitin were obtained from Sigma-Aldrich (St. Louis, MO), 68-kDa dextran was obtained from Fisher Scientific (Pittsburg, PA). Ficoll was purchased from GE Healthcare Biosciences (Piscataway, NJ).

Sample preparation. The T_m of ubiquitin at pH $>$ 4 is higher than 90 °C [302]. Therefore, all experiments were performed at pH* 1. This pH condition was chosen because it is known to effectively decrease the thermal stability of ubiquitin without inducing any structural changes to the folded state [304]. An appropriate amount of crowding agent was weighed out on an analytical balance (Denver Instrument, M-220,

Bohemia, NY) and dissolved in an appropriate amount of DCl solution (∼125 mM in D_2O), dispensed via a pipette (EP2100-100R, Eppendorf, Hauppauge, NY). The amount of DCl solution to be added was determined as follows: In separate experiments, 1.0, 1.5, and 2.0 g of crowding agent was added to a 5 mL volumetric flask (precision ±0.02 mL) and the volume of Millipore water needed to get a total volume of 5 mL was determined via an Eppendorf pipette. The volume of added water was then subtracted from the total volume of 5 mL to calculate the volume occupancy of the polymer. The volume fraction of polymer ($V_{polymer}$ in %) was then plotted against the concentration of the polymer ($conc_{polymer}$ in g/L), and the data was fit by a linear equation, $V_{polymer} = conc_{polymer} \cdot m$, which was used to calculate the amounts of DCl solution needed to prepare the protein samples. The slopes m were found to be 0.0651, 0.0633, 0.0632, and 0.0659 for dextran 6, dextran 40, dextran 70, and ficoll 70, respectively. The polymer solutions were vortexed and sonicated for about 10 min, then added to dry, deuterium exchanged ubiquitin powder. The resultant protein solution was then vortexed and sonicated for another 5 min before use.

Fourier transform infrared (FTIR) measurements. FTIR spectra were collected on a Nicolet Magna-IR 860 spectrometer with 2 cm^{-1} resolution. The details of the sample cell (56 μm optical pathlength), temperature control, and data acquisition protocol have been described elsewhere [305]. The corresponding temperature-dependent FTIR spectra (in the range of 1550-1800 cm^{-1}) obtained under a specific

solution condition were analyzed using the singular value decomposition (SVD) [306] function within MATLAB. Based on the absorbance at 1624 cm^{-1}, the reversibility of the thermally induced unfolding transitions of ubiquitin investigated herein was estimated to be $> 90\%$.

6.5 Acknowledgement

We gratefully acknowledge the financial support from the National Institutes of Health (GM-065978 and RR-01348).

6.6 Original Publication

This Chapter has been published: Matthias M. Waegele and Feng Gai, *Journal of Chemical Physics* (2010) 134, 095104-6. DOI: 10.1063/1.3556671

Table 6.1: Midpoints (T_m) and widths (W) of the thermal denaturation curves of ubiquitin obtained under different crowding conditions by fitting the experimental IR data to Eq. 6.2. The standard deviations for all reported melting temperatures and widths are estimated to be ± 1.1 and ± 0.3 °C, respectively. These estimates are based on five independent measurements of the denaturation of ubiquitin in dilute aqueous solution.

Crowder	Concentration (g/L)	T_m (°C)	W (°C)
None	0.0	66.2	4.0
Dextran 6	200	67.9	3.5
Dextran 6	250	71.5	3.8
Dextran 6	300	70.7	3.6
Dextran 6	350	73.3	3.7
Dextran 6	400	73.2	3.9
Dextran 40	200	68.0	3.8
Dextran 40	300	70.9	4.4
Dextran 40	400	74.7	4.3
Dextran 70	200	67.8	3.8
Dextran 70	300	70.3	3.6
Dextran 70	400	73.5	3.5
Ficoll 70	200	67.2	4.0
Ficoll 70	300	67.7	3.3
Ficoll 70	400	69.3	3.6

Table 6.2: Parameters obtained from fitting the data in Fig. 6.3 to Eq. 6.1, i.e., $T_m(\phi) = T_m^d + C\phi^{\alpha/3}$, where T_m^d is the thermal melting temperature of ubiquitin in D_2O. The standard deviations of the fits were estimated using the jackknife method (Ref. [310]) and R is the Pearson product moment correlation coefficient.

Crowder	C	α	R^2
Dextran 6	48.5 ± 31.5	4.1 ± 1.2	0.89
Dextran 40	164.0 ± 43.5	6.5 ± 0.4	1.00
Dextran 70	133.7 ± 48.2	6.3 ± 0.6	1.00
Ficoll 70	41.6 ± 36.7	5.9 ± 2.3	0.98

Figure 6.1: Temperature dependent FTIR spectra (a) and difference FTIR spectra (b) of ubiquitin at pH* 1 in D_2O. The lowest and highest temperatures in this case were 1.6 and 82.6 °C, respectively.

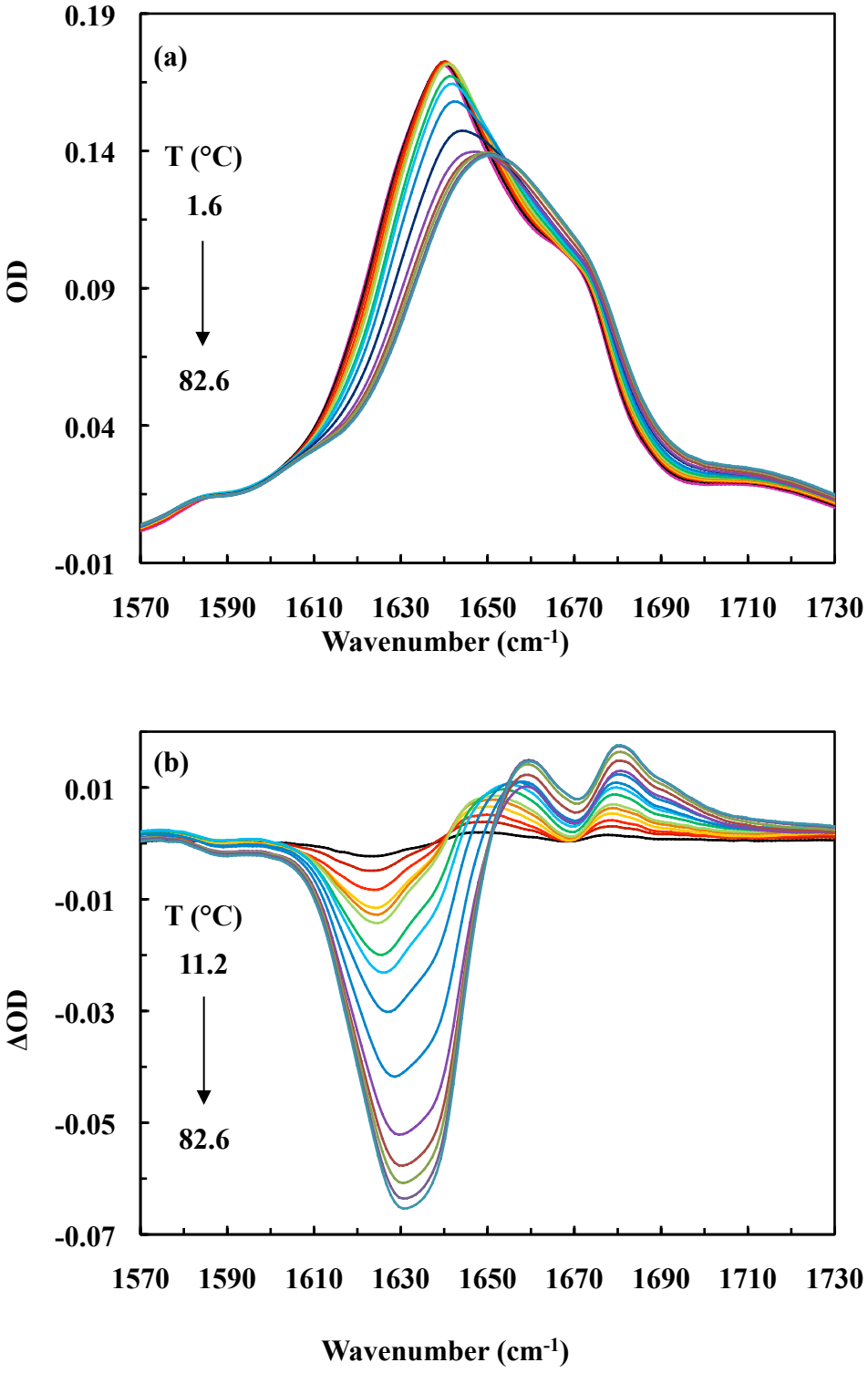

(a)
T (°C)
1.6
82.6
OD
0.19
0.14
0.09
0.04
-0.01
1570 1590 1610 1630 1650 1670 1690 1710 1730
Wavenumber (cm-1)
(b)
T (°C)
11.2
82.6
ΔOD
0.01
-0.01
-0.03
-0.05
-0.07
1570 1590 1610 1630 1650 1670 1690 1710 1730
Wavenumber (cm-1)

Figure 6.2: Thermal melting curves of ubiquitin constructed from the normalized second-component coefficients obtained by singular value decomposition analysis of the temperature dependent FTIR spectra of ubiquitin under different solution conditions, as indicated. The data points below 30 °C are not shown for clarity. Lines are fits of these data to Eq. 6.2 and the resultant fitting parameters are listed in Table 6.1.

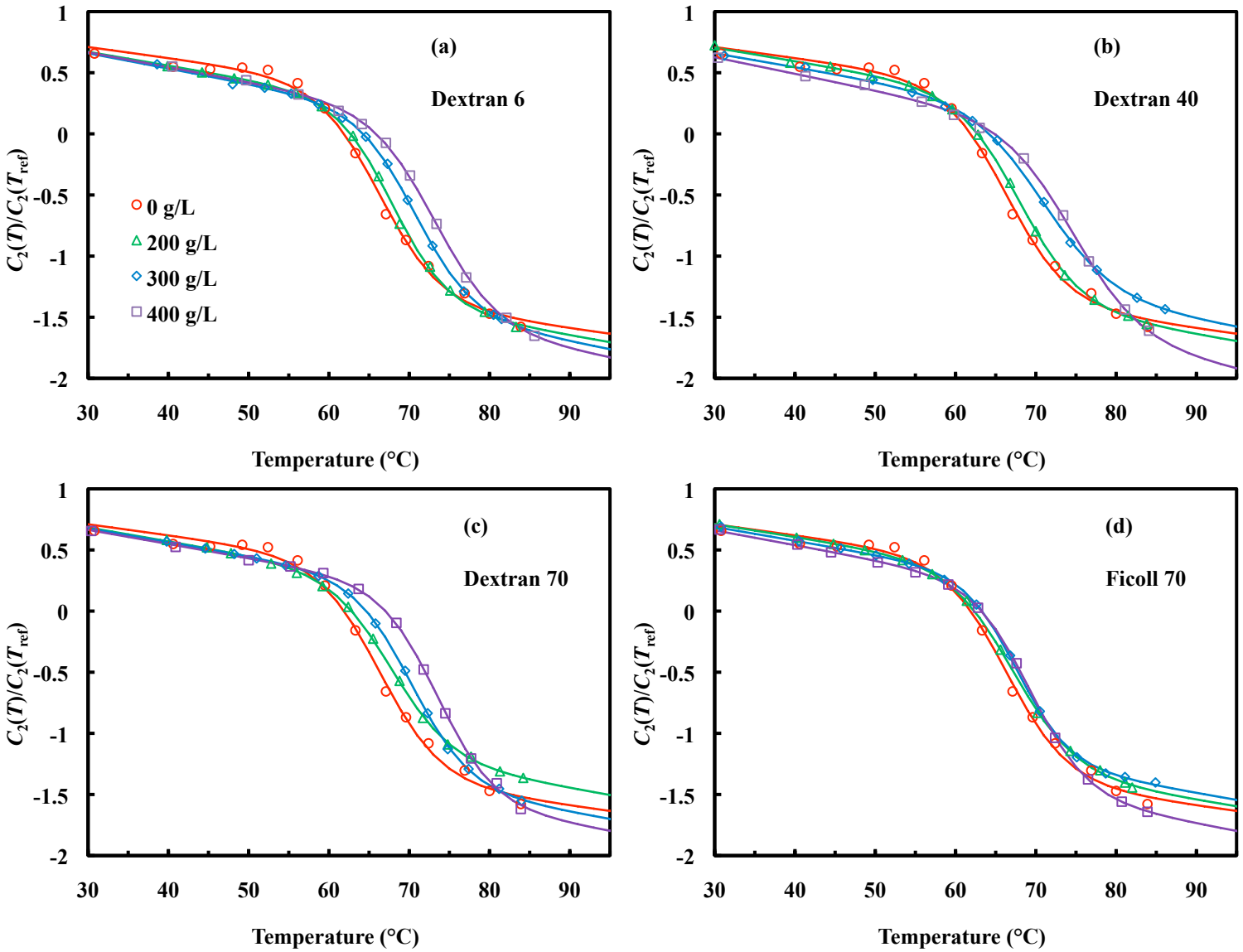
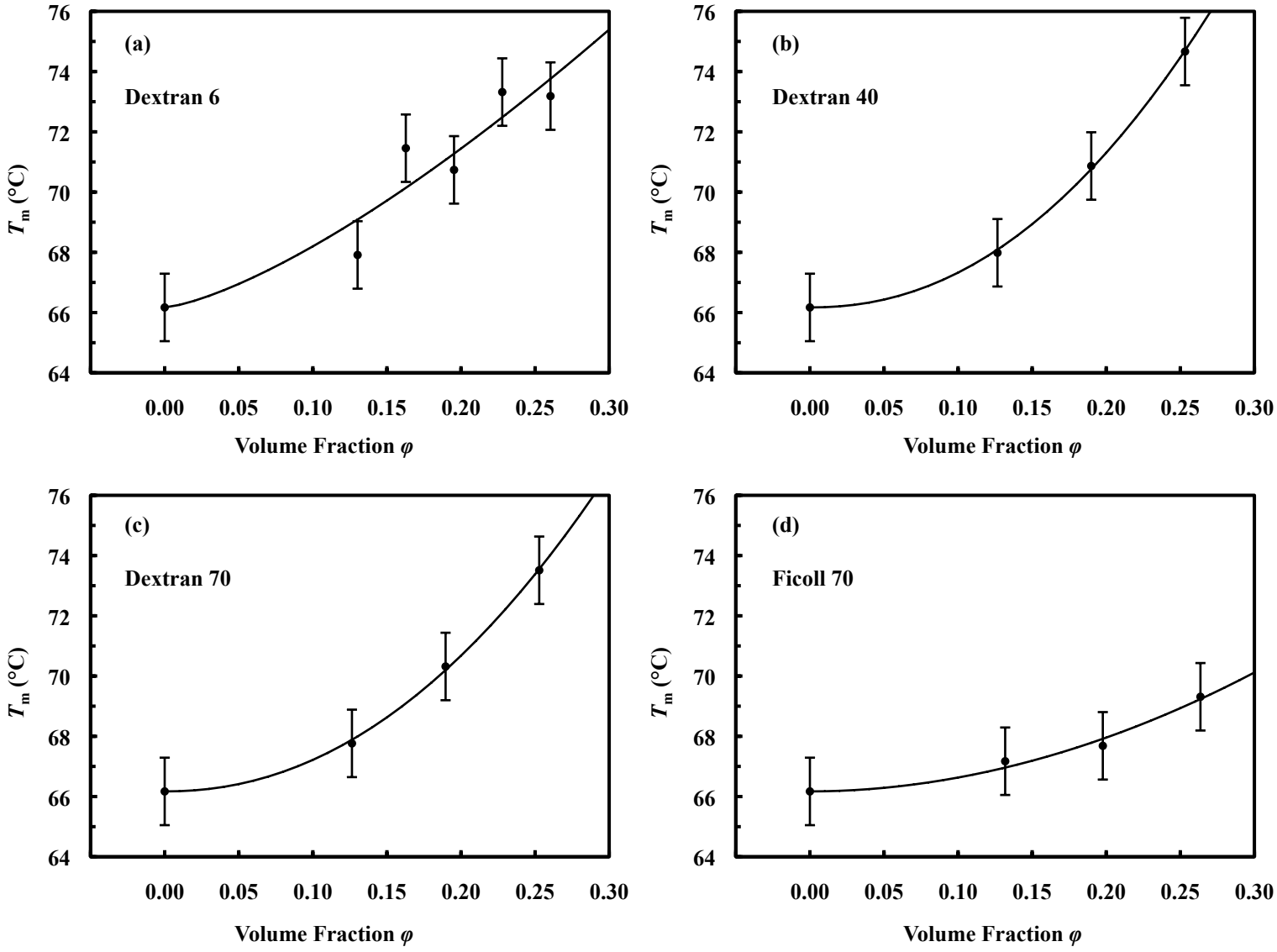

(a)
Dextran 6
0 g/L
200 g/L
300 g/L
400 g/L
(b)
Dextran 40
(c)
Dextran 70
(d)
Ficoll 70
$C_2(T)/C_2(T_{ref})$
Temperature (°C)
1
0.5
0
-0.5
-1
-1.5
-2
30
40
50
60
70
80
90

Figure 6.3: Thermal melting temperature (T_m) of ubiquitin as a function of the volume fraction (ϕ) of different crowding agents, as indicted. Lines are fits of these data to Eq. 6.1, i.e., $T_m(\phi) = T_m^d + C\phi^{\alpha/3}$, and the resultant fitting parameters are listed in Table 6.2. The standard deviations of the data points were assumed to be the same as those for the melting temperature in dilute solution (see Table 6.1.).


(a)
Dextran 6
(b)
Dextran 40
(c)
Dextran 70
(d)
Ficoll 70
T_m (°C)
Volume Fraction φ
76
74
72
70
68
66
64
0.00
0.05
0.10
0.15
0.20
0.25
0.30

Chapter 7

Conformational Distribution of an Unstructured HP-Model Based Polypeptide

7.1 Introduction

Forming the starting point of any protein folding reaction, the denatured state is an integral part of the protein folding problem. The detailed characterization of its physical properties, however, remains a formidable experimental challenge, mainly due to the small population of protein in the denatured state under native conditions and its high degree of structural heterogeneity. In addition to its relevance in the folding problem, knowledge of the physical properties of the unstructured polypeptide is vitally important in elucidating the structure-function relationship of intrinsically disordered proteins. In recent years, single-molecule fluorescence energy transfer spectroscopy has emerged as a popular technique for the study of denatured states [311, 312, 313, 314, 315, 316] as the spectroscopic observables of the folded and unfolded state populations are frequently separated in these experiments, even under native or near-native conditions. However, providing limited structural information, fluorescence techniques largely rely on polymer models for the extraction of structural information of the denatured state. To this end, the freely-jointed chain model [33] has found widespread application to obtain physical properties, such as radius of gyration and chain stiffness from single-molecule measurements [313, 314, 315, 316]. However, doubt exists about the validity of this analysis as the freely-jointed chain model neglects self-exclusion of the polypeptide. For example, a large body of work on peptides with low-complexity sequences,

where intrachain energy transfer efficiencies and/or intrachain diffusion rates were measured [317, 318, 319, 320, 321, 322, 323, 324, 325, 326], shows that unstructured polypeptide chains frequently deviate from freely-jointed chain behavior even for long chains with $N > 20$ peptide bonds, indicating that excluded volume effects are typically non-negligible. This view is supported by extensive small-angle x-ray scattering studies [327], which demonstrated that chemically unfolded proteins follow apparent random coil behavior with $R_g \sim N^{0.6}$, where R_g is the radius of gyration and N is the number of monomers in the chain. Moreover, for short polypeptide chains with a number of $N = 3 - 20$ peptide bonds, studies showed that correlations in the propagation direction of bond vectors are significant [317, 318, 319, 320, 321, 194, 328, 322, 323, 324, 325, 326]. In many of these studies, the Kratky-Porod chain [34], which takes into account this correlation, was therefore found to provide an adequate description of the data. However, with a few exceptions [194, 328] all these studies employed peptides with sequences of low complexity, e.g poly(GlySer), poly(Gln), poly(AlaGlyGln) etc. Some of these sequences were found to be stiffer than what would be expected on theoretical grounds or compared to other experimental findings on peptides with simple sequences. These observations were attributed to specific interactions between side chains of neighboring amino acids [320, 325]. These findings, however, raise the question, if the results for peptides with low-complexity sequences are biased and thus not necessarily transferable to natural

peptides or proteins [329]. Here, using a 29-residue peptide, which is unstructured by design, we characterized its conformational distribution by measuring energy transfer efficiencies between 7 different monomer pairs under mild denaturing conditions. The peptide has a complex sequence and is thus a good model for natural unstructured proteins and peptides. In agreement with previous studies, we find that due to correlations between bond vectors, the freely-jointed chain model does not provide an adequate description of the polypeptide on this lengthscale. In agreement with previous studies [319, 194, 328, 322, 324, 325, 326], we found that the Kratky-Porod model gives a physically reasonable description of the polypeptide with a persistence length of ~ 4.2 Å. Comparison of the measured transfer efficiencies with a random coil model [330] showed that the polymer conformation is somewhat more compact compared to the model's prediction. This observation indicates that long-range attactive interactions might be dominant, giving rise to a relatively short apparent measured persistence length.

7.2 Results

Peptide design and properties. A short, water-soluble peptide with low secondary structure propensity consisting of randomly chosen polar (P) and nonpolar (H) amino acids was designed as follows: A pattern of 27 residues, which was taken from simulations by Yue and Dill [331] formed the starting point of our design strategy. We then

strategically position tryptophan and phenylalanine residues within this HP pattern so that the conformational distribution of the polypeptide chain can later be determined via energy transfer between *p*-cyanophenylalanine and tryptophan [194]. We randomly filled in the remaining sites in the pattern with polar and nonpolar residues, excluding amino acids that are known to be quenchers of *p*-cyanophenylalanine fluorescence [193, 332]. Amino acids were categorized as polar and nonpolar according to the scale of Kyte and Doolittle [333]. Two lysine residues were added at the C-terminal end to ensure solubility. The resulting 29-resiude peptide has the sequence: **F**DTG**F**-LA**F**TI-KP**W**TS-DAT**F**V-E**F**RRG-**F**GKK-NH_2. Six different versions of the peptide were synthesized, in each version a single phenylalanine residue (bold letters) was replaced by Phe_{CN}. In the following, we will refer to the peptides as PepCN1, PepCN5, PepCN8, PepCN19, PepCN22 and PepCN26, where the number indicates to the position of the Phe_{CN} label in the sequence. In addition, we synthesized a peptide of the sequence: $\mathbf{F}_{CN}$DTGF-LAFTI-KPFTS-DATFV-EFRRG-FGKK**W**-NH_2, where the tryptophan at position 13 was mutated to phenylalanine and an additional tryptophan residue was placed at the C-terminus, allowing the determination of the end-to-end distance of the peptide. This version will be referred to as PepE. As shown in Fig. 7.1, hydrophobic and hydrophilic residues as well as the charged residues are fairly evenly distributed along the generated sequence. The mean hydrophobicity of the sequence is -0.3, indicating that the peptide more hy-

drophilic than hydrophobic according to this scale. As shown in Fig. 7.2, the CD spectra of the peptide in aqueous phosphate buffer at pH 3 confirm the absence of any appreciable secondary structure.

Förster energy transfer measurements. As shown in Fig. 7.3, the fluorescence of the $\mathrm{Phe_{CN}}$ residues in the six peptides is quenched to different extent, depending on the proximity of a residue to tryptophan. In order to analyze the observed quenching quantitatively, we deconvoluted the spectra by fitting the sum of the fluorescence spectra of a reference peptide, $\mathrm{GKF_{CN}TV}$, and of free tryptophan to the fluorescence spectra of the peptides, varying only the magnitudes of the $\mathrm{GKF_{CN}TV}$ and tryptophan spectra, as well as a factor that shifts the tryptophan spectrum by < 3 nm. As shown in Fig. 7.4, the recovered normalized integrated areas of $\mathrm{Phe_{CN}}$/Trp increase/decrease with increasing peptide bond separation. The fairly smooth variation with peptide bond separation is not necessarily expected as the sequence composition of the chain segments of the heteropolymer across which the energy transfer takes place varies greatly. Based on the ratios of the integrated $\mathrm{Phe_{CN}}$ fluorescence of each peptide to that of the $\mathrm{GKF_{CN}TV}$ reference peptide, the efficiency of energy transfer can further be calculated by

$$\langle E \rangle = 1 - \frac{I_{\mathrm{peptide}}}{I_{\mathrm{ref}}} \tag{7.1}$$

Although the transfer efficiencies exhibit minor deviations from a monotonic decrease with peptide bond separation (Table 7.1), with exception of PepE (*vide infra*), there

is a strong correlation (R^2=0.85) between transfer efficiency and peptide bond separation. Thus, an attempt to describe the polymer chain by a simple polymer model seems to be justified. In the following, we will investigate the following questions: (1) To what extent can the energy transfer efficiencies described by various polymer models? (2) What are the possible origins of the deviations of the efficiencies from such models?

Relationship between energy transfer and polymer models. The measured efficiency of energy transfer, $\langle E \rangle$, between donor and acceptor molecules are related to the distance distribution, $P(r)$, between the molecules, i.e.

$$\langle E \rangle = 1 - \left\langle \frac{r^6}{r^6 + R_0^6} \right\rangle = \int_0^\infty \frac{r^6}{r^6 + R_0^6} P(r) \mathrm{d}r \tag{7.2}$$

In the present case, where donor and acceptor are attached to a flexible polymer chain, the distance distribution between them is determined by the sequence separation of donor and acceptor, the intrinsic polypeptide properties and the complex interplay between the properties of the polymer and those of the solvent. Intrapolymer interactions are typically partitioned into local and nonlocal interactions, where local interactions describe steric repulsion between neighboring monomers in the polymer chain, i.e. those interactions measure the stiffness of the polymer chain [33, 330]. Nonlocal interactions refer to repulsive or attractive interaction between monomers that are distant in sequence space. The repulsive or attractive nature of those interactions critically depends on the solvent properties in relation to those of the polymer. Under

conditions in which solvent-polymer interactions are stronger than intra-polymer interactions, monomer-monomer interactions are repulsive in nature (excluded volume effect). Such a solvent is referred to as a good solvent for the polymer [33, 330]. In the case of polypeptides, good solvents are aqueous solutions of urea or guanidine hydrochloride or alcohols [334]. In the case where intra-polymer interactions are stronger than polymer-solvent interactions, a description of a heteropolymers dimensions in terms of a theoretical model is formidable as those depend on the relative strength of the intra-polymer interactions. In the simplest case, solvent-polymer and nonlocal interactions are completely neglected. In this case, the polymers properties are solely determined by local interactions, i.e. the intrinsic stiffness of the chain. Two popular models that describe such a scenario are the freely-jointed chain model [33] and Kratky-Porod (worm-like) chain model [330]. Obviously, neglecting polymer-solvent interactions is unrealistic, however, apparent ideal (freely-jointed) chain behavior is observed under Θ-conditions, i.e. under conditions where the excluded volume interactions are compensated for to the second order by intra-polymer attractive forces [33, 330]. Water is typically considered to be a poor solvent for foldable proteins, providing a driving force for collapse. However, in the present case, for the purpose of limiting diffusion enhance FRET [335], the experiments where carried out in a 90%/10% (v/v) mixture of 1,2-propandediol (propylene glycol) and water. The dielectric constant of 1,2,-propanediol is significantly smaller than that

of water, thus the tendency to form hydrophobic contacts is substantially reduced in this solvent compared to water. Furthermore, according to the criteria of Kyte and Doolittle [333], the sequence of the present peptide is hydrophilic, while the charges are fairly evenly distributed along the sequence (Fig. 7.1). As judged by CD, there is no secondary structure present in the peptide (Fig. 7.2) in aqueous medium. These observations suggest that it is not unreasonable to characterize the current conditions as good or near Θ-conditions for the peptide under investigation. However, as judged from the CD spectrum, we note that the 1,2-propanediol/water mixture induces secondary structure formation in the peptide (data not shown) with a helical content of roughly 30-35%.

Freely-jointed chain. Assuming Θ-conditions, a polymer chain may be described as a freely-jointed chain [33, 330], whose end-to-end distribution function is Gaussian, i.e.

$$P(r) = 4\pi r^2 \left(\frac{3}{2\pi \left\langle r^2 \right\rangle}\right)^{3/2} \exp\left(-\frac{3r^2}{2\left\langle r^2 \right\rangle}\right) \tag{7.3}$$

$$\left\langle r^2 \right\rangle^{1/2} = N^{1/2} b \tag{7.4}$$

Here, $\langle r^2 \rangle^{1/2}$, b and N are the mean root mean square distance between donor and acceptor, the Kuhn length and the number of peptide bonds separating donor and acceptor, respectively. This model provides a coarse-grained description of the polymer with contour length l_c in terms of uncorrelated (freely-jointed) bonds of length b, provided that the $l_c \gg b$. Though the freely-jointed chain model neglects excluded

volume interactions, the model was found to provide an appropriate description in a number of cases [313, 314, 315, 316]. As shown in Fig. 7.5, a fit of Eqs. 7.2 and 7.3 to the observed transfer efficiencies provides a poor description of the data. In particular, the value for PepE drastically deviates from Gaussian behavior. It is also evident that the transfer efficiency for the PepCN1 ($N = 12$) peptide is significantly higher than what would be expected based on the value for PepCN26 ($N = 13$). Though these observations might be due to intrinsic polymer properties of the sequence, we note that in both PepE and PepCN1, the Phe_{CN} label is at the N-terminal. Though the protonation state of the N-terminal is expected to only marginally to moderately affect the quantum efficiency of Phe_{CN} [332, 193, 192], we suspected that a modulation of the quantum efficiency might be the origin of the systematic positive deviations of the transfer efficiencies. Thus, we synthesized a version of PepCN1, denoted $\text{PepCN1}_{\text{W13F}}$ in which the tryptophan residue was mutated to phenylalanine. Indeed, we found that the integrated area of the $\text{PepCN1}_{\text{W13F}}$ fluorescence is 36% lower than that of the $\text{GKF}_{\text{CN}}\text{TV}$ peptide. Moreover, in aqueous solvent, the fluorescence spectrum of $\text{PepCN1}_{\text{W13F}}$ exhibits a pronounced shoulder on the red side of the major peak (Fig. 7.6). This shoulder is also present when the peptide is dissolved in the 1,2-propandediol/water mixture, but it is less pronounced in that solvent. The fluorescence excitation spectrum confirms that the shoulder does not arise from a possible contamination of the sample with tryptophan (Fig. 7.6). At this time, we do

not know the origin of this feature in the fluorescence spectrum. Using $\text{PepCN1}_{\text{W13F}}$ as a reference for the PepCN1 and PepE peptides, the energy transfer efficiencies were found to be 0.472 and 0.293, respectively (Table 7.1, values in parentheses). Fitting the freely-jointed chain model to all seven transfer efficiencies provides a value of 5.5 Å for b. The description of the data seems to be satisfactory, showing only small deviations from the fit (Fig 7.7).

Kratky-Porod chain. An alternative description of the distance distribution between points on a polymer chain is given by the Kratky-Porod (worm-like chain) model [34, 330]. Unlike the freely-jointed chain model, where there is no correlation between the orientations of Kuhn segments, the Kratky-Porod chain takes into account correlations between bond vectors. In the framework of this model, the strength of this correlation is measured by the persistence length l_p. For the asymptotic limit, $l_c/l_p > 10$, Gobush *et al.* [336] derived the following expression for the distance distribution:

$$P(r) = 4\pi r^2 \left(\frac{3}{4\pi l_p l_c}\right)^{3/2} \exp\left(-\frac{3r^2}{4l_c l_p}\right) \times \left(1 - \frac{5l_p}{4l_c} + \frac{2r^2}{l_c^2} - \frac{33r^4}{80l_p l_c^3} - \frac{79l_p^2}{160l_c^2} - \frac{329r^2 l_p}{120l_c^3} + \frac{6799r^4}{1600l_c^4} - \frac{3441r^6}{2800l_p l_c^5} + \frac{1089r^8}{12800l_p^2 l_c^6}\right) \quad (7.5)$$

Here, $l_c = Na$ is the contour length and l_p is the persistence length of the polymer, which provides a measure of the stiffness of the polypeptide chain. The constant a is the bond length, which is 3.8 Å in the case of peptides. The Kuhn length is twice the persistence length, i.e. $b = 2l_p$. A fit of Eqs. 7.2 and 7.5 to all seven data points

results in a persistence length of 4.2 Å (Fig. 7.7).

Excluded-volume chain. As discussed earlier, both the freely-jointed and Kratky-Porod chain models neglect volume exclusion of the polymer [330]. However, a large number of studies, using intrinsic viscosity measurements [337], measurements of hydrodynamic radii [338, 339] small-angle neutron and x-ray scattering studies [327], suggest that a large number of chemically denaturants provide good solvent conditions for polypeptide chains, i.e. conditions in which solvent-polymer interactions are stronger than intra-polymer interactions. Thus, a model for a volume excluded random coil polymer under good solvent conditions may be more appropriate for the description of the conformational distribution. According to the scaling laws of polymer physics, the probability distribution of the length of the vector connecting the two ends of a random coil polymer is given by [330]

$$P(R) = \frac{1}{X^3} f\left(R/X\right) \tag{7.6}$$

X is related to the mean-square distance between the termini

$$X = \sqrt{\frac{\langle R^2 \rangle}{3}} \tag{7.7}$$

and $\langle R^2 \rangle$ scales according to

$$\left\langle R^2 \right\rangle^{1/2} = qN^{\nu} \tag{7.8}$$

where q is a constant, N is the number of peptide bonds and ν is the Flory coefficient.

Defining $y = R/X$ and $s = R_0/X$, Eq. 7.2 may be written as

$$\langle E \rangle = 1 - \int_0^\infty 4\pi y^2 f(y) \frac{y^6}{y^6 + s^6} \mathrm{d}y \tag{7.9}$$

The function $f(y)$ may be approximated by the interpolation function [330]

$$f(y) = f_1 y^\theta \exp\left[-D y^{1/(1-\nu)}\right] \tag{7.10}$$

where $\theta = (\gamma - 1)/\nu$ and D and f_1 are respectively given by

$$D = \left[\frac{(4 - 5\nu + \theta - \nu\theta)!}{3(2 - 3\nu + \theta - \nu\theta)!}\right]^{1/2(1-\nu)} \tag{7.11}$$

$$f_1 = \frac{1}{4\pi(1-\nu)} \frac{\left[\frac{1}{3}(4 - 5\nu + \theta - \nu\theta)!\right]^{(2+\nu)/2}}{[(2 - \nu + \theta - \nu\theta)!]^{(5+\nu)/2}} \tag{7.12}$$

A value of 1.1615 was assumed for γ [330]. Eq. 7.9 was derived under the assumption that donor and acceptor are attached to the ends of the polymer chain. As this is not the case in our study, application of Eq. 7.9 for the analysis of the measured transfer efficiencies in the present case needs to be justified. In particular, previous lattice simulation studies found that the end-to-end distance of a subchain with a number of Y bonds of a chain of length N, exhibits an increased end-to-end distance compared to a chain of length Y [340]. This swelling of a subchain is dependent on the position of the segment in the chain, being most pronounced when the segment is symmetrically positioned in the center of the chain of length N. Thus, no single set of parameters q and ν are expected to describe the end-to-end distance of different subchains according to Eq. 7.8. To estimate to what extent the swelling of subchains occurs in the present

case, we performed self-avoiding walk simulations on a cubic lattice employing the Rosenbluth sampling scheme [341]. The mean square distance between donor and accpetor positions of the six subchains of the $N = 28$ self-avoiding walk are compared to the end-to-end distance of the walks of a length corresponding to the length of the respective subchain. As shown in Table 7.2, the mean square monomer-to-monomer distances of the subchains are indeed larger than the end-to-end distances of the corresponding chain. However, with exception of the first subchain, all subchains show a fairly consistent swelling of 10%. Fitting the mean square distances of the subchains to Eq. 7.8 a very good fit with $q^2 = 1.2$ and $2\nu = 1.17$. Based on these results, we feel that in the present case, it is justified to use Eq. 7.9 for the analysis of the measured energy transfer efficiencies between monomers within the chain, though the shape of the distributions might deviate somewhat from Eq. 7.10 [342]. Indeed, fitting Eq. 7.9 (Fig. 7.7) to all seven measured transfer efficiencies provides a fitting parameter of $q = 4.6$ Å, where a value of $\nu = 0.588$ was assumed [327].

7.3 Discussion

As evident from Fig. 7.2, both the freely-jointed chain model as well as the Kratky-Porod model provide reasonable fits to the measured energy transfer efficiencies. However, whereas the Kuhn statistical segment length, b, was found to be 5.5 Å by fitting the freely-jointed model to the data, a fit of the Kratky-Porod model provides an

apparent persistence length of 4.2 Å, implying $b = 2l_p = 8.4$ Å. This stark disagreement between the estimated lengths for the Kuhn statistical segment raises the question, which, if any, of the models provides a realistic value. While the value for the persistence length derived from the Kratky-Porod model falls into the range of previously reported values of 4-13 Å [318, 319, 320, 194, 328, 322, 324, 325, 326], the value derived from the freely-jointed chain model ($l_p = b/2 = 2.25$ Å) is clearly too small (shorter than a peptide bond length). Thus, we conclude that a polypeptide does not behave as a freely-jointed chain on the lengthscales ($N = 5 - 28$) considered in this work. This conclusion is supported by a large number of studies, where the loop closure kinetics of polypeptides with low-complexity sequences were studied [317, 318, 319, 320, 321, 322, 324, 325, 326]. For example, Kiefhaber and coworkers found that poly(Gly-Ser) chains behave as random coils for $N > 20$ [320]. Similarly, Eaton and coworkers showed that the Kratky-Porod chain, but not the freely-jointed chain model, appropriately describes their measured end-to-end diffusion rates of Cys-(Ala-Gly-Gln)$_j$Trp ($j = 1 - 6$) [319, 322]. However, the question arises how well the Kratky-Porod chain describes the conformational behavior of a polypeptide with a more complex sequence, like the one considered in this work. For comparison to previous work, the persistence length may be converted into an alternative measure of chain stiffness, i.e. the asymptotic value of Flory's characteristic ratio, which is

given by

$$C_\infty = \frac{2l_p}{a} \tag{7.13}$$

Using Eq. 7.14, the characteristic ratio based on the Kratky-Porod model is ~ 2.2. Theoretical considerations and experimental observations estimate for polyglycine $C_\infty \sim 2$, and $C_\infty \sim 9$ for all other amino acids, with the exception of proline, for which $C_\infty \sim 100$ [343, 344]. For a polypeptide chain with a content of 10% glycine, Miller *et al.* predicted $C_\infty \sim 6.8$ [345]. For the Cys-(Ala-Gly-Gln)$_j$Trp peptide, Lapidus *et al.* found an apparent $C_\infty \sim 3.4$, [319] which is closer to our value. Discrepancy between different estimates might arise from different experimental conditions: Within the framework of the theory, the persistence length parameter arises only from local interactions, i.e. covalent chemical bonding and rotational potentials around bonds, which include steric effects between neighboring monomers. According to the theory, it is only those effects, together with the chain length that determine the average distance between two points of the chain and thus the experimental observable (e.g. $\langle E \rangle$). However, experimentally, this is only the case under Θ-conditions [344]. If the solvent is a poor solvent, long-range attractive intrapolymer interactions give rise to a collapse of chain and thus an apparent lowering of the apparent persistence length. If the solvent is a good solvent, excluded volume effects give rise to a swelling of the polymer and therefore increase the apparent persistence length. This behavior was clearly shown in recent experiments by Schuler and coworkers, who showed that the

apparent persistence length of the small cold shock protein CspTm varies from 3.9 Å to 11 Å when going from 0 M GndHCl to 8 M GndHCl [313]. The groups of Kiefhaber and Eaton observed the opposite effect for low complexity sequences that lack hydrophobic side chains, i.e. the persistence length tends to decrease with increasing concentration of denaturant [320, 321, 322]. This observation was attributed to interactions between side chains of neighboring residues, which are suppressed at high concentrations of denaturant, thus making the chain more flexible. It is interesting to compare the current results with those of Nau and coworkers [346], who measured the persistence length of poly(GlySer) peptides under similar solvent conditions (100% 1,2-propandiol) as ours. The recover persistence length is on the order of $\sim$ 10 Å. Similarly, Lapidus and coworkers found a persistence length of $\sim$ 13 Å for polyglutamine peptides [325]. The fact that the apparent persistence length measured for the peptide under investigation here is significantly smaller than that for the poly(GlySer) peptide is not necessarily due to poor solvent conditions which would give rise to a compact chain dimensions, but rather a reflection of the specific interactions that are present in poly(GlySer). For example, kinetic studies of Krieger *et al*. showed that poly(GlySer) in aqueous solution are significantly stiffer than the theoretically predicted value of $\sim$ 5.7 Å [343]. Zhou showed that the end-to-end distributions of loops in proteins can be modeled by a Kratky-Porod chain with a persistence length of 3 Å [347]. However, our value of 4.2 Å reflects the presence of long-range interactions as

evidenced by the CD spectrum of the peptide in 1,2-propanediole (data not shown), which indicates a helical content of about 30-35% under these conditions. Under such conditions, the polypeptide could adopt a compact and/or molten globule like conformation [348]. If this is the case, a description with a simple excluded volume model should fail. As seen in Fig. 7.7, such a model indeed provides poor description of the data. Flory's exponent ν in Eq. 7.8 indicates how the root mean square end-to-end distance varies with chain length N. Under poor solvent conditions $\nu < 1/2$, under Θ-conditions $\nu = 1/2$, and under good solvent conditions, $\nu \sim 5/3$. Thus, provided the Flory exponent can be measured, it will provide information about the solvent conditions. Though, we have no means in our laboratory to measure this quantity independently, fitting the excluded volume model to the measured FERT efficiencies, provides some insight into the conformational distribution of the peptide under the current solvent conditions. The excluded volume model described above was derived for good solvent conditions. Thus, the model should provide a reasonable description if $\nu \sim 5/3$. Furthermore, the model is only valid in the limit of large N, i.e. $aN > l_p$. In the present case, the shortest sequence distance between donor and acceptor is $aN = 19$ Å, suggesting that the $aN > l_p$ case applies here. Interestingly, a fit of the model to the data provides results in $q = 4.6$ Å, where $\nu = 0.588$ was set to the value determined by Kohn *et al.* [327]. Using this value, together with the parameter

$q = 4.6$ Å, we calculated the radius of gyration of the peptide by

$$\left\langle R_g^2 \right\rangle^{1/2} = \sqrt{\frac{\aleph}{6} \left\langle R^2 \right\rangle} \tag{7.14}$$

where the constant $\aleph$ is close to unity [330]. It is $R_g = 13.3$ Å. This is an interesting result as the relation given by Kohn *et al.* [327] predicts a value of 14.1 Å. However, as can be seen from Fig. 7.7, this fit is comparatively poor, but still acceptable. Taken together, these results indicate that the peptide is somewhat more collapsed than what would be expected for a volume excluded random coil, consistent with the CD spectrum under these conditions (not shown), which indicates the presence of helical structure. Thus the apparent persistence length recovered from fitting the Kratky-Porod chain reflects the presence of this helical structure, which gives rise to a more compact state. However, its value depends on the assumption that Eq. 7.5 also applies to $l_c/l_p \sim 4.5$, which is not necessarily the case and the usage of a switching function [349] might be more appropriate.

7.4 Materials and Methods

Materials. Natural Fmoc-protected amino acids were purchased from Advaned Chem-Tech (Louisville, KY) and Fmoc-Phe_{CN} was obtained from Bachem (Torrance, CA). Millipore water (Billerica, MA) was used to prepare aqueous solutions. 1,2-propanediol (Acros Organics, 99% p.a.) was obtained from Fisher Scientific. Peptides were synthesized using the standard Fmoc-protocol employing Rink resin. All peptide sam-

ples were purified to homogeneity by reverse-phase chromatography and characterized by mass-spectrometry. Following purification, peptides solutions were lyophilized and subsequently re-dissolved in 20 mM phosophate buffer at pH 3.0 or in a mixture of 10% water and 90% 1,2-propandediol (v/v) for spectroscopic measurements.

Absorption measurements. The concentrations of the peptides were determined optically via the sample absorbance at 280 nm, which was measured on a PerkinElmer Lambda 25 UV-Vis spectrometer (Fremont, CA). The molar absorption coefficients of tryptophan and *p*-cyanophenylalanine are $\epsilon_{280\text{nm}} = 5500\ \text{M}^{-1}\ \text{cm}^{-1}$ and $\epsilon_{280\text{nm}} = 850\ \text{M}^{-1}\ \text{cm}^{-1}$ [194], respectively.

Fluorescence and circular dichroism spectroscopy. The fluorescence spectra were obtained at 20 °C on a Fluorolog 3.10 spectrofluorometer (Jobin Yvon Horiba, Edison, NJ) with a 1 cm quartz sample holder. Temperature was regulated using a TLC 50 Peltier temperature controller (Quantum Northwest, Spokane, WA). An integration time of 2.0 s/nm was used in all measurements. The concentrations of all samples were $\sim$ 20 μM. The peptide solutions were stirred during measurements using a magnetic stirrer to avoid photo-bleaching. The far-UV circular dichroism spectrum of PepCN5 was measured on Aviv Model 410 instrument (Aviv Biomedical, Lakewood, NJ) using a 1 mm cuvette at 20 °C. The peptide concentration was $\sim$ 20 μM for all measurements.

Determination of the Förster radius. The Förster radius of the Phe_{CN}-Trp chro-

mophore pair in a mixture of 10% water and 90% 1,2-propandediol (v/v) was determined according to [350]

$$R_0^6 = \left(\frac{9000(\ln 10)\kappa^2 Q_D}{128\pi^5 N \eta^4} \right) J(\lambda) \tag{7.15}$$

κ^2, Q_D, η, and $J(\lambda)$ are the orientation factor, quantum yield, refractive index of the solvent and overlap integral, respectively. The overlap integral is defined as

$$J(\lambda) = \int_0^\infty F_D(\lambda)\epsilon_A(\lambda)\lambda^4 \mathrm{d}\lambda \tag{7.16}$$

where F_D and $\epsilon_A(\lambda)$ are the peak-normalized emission spectrum of the donor and the molar absorption coefficient of the acceptor, respectively. The quantum yield of $\mathrm{Phe_{CN}}$ was calculated according to

$$\Phi_x = \Phi_{ref} \frac{I_x}{I_{ref}} \frac{A_{ref}}{A_x} \frac{\eta_x^2}{\eta_{ref}^2} 10^{\Delta OD/2} \tag{7.17}$$

where I and A represent the integrated fluorescence intensity and absorbance of $\mathrm{Phe_{CN}}$ at 280 nm, respectively. The quantum yield of $\mathrm{Phe_{CN}}$ in water ($\Phi_{ref} = 0.11$) served as a reference [192]. The last factor in Eq. 7.17 is the "inner-filter" correction factor [351]. The orientation factor was assumed to be equal to 2/3 and the refractive index assumed to be 1.42 for the solvent mixture and 1.333 for water. Φ_x was found to be 0.13, slightly larger than what would be expected based on previous results [192]. However, these small differences as well as assuming a refractive index of 1.42 for the solvent mixture affect the calculated Förster radius by less than half an ångström, which was found to be about 17 Å.

7.5 Acknowledgement

We gratefully acknowledge financial support from the National Institutes of Health (GM-065978 and RR-01348). M.M.W. thanks David Lee for assisting in the preparation of peptide samples.

Table 7.1: Calculated FRET efficiencies, $\langle E \rangle$, (standard deviation= ± 0.032) according to Eq. 7.1 using $GKF_{CN}TV$ fluorescence as a reference. Values in parentheses were determined using the $PepCN1_{W13F}$ peptide as a reference (see discussion in text). N is the number of peptide bonds separating donor and acceptor.

Peptide	N	$\langle E \rangle$
PepCN8	5	0.760
PepCN19	6	0.830
PepCN5	8	0.661
PepCN22	9	0.667
PepCN1	12	0.641 (0.472)
PepCN26	13	0.471
PepE	29	0.518 (0.292)

Table 7.2: Comparison of the mean square distance, $\langle r^2 \rangle$, of subchains of length N for a self-avoiding walk of 28 links to the mean square end-to-end distance, $\langle R^2 \rangle$, of walks of length N. The end positions of the subchains in the chain of 28 links correspond to the positions of $\mathrm{Phe_{CN}}$ and Trp in the first six peptides listed in Table 7.1.

Number of links, N	$\langle r^2 \rangle$ for subchain of length N	$\langle R^2 \rangle$ for chain of length N
5	8.01	7.19
6	9.97	9.07
8	14.26	12.86
9	16.44	14.84
12	22.18	21.01
13	25.24	23.14

Figure 7.1: Hydrophobicity and charge plot for of the designed sequence.

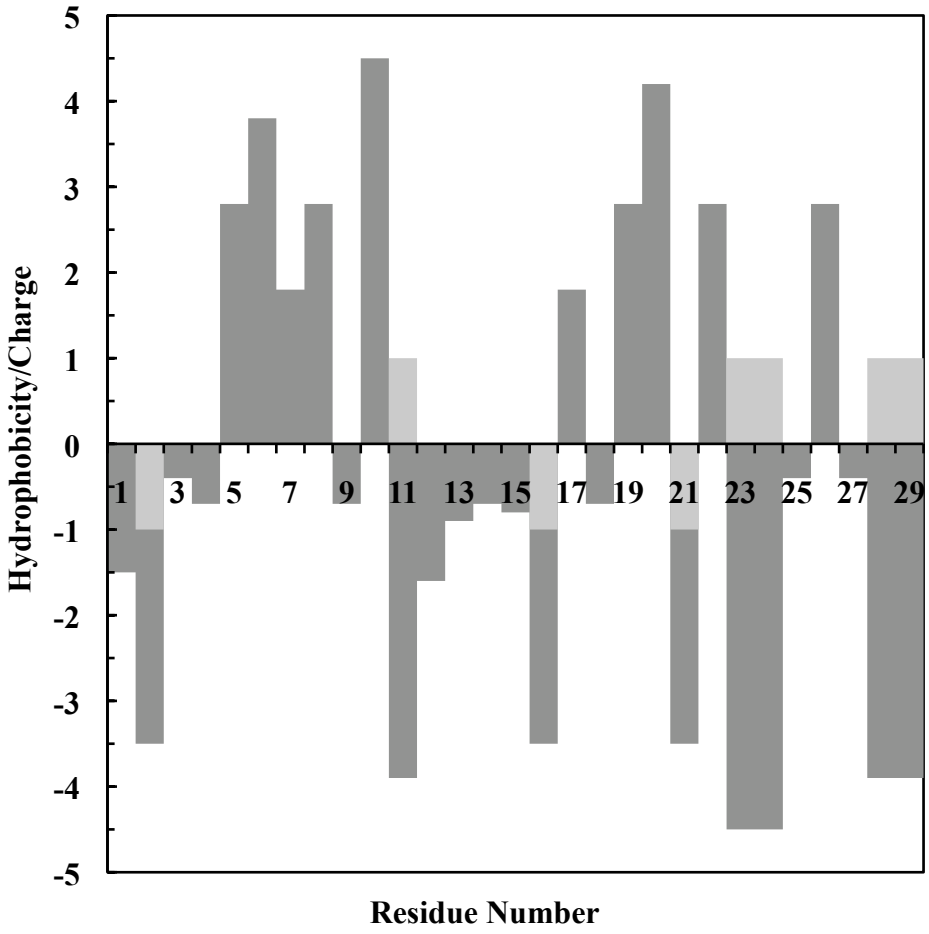

Hydrophobicity/Charge
5
4
3
2
1
0
-1
-2
-3
-4
-5
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
Residue Number

Figure 7.2: CD spectrum of PepCN5 in 20 mM phosphate buffer at pH 3.0 and 20 °C. The concentration of the sample was ~ 31 μM.

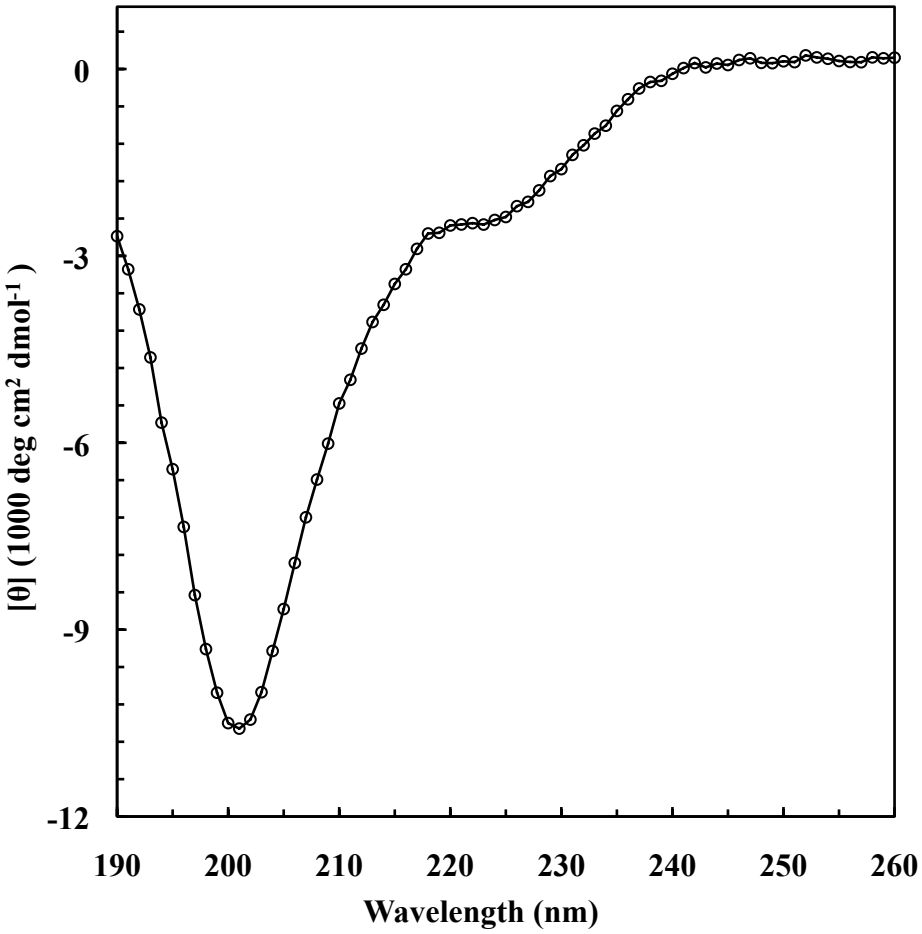

0
-3
-6
-9
-12
[θ] (1000 deg cm² dmol⁻¹)
190
200
210
220
230
240
250
260
Wavelength (nm)

Figure 7.3: Fluorescence spectra of the peptides in a mixture of 10% water and 90% 1,2-propandediol. $\lambda_{ex} = 240$ nm. The concentrations of all samples was ~ 20 μM. Each spectrum is corrected for its respective sample OD at 280 nm.

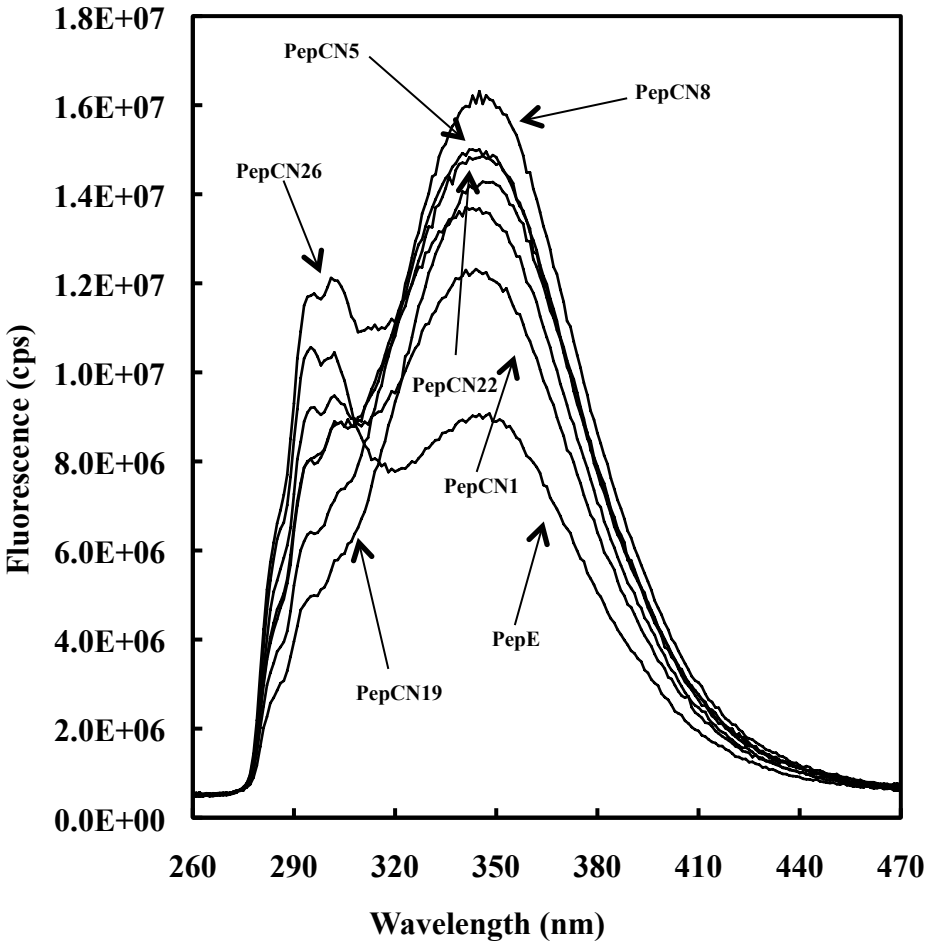

1.8E+07
1.6E+07
1.4E+07
1.2E+07
1.0E+07
8.0E+06
6.0E+06
4.0E+06
2.0E+06
0.0E+00
Fluorescence (cps)
260
290
320
350
380
410
440
470
Wavelength (nm)
PepCN5
PepCN8
PepCN26
PepCN22
PepCN1
PepE
PepCN19

Figure 7.4: Normalized integrated areas of Phe_{CN} fluorescence ($\triangle$), Trp fluorescence ($\bigcirc$), and their ratios ($\diamond$).

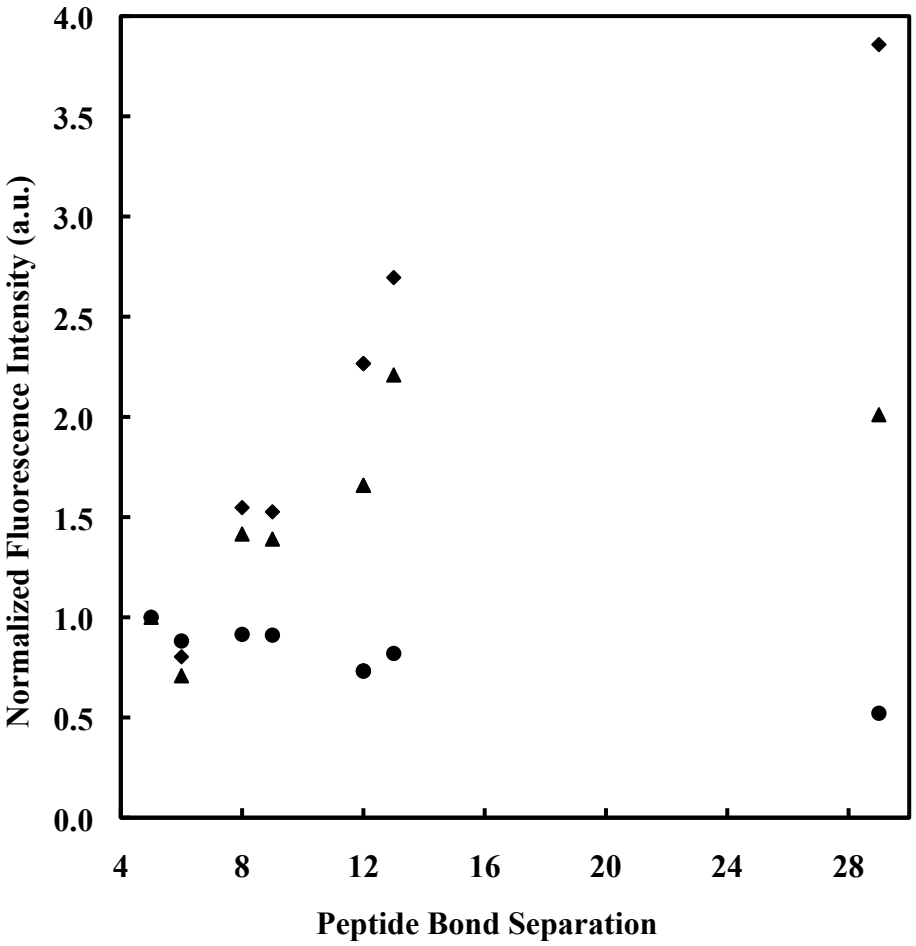

Normalized Fluorescence Intensity (a.u.)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
4
8
12
16
20
24
28
Peptide Bond Separation

Figure 7.5: Calculated efficiencies, $\langle E \rangle$, (based on Eq. 7.1 and $GKF_{CN}TV$ peptide fluorescence as a reference) as a function of peptide bond separation. The curve is a fit to the freely-jointed chain model.

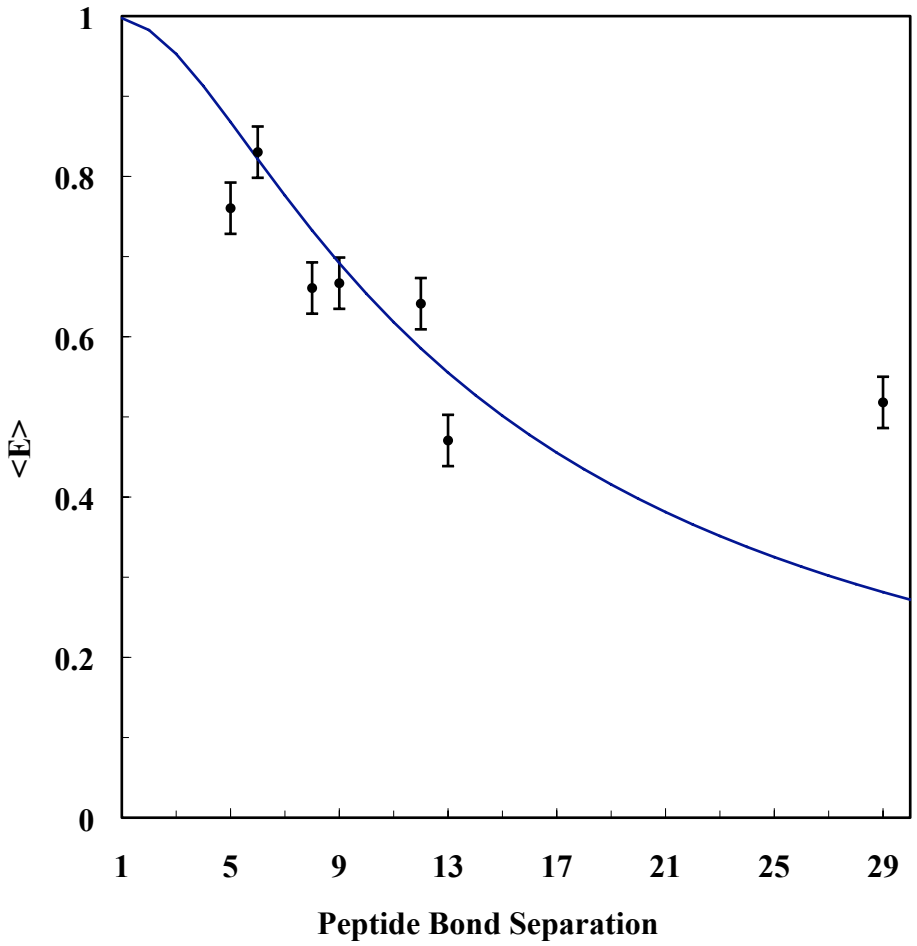

1
0.8
0.6
0.4
0.2
0
<E>
1
5
9
13
17
21
25
29
Peptide Bond Separation

Figure 7.6: Peak normalized fluorescence spectra of the $GKF_{CN}TV$ (black) and the $PepCN1_{W13F}$ (red) peptides in 20 mM phosphate buffer pH 3. The peak ratio $GKF_{CN}TV/PepCN1_{W13F}$ was found to be ~ 4.3 under these conditions. $\lambda_{ex} = 240$ nm. The fluorescence excitation spectrum of $PepCN1_{W13F}$ is shown in the inset. $\lambda_{em} = 310$ nm.

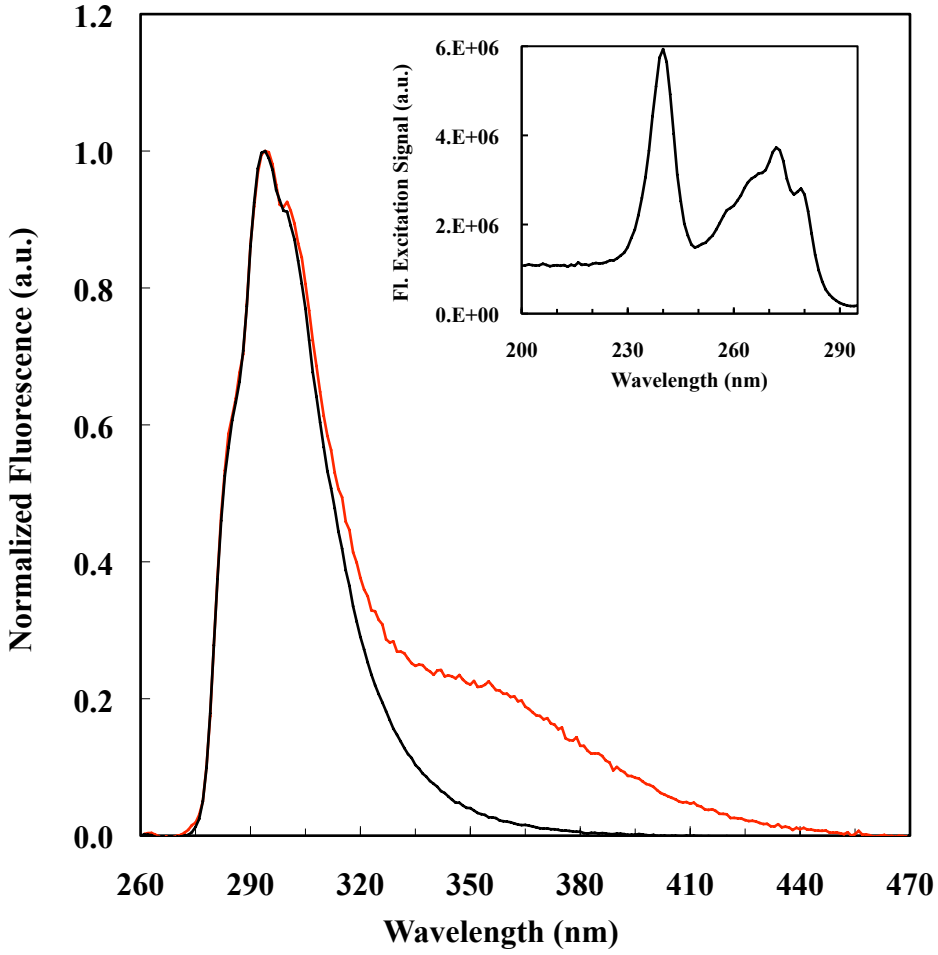
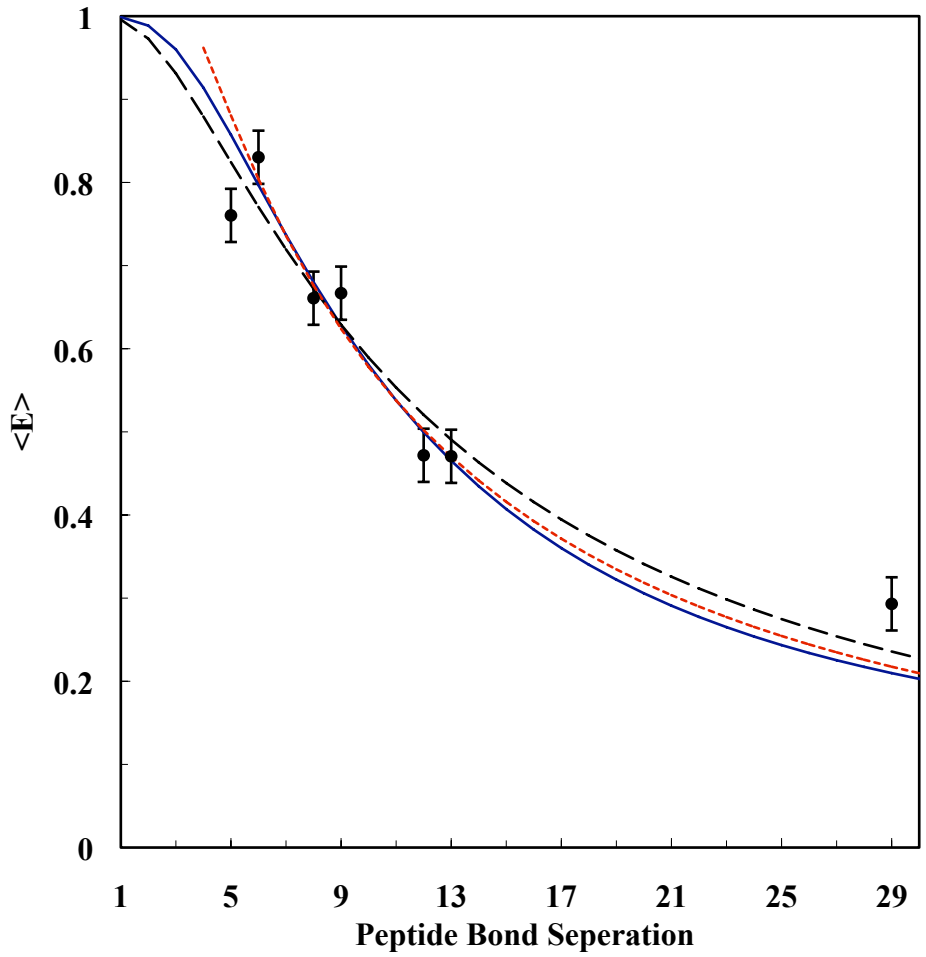

Normalized Fluorescence (a.u.)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
260
290
320
350
380
410
440
470
Wavelength (nm)
Fl. Excitation Signal (a.u.)
6.E+06
4.E+06
2.E+06
0.E+00
200
230
260
290
Wavelength (nm)

Figure 7.7: Freely-jointed (black, dashed line), Kratky-Porod (red, dashed line), and volume-excluded (blue, solid line) chain models fit to calculated efficiencies, $\langle E \rangle$. All $\langle E \rangle$ were calculated according to Eq. 7.1 using $\mathrm{GKF_{CN}TV}$ as a reference, except the efficiencies corresponding to PepCN1 and PepE, for which the $\mathrm{PepCN1_{W13F}}$ mutant was used as a reference.


1
0.8
0.6
0.4
0.2
0
<E>
1
5
9
13
17
21
25
29
Peptide Bond Seperation

Chapter 8

5-Cyanotryptophan as an Infrared Probe of Local Hydration Status of Proteins

8.1 Introduction

The stretching frequency of a nitrile (C≡N) group is sensitive to its environment, determined jointly by the solvent and the parent molecule [138, 139, 171, 352, 149, 151, 160, 164, 165, 146, 147]. In addition, the C≡N stretching vibration is largely decoupled from other vibrational modes of biological molecules and thus, has been used to probe various properties of a wide variety of biological systems. For example, it has been used to (a) report on the orientation and hydration status of individual sidechains of a peptide that is bound to lipid membranes [353, 354] or AOT reverse micelles [355], (b) probe the structure of amyloid fibers [356], (c) examine the binding of a ligand [172] or anesthetic halothane [357, 358] to a protein, (d) probe the dynamics of an HIV-1 reverse transcriptase inhibitor via two-dimensional infrared spectroscopy [359], (e) investigate local electric fields in proteins by means of vibrational Stark spectroscopy [141, 178, 360], and (f) study the local environments of DNA fragments [177, 361, 362].

For studies involving peptides and/or proteins, a nitrile group is typically introduced by incorporating a nitrile-derivatized amino acid into the targeted polypeptide sequence using either an *in vitro* [171, 178] or *in vivo* synthesis method [172]. So far, three nitrile-derivatized amino acids, i.e., *p*-cyanophenylalanine ($\mathrm{Phe_{CN}}$) [171, 172], cyanoalanine [171] and thiocyanatoalanine [178], have been used for this purpose. However, in light of concerns that substitution of a native residue with a non-native

amino acid, especially a non-natural amino acid, could lead to undesirable structural perturbations to the native fold, it would be beneficial to have other nitrile-derivatized amino acids available to replace a targeted native residue with the respective nitrile-containing derivative in order to minimize any potential structural perturbations. Herein, we investigate the potential utility of 5-cyanotryptophan (Trp_{CN}) as an infrared reporter of local environment of proteins and/or peptides. Our motives for choosing this non-natural amino acid include the following: (a) tryptophan (Trp) is often located at sites that are important for either function or stability of proteins (b) the Stark spectroscopic study of Boxer and coworkers [141] suggests that the nitrile group in Trp_{CN} might be a more sensitive reporter of the local electrostatic environment than those in other nitrile-derivatized amino acids, and (c) the molar absorption coefficient of the C≡N stretching vibration of Trp_{CN} is expected to be about 1.5 times larger than that of Phe_{CN} [141], an added advantage for studies wherein relatively low peptide/protein concentration is required. Finally, besides its potential usefulness as an infrared reporter, Trp_{CN} could exhibit interesting fluorescence properties, akin to its structure analog, 5-hydroxytryptophan [363]. To test the utility of Trp_{CN} as an infrared probe of local environment, we investigate how the C≡N stretching band of the free amino acid changes with solvent and how that of two Trp_{CN} containing mutants of a cationic antimicrobial peptide [364, 365], indolicidin (sequence: ILPWKWPWWPWRR-NH_2), varies in response to the binding of the peptide to

DPC micelles and POPG lipid bilayers. Our results show that the bandwidth of the C≡N stretching vibration of Trp_{CN} is especially sensitive to the degree of hydration of the amino acid, thus making it a very attractive infrared probe of local hydration status. In addition, our results indicate that while the fluorescence quantum yield of Trp_{CN} is approximately one order of magnitude smaller than that of Trp, rendering it less useful as a stand-alone fluorescent probe, it can still be used together with Phe_{CN} [191] for florescence energy transfer applications.

8.2 Results and Discussion

5-Cyanoindole and 5-cyanotryptophan. The FTIR spectra of 5-cyanoindole and the amine protected form of Trp_{CN} (Fmoc-L-Trp_{CN}) were measured in H_2O, tetrahydrotetrahydrofuran (THF) and THF-H_2O mixtures. Here, THF was used to mimic the hydrophobic interior of proteins because its dielectric constant is roughly 10 times smaller than that of H_2O. As shown (Fig. 8.1), the peak position of the C≡N stretching band of 5-cyanoindole in THF is centered at about 2220 cm^{-1}, which is only about 4 cm^{-1} apart from that in H_2O. However, the bandwidth (FWHM) of the C≡N stretching vibration in THF (8.7 cm^{-1}) is significantly narrower than that in H_2O (15.6 cm^{-1}). As indicated (Fig. 8.2), the C≡N stretching vibration of Fmoc-L-Trp_{CN} shows a similar trend. Moreover, in agreement with the study of Suydam and Boxer [141], the vibrational transition dipole moment of Fmoc-L-Trp_{CN} is estimated

to be a factor of about 1.8 larger than that of Fmoc-L-Phe$_{CN}$ [171], calculated using the integrated areas of the respective C≡N stretching bands. Taken together, these results indicate that the C≡N stretching vibration of Trp$_{CN}$, especially its bandwidth, is sensitive to solvent and thus, could serve as an infrared reporter of the local hydration status of peptides and proteins. Suydam and Boxer have shown that the C≡N stretching vibration of 5-cyanoindole has a larger Stark tuning rate than that of *p*-tolunitrile [141]. Thus, the seemingly low sensitivity of the peak frequency of the nitrile stretching band of 5-cyanoindole to solvent is somewhat surprising. However, this finding is consistent with the notion put forward by Webb and Boxer [360], as well as Cho and coworkers [160, 161] that the Stark tuning rate is not necessarily a quantitative indicator of how the C≡N stretching vibration is affected by changes in solvation, due to the complexity of the interactions between solvent molecules and the C≡N moiety. For example, Cho and coworkers found that the C≡N stretching frequency of cyanomethane (CH_3CN) is determined by the precise configuration of water molecules surrounding the C≡N group, namely, a linear arrangement of CH_3CN-H_2O (termed σ-bonding with nitrogen lone-pair) will result in a blue-shift, whereas interactions of water molecules with the *p*-orbital of the nitrile moiety (termed π-bonding) will lead to a red-shift [160]. In other words, the solvent-induced vibrational frequency shift for aliphatic nitriles is determined by an intricate balance between the σ-bonding and π-bonding interactions. In the present case the indole ring is an ad-

ditional, and perhaps more important, factor that needs to be considered. As shown (Fig. 8.1), the bandwidth (FWHM) of the C≡N stretching vibration of 5-cyanoindole in water is about 7 cm^{-1} broader than that of *p*-tolunitrile [171], indicating that the structure of the parent molecule is an important determinant of the bandwidth (and peak frequency) of the C≡N stretching vibration. In other words, any interactions between the parent and solvent molecules that can affect the C≡N bond strength would result in a change in its vibrational frequency. Therefore, it is not surprising that an indole ring (in 5-cyanoindole and Trp_{CN}), which is large and aromatic and thus capable of forming various water-indole complexes with different C≡N stretching frequencies, can lead to a broad C≡N stretching vibrational band. On the other hand, a smaller phenyl group (in *p*-tolunitrile and Phe_{CN}) would support a smaller number of static solute-solvent complexes on the time scale of the C≡N stretching vibration compared to the indole ring, thus a narrower C≡N stretching vibrational band. While a change in the lifetime of the excited vibrational state can also lead to spectral broadening/narrowing, the reported lifetimes of the first excited state of the C≡N stretching vibration of various nitriles [146, 147] suggest that lifetime broadening is not a dominating factor in the current case.

Indolicidin. To test the utility of Trp_{CN} as an infrared reporter of local hydration environment in the context of a biological system, we used it to probe how an individual Trp residue in indolicidin changes its hydration status upon peptide binding to

micelles or membranes from the aqueous phase. Indolicidin is a Trp rich antimicrobial peptide, which is unstructured in aqueous solution but adopts a wedge-shaped conformation when bound to the membrane-water interface region of DPC micelles or POPG lipid bilayers [365]. Thus, indolicidin constitutes a good model system to test the applicability of Trp_{CN} in monitoring changes in local hydration status at the single-residue level. Specifically, we prepared two indolicidin mutants, namely Trp11/Trp_{CN} and Trp9/Trp_{CN}, and measured their C≡N stretching bands in the presence or absence of those model membranes. As shown (Fig. 8.3), the C≡N stretching band of Trp9/Trp_{CN} in water is very broad (the FWHM is approximately 34 cm^{-1}). This result is consistent with the aforementioned picture that in aqueous solution the bandwidth of the C≡N stretching vibration of Trp_{CN} is dominated by the mechanism of inhomogeneous broadening and also the fact that the unstructured indolicidin polypeptide chain molecule can sample a vastly large number of conformations. As shown (Fig. 8.3), the C≡N stretching band of Trp9/Trp_{CN} becomes significantly narrower upon binding to DPC micelles, indicating that the corresponding Trp sidechain (i.e., Trp9) is less hydrated in the micelle-bound state of indolicidin. Similarly, the C≡N stretching band of Trp11/Trp_{CN} also indicates that Trp11 becomes less exposed to water upon peptide-micelle association. In fact, the C≡N stretching vibrational bands of the micelle-bound Trp11/Trp_{CN} and Trp9/Trp_{CN} are practically superimposable (Fig. 8.4). Taken together, these results indicate that on average Trp9 and Trp11

sample a similar environment (in terms of hydration), even though they may show different orientations with respect to the peptide backbone and depths of membrane insertion according to molecular dynamics simulations [366, 367]. While it is not a direct proof, the NMR study by Hancock and coworkers [365], which showed that a spin label residing mostly at the surface of the DPC micelles reduces the ^{1}H NMR signal intensity of Trp9 and Trp11 by 35% and 22%, respectively, is nevertheless consistent with our findings. To verify that the Trp-to-Trp$_{CN}$ mutation does not cause significant conformational changes of the micelle-bound peptides, we further carried out CD studies. The CD spectrum of Trp11/Trp$_{CN}$ in the presence of DPC micelles (data not shown) shows that the peptide backbone gives rise to a negative CD band centered at 205 nm, as observed for the wild-type indolicidin [365], suggesting that the structural perturbation arising from the Trp$_{CN}$ mutation is insignificant. Thus, given the rather distinct role of Trp in many antimicrobial peptides and membrane proteins, Trp$_{CN}$ may prove to be a valuable infrared reporter that complements other probes for studying peptide-membrane interactions. In addition, these results validate the utility of Trp$_{CN}$ as an infrared probe of local hydration status or proteins, a subject of great importance [368, 356]. It has been shown that indolicidin binds more strongly to lipid bilayers than to micelles [364, 365]. Thus, we further measured the C≡N stretching bands of both indolicidin mutants in the presence of POPG bilayers (Fig. 8.3). Indeed, the C≡N stretching band of the POPG bilayer-bound peptides

exhibits a measurable decrease (about 6 cm^{-1}) in its bandwidth compared to DPC. This result is in agreement with the NMR study indicating that the Trp sidechains of the membrane-bound peptide are less disordered [365], further underscoring the sensitivity of the respective infrared probe. Finally, it is worth noting that the C≡N stretching band of these peptides measured in the aforementioned micelle/membrane environment is broader than that measured in THF, suggesting that the corresponding Trp sidechains still sample a rather heterogeneous environment, consistent with the fact that the membrane-bound indolicidin molecules lie at the interfacial region of the membrane and, as a result, the Trp residues are situated in a partially hydrated and dynamic environment.

5-Cyanotryptophan as a fluorescence quencher of p-cyano-phenylalanine. 5-Hydroxytryptophan, an analog of Trp_{CN}, has been widely used as a fluorescence probe of protein structure and dynamics. This is because the fluorescence quantum yield of 5-hydroxytryptophan is higher than that of tryptophan and is also much less sensitive to solvent polarity [363]. In light of the usefulness of Trp_{CN} as an infrared probe, it would be beneficial to further explore whether this chromophore exhibits similar fluorescent properties as 5-hydroxytryptophan. To our surprise, the acetylated and aminated form of Trp_{CN} (i.e., Acetyl-Trp_{CN}-NH_2) is only weakly fluorescent in aqueous solution (Fig. 8.5), making it less attractive to be used as a fluorescence reporter. However, the absorption spectrum of Trp_{CN} overlaps significantly with the

fluorescence emission spectrum of Phe_{CN} (Fig. 8.5), suggesting that it is a viable candidate to serve as a FRET acceptor to Phe_{CN}. Indeed, the fluorescence spectrum of an amidated tripeptide, Phe_{CN}-Ala-Trp_{CN}-NH_2, shows that the Phe_{CN} emission is almost completely quenched by Trp_{CN} (Fig. 8.5). Assuming that the mechanism of the observed fluorescence quenching can be described within the framework of Förster theory and that at 280 nm the molar absorptivity of Trp_{CN} is equal to that of Trp, we estimated the Förster radius of this pair to be 17.4 ± 1.0 Å. Previously, we have shown that Phe_{CN} is not only a useful fluorescent probe [191, 369, 370], but can also be used together with Trp for FRET applications where Trp serves as the acceptor [191, 195]. Compared to the Phe_{CN}-Trp FRET pair, whose fluorescence spectra overlap, the Phe_{CN}-Trp_{CN} FRET system is more convenient to use. This is because their fluorescence emissions are well separated and thus, in conjunction with the low fluorescence quantum yield of Trp_{CN}, making the extraction of the donor fluorescence intensity from a measured FRET spectrum straightforward, a feature particularly useful in time-resolved FRET measurements.

8.3 Conclusions

In summary, we demonstrate that the C≡N stretching vibration of Trp_{CN} is a useful infrared probe of local environment. Measurements involving binding of indolicidin to model membranes indicate that the bandwidth of this vibrational mode is partic-

ularly sensitive to the degree of hydration of the respective Trp sidechains. Thus, we believe that this infrared reporter is especially useful in probing biological processes, such as protein folding and binding, wherein a Trp residue is becoming more or less hydrated. In addition, we show that, similar to the Phe_{CN}-Trp FRET pair, Phe_{CN}-Trp_{CN} also constitutes a useful fluorescence donor-acceptor pair. The advantage of the Phe_{CN}-Trp_{CN} pair is that their fluorescence emissions are well separated, thus making decomposition of a FRET spectrum into its constituent donor and acceptor contributions straightforward.

8.4 Materials and Methods

Materials. 5-Cyanoindole (99%) was obtained from Aldrich (St. Louis, MO). Fmoc-L-Trp_{CN} (98%, 99% ee 2S) was purchased from RSP Amino Acids (Shirley, MA). Chloroform solutions of dodecylphosphocholine (DPC) and the sodium salt of 1-hexadecanoyl- 2-(9Z-octadecenoyl)- sn-glycero- 3-phospho- (10-sn-glycerol) (POPG) were purchased form Avanti Polar Lipids (Alabaster, AL). Millipore water was used to prepare aqueous solutions. Peptides were synthesized using standard Fmoc-protocols employing Rink resin. All peptide samples were purified to homogeneity by reverse-phase chromatography and characterized by mass-spectrometry. Following purification, peptide solutions were lyophilized and subsequently re-dissolved in H_2O and titrated to pH 7.0. Chloroform solutions of DPC and POPG were dried under a

stream of nitrogen, followed by dissication under vacuum to remove residual solvent. Dried DPC or POPG samples were re-dissolved in 20 mM phosphate buffer solution (pH 7.0) and then sonicated for about 10 min. Lyophilized peptides were then added to the resultant micelle or vesicle solution, where the total DPC or POPG concentration was 250 mM. The mixture was further sonicated for about 5 min and then used in infrared measurements involving micelle- or membrane-bound peptides. The concentration of free Trp_{CN} amino acid was estimated by weight, whereas that of the two indolicidin mutants was determined optically via the sample absorbance at 280 nm, which was measured on a PerkinElmer Lambda 25 UV-Vis spectrometer (Fremont, CA). The molar absorption coefficient of Trp_{CN} was assumed to be the same as that of Trp (i.e., $\epsilon(280\text{ nm}) = 5500\text{ M}^{-1}\text{ cm}^{-1}$).

FTIR spectroscopy. FTIR spectra were collected on a Nicolet Magna-IR 860 spectrometer using 1 cm^{-1} resolution. A CaF_2 sample cell that is divided into two compartments using a Teflon spacer (56 μm unless explicitly specified) was employed to allow the separate measurements of the sample and reference (i.e., buffer) under identical conditions. The temperature was maintained at 20 °C for all measurements. To correct for slow instrument drift, an automated translation stage was used to move both the sample and reference sides in and out of the IR beam alternately, and each time a single-beam spectrum corresponding to an average of 16 scans was collected. The final result was usually an average of 64-128 such spectra. The spectra shown

correspond to the net absorbance of the solutes, which were obtained by subtracting a linear background in the C≡N stretching spectral region.

Fluorescence and circular dichroism measurements. Fluorescence spectra were obtained at 20 °C on a Fluorolog 3.10 spectrofluorometer (Jobin Yvon Horiba, Edison, NJ) with a 1 cm^{-1} quartz sample holder. Temperature was regulated using a TLC 50 Peltier temperature controller (Quantum Northwest, Spokane, WA). An integration time of 1.5 s/nm was used in all measurements. Far-UV circular dichroism (CD) spectra were collected on an Aviv Model 410 CD spectrometer (Aviv Biomedical, NJ).

8.5 Acknowledgement

We gratefully acknowledge financial support from the National Institutes of Health (GM-065978 and RR-01348). M.M.W. further thanks Matthew Tucker, who collected the spectra of 5-cyanoindole, for fruitful collaboration.

8.6 Original Publication

This Chapter has been published: Matthias M. Waegele, Matthew J. Tucker, Feng Gai, *Chemical Physics Letters* (2009) 478, 249-253. DOI: 10.1016/j.cplett.2009.07.058

Figure 8.1: C≡N stretching bands of 5-cyanoindole in THF and a water/methanol mixture (95/5, v/v), as indicated.

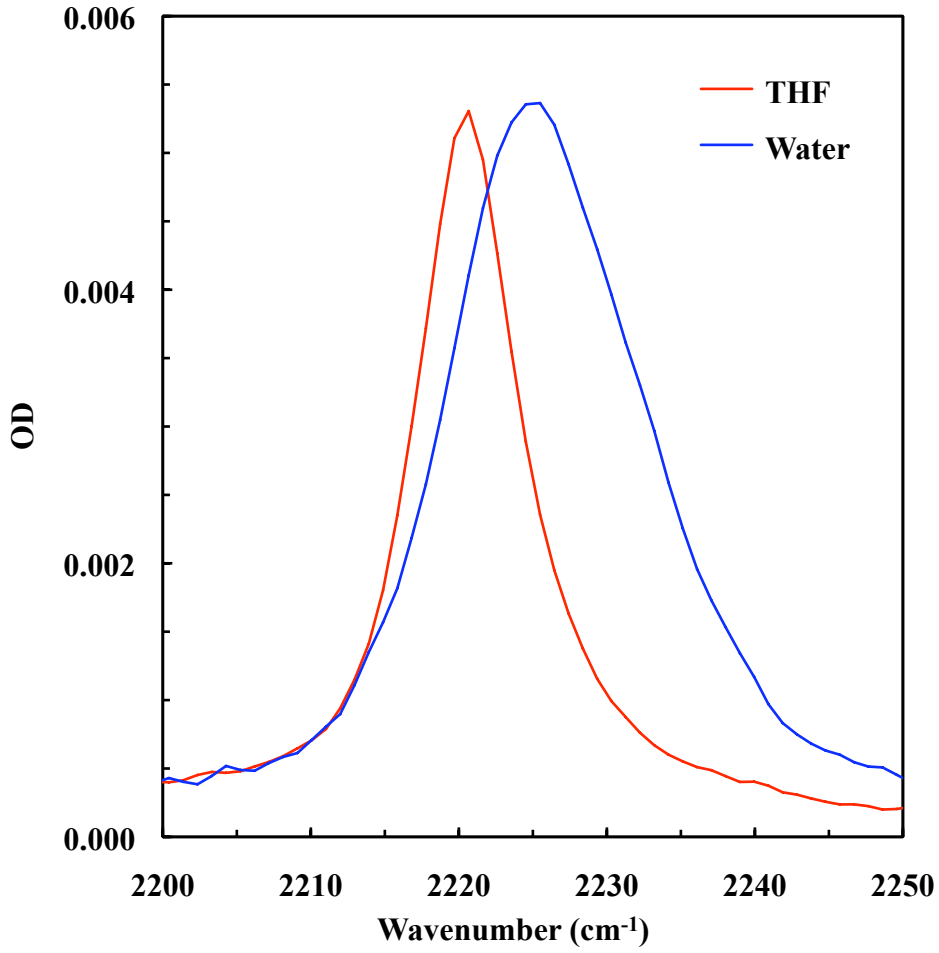

0.006
0.004
0.002
0.000
OD
THF
Water
2200
2210
2220
2230
2240
2250
Wavenumber (cm-1)

Figure 8.2: C≡N stretching bands of Fmoc-L-Trp_{CN} in THF and a water/THF mixture (60/40, v/v), as indicated. The concentration of Fmoc-L-Trp_{CN} was estimated to be 3.3 mM based on weight, and the optical pathlength was about 130 μm. For easy comparison, the spectrum measured in water/THF mixture has been scaled by a factor of 1.3.

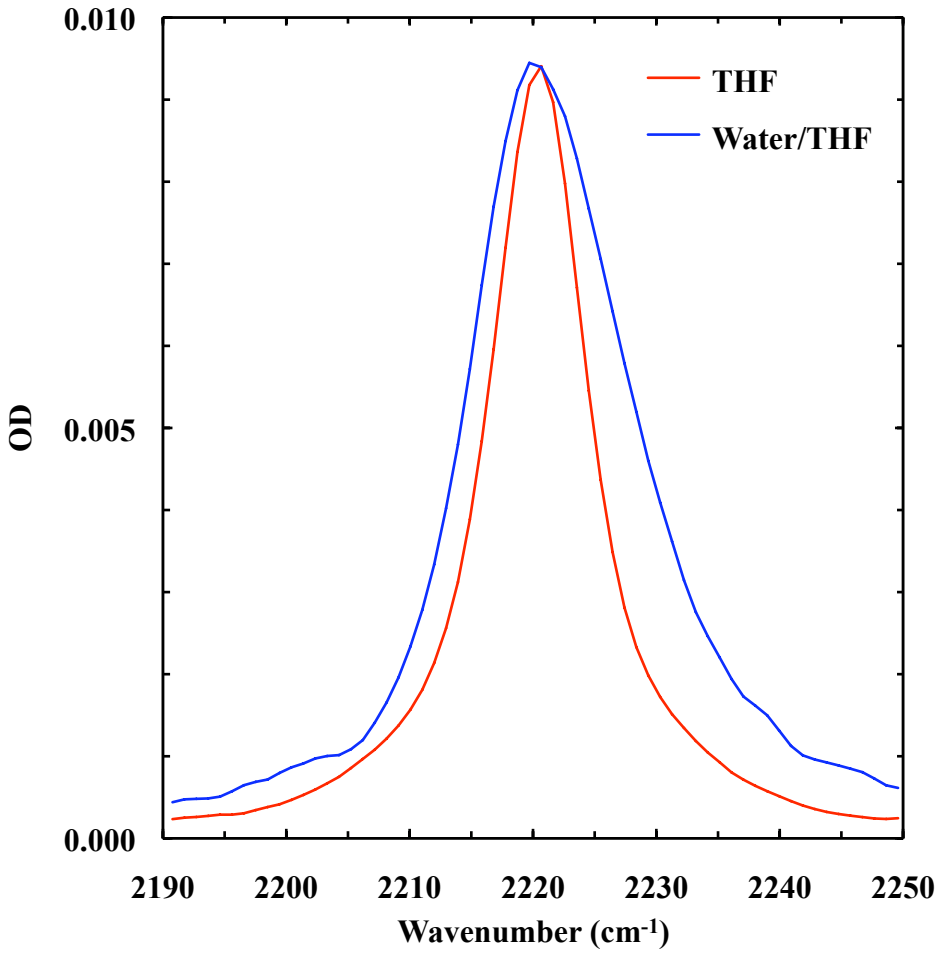

0.010
0.005
0.000
OD
THF
Water/THF
2190
2200
2210
2220
2230
2240
2250
Wavenumber (cm-1)

Figure 8.3: C≡N stretching bands of Trp9/Trp_{CN} in THF, water, POPG membranes and DPC micelles, as indicated. The peptide concentration was about 2 mM in all solvents, except water and POPG where the concentration was not determined. For easy comparison, the spectra obtained in water and DPC micelles have been scaled, by a factor of 2.2 and 0.62, respectively.

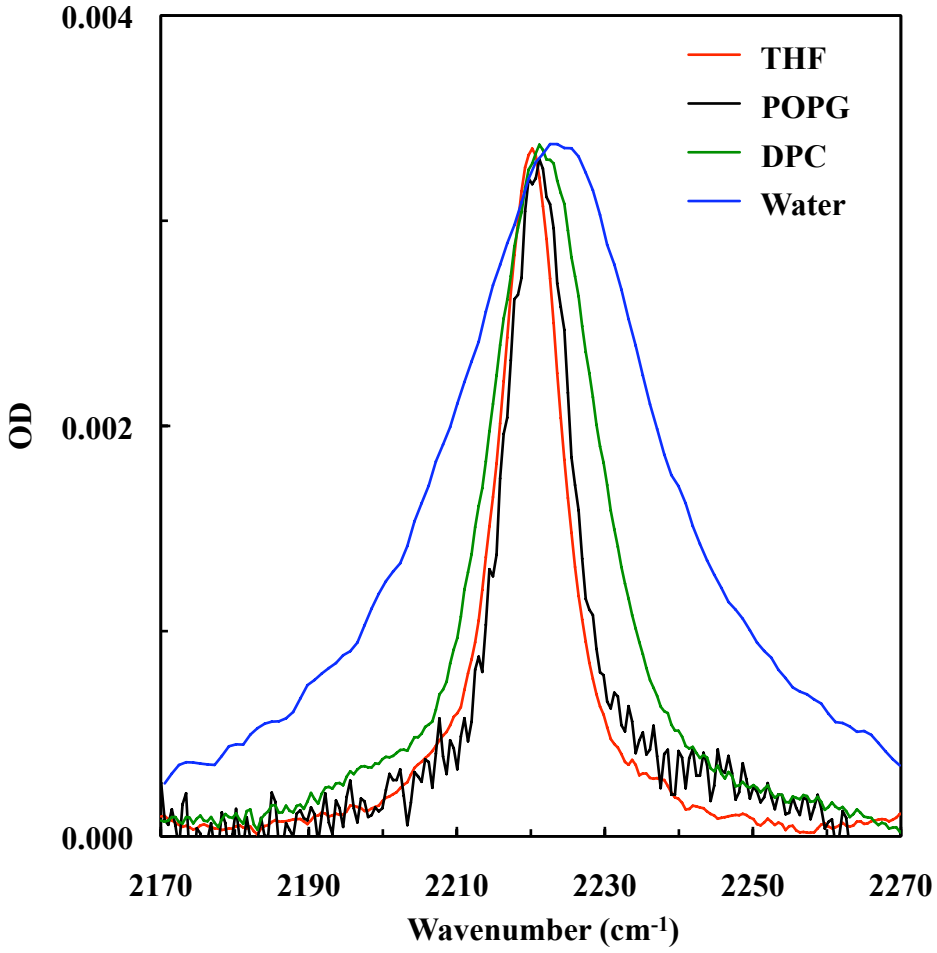

0.004
0.002
0.000
OD
2170
2190
2210
2230
2250
2270
Wavenumber (cm-1)
THF
POPG
DPC
Water

Figure 8.4: Comparison of the C≡N stretching bands of Trp11/Trp_{CN} (red) and Trp9/Trp_{CN} (blue) in DPC micelles. The Trp9/Trp_{CN} data are identical to those used in Fig. 8.3 and the Trp11/Trp_{CN} spectrum was scaled by a factor of 1.3.

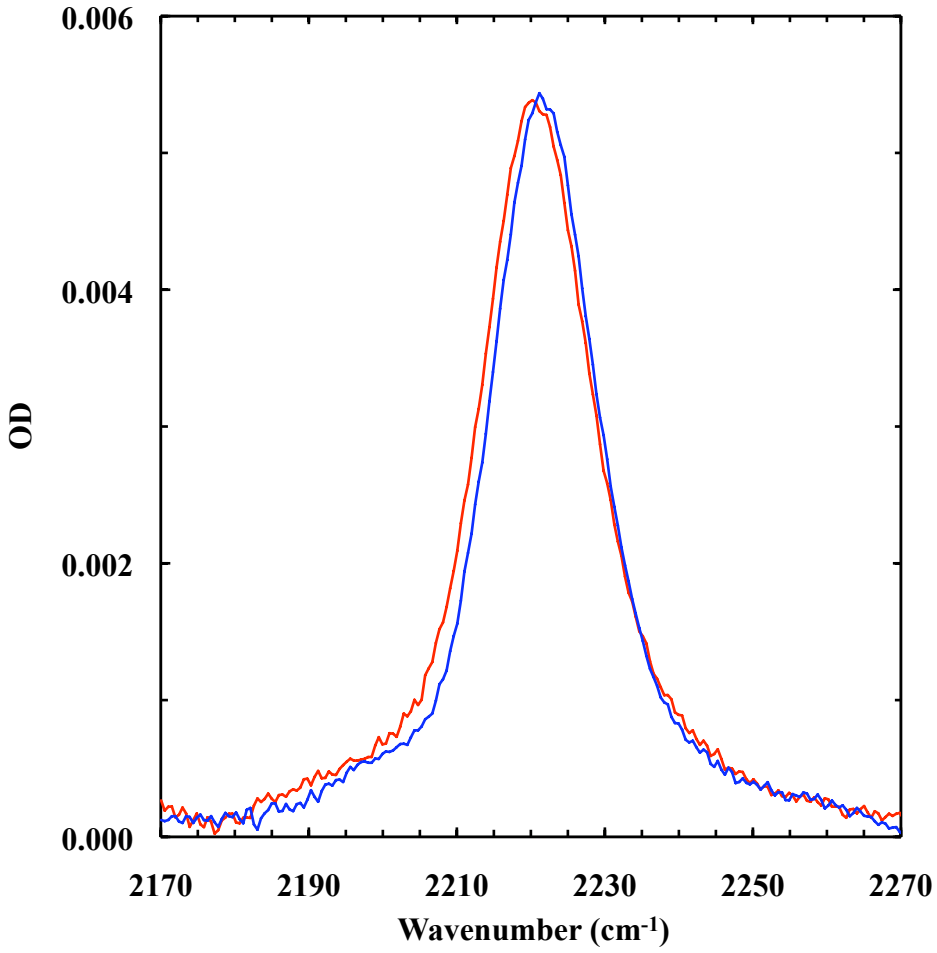
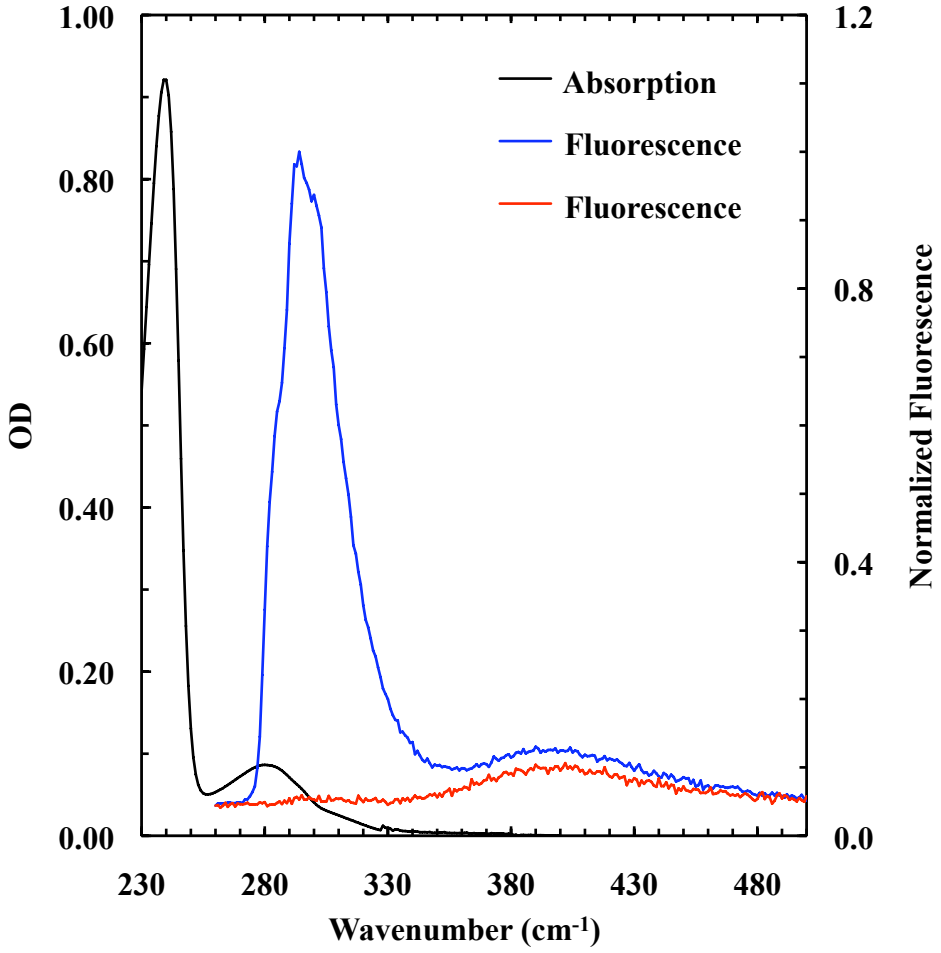

0.006
0.004
0.002
0.000
OD
2170
2190
2210
2230
2250
2270
Wavenumber (cm-1)

Figure 8.5: Absorption spectrum of Ac-Trp_{CN}-NH_2 in water (black), fluorescence spectrum of an equimolar ($\sim$ 20 μM) aqueous solution of Phe_{CN} and Ac-Trp_{CN}-NH_2 (blue), and fluorescence spectrum of Phe_{CN}-Ala-Trp_{CN}-NH_2 (red) in water ($\sim$ 16 μM). For fluorescence measurements, the excitation wavelength was 232 nm.


1.00
0.80
0.60
0.40
0.20
0.00
OD
1.2
0.8
0.4
0.0
Normalized Fluorescence
Absorption
Fluorescence
Fluorescence
230
280
330
380
430
480
Wavenumber (cm-1)

Chapter 9

Computational Modeling of the Nitrile Stretching Vibration of 5-Cyanoindole in Water

9.1 Introduction

It is well known that the C≡N stretching vibration of a nitrile group is sensitive to its environment [371, 138, 166, 143]. Thus, several nitrile-derivatized, non-natural amino acids, including *p*-cyanophenylalanine, cyanoalanine, and cyanylated cyteine, have been used as infrared (IR) reporters of local conformation and/or electrostatics of peptides [171, 353, 355, 358, 357] and proteins [178, 172, 179]. These applications are primarily based on the premise that the C≡N stretching frequency shifts upon a change in the immediate environment of the corresponding non-natural amino acid. However, recently we have shown that in the case of 5-cyanotryptophan it is the bandwidth of the C≡N stretching vibrational band, instead of its peak frequency, that is more sensitive to the hydration status of the IR reporter [176]. In fact, the bandwidth of the C≡N stretching vibrational band of 5-cyanoindole (Fig. 9.1), which is the sidechain of 5-cyanotryptophan, is increased from ∼8.7 cm^{-1} to ∼15.6 cm^{-1} when the solvent is changed from tetrahydrofuran (THF) to water, whereas the band frequency shows only a slight blue shift (∼4 cm^{-1}). In addition, the C≡N stretching band of 5-cyanoindole in water is also significantly broader than those of other model nitrile compounds (i.e., acetonitrile [371, 138, 171], *p*-tolunitrile [171], and methyl cyanate [151, 159]. Given the range of lifetimes (∼1-5 ps) measured for the C≡N stretching vibration of several related nitrile compounds [146, 147], it seems unlikely that lifetime broadening plays a dominant role in the broadening of the C≡N

stretching band of 5-cyanoindole. Instead, it seems more likely that the broadening is caused by the heterogeneous interactions between the bulky indole ring and water molecules. Such a picture is supported by the studies of the groups of Eaton, Reimers and Cho, which showed that even for acetonitrile (CH_3CN) the C≡N stretching frequency is determined by the precise configuration of water molecules surrounding the nitrile group [371, 138, 159, 160]. For example, a linear arrangement of CH_3CN–HOH (termed σ-bonding with nitrogen lone-pair) results in a blue-shift of the C≡N stretching frequency as compared to its gas phase value, whereas interactions of water molecules with the π-orbital of the C≡N group (termed π-bonding) lead to a red-shift. However, for 5-cyanoindole, besides those interactions between water molecules and the C≡N group, one might expect that the solvation status of the parent molecule, which is comparatively much larger than a methyl group and also contains an extended π-electron system that is electronically coupled to the triple bond of the nitrile group, also plays an important role in determining the peak frequency and band shape of the C≡N stretching vibration. In order to provide a more quantitative description, herein we carry out a theoretical investigation of how interactions of water molecules with 5-cyanoindole influence its C≡N stretching vibration.

9.2 Results and Discussion

The details of the frequency calculations, which are based on methods used in similar studies [166, 159, 160, 158, 93, 163, 79, 372, 373, 80, 374, 96, 81, 164, 165], are described in Section 9.4. Briefly, a molecular mechanics (MM) model of 5-cyanoindole was first developed, and subsequently used in a molecular dynamics (MD) simulation of 5-cyanoindole in water. The Born-Oppenheimer potentials of the C≡N stretch were then determined for a set of 5-cyanoindole-water clusters, which were randomly selected from a MD trajectory. These potentials were then fitted to a Morse potential function in order to obtain their quantum mechanical (QM) vibrational frequencies. Finally, an empirical model, which relates the C≡N stretching frequency to its electrostatic environment, was developed. Specifically, the atomic charges of 5-cyanoindole (Fig. 9.1) were parameterized by achieving a satisfactory agreement between the MM and QM solvent-solute interaction energies. To do so, the internal parameters of 5-cyanoindole (i.e., bond-, angle-, and dihedral-force constants, as well as Urey-Bradley parameters) were based on those of its structural analogs, indole [375] and 3-cyanopyridine [169]. Its bond lengths were determined based on the MPW1B95/6-31++G(d,p) optimized 5-cyanoindole molecule, while its Lennard-Jones parameters were taken from those of indole [375] and acetonitrile [170]. Using the Gaussian 03 software package [376], the initial MM atomic charges were then obtained via a CHelpG [377] analysis of the electron density of 5-cyanoindole at the MPW1B95/6-

31++G(d,p) level of theory, under the constraint that the atomic charges are to reproduce the QM permanent dipole moment of the molecule. The results are summarized in Table 9.1. To further optimize the initially obtained atomic charges so that the MM interaction energies agree with the QM interaction potentials, a fitting procedure using the following harmonic penalty function [378] was then carried out.

$$\chi^2 = \chi^2_{inter} + \chi^2_{penalty} = \sum_{\alpha}(\mathrm{IE}^{\mathrm{MM}}_{\alpha} - \mathrm{IE}^{\mathrm{QM}}_{\alpha})^2 + \rho\sum_{i}(q^{\mathrm{MM}}_{i} - q^{\mathrm{QM}}_{i})^2 \qquad (9.1)$$

Where $\mathrm{IE}^{\mathrm{QM}}_{\alpha}$ and $\mathrm{IE}^{\mathrm{MM}}_{\alpha}$ represent the MM and QM interaction energies at geometry α of the interaction potential and q^{MM}_{i} and q^{QM}_{i} are the MM charge of atom i and the corresponding CHelpG reference charge, respectively. The weighting factor, ρ, of the penalty function was set to a value of 15. Three QM interaction potentials, defined as IE=E(complex)-E(5-cyanoindole)-E(water), were calculated at the MPW1B95/6-311++G(3df,2dp) level of theory. Specifically, the optimization was achieved by considering the following three interaction potentials: (A) A water molecule was placed 2.0 Å away from the nitrile nitrogen with one H-O bond pointing towards the C≡N bond and the other pointing out of the indole ring plane (Fig. 9.2d). The C≡N–H angle was then stepped in 10° increments, moving the water out of the indole ring plane. (B) A water molecule was placed 2.0 Å away from the nitrile nitrogen, within the plane defined by the indole ring (Fig. 9.2e) and with one H-O bond pointing towards the C≡N bond. The C≡N–H angle was then stepped in 10° increments, moving the water within the indole ring plane. (C) A water molecule was placed 1.6 Å

from the nitrile nitrogen with one H-O bond pointing towards the C≡N bond and the other pointing out of the indole ring plane. The C≡N–HOH distance was than stepped in 0.05 Å increments from 1.60 to 2.65 Å. Once the charges on the water molecule (O = -0.834 e; H = 0.417 e) and the Lennard-Jones term are given, the Coulomb part of the interaction potential can be optimized by adjusting the MM atomic charges of 5-cyanoindole via minimization of the χ^2 (Eq. 9.1). As shown (Fig. 9.2), the new MM interaction energies of the optimized force field match the QM interaction potential significantly better than the CHelpG based force field energies, while the optimized charges deviate only slightly from those obtained via QM calculations (Table 9.1). Thus, this optimization procedure (i.e., Eq. 9.1) produced a set of charges that is expected to better describe the interactions between water and the C≡N moiety, a pivotal requirement for the correct prediction of its stretching frequency shift upon hydration [159, 160, 164, 165]. Having developed a MM model of 5-cyanoindole, we next sought to establish an empirical model that allows us to determine the localized C≡N stretching frequency in water from MD simulations based on the notion that the vibrational frequency of a localized vibrational mode is related to the local electric field (or potential) [159, 160, 158, 93, 163, 79, 372, 373, 80, 374, 96, 81]. Specifically, we employed a computational procedure that is similar to those used by Skinner and coworkers [163, 79, 372] as well as by Corcelli and coworkers [164, 165]. First, we randomly selected 40 5-cyanoindole-water clusters from a 1 ns MD simulation at

300 K and 1 atm (details of the NAMD simulation [379] are given in Section 9.4). As shown (Fig. 9.3), these clusters contain 5-17 water molecules (12 on average) and were further divided into two subsets, Set-A and Set-B. For Set-A, wherein the indole ring is considered dehydrated, only those water molecules that have at least one atom located within 4.5 Å of the NN1 atom of the indole ring are included in each cluster. On the other hand, the clusters in Set-B are composed of water molecules that have at least one atom located within 3.5 Å of the NN1, NE1, or CD2 atom of the indole ring (atom names are given in Fig. 9.1). Thus, the indole ring in Set-B is regarded as partially hydrated. In the following we will consider Set-A and Set-B as a common set. Second, DFT calculations were carried out on these clusters using the Zhao-Truhlar hybrid meta functional MPW1B95 [380] in conjunction with the 6-31++G(d,p) basis set. It has been shown that this level of theory adequately describes the weak interactions between water and the indole ring [381]. For each cluster, the potential energy of the system as a function of the C≡N bond length (from 1.0 to 1.4 Å in increments of 0.05 Å) was determined via DFT calculations wherein all other atomic coordinates were fixed. Third, the obtained potential energy curves were then fitted to a Morse oscillator model, allowing analytical determination of the underlying anharmonic frequencies. Fourth, an empirical model relating the vibrational frequency to the electrostatic field components on the nitrile group was obtained by least-square fitting the following equation to the *ab initio* frequencies:

$\omega_{\text{emp}} = \sum_{a,i} c_i^a E_i^a + \omega_{\text{gas}}$, where E_i^a is the Cartesian electric field component i at a site a on the molecule, c_i^a is a constant of proportionality, and ω_{gas} corresponds to the gas phase value of the C≡N stretching frequency of 5-cyanoindole, which was calculated to be 2392.3 cm^{-1} using an isolated 5-cyanoindole molecule and the procedure described above. Finally, the electric fields arising from the partial charges of water were calculated using an in-house written code within the VMD software environment [382], and the proportionality constants (c_i^a) were determined by the multivariate least-square fitting method. As shown (Fig. 9.4a), fitting the 40 *ab initio* frequencies obtained from the vibrational analysis of the clusters with a model that only includes the electric field components on the nitrile group in the direction of the vibrational transition dipole, i.e.,

$$\omega_{\text{emp}} = c_z^{\text{CN1}} E_z^{\text{CN1}} + c_z^{\text{NN1}} E_z^{\text{NN1}} + \omega_{\text{gas}} \tag{9.2}$$

yields a relatively poor correlation coefficient (R^2) of 0.64 and a root mean square deviation (rmsd) of 7.6 cm^{-1} between empirical and *ab initio* frequencies. However, when the model was refined to include the z-component of the electric field at the center of the six-membered ring, E_z^{NN1}, i.e.,

$$\omega_{\text{emp}} = c_z^{\text{CN1}} E_z^{\text{CN1}} + c_z^{\text{NN1}} E_z^{\text{NN1}} + c_z^{\text{6m}} E_z^{\text{6m}} + \omega_{\text{gas}} \tag{9.3}$$

the resultant correlation (Fig. 9.4b) is significantly improved, yielding a correlation coefficient of 0.81 and a rmsd of 5.8 cm^{-1} with c_z^{CN1} = 4554.4 cm^{-1}/a.u., c_z^{NN1} = -1652.2 cm^{-1}/a.u., and c_z^{6m} = 1989.7 cm^{-1}/a.u. Taken together, these results indicate

that the electric field component at a point on the ring but distant from the nitrile group could have a significant effect on its vibrational frequency. While a model including only the electric field localized at the C≡N group could also yield a good correlation, it requires at least six electric field components to produce a comparable correlation (Table 9.2). Thus, these findings show that by considering the electric field away from the C≡N group it is possible to fit the *ab initio* frequencies with a minimum number of parameters.

To further verify that the indole ring indeed plays a significant role in mediating the stretching frequency of the nitrile group, we have investigated the specific effect of one or two water molecules placed at different positions with respect to the ring, as shown in Fig. 9.5. As expected, we found that the C≡N stretching frequency of all three clusters shows a significant shift (i.e., $\Delta\omega = \omega_{\text{cluster}} - \omega_{\text{gas}} \neq 0$) from that of the unsolvated ring with $\Delta\omega$ = 3.9, 8.8 , and -4.0 cm^{-1} for (a), (b) and (c), respectively. Thus, these results, which indicate that $\Delta\omega$ is sensitive to the position and orientation of the water molecule(s), corroborate the notion that the solvent-indole interaction is a significant determinant of the C≡N stretching frequency. Having determined a set of parameters (c_i^a), Eq. 9.3 was then used to calculate the C≡N stretching frequency of individual snapshots along the 1 ns MD trajectory of 5-cyanoindole in water. The resultant frequencies show a Gaussian distribution with a standard deviation of 13.5 cm^{-1} and a mean that is -9.7 cm^{-1} red-shifted with respect to ω_{gas}.

The latter suggests that the current model correctly predicts the trend of the frequency shift of the C≡N stretching vibration upon hydration as the nitrile stretching frequency is typically red-shifted with respect to the gas-phase value. For example, the C≡N stretching frequency of benzonitrile is ∼2244 cm^{-1} in the gas-phase [383], which shifts to ∼2235 cm^{-1} in water [146]. Unfortunately, for the present case a more quantitative assessment of the predicted shift cannot be made at this time because an experimental value for ω_{gas} for 5-cyanoindole is currently not available.

Finally, we calculated the IR band of the C≡N stretching vibration from the frequency-frequency correlation function, $M(t) = \langle \delta\Omega(t)\delta\Omega(0)\rangle$ with $\delta\Omega(t) = 2\pi c\{\delta\tilde{\nu}(t) - \langle\delta\tilde{\nu}(0)\rangle\}$. As shown (Fig. 9.6), the (normalized) autocorrelation function decays rapidly in a bimodal fashion, similar to that observed for other nitrile compounds [159, 146, 147, 165]. The vibrational pure dephasing constant, $T_2^* = 1/\int_0^\infty \mathrm{d}t\, M(t)$, was estimated to be 0.25 ps. The variance, $\Delta \equiv [M(0)]^{1/2}$, and correlation time, $\tau \equiv \int_0^\infty \mathrm{d}t\, M(t)/M(0)$, of the C≡N stretching frequency fluctuations of 5-cyanoindole were found to be 2.53 ps^{-1} and 0.61 ps, respectively. Thus, the spectrum is neither in the homogeneous nor in the inhomogeneous regime. Therefore, the vibrational spectrum was calculated using a semiclassical approximation [98, 101, 90] to $I(\omega) \sim \int_{-\infty}^{\infty} \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t}\, \langle \mu(t)\cdot\mu(0)\rangle$, where μ is the quantum mechanical dipole operator (see Section 9.4 for details). As shown (Fig. 9.7), the bandwidth of the simulated spectrum is in very good agreement with that determined experimentally. However,

this agreement must be cautiously interpreted. First, the vibrational lifetime of 5-cyanoindole is currently not available. Thus, we have assumed a vibrational lifetime of 5 ps, which is a typical value for the vibrationally excited state of nitriles [147]. Second, the production run of the simulation was carried out in the NPT ensemble, which may adversely affect the quality of the frequency-frequency time correlation function. Moreover, it is worth noting that the present electrostatic map may not be applicable to 5-cyanotryptophan, especially when it is located in a heterogeneous environment of proteins or membranes, as this map incorporates interactions that are specific to water. Although such "solvent-specific" maps have been used within the context of peptides [91], alternative approaches such as those developed by the groups of Corcelli [164, 165], Mukamel [80, 373, 374], Jansen [81] and Kurnikova [384] may be more appropriate.

9.3 Conclusions

In conclusion, we demonstrate that interactions between water and the indole ring make an important contribution to the broadening of the C≡N stretching band of 5-cyanoindole in aqueous solution. Thus, this study provides a molecular basis to support the notion that 5-cyanotryptophan is a useful IR probe of local hydration status of proteins, and also indicates that the chemical moiety to which the C≡N group is attached may need to be considered in the development of electrostatic

maps for calculation of the C≡N stretching frequencies.

9.4 Methods

Molecular dynamics simulation. After immersion of 5-cyanoindole in a box of 727 TIP3P waters [385], the energy of the system was first minimized for 2000 steps and then the system was allowed to equilibrate for 1 ns at 300 K and 1 atm. After equilibration, a 1 ns production run was carried out in the NPT ensemble under identical conditions. While the Berendsen barostat was used to relax the system towards equilibrium, the pressure in the production run was maintained at 1 atm using the Nosé-Hoover Langevin piston method. Temperature was controlled by Langevin dynamics; and periodic boundary conditions were used to reduce edge effects. The SHAKE algorithm was employed to constrain all bonds involving the hydrogen atom. A cut-off of 12 Å was used for nonbonded interactions, which were switched to zero between 10 and 12 Å. Full electrostatics were calculated every second step using the particle-mesh Ewald (PME) method. A 2 fs time step was used to integrate the equations of motion and coordinates were saved every 4 fs for analysis.

Electronic structure calculation. Since 5-cyanoindole was not treated as a rigid body in the MD simulation, it undergoes structural fluctuations due to the finite temperature of the system. Such thermally induced fluctuations could modulate the frequency of the nitrile stretching vibration and also lead to uncertainties in the def-

inition of the molecular coordinate system. Thus, in order to eliminate this thermal noise, the structurally distorted molecules in the 5-cyanoindole-water clusters from MD simulation were replaced by the optimized structure of 5-cyanoindole. The C≡N bond of the geometrically optimized molecule was then stretched in a step-wise manner (from 1.0 to 1.4 Å in 0.05 Å steps) by moving both the nitrile carbon and nitrogen, while keeping the center of mass of the nitrile group fixed, allowing determination of the potential energy of the 5-cyanoindole-water clusters as a function of the nitrile bond length (e.g., the one in Fig. 9.8). For all electronic structure calculations, an ultrafine numerical integration grid having 99 radial shells and 590 angular points per shell was used. Tight SCF convergence criteria as defined in the Gaussian 03 software were used for all single point energy calculations. To determine the anharmonic frequency of the C≡N stretching vibration, those point energies obtained above were then fit to a Morse oscillator function:

$$V(r) = D(1 - \exp[-\alpha(r - r_0)]) + V_0 \tag{9.4}$$

where D is the bond dissociation energy, α is a constant that determines the width of the potential well, r_0 is the equilibrium bond length, and V_0 is the equilibrium potential energy of the system. By stretching the C≡N bond of the isolated 5-cyanoindole molecule from 1.0 to 12.5 Å and fitting the points to the above equation, the bond dissociation energy D was estimated to be 326 kcal/mol. The ground and

first excited vibrational state energies were then calculated according to:

$$E_n = DB\left(n + \frac{1}{2}\right)\left[2 - B\left(n + \frac{1}{2}\right)\right] \tag{9.5}$$

where $n = 0$ or 1 for the ground and first excited state, respectively. B is given by

$$B = \frac{\alpha\hbar}{\sqrt{2\mu D}} \tag{9.6}$$

where μ is the reduced mass of the nitrile group, i.e. $(12 \times 14)/(12 + 14)$ amu.

Band shape calculation. The electric field on site i of 5-cyanoindole arising from n water molecules was calculated according to:

$$\vec{\mathbf{E}}_i = \sum_{j=1}^{3n} \frac{q_j}{r_{ij}^2}\hat{\mathbf{r}}_{ij} \tag{9.7}$$

Index j runs over the three atomic sites of the TIP3P water molecule, r_{ij} corresponds to the distance between the ith site on 5-cyanoindole and the jth site on TIP3P water. is the unit vector pointing from atomic site j to site i. The frequency trajectory was obtained by including all water molecules in the box that were within 12 Å of any atom of the 5-cyanoindole molecule, which was placed in the center of the box. Within the framework of a semiclassical picture, the band shape of the C≡N stretching vibration was then calculated using the following equation:

$$I(\omega) \sim \int_{-\infty}^{\infty} \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t}\exp[-g(t) - |t|/2T_1] \tag{9.8}$$

where T_1 is the vibrational lifetime and $g(t)$ is the line shape broadening function

given by

$$g(t) = \frac{1}{\pi}\int_0^\infty \mathrm{d}\omega \coth\left(\frac{\hbar\omega}{2k_BT}\right)\tilde{\xi}_I(\omega)\frac{1-\cos(\omega t)}{\omega^2} + \frac{\mathrm{i}}{\pi}\int_0^\infty \mathrm{d}\omega\,\tilde{\xi}_I(\omega)\frac{\sin(\omega t)-\omega t}{\omega^2} \quad (9.9)$$

Within a semiclassical picture, $\tilde{\xi}_I(\omega)$ may be approximated by

$$\tilde{\xi}_I(\omega) = M(0)\rho(\omega) \quad (9.10)$$

where $\rho(\omega) \equiv \tanh(\hbar\omega/2k_BT)\tilde{M}_R(\omega)$ is the spectral distribution function, which is shown in Fig. 9.9. $\tilde{M}_R(\omega)$ is the real part of the Fourier transform of the normalized frequency autocorrelation function $M(t)$:

$$\tilde{M}_R(\omega) \equiv 2\int_0^\infty \mathrm{d}t\cos(\omega t)[M(t)/M(0)] \quad (9.11)$$

For the numerical evaluation of Eq. 9.9, spline interpolation on the first 2000 points (8 ps) of the frequency-frequency correlation function obtained from the MD trajectory was performed to generate 800,000 data points. Eq. 9.9 was integrated with a step size 1 cm^{-1}. The upper integration limit was 1042.4 cm^{-1}.

9.5 Acknowledgement

We gratefully acknowledge financial support from the National Institutes of Health (GM-065978 and RR-01348) and also the Computer Facility of the Department of Chemistry at the University of Pennsylvania. We thank Professor Jeffery G. Saven for helpful discussions and comments, and Christopher MacDermaid for his assistance in setting up the computer system.

9.6 Original Publication

This Chapter has been published: Matthias M. Waegele and Feng Gai, *Journal of Physical Chemistry Letters* (2010) 1, 781-786. DOI: 10.1021/jz900429z

Table 9.1: Charges derived from a CHelpG analysis of the electron charge density (Old MM) and those obtained from fitting of the QM interaction potentials (New MM). Atom names are defined in Fig. 9.1.

Atom name	New MM	Old MM
CH2	-0.1839	-0.142679
CZ2	-0.2755	-0.212108
CE2	0.1938	0.207008
CZ3	0.0216	-0.007346
CE3	-0.2753	-0.286345
NE1	-0.4721	-0.470295
CD1	0.0533	0.028534
CG	-0.3405	-0.364845
CD2	0.2657	0.252653
HG	0.1980	0.163512
HD1	0.1685	0.128804
HE1	0.3661	0.375028
HE3	0.1202	0.166148
CN1	0.4616	0.406976
NN1	-0.4772	-0.511574
HH2	0.1535	0.129601
HZ2	0.0222	0.136927

Table 9.2: Various possible empirical relationships, $\omega_{\mathrm{emp}} = \sum_{a,i} c_i^a E_i^a + \omega_{\mathrm{gas}}$, and the corresponding correlation coefficients R^2 and root mean square deviations (rmsd) from the *ab initio* frequencies. For the x-components, the magnitude of the field was used due to the plane of symmetry in the indole ring plane. Entries $E_z^{\mathrm{CN1-NN1}}$ and $E_z^{\mathrm{CZ3-CN1}}$ refer to the z-components of the fields at the center of the respective bonds.

Model	Electric Field Components Used	R^2	rmsd (cm^{-1})
1	E_z^{CN1}, E_z^{NN1},E_z^{6m}	0.81	5.8
2	E_z^{CN1},E_z^{NN1}	0.64	7.6
3	E_z^{CN1}, E_z^{NN1}, E_y^{CN1}, E_y^{NN1}	0.71	6.9
4	E_z^{CN1}, E_z^{NN1}, E_y^{CN1}, E_y^{NN1}, $\lvert E_x^{CN1}\rvert$, $\lvert E_x^{NN1}\rvert$	0.81	5.7
5	E_z^{CN1}, E_z^{NN1}, $E_z^{CN1-NN1}$	0.63	7.6
6	E_z^{CN1}, E_z^{NN1}, $E_z^{CZ3-CN1}$	0.67	7.5
7	E_z^{CN1}, E_z^{NN1}, E_z^{CZ3}	0.74	6.9

Figure 9.1: Atom names in 5-cyanoindole and definition of molecular coordinate system. The x unit vector (not shown) is pointing out of the molecular plane, towards the reader.

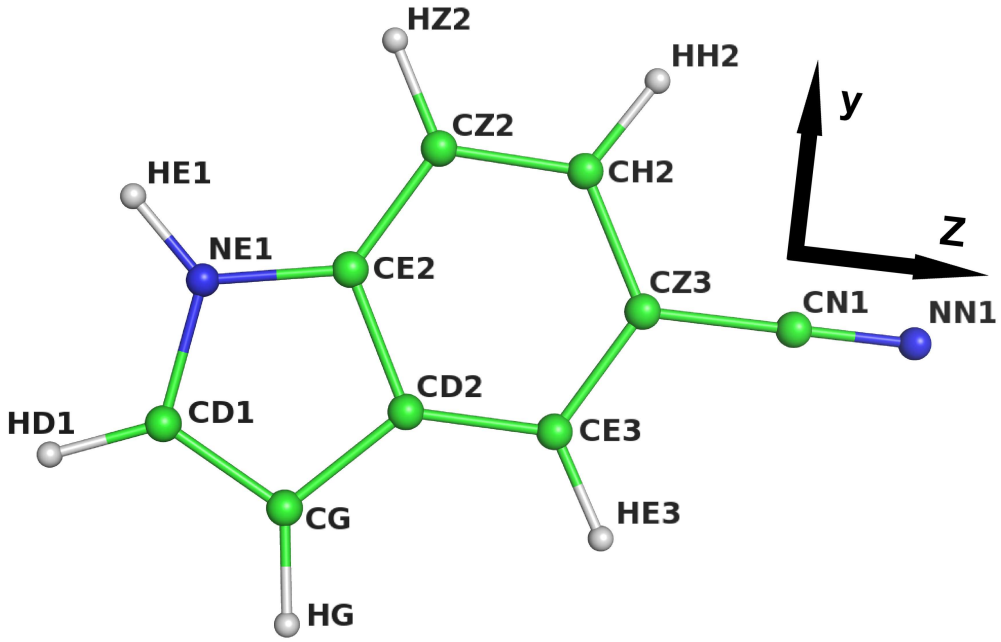
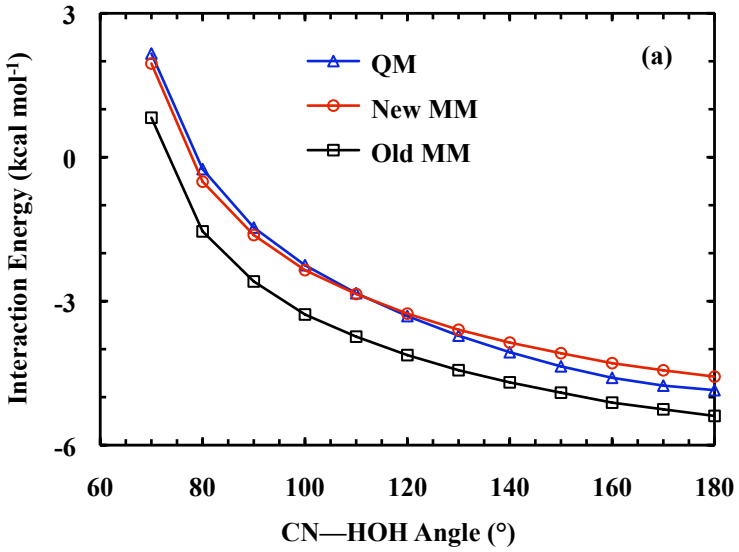
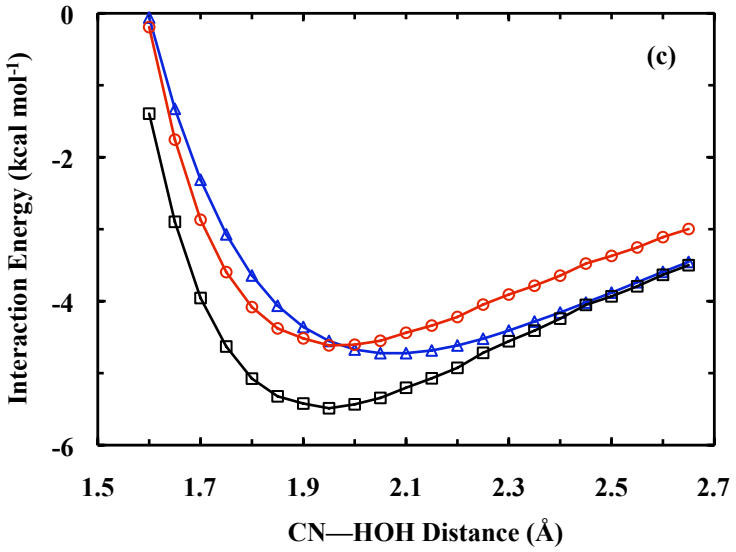
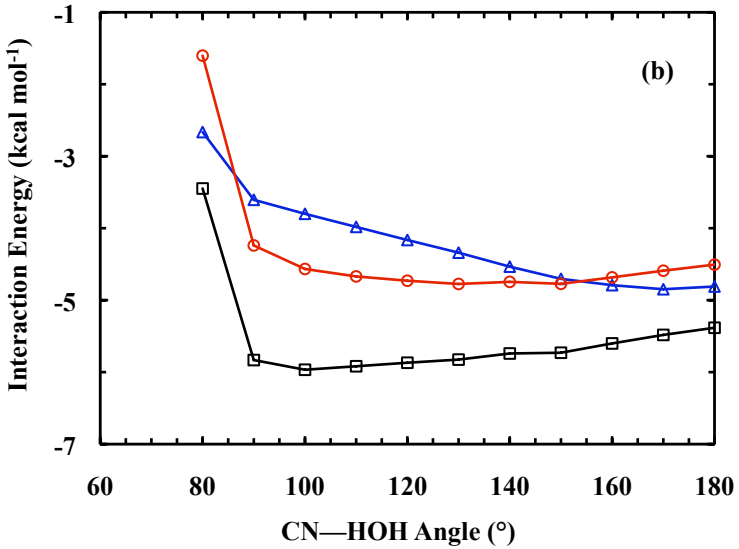

HZ2
HH2
y
CZ2
HE1
CH2
Z
NE1
CE2
CZ3
CN1
NN1
CD2
HD1
CD1
CE3
HE3
CG
HG

Figure 9.2: (a) QM and MM interaction energies between water and 5-cyanoindole as a function of the CN–H angle when the water is bent out of the indole plane (b) and when the water is kept in the indole ring plane. (c) QM and MM interaction energies as a function of the CN–HOH distance with one of the H-O bonds aligned with the nitrile group. Example geometries (CN–H angle = 130°) corresponding to the scans in (a) and (b) are shown in panels (d) and (e), respectively.


(a)
QM
New MM
Old MM
Interaction Energy (kcal mol-1)
CN—HOH Angle (°)
3
0
-3
-6
60
80
100
120
140
160
180


(c)
Interaction Energy (kcal mol-1)
CN—HOH Distance (Å)
0
-2
-4
-6
1.5
1.7
1.9
2.1
2.3
2.5
2.7


(b)
Interaction Energy (kcal mol-1)
CN—HOH Angle (°)
-1
-3
-5
-7
60
80
100
120
140
160
180

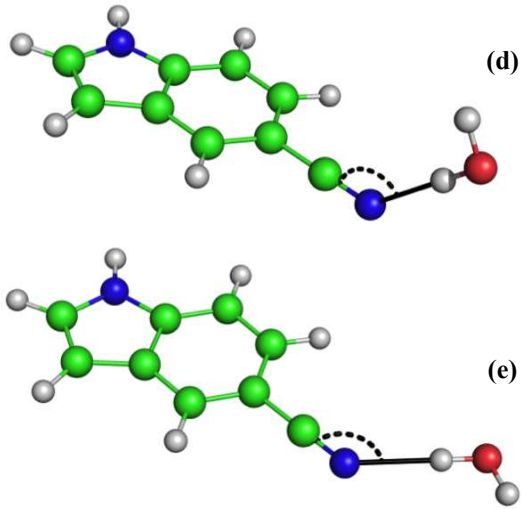
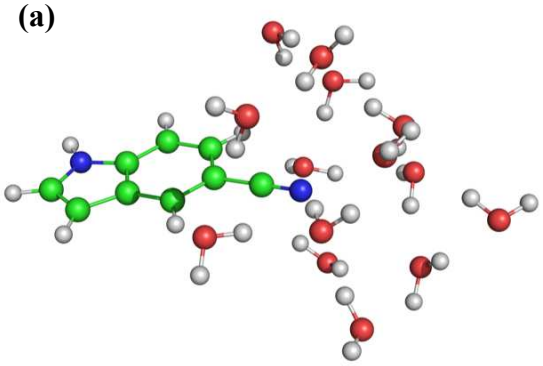

(d)
(e)

Figure 9.3: Representative snapshots of clusters in Set-A (a) and Set-B (b), respectively.

(a)



(b)

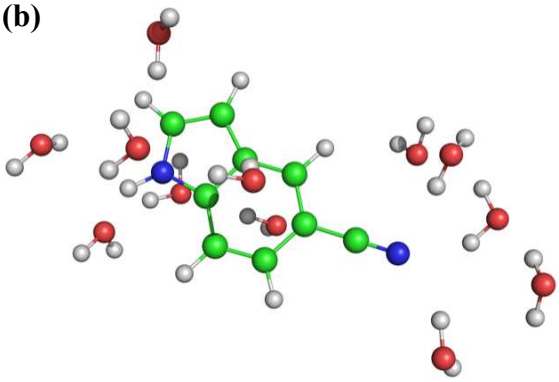


Figure 9.4: *Ab initio* frequency versus the empirical frequency determined from (a) the two-component and (b) the three-component empirical models. The correlation coefficients (R^2) are 0.64 and 0.81 for the data presented in (a) and (b), respectively.

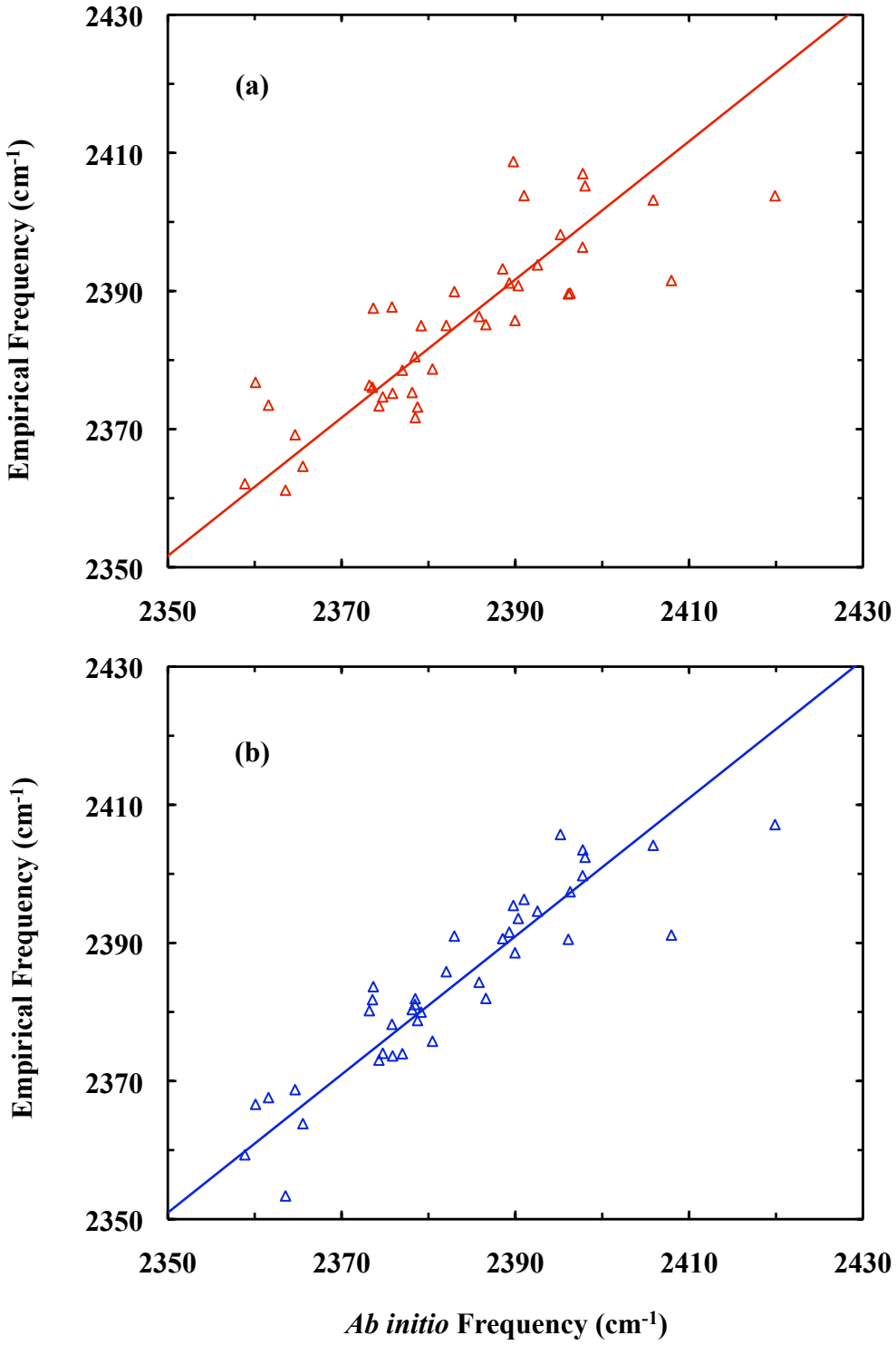
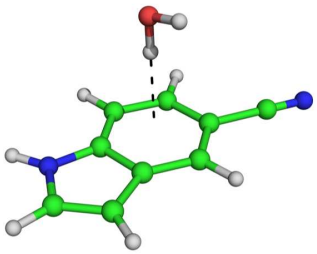
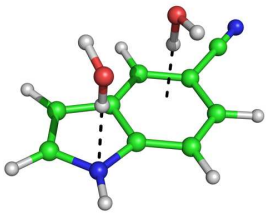

(a)
(b)
Empirical Frequency (cm-1)
Empirical Frequency (cm-1)
Ab initio Frequency (cm-1)
2350
2370
2390
2410
2430

Figure 9.5: Three model 5-cyanoindole-water clusters. In (a) and (b), the distance between the closest water hydrogen and the indole ring plane is 2.4 Å, whereas in (c) the H_2O–HN distance is 2.0 Å.

(a)



(b)



(c)

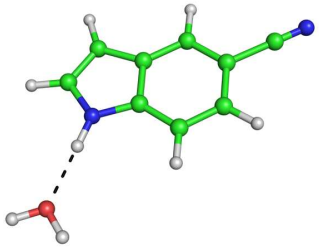


Figure 9.6: Normalized frequency-frequency correlation function, calculated from the C≡N frequency trajectory. The first 600 fs of the function are shown in the inset.

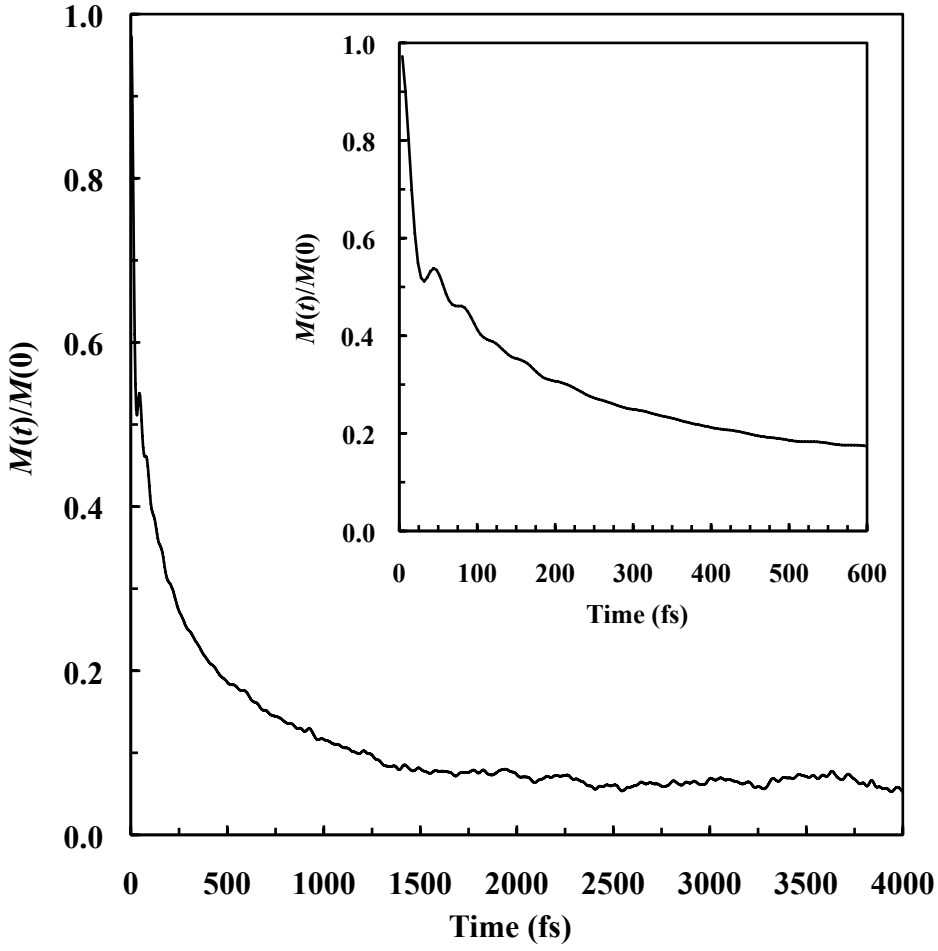

1.0
0.8
0.6
0.4
0.2
0.0
M(t)/M(0)
0
500
1000
1500
2000
2500
3000
3500
4000
Time (fs)
1.0
0.8
0.6
0.4
0.2
0.0
M(t)/M(0)
0
100
200
300
400
500
600
Time (fs)

Figure 9.7: Comparison between the experimentally measured (red) and simulated (blue) C≡N stretching bands of 5-cyanoindole. In the calculation, the lifetime of the C≡N stretching vibration was assumed to be 5 ps.

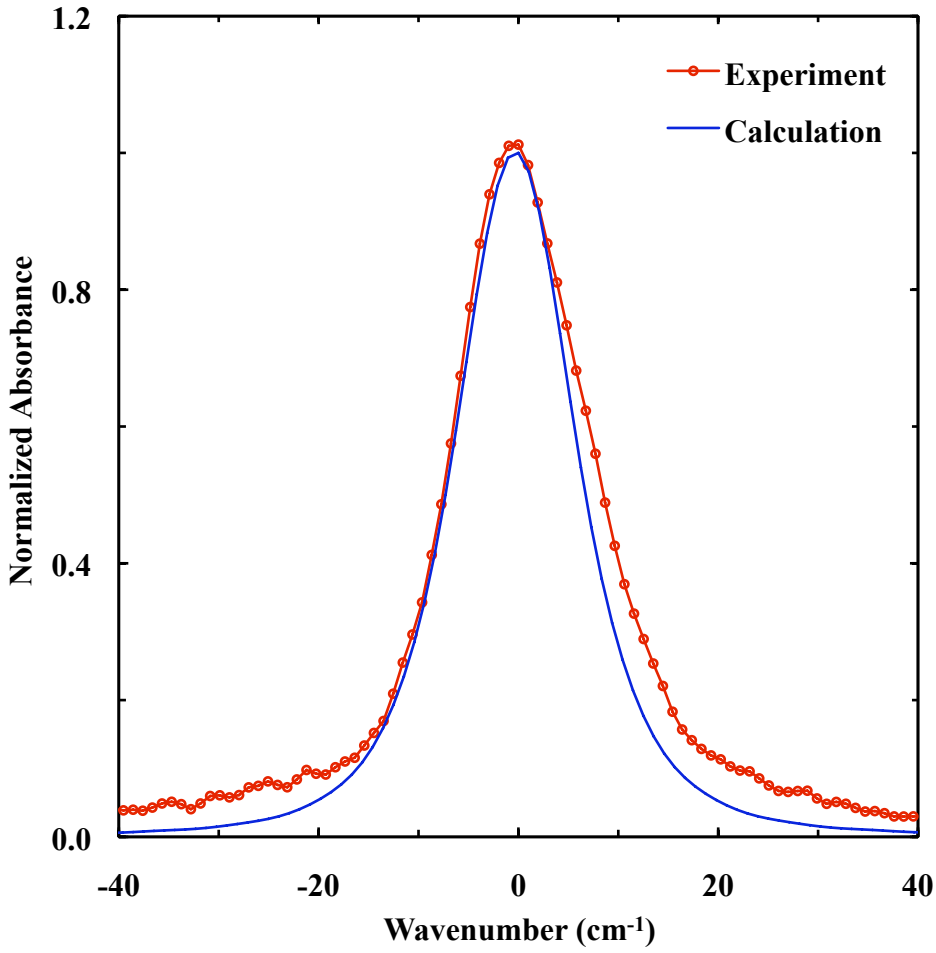

1.2
0.8
0.4
0.0
Normalized Absorbance
-40
-20
0
20
40
Wavenumber (cm-1)
Experiment
Calculation

Figure 9.8: A representative potential energy surface of 5-cyanoindole as a function of the C≡N bond length. The dots correspond to the *ab initio* single point energies, while the line is the best fit of these data to Eq. 9.4 using $D = 326$ kcal mol^{-1}.

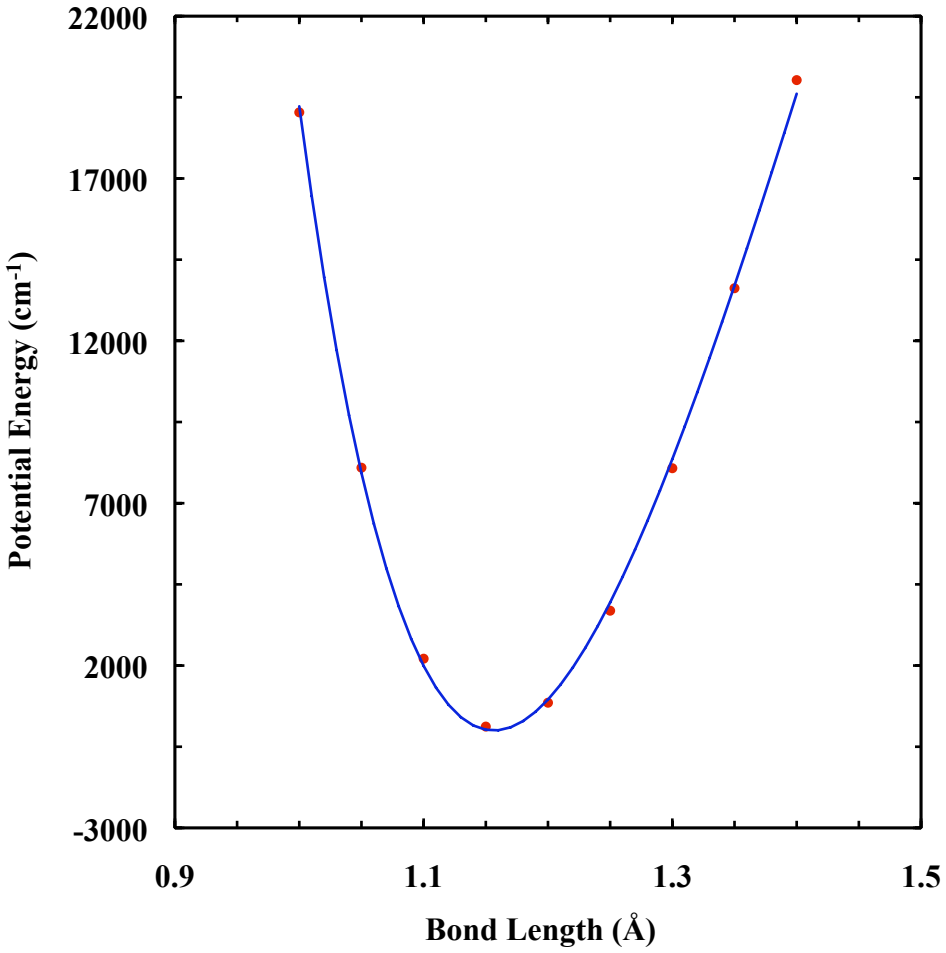
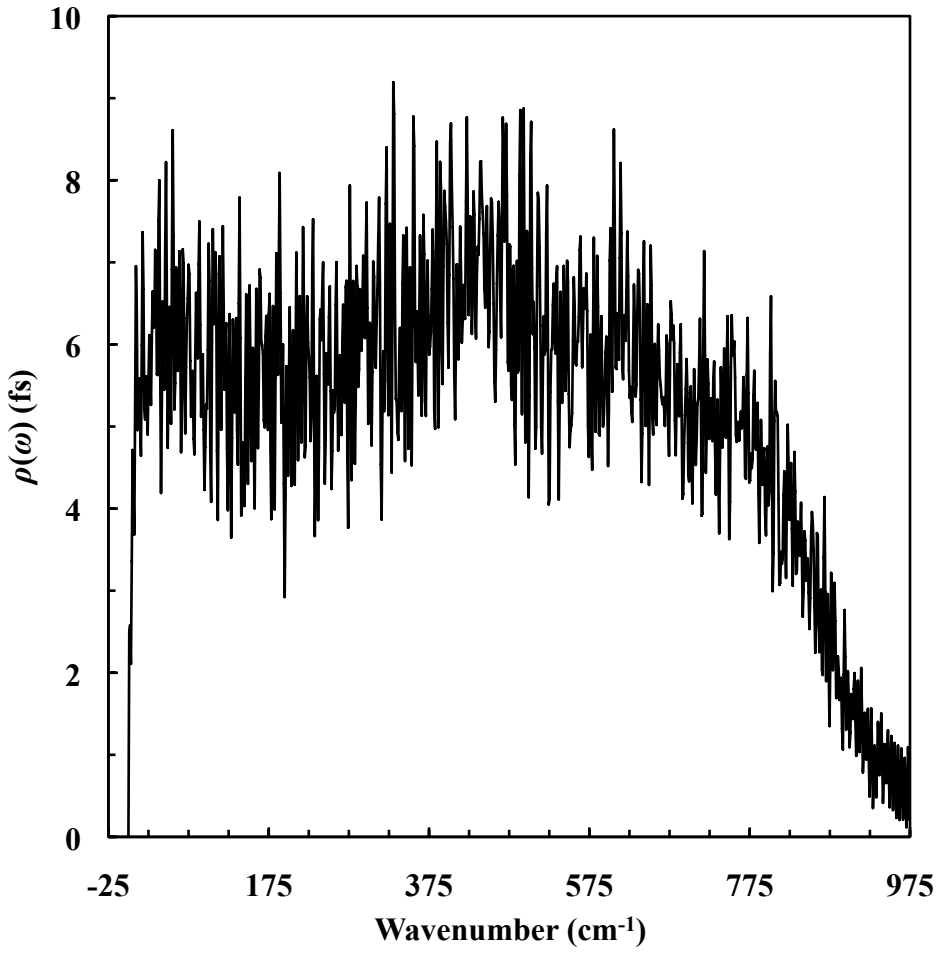

22000
17000
12000
7000
2000
-3000
Potential Energy (cm-1)
0.9
1.1
1.3
1.5
Bond Length (Å)

Figure 9.9: Spectral distribution, $\rho(\omega)$, of 5-cyanoindole in water.


10
8
6
4
2
0
ρ(ω) (fs)
-25
175
375
575
775
975
Wavenumber (cm-1)

Chapter 10

Using Tryptophan-to-disulfide Electron-transfer for the Initiation of Protein Folding Reactions

10.1 Introduction

The sudden perturbation of the thermodynamic equilibrium between folded and unfolded states of a protein is fundamental to any ensemble-level study of protein folding mechanisms. Typically, this perturbation is achieved by rapidly changing a thermodynamic variable that is readily controlled, such as temperature [17, 18, 19, 20], pressure [49], or solvent environment [50]. These well-established techniques greatly expanded our understanding of the mechanism of protein folding. However, these methods have certain limitations on the conditions under which protein folding can be studied. For example, the techniques provide no or limited control over the structures of the unfolded state ensemble. However, because the ensemble of unfolded states plays an integral part in the protein folding problem as it forms the starting point of any folding reaction, it would be desirable to have a technique that enables the experimenter to impose a certain structure or ensemble of structures on the starting point of the folding reaction. One promising method is to constrain a protein of interest in some non-native state by chemically bonding two residues together. In principle, this chemical bond can be introduced between any two residues in the peptide, thus giving the experimenter control over the initial structural ensemble of the protein. Then, by lifting the covalent constraint, the protein is allowed to adopt its native conformation. Previously, Lu *et al.* [52] have proposed a disulfide based phototrigger for protein folding reactions. In this approach, a disulfide bond is used

to link two residues of a peptide together. Then, by directly cleaving the bond by an ultraviolet (UV) laser pulse, the ensuing folding reaction was monitored. However, this approach suffers from geminate recombination of the radicals that were produced by cleaving the bond, i.e. a large fraction of the highly reactive radicals produced by the cleavage of the bond react to re-form the covalent link before any conformational relaxation can be observed. Moreover, this trigger requires the incorporation of aryl disulfides into the peptide of interest and thus relies on elaborate synthesis.

Here, we propose an optical trigger for the initiation of protein folding reactions that is solely based on natural amino acids and is therefore easily incorporated into larger proteins. Briefly, a disulfide bond is used to constrain a protein in a non-native state. The disulfide bond is cleaved by light-induced electron-transfer from an adjacent tryptophan residue. Upon photoexcitation, we observed transients on the μs timescale at 1631 cm^{-1} and 1658 cm^{-1}, which we tentatively interpret as evidence of helix formation.

10.2 Results and Discussion

It was found that 90% of thiyl radicals, formed by direct excitation of disulfide bond electrons via a UV pulse typically undergo geminate recombination on a timescale of $\sim 2-5$ ns [52, 51], thus rendering this method unsuitable as a trigger for peptide or protein folding reactions, which occur on much longer timescales. Recently,

Hamm and coworkers [386] found that peptide backbone dynamics can be observed following direct cleavage of a disulfide bridge, provided sufficient strain is put onto the bond, thus reducing the probability of geminate recombination. However, sufficiently large strains can only be imposed as long as the two bridged cysteines are close together in sequence, limiting this approach to the study of small peptides. Here, we seek to enhance the yield of the cleavage reaction by exciting a tryptophan residue, resulting in the transfer of an electron to an adjacent disufide bridge, which is irreversibly disrupted. In equilibrium experiments, following photoexcitation of tryptophan residues adjacent to a disulfide bridge in a cyclic model peptide (sequence: $CW(AAAAK)_3WC\text{-}NH_2$), we found a readily detectable increase in helicity, associated with the folding of the peptide into a more helical structure upon removal of the disulfide bond constraint. In time-resolved infrared spectroscopy experiments, we observed transients on the μs timescale. Though we found evidence that these transients most likely arise from helix-formation, their interpretation is complicated due to significant peptide aggregate formation and sample heating associated with UV irradiation.

Secondary structure content of the cyclic and linear peptides.

As shown in Fig. 10.1, the CD spectrum of the linear peptide at 4 °C in 20% TFE at pH* 7 exhibits the characteristic double minima of helical peptides, as previously observed for related alanine based peptides [387]. In contrast, there is no clear double

minimum in the CD spectrum of the cross-linked peptide, showing that the population of the helical conformation is significantly reduced. Based on this equilibrium CD study, we can expect to observe measurable changes in the spectral signature of this peptide upon efficient cleavage of the disulfide bond via UV light.

Evidence for disulfide bond cleavage. Having established a suitable model system, we further investigated the feasibility of cleaving the disulfide bond via the transfer of an electron from photoexcited tryptophan. Cleavage of disulfide bonds via light-induced electron transfer from adjacent tryptophan residues has been observed in various proteins [388, 389, 390, 391, 392, 393, 394]. As shown in Fig. 10.2, prolonged exposure of both peptides to light of a wavelength of 280 nm results in a decrease in the fluorescence band of tryptophan, indicating that the tryptophan is photo-bleached. Though we observed significant deviations in the rate of photo-bleaching for different samples of the cyclic peptide, the photo-bleaching process was always significantly faster in the cyclic peptide than in the linear peptide (Fig. 10.2). We also note that the initial fluorescence intensity (before irradiation) is substantially lower for the cyclic peptide relative to the linear one, indicating that tryptophan fluorescence is quenched in the presence of a disulfide bond, as observed previously by others [388, 389, 390, 391, 392, 393, 394]. The increased rate of tryptophan photo-bleaching in the cyclic peptide is rationalized as follows: It is well known that the triplet state of tryptophan can be quenched by an electron transfer reaction to a

disulfide moiety [395], thus converting the tryptophan to a non-fluorescent species. Another possible channel for tryptophan photo-bleaching is the generation of a free electron following photoexcitation. The quantum yield of generating a free electron by excitation of tryptophan that is incorporated into a peptide bond is about 0.04-0.06 [395], depending on the pH of the solution. This free electron can undergo geminate recombination or react with another chemical moiety, such as a disulfide bond. In either case, the formed disulfide radical anions can undergo dissociation [396]. Taken together, the increased rate of photo-bleaching of the tryptophan residues of the cyclic peptide most likely originates from the transfer of an electron from photoexcited tryptophan to the disulfide bridge.

Photoproducts. To establish the photoproducts of this reaction, the absorption spectra of the linear and cyclic peptides before and after irradiation with light of a wavelength of 280 nm were measured (Figs. 10.3 and 10.4). An oxidation product of tryptophan, *N*-formylkynurenine, has a smaller absorption at 280 nm and larger absorption at 300 and 260 nm compared to that of tryptophan [397]. Such a change in the absorption profile is observed for the linear peptide after irradiation, indicating the formation of this compound (Fig. 10.3). For the cross-linked version (Fig. 10.4), a similar increase in absorbance is observed at around 320 nm, however, the absorbance around 280 nm also increases, inconsistent with the absorption profile of *N*-formylkynurenine [397], suggesting the formation of a different oxidation product.

Vanhooren *et al.* [389] reported similar changes in the absorption spectrum of irradiated goat α-lactalbumin, a protein which contains a tryptophan residue in immediate vicinity to a disulfide bond. Based on the analysis of mass spectra of peptide fragments of goat α-lactalbumin, they found evidence that the absorption spectrum is the spectral signature of sulfenylated tryptophan, resulting from the reaction of the tryptophan and thiyl radicals.

$$\mathrm{Trp}^{\bullet} + \mathrm{RS}^{\bullet} \rightarrow \mathrm{Trp} - \mathrm{SR} \tag{10.1}$$

Taken together, the absorption spectra of the linear and cyclic peptides before and after irradiation suggest that whereas *N*-formylkynurenine is the major photoproduct in the case of the linear peptide, the tryptophan and disulfide radical react to from sulfenylated tryptophan.

Kinetics of disulfide bond cleavage. In order to further characterize the utility of the tryptophan to disulfide bond electron transfer process as a phototrigger for protein folding reactions, we investigated the kinetics of the disulfide bond cleavage reaction following a UV light pulse, using the formation of helical structure, probed via the amide I' band of the peptide backbone. As shown in Fig. 10.5, the kinetics traces consist of a pump induced spike (possibly due to heating) and a slower, single-exponential rise at 1631 cm^{-1}. This observation is consistent with the formation of helical structure as helices show strong absorbance at this wavelength. To further confirm that the transient arises from helix formation, we also probed the signal at

1658 cm^{-1}, which falls into a spectral region that is associated with disordered structure. Indeed, the signal at 1658 cm^{-1} decreases following a UV light pulse. Fitting the slow phase of both transients with a single exponential gives rate constants of $(1.1\ \mu s)^{-1}$ and $(0.8\ \mu s)^{-1}$ for the 1631 cm^{-1} and 1658 cm^{-1} transients, respectively. The similar rates for both transients further support the interpretation that the signals are due to helix formation following the cleavage of the disulfide bridge. We note that the formation of the helix is expected to be faster than the rates found here [398]. Indeed, the observed transients show no temperature dependence (data not shown), indicating that the dissociation of the disulfide radical anion is always slower than the formation of helical structure. Thus, the rates reported inhere are most likely a convolution of disulfide radical anion and backbone relaxation kinetics. The observed rates fall into the range of the microsecond dynamics of disulfide radical anions reported by others [396, 394].

10.3 Conclusions

Taken together, the time-resolved experiments suggest that the yield of dissociation of a disulfide bond, induced by the transfer of an electron from photoexcited tryptophan, is possibly sufficiently high to readily observe peptide or protein conformational dynamics. However, the assignment of the transients to helix formation must remain tentative at this time as heating of aggregates or other molecular species may also

contribute to the signal. If the transients indeed arise from helix formation, the application of this novel trigger would be useful in the context of proteins that fold on a timescale of hundredths of microseconds and slower, where the dynamics of the disulfide radical anion and protein are well separated. In particular, the trigger would allow constraining proteins in various non-native conformational ensembles, opening an avenue to a new set of experiments. However, the reactivity of the generated radicals may lead to aggregation or undesirable side-reactions like the one described by Eq. 10.1. However, recent demonstration of a related disulfide based phototrigger in the context of enzyme kinetics [399] shows that the proposed triggering method described herein may also be extended to more complex protein folding scenarios.

10.4 Materials and Methods

Materials. All materials were used as received. D_2O (D, 99.96%) was purchased from Cambridge Isotope Laboratories (Andover, MA), deuterated 2,2,2-trifluroethanol-D_3 (D, 99.5%) was obtained from Sigma-Aldrich (St. Louis, MO). Amino acids for peptide synthesis were obtained from Advanced ChemTech (Louisville, KY).

Sample preparation. The peptides were synthesized using standard flouren-9-ylmethoxycarbonyl (Fmoc)-chemistry on a PS3 peptide synthesizer from Protein Technologies, (Tucson, AZ) and were purified by reversed-phase HPLC (Agilent Technologies, Santa Clara, CA). The mass of each peptide was verified by matrix-assisted

laser desorption ionization mass spectrometry on a Burker UltrafeXtreme instrument (Bruker Daltonics, Billerica, MA). Buffer solutions were prepared by adding the appropriate amounts of D_2O and 2,2,2-trifluoroethanol-D_3 (20 volume%) to buffer salt. The 20 mM buffer solution was then added to dry, deuterium exchanged peptide sample to give a $\sim 2-3$ mM solution of peptide for infrared measurements.

Spectroscopic measurements. Circular dichroism (CD) data were collected on an Aviv Model 410 instrument (Aviv Biomedical, Lakewood, NJ) using a 1 mm cuvette. The sample concentration was $\sim$ 36.5 μM in the presence of 20 mM buffer pH 6. UV-vis spectra were collected on a Perkin-Elmer Lambda 25 UV-Vis spectrometer (Fremont, CA). Fluorescence spectra were obtained at 20 °C on a Fluorolog 3.10 spectrofluorometer (Jobin Yvon Horiba, Edison, NJ) with a 1 cm quartz sample holder. Temperature was regulated using a TLC 50 Peltier temperature controller (Quantum Northwest, Spokane, WA). An integration time of 1.5 s/nm was used in all measurements. The samples were not stirred during measurements.

The transient IR spectroscopy setup consisted of a Nd:YAG laser, whose fundamental (1064 nm, 10 Hz repetition rate, 3 ns pulse width) was frequency quadrupled to yield an excitation wavelength of 266 nm. The excitation wavelength was separated from the fundamental and second harmonic by a collection of prisms and weakly focused (focal spot $\sim$ 1 mm^2) onto the sample cell, consisting of two CaF_2 windows, separated by a 100 μm thick Teflon spacer. The excitation power was estimated to

be on the order of $\sim 10 - 20$ μJ per pulse. The transient IR signals were probed by a tunable quantum cascade laser (Daylight Solutions, CA), which was focused onto the sample (focal diameter < 300 μm), and detected by a HgCdTe detector (Kolmar Technologies, Newburyport, MA). The temperature of the sample was maintained by a water bath. The sample cell holder was constantly horizontally moved through the excitation beam during data collection to avoid excessive aggregation of the photoexcited peptides.

10.5 Acknowledgement

We gratefully acknowledge financial support from the National Institutes of Health (GM-065978 and RR-01348).

Figure 10.1: CD spectra of the linear (△) and cyclic (◯) peptides at 4 °C in 20 mM phosphate buffer pH* 7 and 20% TFE.

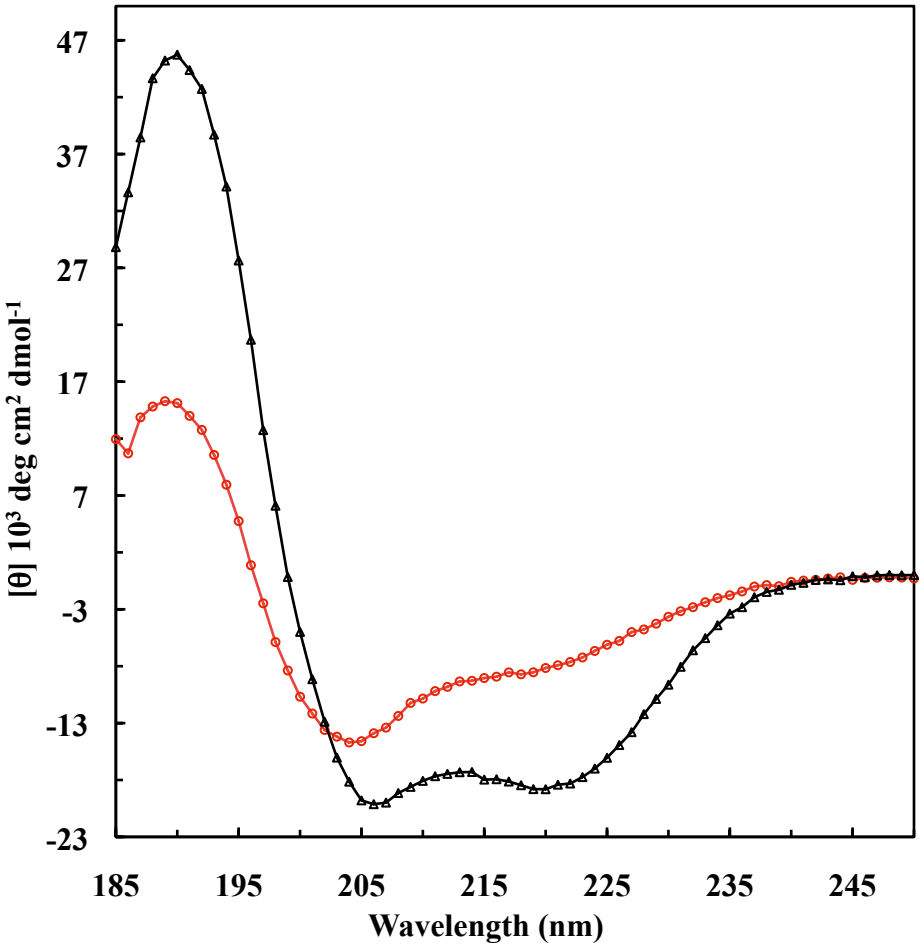

47
37
27
17
7
-3
-13
-23
[θ] 10³ deg cm² dmol⁻¹
185
195
205
215
225
235
245
Wavelength (nm)

Figure 10.2: Decrease of the fluorescence intensity of the tryptophan fluorescence band as a function of irradiation time at 280 nm for the linear ($\triangle$) and cyclic ($\bigcirc$) peptides.

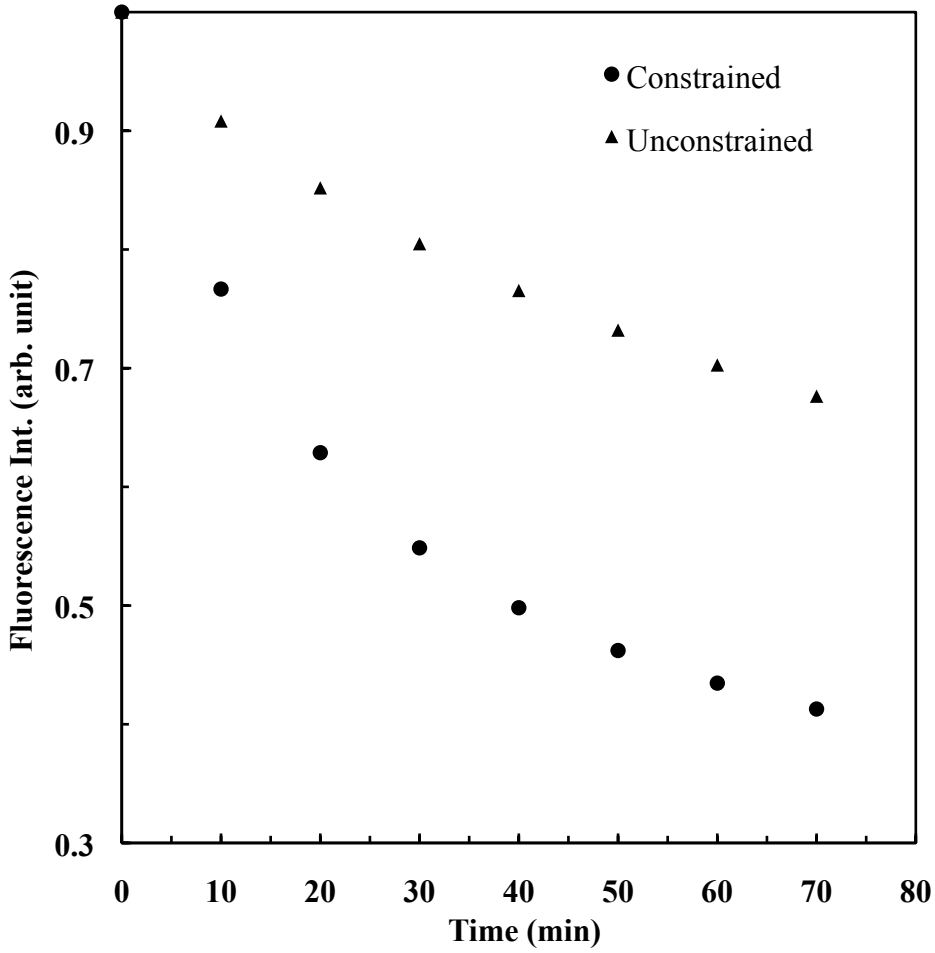

Constrained
Unconstrained
Fluorescence Int. (arb. unit)
0.9
0.7
0.5
0.3
0
10
20
30
40
50
60
70
80
Time (min)

Figure 10.3: Absorption spectra of the linear peptide before (red) and after (black) irradiation for 70 min at 280 nm.

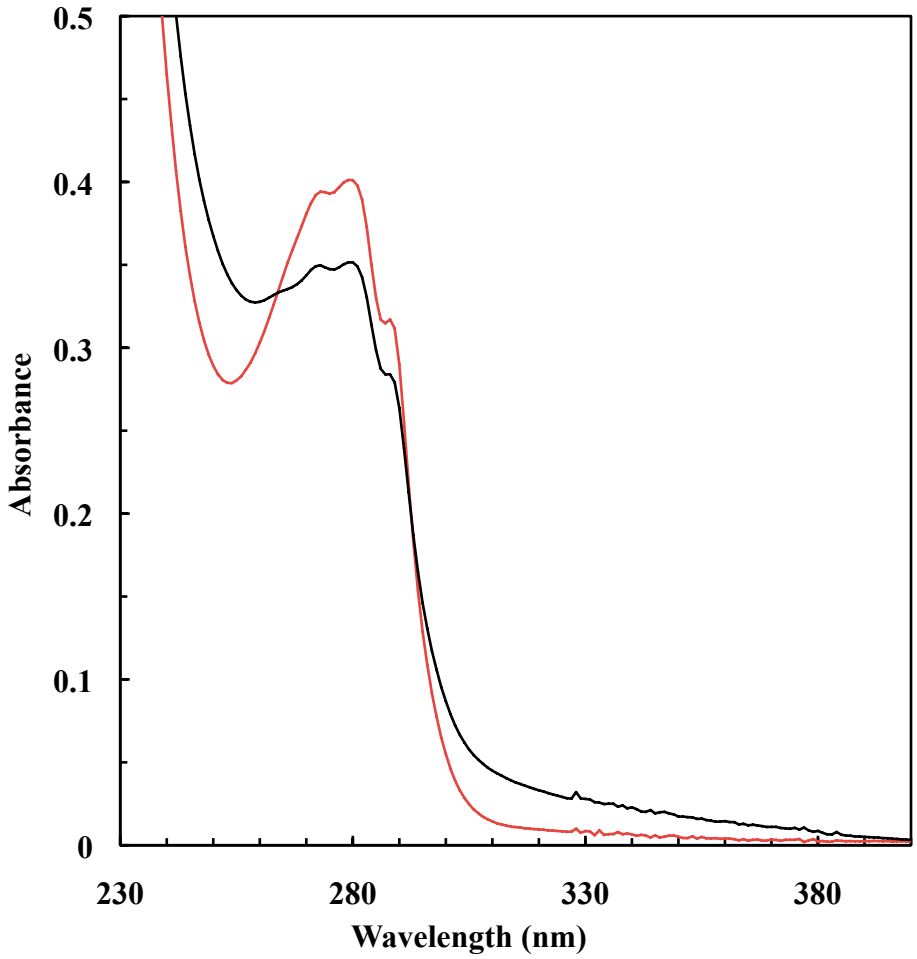

0.5
0.4
0.3
0.2
0.1
0
Absorbance
230
280
330
380
Wavelength (nm)

Figure 10.4: Absorption spectra of the cyclic peptide before (red) and after (black) irradiation for 70 min at 280 mn.

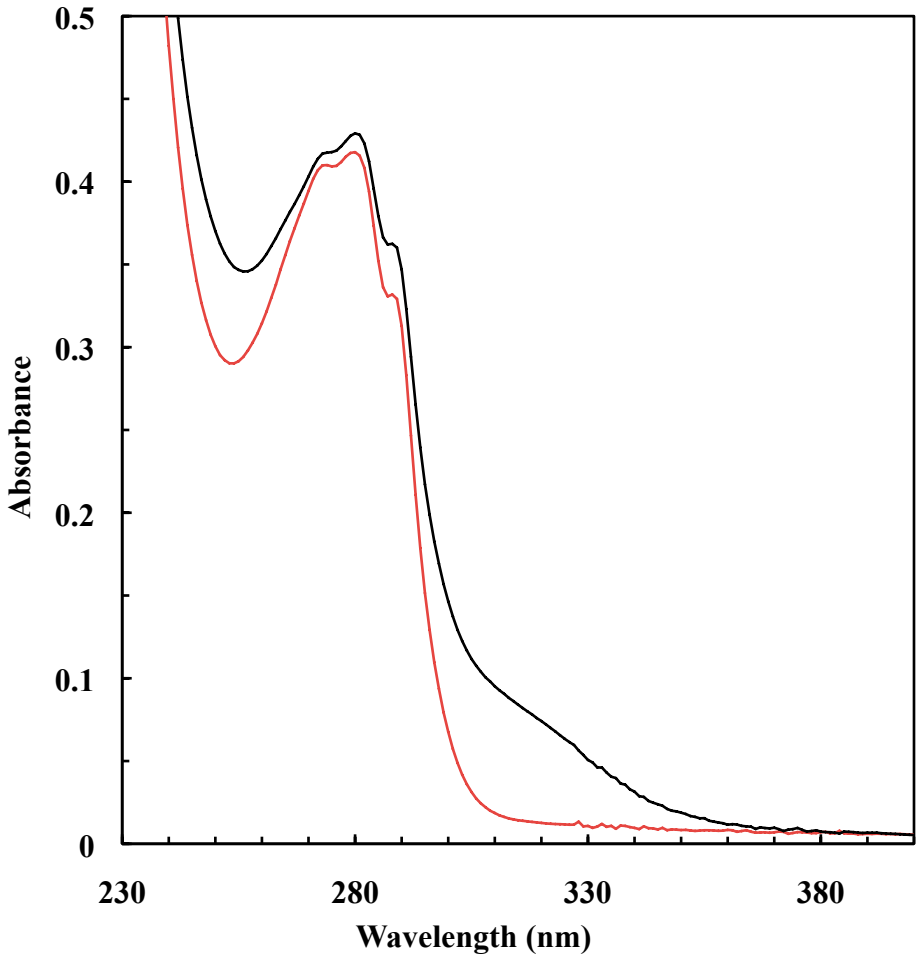
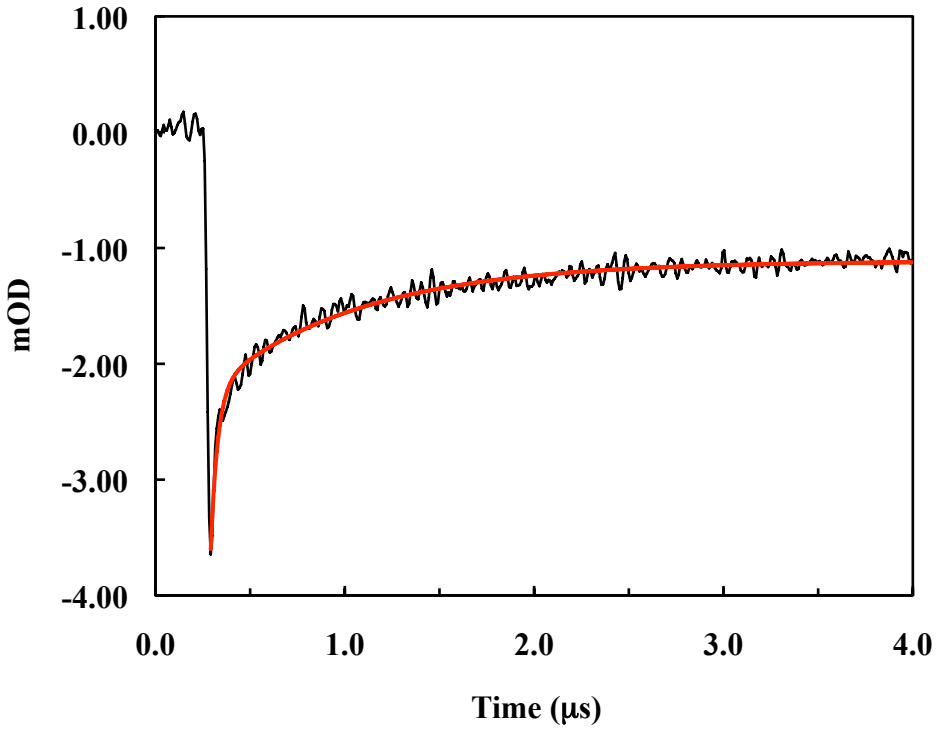
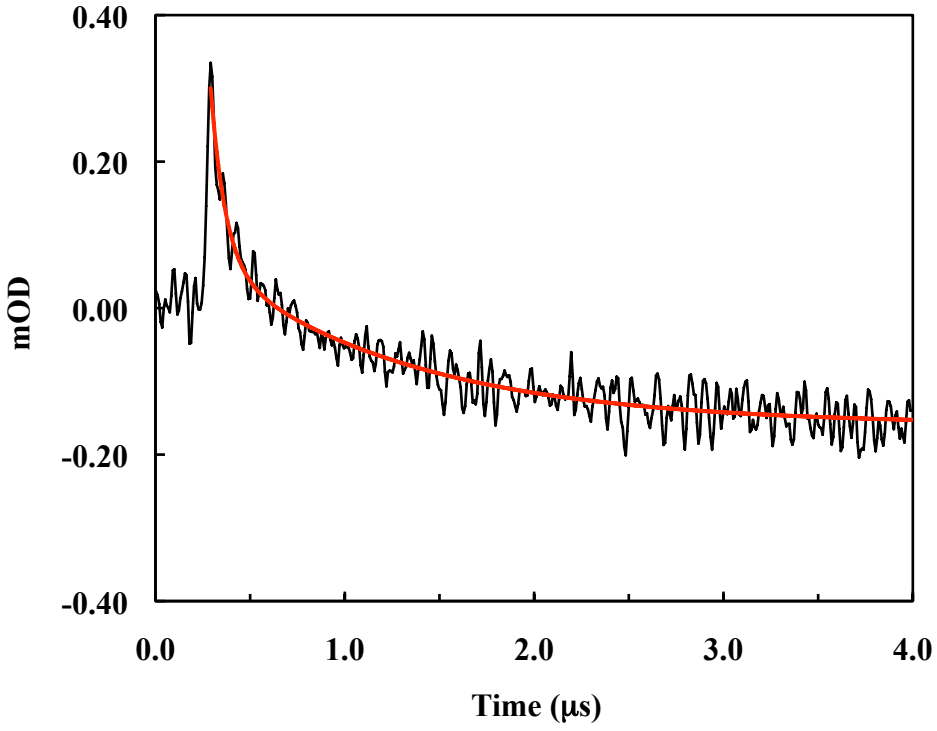

0.5
0.4
0.3
0.2
0.1
0
Absorbance
230
280
330
380
Wavelength (nm)

Figure 10.5: Relaxation trace observed at 1631 cm^{-1} (upper panel) and 1658 cm^{-1} (lower panel), consisting of a fast, pump-induced shift ($\tau_1 = 0.04$ μs) / ($\tau_1 = 0.08$ μs) and a slow phase ($\tau_2 = 0.8$ μs) / ($\tau_2 = 1.1$ μs) for 1631 cm^{-1} / 1658 cm^{-1}, respectively. The slow phase is tentatively interpreted as a convolution of the helix folding and the bond dissociation kinetics. The temperature was 15 °C.


1.00
0.00
-1.00
-2.00
-3.00
-4.00
mOD
0.0
1.0
2.0
3.0
4.0
Time (μs)


0.40
0.20
0.00
-0.20
-0.40
mOD
0.0
1.0
2.0
3.0
4.0
Time (μs)

Chapter 11

Future Directions

In Chapter 4, we offered insights into the folding kinetics of a naturally occurring helical hairpin. One interesting extension of this project would be to determine the generality of the proposed mechanism. The study of related peptides, like the helical hairpin of Engrailed Homeodomain, would be an interesting candidate for future studies.

In Chapters 5 and 6, we investigated the effects of macromolecular crowding on peptide and protein folding. Most common crowding models picture the macromolecular crowding system as an aqueous suspension of hard spheres (i.e. the crowders), the protein being confined to the space in between these crowders. Based on qualitative agreement between the model's predictions and experimental results, there seems to be some justification for this picture. However, for more quantitative predictions, the characteristics of common crowder systems need to be known more precisely. For instances, there is no estimate for the average size of the aforementioned interstitial spaces for a given crowder system. One could think of studies, reminiscent to those of Masuda *et al.* [258, 259], that would provide an improved understanding of the microcompartmentalization of the solvent by the macromolecular crowders.

Furthermore, we have only tested the crowder model of Thirumalai and Cheung [29] using one model protein, i.e. ubiquitin. In our study, due to the properties of ubiquitin, we used IR spectroscopic measurements, which are more cumbersome to perform than, for example, circular dichroism measurements. It would be insightful

to choose a large number of proteins, e.g. 30 or more, that are cheap, commercially available, not prone to aggregation, and have well defined, readily measurable simple (ideally, two-state) circular dichroism unfolding transitions. Testing various models against this set of protein unfolding data under different crowding conditions would provide tremendous insights and would allow to draw much more general conclusions.

In Chapter 7, we attempted to test various polymer models for the description of the conformational ensemble of an unstructured polypeptide with a complex sequence. However, this test cannot be performed without bias as the shape of the conformational distribution function has to be assumed *a priori*, though the experiment measures only the mean of the distribution. As distribution functions of different shapes might have similar or identical averages, the experiment described in Chapter 7 and related experiments preformed by other investigators are not rigorous tests of the polymer models. A more direct measurement should be carried out. One could, for example, perform single-molecule measurements of the energy transfer efficiencies where the polymers are dissolved in frozen matrices. Freezing out any polymer dynamics and measuring on the single-molecule level would allow for the determination of the energy transfer efficiency and also the distance as, in this case, the distribution is a delta function for the single-molecule. Building up a histogram of distance distributions, one could preform a rigorous test of any polymer model. Furthermore, the photophysical properties of some of the Phe_{CN} containing peptides used

in Chapter 7 are not well understood. It would be worthwhile to investigate some of the observations related to Phe_{CN} fluorescence described in Chapter 7 in more detail to allow for a rigorous interpretation of the Phe_{CN} based FRET efficiencies.

In Chapters 8 and 9, a new infrared probe, i.e. the nitrile stretching band of 5-cyanotryptophan, was proposed. An interesting observation is that the center frequency of the C≡N stretch of 5-cyanotryptophan is $\sim 30 - 40$ cm^{-1} lower than the one of *p*-cyanophenylalanine. It is therefore conceivable that coupling between the vibrational transition dipoles of the two probes could occur [361], if they are spatially brought close together. If such coupling would occur, this could provide a novel means to measure short distances between the probes in protein systems. An appropriately designed 2D IR experiment could reveal such coupling.

In Chapter 10, we proposed a new triggering method based on tryptophan-mediated reductive cleavage of disulfide bonds. Our results seem to suggest that this method could be applied to more complex folding problems, provided that undesired side-reactions of the reactive radical species are minimized. An interesting candidate for a follow-up study would be β-lactoglobulin A.

Bibliography

[1] Kendrew, J. C, Bodo, G, Dintzis, H. M, Parrish, R. G, Wyckoff, H, & Phillips, D. C. (1958) A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **181**, 662–666.

[2] Muirhead, H & Perutz, M. F. (1963) Structure of Haeemoglobin: A Three-Dimensional Fourier Synthesis of Reduced Human Haemoglobin at 5.5 Å Resolution. *Nature* **199**, 633–638.

[3] Sela, M, White, F. H, & Anfinsen, C. B. (1957) Reductive Cleavage of Disulfide Bridges in Ribonuclease. *Science* **125**, 691–692.

[4] Anfinsen, C. B. (1973) Principles that Govern the Folding of Protein Chains. *Science* **181**, 223–230.

[5] Epstein, H. F, Schechter, A. N, & Cohen, J. S. (1971) Folding of Staphylococcal Nuclease - Magnetic Resonance and Fluorescence Studies of Individual Residues. *Proceedings of the National Academy of Sciences of the United States of America* **68**, 2042–2046.

[6] Epstein, H. F, Schechter, A. N, Chen, R. F, & Anfinsen, C. B. (1971) Folding of Staphylococcal Nuclease - Kinetic Studies of 2 Processes in Acid Renaturation. *Journal of Molecular Biology* **60**, 499–508.

[7] Ikai, A & Tanford, C. (1971) Kinetic Evidence for Incorrectly Folded Intermediate States in Refolding of Denatured Proteins. *Nature* **230**, 100–102.

[8] Teipel, J. W & Koshland, D. E. (1971) Kinetic Aspects of Conformational Changes in Proteins. I. Rate of Regain of Enzyme Activity from Denatured Proteins. *Biochemistry* **10**, 792–798.

[9] Teipel, J. W & Koshland, D. E. (1971) Kinetic Aspects of Conformational Changes in Proteins. II. Structural Changes in Renaturation of Denatured Proteins. *Biochemistry* **10**, 798–805.

[10] Lewis, P. N, Momany, F. A, & Scheraga, H. A. (1971) Folding of Polypeptide Chains in Proteins: A Proposed Mechanism for Folding. *Proceedings of the National Academy of Sciences of the United States of America* **68**, 2293–2297.

[11] Wetlaufer, D. B. (1973) Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **70**, 697–701.

[12] Matheson, R. R & Scheraga, H. A. (1978) A Method for Predicting Nucleation

Sites for Protein Folding Based on Hydrophobic Contacts. *Macromolecules* **11**, 819–829.

[13] Rose, G. D & Roy, S. (1980) Hydrophobic Basis of Packing in Globular Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 4643–4647.

[14] Lesk, A. M & Rose, G. D. (1981) Folding Units in Globular Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 4304–4308.

[15] Oka, M, Montelione, G. T, & Scheraga, H. A. (1984) Chain-Folding Initiation Structures in Ribonuclease A: Conformational Free Energy Calculations on Ac-Asn-Pro-Tyr-NHMe, Ac-Tyr-Pro-Asn-NHMe, and Related Peptides. *Journal of the American Chemical Society* **106**, 7946–7958.

[16] Schwarz, G. (1965) On Kinetics of Helix-Coil Transition of Polypeptides in Solution. *Journal of Molecular Biology* **11**, 64–77.

[17] Dyer, R. B, Gai, F, & Woodruff, W. H. (1998) Infrared Studies of Fast Events in Protein Folding. *Accounts of Chemical Research* **31**, 709–716.

[18] Gruebele, M, Sabelko, J, Ballew, R, & Ervin, J. (1998) Laser Temperature Jump Induced Protein Refolding. *Accounts of Chemical Research* **31**, 699–707.

[19] Callender, R & Dyer, R. B. (2002) Probing Protein Dynamics Using Temperature Jump Relaxation Spectroscopy. *Current Opinion in Structural Biology* **12**, 628–633.

[20] Kubelka, J. (2009) Time-Resolved Methods in Biophysics. 9. Laser Temperature-Jump Methods for Investigating Biomolecular Dynamics. *Photochemical & Photobiological Sciences* **8**, 499–512.

[21] Huffman, J. L & Brennan, R. G. (2002) Prokaryotic Transcription Regulators: More Than Just the Helix-Turn-Helix Motif. *Current Opinion in Structural Biology* **12**, 98–106.

[22] Aravind, L, Anantharaman, V, Balaji, S, Babu, M, & Iyer, L. (2005) The Many Faces of the Helix-Turn-Helix Domain: Transcription Regulation and Beyond. *FEMS Microbiology Reviews* **29**, 231–262.

[23] Starovasnik, M. A, Braisted, A. C, & Wells, J. A. (1997) Structural Mimicry of a Native Protein by a Minimized Binding Domain. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 10080–10085.

[24] Zondlo, N. J & Schepartz, A. (1999) Highly Specific DNA Recognition by a Designed Miniature Protein. *Journal of the American Chemical Society* **121**, 6938–6939.

[25] Hodges, A. M & Schepartz, A. (2007) Engineering a Monomeric Miniature Protein. *Journal of the American Chemical Society* **129**, 11024–11025.

[26] Zhou, H.-X, Rivas, G, & Minton, A. P. (2008) Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences. *Annual Review of Biophysics* **37**, 375–397.

[27] Richter, K, Nessling, M, & Lichter, P. (2008) Macromolecular Crowding and its Potential Impact on Nuclear Function. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1783**, 2100 – 2107.

[28] Dix, J. A & Verkman, A. S. (2008) Crowding Effects on Diffusion in Solutions and Cells. *Annual Review of Biophysics* **37**, 247–263.

[29] Cheung, M. S, Klimov, D, & Thirumalai, D. (2005) Molecular Crowding Enhances Native State Stability and Refolding Rates of Globular Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 4753–4758.

[30] Tompa, P. (2002) Intrinsically Unstructured Proteins. *Trends in Biochemical Sciences* **27**, 527–533.

[31] Uversky, V. N. (2002) Natively Unfolded Proteins: A Point where Biology Waits for Physics. *Protein Science* **11**, 739–756.

[32] Dyson, H. J & Wright, P. E. (2005) Intrinsically Unstructured Proteins and Their Functions. *Nature Reviews Molecular Cell Biology* **6**, 197–208.

[33] Tanford, C. (1961) *Physical Chemistry of Macromolecules.* (John Wiley & Sons, New York, NY).

[34] Kratky, O & Porod, G. (1949) Röntgenuntersuchung gelöster Fadenmoleküle. *Recueil Des Travaux Chimiques Des Pays-Bas-Journal of the Royal Netherlands Chemical Society* **68**, 1106–1122.

[35] Kauzmann, W. (1959) Some Factors in the Interpretation of Protein Denaturation. *Advances in Protein Chemistry* **14**, 1–63.

[36] Privalov, P. L. (1979) Stability of Proteins: Small Globular Proteins. *Advances in Protein Chemistry* **33**, 167–241.

[37] Schellman, J. A. (1987) The Thermodynamic Stability of Proteins. *Annual Review of Biophysics and Biophysical Chemistry* **16**, 115–137.

[38] Becktel, W. J & Schellman, J. A. (1987) Protein Stability Curves. *Biopolymers* **26**, 1859–1877.

[39] Rose, G & Wolfenden, R. (1993) Hydrogen-Bonding, Hydrophobicity, Packing, and Protein-Folding. *Annual Review of Biophysics and Biomolecular Structure* **22**, 381–415.

[40] Robertson, A. D & Murphy, K. P. (1997) Protein Structure and the Energetics of Protein Stability. *Chemical Reviews* **97**, 1251–1267.

[41] Jaenicke, R & Böhm, G. (1998) The Stability of Proteins in Extreme Environments. *Current Opinion in Structural Biology* **8**, 738–748.

[42] Atkins, P & de Paula, J. (2002) *Physical Chemistry.* (Oxford University Press, Oxford, U.K.).

[43] Rees, D. C & Robertson, A. D. (2001) Some Thermodynamic Implications for the Thermostability of Proteins. *Protein Science* **10**, 1187–1194.

[44] Edsall, J. T. (1935) Apparent Molal Heat Capacities of Amino Acids and Other Organic Compounds. *Journal of the American Chemical Society* **57**, 1506–1507.

[45] Sturtevant, J. M. (1977) Heat-Capacity and Entropy Changes in Processes Involving Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 2236–2240.

[46] Spolar, R. S, Ha, J. H, & Record, M. T. (1989) Hydrophobic Effect in Protein Folding and Other Noncovalent Processes Involving Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 8382–8385.

[47] Spolar, R. S, Livingstone, J. R, & Record, M. T. (1992) Use of Liquid-Hydrocarbon and Amide Transfer Data to Estimate Contributions to Ther-

modynamic Functions of Protein Folding from the Removal of Nonpolar and Polar Surface from Water. *Biochemistry* **31**, 3947–3955.

[48] McQuarrie, D. A & Simon, J. D. (1997) *Physical Chemistry, A Molecular Approach*. (University Science Books, Sausalito, CA).

[49] Silva, J. L, Foguel, D, & Royer, C. A. (2001) Pressure Provides New Insights into Protein Folding, Dynamics and Structure. *Trends in Biochemical Sciences* **26**, 612 – 618.

[50] Roder, H, Maki, K, Cheng, H, & Shastry, M. C. R. (2004) Rapid Mixing Methods for Exploring the Kinetics of Protein Folding. *Methods* **34**, 15 – 27.

[51] Volk, M, Kholodenko, Y, Lu, H. S. M, Gooding, E. A, Degrado, W. F, & Hochstrasser, R. M. (1997) Peptide Conformational Dynamics and Vibrational Stark Effects Following Photoinitiated Disulfide Cleavage. *The Journal of Physical Chemistry B* **101**, 8607–8616.

[52] Lu, H. S. M, Volk, M, Kholodenko, Y, Gooding, E, Hochstrasser, R. M, & Degrado, W. F. (1997) Aminothiotyrosine Disulfide, an Optical Trigger for Initiation of Protein Folding. *Journal of the American Chemical Society* **119**, 7173–7180.

[53] Fersht, A. R, Matouschek, A, & Serrano, L. (1992) The Folding of an Enzyme.

I. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding. *Journal of Molecular Biology* **224**, 771–782.

[54] Evans, M. G & Polanyi, M. (1935) Some Applications of the Transition State Method to the Calculation of Reaction Velocities, Especially in Solution. *Transactions of the Faraday Society* **31**, 875–894.

[55] Eyring, H. (1935) The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics* **3**, 107–115.

[56] Laidler, K. J & King, M. C. (1983) Development of Transition-State Theory. *The Journal of Physical Chemistry* **87**, 2657–2664.

[57] Kramers, H. A. (1940) Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions. *Physica* **7**, 284 – 304.

[58] Talkner, P & Hänggi, P, eds. (1995) *New Trends in Kramers' Reaction Theory.* (Kluwer Academic Publishers, Dordrecht, The Netherlands).

[59] Jas, G. S, Eaton, W. A, & Hofrichter, J. (2001) Effect of Viscosity on the Kinetics of α-Helix and β-Hairpin Formation. *Journal of Physical Chemistry B* **105**, 261–272.

[60] Chung, H. S, Khalil, M, Smith, A. W, Ganim, Z, & Tokmakoff, A. (2005) Conformational Changes during the Nanosecond-to-Millisecond Unfolding of

Ubiquitin. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 612–617.

[61] Ma, H & Gruebele, M. (2005) Kinetics Are Probe-Dependent during Downhill Folding of an Engineered λ6–85 Protein. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2283–2287.

[62] Muñoz, V. (2007) Conformational Dynamics and Ensembles in Protein Folding. *Annual Review of Biophysics and Biomolecular Structure* **36**, 395–412.

[63] Chung, H. S & Tokmakoff, A. (2008) Temperature-Dependent Downhill Unfolding of Ubiquitin. I. Nanosecond-to-Millisecond Resolved Nonlinear Infrared Spectroscopy. *Proteins-Structure Function and Bioinformatics* **72**, 474–487.

[64] Kubelka, J, Hofrichter, J, & Eaton, W. A. (2004) The Protein Folding 'Speed Limit'. *Current Opinion in Structural Biology* **14**, 76 – 88.

[65] Förster, T. (1965) *Delocalized Excitation and Excitation Transfer*, Modern Quantum Chemistry ed. Sinanoglu, O. (Academic Press Inc) Vol. 3.

[66] Tokmakoff, A. (2009) Introductory Quantum Mechanics II. *Lecture Notes.* http://ocw.mit.edu/courses/chemistry/5-74-introductory-quantum-mechanics-ii-spring-2009/.

[67] Schulten, K, Ritz, T, Damjanovic, A,

& Sener, M. (2000) Excitation Transfer in Photosynthetic Systems. *Lecture Notes*. http://www.ks.uiuc.edu/Research/psu/lecture.html.

[68] Förster, T. (1948) Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik* **437**, 55–75.

[69] Dirac, P. A. M. (1927) The Quantum Theory of the Emission and Absorption of Radiation. *Proceedings of the Royal Society of London. Series A* **114**, 243–265.

[70] Greiner, W. (1994) *Quantum Mechanics - An Introduction*. (Springer Verlag, New York, NY).

[71] Baym, G. A. (1969) *Lectures on Quantum Mechanics*. (Westview Press, Boulder, CO).

[72] Scholes, G. D. (2003) Long-Range Resonance Energy Transfer in Molecular Systems. *Annual Review of Physical Chemistry* **54**, 57–87.

[73] Bandekar, J. (1992) Amide Modes and Protein Conformation. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1120**, 123 – 143.

[74] Barth, A & Zscherp, C. (2002) What Vibrations Tell us about Proteins. *Quarterly Reviews of Biophysics* **35**, 369–430.

[75] Decatur, S. M. (2006) Elucidation of Residue-Level Structure and Dynamics of Polypeptides via Isotope-Edited Infrared Spectroscopy. *Accounts of Chemical Research* **39**, 169–175.

[76] Ganim, Z, Chung, H. S, Smith, A. W, DeFlores, L. P, Jones, K. C, & Tokmakoff, A. (2008) Amide I Two-Dimensional Infrared Spectroscopy of Proteins. *Accounts of Chemical Research* **41**, 432–441.

[77] Oh, K.-I, Lee, J.-H, Joo, C, Han, H, & Cho, M. (2008) β-Azidoalanine as an IR Probe: Application to Amyloid Aβ(16-22) Aggregation. *Journal of Physical Chemistry B* **112**, 10352–10357.

[78] Ham, S & Cho, M. (2003) Amide I Modes in the N-Methylacetamide Dimer and Glycine Dipeptide Analog: Diagonal Force Constants. *The Journal of Chemical Physics* **118**, 6915–6922.

[79] Schmidt, J. R, Corcelli, S. A, & Skinner, J. L. (2004) Ultrafast Vibrational Spectroscopy of Water and Aqueous N-Methylacetamide: Comparison of Different Electronic Structure/Molecular Dynamics Approaches. *Journal of Chemical Physics* **121**, 8887–8896.

[80] Hayashi, T, Zhuang, W, & Mukamel, S. (2005) Electrostatic DFT Map for the Complete Vibrational Amide Band of NMA. *Journal of Physical Chemistry A* **109**, 9747–9759.

[81] Jansen, T. L & Knoester, J. (2006) A Transferable Electrostatic Map for Solvation Effects on Amide I Vibrations and Its Application to Linear and Two-Dimensional Spectroscopy. *Journal of Chemical Physics* **124**, 044502–11.

[82] Ham, S, Cha, S, Choi, J.-H, & Cho, M. (2003) Amide I Modes of Tripeptides: Hessian Matrix Reconstruction and Isotope Effects. *The Journal of Chemical Physics* **119**, 1451–1461.

[83] Krimm, S & Abe, Y. (1972) Intermolecular Interaction Effects in the Amide I Vibrations of β Polypeptides. *Proceedings of the National Academy of Sciences of the United States of America* **69**, 2788–2792.

[84] Cheam, T & Krimm, S. (1984) Transition Dipole Interaction in Polypeptides: *AbInitio* Calculation of Transition Dipole Parameters. *Chemical Physics Letters* **107**, 613 – 616.

[85] Torii, H & Tasumi, M. (1992) Model Calculations on the Amide-I Infrared Bands of Globular Proteins. *The Journal of Chemical Physics* **96**, 3379–3387.

[86] Krimm, S & Reisdorf, W. C. (1994) Understanding Normal Modes of Proteins. *Faraday Discussions* **99**, 181–197.

[87] Hamm, P, Lim, M, & Hochstrasser, R. M. (1998) Structure of the Amide I Band of Peptides Measured by Femtosecond Nonlinear-Infrared Spectroscopy. *The Journal of Physical Chemistry B* **102**, 6123–6138.

[88] Steinfeld, J. I. (2005) *Molecules and Radiation.* (Dover, Mineola, NY).

[89] Gnanakaran, S, Hochstrasser, R. M, & García, A. E. (2004) Nature of Structural Inhomogeneities on Folding a Helix and their Influence on Spectral Measurements. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9229–9234.

[90] Kwac, K & Cho, M. (2003) Molecular Dynamics Simulation Study of *N*-Methylacetamide in Water. I. Amide I Mode Frequency Fluctuation. *Journal of Chemical Physics* **119**, 2247–2255.

[91] Ganim, Z & Tokmakoff, A. (2006) Spectral Signatures of Heterogeneous Protein Ensembles Revealed by MD Simulations of 2DIR Spectra. *Biophysical Journal* **91**, 2636–2646.

[92] Sandeman, I. (1955) Amide Bands in Infra-Red Spectra: The Direction of the Transition Moments of Bands in *N*,*N*'-Diacetylhexamethylenediamine. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **232**, 105–113.

[93] Bour, P & Keiderling, T. A. (2003) Empirical Modeling of the Peptide Amide I Band IR Intensity in Water Solution. *Journal of Chemical Physics* **119**, 11253–11262.

[94] Bouř, P & Keiderling, T. A. (2005) Vibrational Spectral Simulation for Peptides of Mixed Secondary Structure: Method Comparisons with the Trpzip Model Hairpin. *The Journal of Physical Chemistry B* **109**, 23687–23697.

[95] Hahn, S, Ham, S, & Cho, M. (2005) Simulation Studies of Amide I IR Absorption and Two-Dimensional IR Spectra of β Hairpins in Liquid Water. *The Journal of Physical Chemistry B* **109**, 11789–11801.

[96] Watson, T. M & Hirst, J. D. (2005) Theoretical Studies of the Amide I Vibrational Frequencies of Leu-Enkephalin. *Molecular Physics* **103**, 1531–1546.

[97] Chung, H. S & Tokmakoff, A. (2006) Visualization and Characterization of the Infrared Active Amide I Vibrations of Proteins. *The Journal of Physical Chemistry B* **110**, 2888–2898.

[98] Gordon, R. G. (1968) Correlation Functions for Molecular Motion. *Advances in Magnetic Resonance* **3**, 1–42.

[99] McQuarrie, D. A. (2000) *Statistical Mechanics*. (University Science Books, Sausalito, CA).

[100] Kubo, R. (1969) *A Stochastic Theory of Line Shape*, Advances in Chemical Physics. (John Wiley & Sons, Inc., Hoboken, NJ), pp. 101–127.

[101] Mukamel, S. (1995) *Principles of Nonlinear Optical Spectroscopy.* (Oxford University Press, Oxford, U.K.).

[102] Czerlinski, G & Eigen, M. (1959) Eine Temperatursprungmethode zur Untersuchung chemischer Relaxation. *Zeitschrift für Elektrochemie, Berichte der Bunsengesellschaft für physikalische Chemie* **63**, 652–661.

[103] Bremer, C, Bergbauer, R, Ruf, H, Bannister, J. J, & Grell, E. (1993) A Laser Temperature Jump Apparatus Based on Commercial Parts Equipped with Highly Sensitive Spectrophotometric Detection. *Measurement Science and Technology* **4**, 1385.

[104] Wyn-Jones, E, ed. (1975) *The Stopped-Flow and Temperature-Jump Techniques - Principles and Recent Advances*, Chemical and Biological Applications of Relaxation Spectrometry. (D. Reidel Puplishing Company, Dordrecht, The Netherlands).

[105] Stoutland, P, Dyer, R, & Woodruff, W. (1992) Ultrafast Infrared Spectroscopy. *Science* **257**, 1913–1917.

[106] Phillips, C. M, Mizutani, Y, & Hochstrasser, R. M. (1995) Ultrafast Thermally Induced Unfolding of RNase A. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 7292–7296.

[107] Beitz, J. V, Flynn, G. W, Turner, D. H, & Sutin, N. (1970) Stimulated Raman Effect. A New Source of Laser Temperature-Jump Heating. *Journal of the American Chemical Society* **92**, 4130–4132.

[108] Geusic, J. E, Marcos, H. M, & Uitert, L. G. V. (1964) Laser Oscillations in Nd-doped Yttrium Aluminum, Yttrium Gallium and Gadolinium Garnets. *Applied Physics Letters* **4**, 182–184.

[109] Coherent Inc. (1998) *Operator's Manual of the Coherent Infinity Nd:YAG Laser System* (Santa Clara, CA).

[110] Paltauf, G & Schmidt-Kloiber, H. (1996) Microcavity Dynamics during Laser-Induced Spallation of Liquids and Gels. *Applied Physics A: Materials Science & Processing* **62**, 303–311.

[111] Wray, W. O, Aida, T, & Dyer, R. B. (2002) Photoacoustic Cavitation and Heat Transfer Effects in the Laser-Induced Temperature Jump in Water. *Applied Physics B: Lasers and Optics* **74**, 57–66.

[112] Woody, R. W. (2005) The Exciton Model and the Circular Dichroism of Polypeptides. *Monatshefte für Chemie / Chemical Monthly* **136**, 347–366.

[113] Johnson, W. C. (1990) Protein Secondary Structure and Circular Dichroism: A Practical Guide. *Proteins: Structure, Function, and Bioinformatics* **7**, 205–214.

[114] Davis, S. P, Abrams, M. C, & Brault, J. W. (2001) *Fourier Transform Spectroscopy.* (Academic Press Inc).

[115] Venkateswarlu, P. (1951) The Rotation-Vibration Spectrum of Methyl Cyanide in the Region 1.6μ-20μ. *Journal of Chemical Physics* **19**, 293–298.

[116] Thompson, H. W & Williams, R. L. (1952) The Infra-Red Spectra of Methyl Cyanide and Methyl Isocyanide. *Transactions of the Faraday Society* **48**, 502–513.

[117] Parker, F. W, Nielsen, A. H, & Fletcher, W. H. (1957) The Infrared Absorption Spectrum of Methyl Cyanide Vapor. *Journal of Molecular Spectroscopy* **1**, 107 – 123.

[118] Nakagawa, I & Shimanouchi, T. (1962) Rotation-Vibration Spectra and Rotational, Coriolis Coupling, Centrifugal Distortion and Potential Constants of Methyl Cyanide. *Spectrochimica Acta* **18**, 513–539.

[119] Wiederkehr, R. R & Drickamer, H. G. (1958) Effect of Pressure on Cyanide and Carbonyl Spectra in Solution. *Journal of Chemical Physics* **28**, 311–316.

[120] Brown, T. L. (1960) A Molecular Orbital Model for Infrared Band Intensities: Functional Group Intensities in Aromatic Compounds. *Journal of Physical Chemistry* **64**, 1798–1804.

[121] Brown, T. L & Kubota, M. (1961) Molecular Addition Compounds of Tin(IV) Chloride. 2. Frequency and Intensity of Infrared Nitrile Absorption in Benzonitrile Complexes. *Journal of the American Chemical Society* **83**, 4175–4177.

[122] Augdahl, E & Klaboe, P. (1963) Spectrocopic Studies of Charge Transfer Complexes. 6. Nitriles and Iodine Monochloride. *Spectrochimica Acta* **19**, 1665–1673.

[123] Evans, J. C & Lo, G. Y. S. (1965) Raman and Infrared Studies of Acetonitrile Complexed with Zinc Chloride. *Spectrochimica Acta* **21**, 1033–1038.

[124] Brewer, D. G & Wong, P. T. T. (1966) Studies of Metal Complexes of Pyridine Derivatives. Effects of Coordination upon Infrared Intensity of Functional Group in β-Cyanopyridine. *Canadian Journal of Chemistry* **44**, 1407–1419.

[125] Purcell, K. F & Drago, R. S. (1966) Studies of Bonding in Acetonitrile Adducts. *Journal of the American Chemical Society* **88**, 919–924.

[126] Kawai, K & Kanesaka, I. (1969) Infrared Spectra of Addition Compounds of Hydrogen and Deuterium Cyanides with Some Inorganic Chlorides. *Spectrochimica Acta Part A-Molecular Spectroscopy* **25**, 1265–1273.

[127] Thomas, B. H & Orvillet, W. J. (1969) Molecular Parameters and Bond Structure. 8. Environmental Effects on ν(C$\equiv$N) Bond Stretching Frequencies. *Journal of Molecular Structure* **3**, 191–206.

[128] Breuillard-Alliot, C & Soussen-Jacob, J. (1974) Etude des mouvements moleculaires en milieu liquide a partir du profil des bandes de vibration dans l'infrarouge et des fonctions de correlation. *Molecular Physics* **28**, 905–920.

[129] Loewenschuss, a & Yellin, N. (1975) Secondary Structure of Some Acetonitrile Vibrational Bands. *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy* **31**, 207–212.

[130] Fini, G & Mirone, P. (1976) Secondary Structure of Some Vibrational Bands of Acetonitrile. *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy* **32**, 439–440.

[131] Kawano, Y, Hase, Y, & Sala, O. (1976) Vibrational-Spectra of Adducts of Acetonitrile with Titanium and Tin Tetrachloride. *Journal of Molecular Structure* **30**, 45–55.

[132] Eaton, G, Penanunez, A. S, Symons, M. C. R, Ferrario, M, & McDonald, I. R. (1988) Spectroscopic and Molecular-Dynamics Studies of Solvation of Cyanomethane and Cyanide Ions. *Faraday Discussions* **85**, 237–253.

[133] Eaton, G, Penanunez, a. S, & Symons, M. C. R. (1988) Solvation of Cyanoalkanes [CH_3CN] and [$(CH_3)_3CCN$] - an Infrared and Nuclear Magnetic-Resonance Study. *Journal of the Chemical Society-Faraday Transactions I* **84**, 2181–2193.

[134] Nyquist, R. A. (1990) Solvent-Induced Nitrile Frequency-Shifts - Acetonitrile and Benzonitrile. *Applied Spectroscopy* **44**, 1405–1407.

[135] Benamotz, D, Lee, M. R, Cho, S. Y, & List, D. J. (1992) Solvent and Pressure-Induced Perturbations of the Vibrational Potential Surface of Acetonitrile. *Journal of Chemical Physics* **96**, 8781–8792.

[136] Fawcett, W. R, Liu, G. J, & Kessler, T. E. (1993) Solvent-Induced Frequency-Shifts in the Infrared-Spectrum of Acetonitrile in Organic-Solvents. *Journal of Physical Chemistry* **97**, 9293–9298.

[137] Wang, H, Borguet, E, Yan, E. C. Y, Zhang, D, Gutow, J, & Eisenthal, K. B. (1998) Molecules at Liquid and Solid Surfaces. *Langmuir* **14**, 1472–1477.

[138] Reimers, J. R & Hall, L. E. (1999) The Solvation of Acetonitrile. *Journal of the American Chemical Society* **121**, 3730–3744.

[139] Andrews, S. S & Boxer, S. G. (2000) Vibrational Stark Effects of Nitriles I. Methods and Experimental Results. *Journal of Physical Chemistry A* **104**, 11853–11863.

[140] Andrews, S. S & Boxer, S. G. (2002) Vibrational Stark Effects of Nitriles II. Physical Origins of Stark Effects from Experiment and Perturbation Models. *Journal of Physical Chemistry A* **106**, 469–477.

[141] Suydam, I. T & Boxer, S. G. (2003) Vibrational Stark Effects Calibrate the Sensitivity of Vibrational Probes for Electric Fields in Proteins. *Biochemistry* **42**, 12050–12055.

[142] Rao, Y, Comstock, M, & Eisenthal, K. B. (2006) Absolute Orientation of Molecules at Interfaces. *Journal of Physical Chemistry B* **110**, 1727–1732.

[143] Aschaffenburg, D. J & Moog, R. S. (2009) Probing Hydrogen Bonding Environments: Solvatochromic Effects on the CN Vibration of Benzonitrile. *Journal of Physical Chemistry B* **113**, 12736–12743.

[144] Kim, Y. S & Hochstrasser, R. M. (2005) Chemical Exchange 2D IR of Hydrogen-Bond Making and Breaking. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 11185–11190.

[145] Kurochkin, D. V, Naraharisetty, S. R. G, & Rubtsov, I. V. (2007) A Relaxation-Assisted 2D IR Spectroscopy Method. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 14209–14214.

[146] Ghosh, A, Remorino, A, Tucker, M. J, & Hochstrasser, R. M. (2009) 2D IR Photon Echo Spectroscopy Reveals Hydrogen Bond Dynamics of Aromatic Nitriles. *Chemical Physics Letters* **469**, 325–330.

[147] Ha, J.-H, Lee, K.-K, Park, K.-H, Choi, J.-H, Jeon, S.-J, & Cho, M. (2009)

Integrated and Dispersed Photon Echo Studies of Nitrile Stretching Vibration of 4-Cyanophenol in Methanol. *Journal of Chemical Physics* **130**, 204509–9.

[148] Beattie, I. R & Gilson, T. (1964) Normal Co-ordinate Analysis of $MeCNBX_3$ and its Relevance to Thermodynamic Stability of Co-ordination Compounds. *Journal of the Chemical Society* pp. 2292–2295.

[149] Kurochkin, D. V, Naraharisetty, S. R. G, & Rubtsov, I. V. (2005) Dual-Frequency 2D IR on Interaction of Weak and Strong IR Modes. *Journal of Physical Chemistry A* **109**, 10799–10802.

[150] Miller, F. A & White, W. B. (1960) Infrared and Raman Spectra of Methyl Thiocyanate and Methyl Isothiocyanate. *Zeitschrift für Elektrochemie* **64**, 701–707.

[151] Maienschein-Cline, M. G & Londergan, C. H. (2007) The CN Stretching Band of Aliphatic Thiocyanate is Sensitive to Solvent Dynamics and Specific Solvation. *Journal of Physical Chemistry A* **111**, 10020–10025.

[152] Groving, N & Holm, A. (1965) Alkyl Cyanates. 4. Infrared and Nuclear Magnetic Resonance Spectra. *Acta Chemica Scandinavica* **19**, 443–450.

[153] Tucker, M. J, Kim, Y. S, & Hochstrasser, R. M. (2009) 2D IR Photon Echo Study of the Anharmonic Coupling in the OCN Region of Phenyl Cyanate. *Chemical Physics Letters* **470**, 80–84.

[154] Lieber, E, Rao, C. N. R, Thomas, A. E, Oftedahl, E, Minnis, R, & Nambury, C. V. N. (1963) Infrared Spectra of Acides, Carbamyl Azides and Other Azido Derivatives - Anomalous Splittings of the N_3 Stretching Bands. *Spectrochimica Acta* **19**, 1135–1144.

[155] Sheinker, Y. N, Senyavin, L. B, & Zheltova, V. N. (1965) Position and Intensity of Absorption Band Due to Antisymmetric Valence Vibration of N_3 Group in Infra-Red Spectra of Organic Azides. *Doklady Akademii Nauk SSSR* **160**, 1339.

[156] Dyall, L. K & Kemp, J. E. (1967) Infrared Spectra of Aryl Azides. *Australian Journal of Chemistry* **20**, 1395–1402.

[157] Nydegger, M. W, Dutta, S, & Cheatum, C. M. (2010) Two-Dimensional Infrared Study of 3-Azidopyridine as a Potential Spectroscopic Reporter of Protonation State. *Journal of Chemical Physics* **133**, 134506–8.

[158] Ham, S, Kim, J. H, Lee, H, & Cho, M. H. (2003) Correlation between Electronic and Molecular Structure Distortions and Vibrational Properties. II. Amide I Modes of NMA-$n$$D_2O$ Complexes. *Journal of Chemical Physics* **118**, 3491–3498.

[159] Oh, K.-I, Choi, J.-H, Lee, J.-H, Han, J.-B, Lee, H, & Cho, M. (2008) Nitrile and Thiocyanate IR Probes: Molecular Dynamics Simulation Studies. *Journal of Chemical Physics* **128**, 154504–10.

[160] Choi, J.-H, Oh, K.-I, Lee, H, Lee, C, & Cho, M. (2008) Nitrile and Thiocyanate IR Probes: Quantum Chemistry Calculation Studies and Multivariate Least-Square Fitting Analysis. *Journal of Chemical Physics* **128**, 134506–8.

[161] Cho, M. (2009) Vibrational Solvatochromism and Electrochromism: Coarse-Grained Models and Their Relationships. *Journal of Chemical Physics* **130**, 094505–15.

[162] Lee, H, Choi, J.-H, & Cho, M. (2010) Vibrational Solvatochromism and Electrochromism of Cyanide, Thiocyanate, and Azide Anions in Water. *Physical Chemistry Chemical Physics* **12**, 12658–12669.

[163] Corcelli, S. A, Lawrence, C. P, & Skinner, J. L. (2004) Combined Electronic Structure/Molecular Dynamics Approach for Ultrafast Infrared Spectroscopy of Dilute HOD in Liquid H_2O and D_2O. *Journal of Chemical Physics* **120**, 8107–8117.

[164] Lindquist, B. A & Corcelli, S. A. (2008) Nitrile Groups as Vibrational Probes: Calculations of the C≡N Infrared Absorption Line Shape of Acetonitrile in Water and Tetrahydrofuran. *Journal of Physical Chemistry B* **112**, 6301–6303.

[165] Lindquist, B. A, Haws, R. T, & Corcelli, S. A. (2008) Optimized Quantum Mechanics/Molecular Mechanics Strategies for Nitrile Vibrational Probes: Ace-

tonitrile and *para*-Tolunitrile in Water and Tetrahydrofuran. *Journal of Physical Chemistry B* **112**, 13991–14001.

[166] Lindquist, B. A, Furse, K. E, & Corcelli, S. A. (2009) Nitrile Groups as Vibrational Probes of Biomolecular Structure and Dynamics: An Overview. *Physical Chemistry Chemical Physics* **11**, 8119–32.

[167] Waegele, M. M & Gai, F. (2010) Computational Modeling of the Nitrile Stretching Vibration of 5-Cyanoindole in Water. *Journal of Physical Chemistry Letters* **1**, 781–786.

[168] Böhm, H. J, McDonald, I. R, & Madden, P. A. (1983) An Effective Pair Potential for Liquid Acetonitrile. *Molecular Physics* **49**, 347–360.

[169] Yin, D. (1997) Ph.D. thesis (University of Maryland, Baltimore, MD).

[170] Grabuleda, X, Jaime, C, & Kollman, P. A. (2000) Molecular Dynamics Simulation Studies of Liquid Acetonitrile: New Six-Site Model. *Journal of Computational Chemistry* **21**, 901–908.

[171] Getahun, Z, Huang, C. Y, Wang, T, DeLeon, B, DeGrado, W. F, & Gai, F. (2003) Using Nitrile-Derivatized Amino Acids as Infrared Probes of Local Environment. *Journal of the American Chemical Society* **125**, 405–411.

[172] Schultz, K. C, Supekova, L, Ryu, Y, Xie, J, Perera, R, & Schultz, P. G. (2006) A

Genetically Encoded Infrared Probe. *Journal of the American Chemical Society* **128**, 13984–13985.

[173] Huang, C. Y, Wang, T, & Gai, F. (2003) Temperature Dependence of the CN Stretching Vibration of a Nitrile-Derivatized Phenylalanine in Water. *Chemical Physics Letters* **371**, 731–738.

[174] Fafarman, A. T & Boxer, S. G. (2010) Nitrile Bonds as Infrared Probes of Electrostatics in Ribonuclease S. *Journal of Physical Chemistry B* **114**, 13536–13544.

[175] Jo, H, Culik, R. M, Korendovych, I. V, DeGrado, W. F, & Gai, F. (2010) Selective Incorporation of Nitrile-Based Infrared Probes into Proteins via Cysteine Alkylation. *Biochemistry* **49**, 10354–10356.

[176] Waegele, M. M, Tucker, M. J, & Gai, F. (2009) 5-Cyanotryptophan as an Infrared Probe of Local Hydration Status of Proteins. *Chemical Physics Letters* **478**, 249–253.

[177] Silverman, L. N, Pitzer, M. E, Ankomah, P. O, Boxer, S. G, & Fenlon, E. E. (2007) Vibrational Stark Effect Probes for Nucleic Acids. *Journal of Physical Chemistry B* **111**, 11611–11613.

[178] Fafarman, A. T, Webb, L. J, Chuang, J. I, & Boxer, S. G. (2006) Site-Specific

Conversion of Cysteine Thiols into Thiocyanate Creates an IR Probe for Electric Fields in Proteins. *Journal of the American Chemical Society* **128**, 13356–13357.

[179] Sigala, P. A, Fafarman, A. T, Bogard, P. E, Boxer, S. G, & Herschlag, D. (2007) Do Ligand Binding and Solvent Exclusion Alter the Electrostatic Character within the Oxyanion Hole of an Enzymatic Active Site? *Journal of the American Chemical Society* **129**, 12104–12105.

[180] McMahon, H. A, Alfieri, K. N, Clark, C. A. A, & Londergan, C. H. (2010) Cyanylated Cysteine: A Covalently Attached Vibrational Probe of Protein-Lipid Contacts. *Journal of Physical Chemistry Letters* **1**, 850–855.

[181] Edelstein, L, Stetz, M. A, McMahon, H. A, & Londergan, C. H. (2010) The Effects of α-Helical Structure and Cyanylated Cysteine on Each Other. *Journal of Physical Chemistry B* **114**, 4931–4936.

[182] Bischak, C. G, Longhi, S, Snead, D. M, Costanzo, S, Terrer, E, & Londergan, C. H. (2010) Probing Structural Transitions in the Intrinsically Disordered C-Terminal Domain of the Measles Virus Nucleoprotein by Vibrational Spectroscopy of Cyanylated Cysteines. *Biophysical Journal* **99**, 1676–1683.

[183] Ye, S, Huber, T, Vogel, R, & Sakmar, T. P. (2009) FTIR Analysis of GPCR Activation Using Azido Probes. *Nature Chemical Biology* **5**, 397–399.

[184] Ye, S, Zaitseva, E, Caltabiano, G, Schertler, G. F. X, Sakmar, T. P, Deupi, X, & Vogel, R. (2010) Tracking G-Protein-Coupled Receptor Activation Using Genetically Encoded Infrared Probes. *Nature* **464**, 1386–1389.

[185] Oh, K.-I, Kim, W, Joo, C, Yoo, D.-G, Han, H, Hwang, G.-S, & Cho, M. (2010) Azido Gauche Effect on the Backbone Conformation of β-Azidoalanine Peptides. *Journal of Physical Chemistry B* **114**, 13021–13029.

[186] Kiick, K. L, Saxon, E, Tirrell, D. A, & Bertozzi, C. R. (2002) Incorporation of Azides into Recombinant Proteins for Chemoselective Modification by the Staudinger Ligation. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 19–24.

[187] Taskent-Sezgin, H, Chung, J, Banerjee, P. S, Nagarajan, S, Dyer, R. B, Carrico, I, & Raleigh, D. P. (2010) Azidohomoalanine: A Conformationally Sensitive IR Probe of Protein Folding, Protein Structure, and Electrostatics. *Angewandte Chemie International Edition* **49**, 7473–7475.

[188] Gai, X. S, Fenlon, E. E, & Brewer, S. H. (2010) A Sensitive Multispectroscopic Probe for Nucleic Acids. *Journal of Physical Chemistry B* **114**, 7958–7966.

[189] Tucker, M. J, Gai, X. S, Fenlon, E. E, Brewer, S. H, & Hochstrasser, R. M. (2011) 2D IR photon echo of azido-probes for biomolecular dynamics. *Physical Chemistry Chemical Physics* **13**, 2237–2241.

[190] Dutta, S, Cook, R. J, Houtman, J. C. D, Kohen, A, & Cheatum, C. M. (2010) Characterization of Azido-NAD^+ to Assess Its Potential as a Two-Dimensional Infrared Probe of Enzyme Dynamics. *Analytical Biochemistry* **407**, 241–246.

[191] Tucker, M. J, Oyola, R, & Gai, F. (2006) A Novel Fluorescent Probe for Protein Binding and Folding Studies: *p*-Cyano-Phenylalanine. *Biopolymers* **83**, 571–576.

[192] Serrano, A. L, Troxler, T, Tucker, M. J, & Gai, F. (2010) Photophysics of a Fluorescent Non-Natural Amino Acid: *p*-Cyanophenylalanine. *Chemical Physics Letters* **487**, 303–306.

[193] Taskent-Sezgin, H, Marek, P, Thomas, R, Goldberg, D, Chung, J, Carrico, I, & Raleigh, D. P. (2010) Modulation of *p*-Cyanophenylalanine Fluorescence by Amino Acid Side Chains and Rational Design of Fluorescence Probes of Alpha-Helix Formation. *Biochemistry* **49**, 6290–6295.

[194] Tucker, M. J, Oyola, R, & Gai, F. (2005) Conformational Distribution of a 14-Residue Peptide in Solution: A Fluorescence Resonance Energy Transfer Study. *Journal of Physical Chemistry B* **109**, 4788–4795.

[195] Glasscock, J. M, Zhu, Y, Chowdhury, P, Tang, J, & Gai, F. (2008) Using an Amino Acid Fluorescence Resonance Energy Transfer Pair to Probe Pro-

tein Unfolding: Application to the Villin Headpiece Subdomain and the LysM Domain. *Biochemistry* **47**, 11070–11076.

[196] Rogers, J. M. G, Lippert, L. G, & Gai, F. (2010) Non-Natural Amino Acid Fluorophores for One- and Two-Step Fluorescence Resonance Energy Transfer Applications. *Analytical Biochemistry* **399**, 182–189.

[197] Lide, D. R. (2009) *CRC Handbook of Chemistry and Physics*. (CRC Press, Boca Raton, FL) Vol. 90th ed.

[198] Zhu, Y. (2006) Ph.D. thesis (University of Pennsylvania, Philadelphia, PA).

[199] Kim, P. S & Baldwin, R. L. (1982) Specific Intermediates in the Folding Reactions of Small Proteins and the Mechanism of Protein Folding. *Annual Review of Biochemistry* **51**, 459–489.

[200] Fersht, A. R. (1997) Nucleation Mechanisms in Protein Folding. *Current Opinion in Structural Biology* **7**, 3–9.

[201] Religa, T. L, Johnson, C. M, Vu, D. M, Brewer, S. H, Dyer, R. B, & Fersht, A. R. (2007) The Helix-Turn-Helix Motif as an Ultrafast Independently Folding Domain: The Pathway of Folding of Engrailed Homeodomain. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 9272–9277.

[202] Woll, M. G & Gellman, S. H. (2004) Backbone Thioester Exchange: A New Approach to Evaluating Higher Order Structural Stability in Polypeptides. *Journal of the American Chemical Society* **126**, 11172–11174.

[203] Neumoin, A, Mares, J, Lerch-Bader, M, Bader, R, & Zerbe, O. (2007) Probing the Formation of Stable Tertiary Structure in a Model Miniprotein at Atomic Resolution: Determinants of Stability of a Helical Hairpin. *Journal of the American Chemical Society* **129**, 8811–8817.

[204] Amunson, K. E, Ackels, L, & Kubelka, J. (2008) Site-Specific Unfolding Thermodynamics of a Helix-Turn-Helix Protein. *Journal of the American Chemical Society* **130**, 8146–8147.

[205] Du, D & Gai, F. (2006) Understanding the Folding Mechanism of an α-Helical Hairpin. *Biochemistry* **45**, 13131–13139.

[206] Keire, D. A, Kobayashi, M, Solomon, T. E, & Reeve, J. R. (2000) Solution Structure of Monomeric Peptide YY Supports the Functional Significance of the PP-Fold. *Biochemistry* **39**, 9935–9942.

[207] Lerch, M, Mayrhofer, M, & Zerbe, O. (2004) Structural Similarities of Micelle-Bound Peptide YY (PYY) and Neuropeptide Y (NPY) Are Related to Their Affinity Profiles at the Y Receptors. *Journal of Molecular Biology* **339**, 1153–1168.

[208] Du, D, Tucker, M. J, & Gai, F. (2006) Understanding the Mechanism of β-Hairpin Folding Via ϕ-Value Analysis. *Biochemistry* **45**, 2668–2678.

[209] Du, D, Zhu, Y. J, Huang, C. Y, & Gai, F. (2004) Understanding the Key Factors that Control the Rate of β-Hairpin Folding. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15915–15920.

[210] Fezoui, Y, Weaver, D. L, & Osterhout, J. J. (1994) De Novo Design and Structural Characterization of an α-Helical Hairpin Peptide: A Model System for the Study of Protein Folding Intermediates. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 3675–3679.

[211] Huang, C. Y, Getahun, Z, Zhu, Y. J, Klemke, J. W, DeGrado, W. F, & Gai, F. (2002) Helix Formation Via Conformation Diffusion Search. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 2788–2793.

[212] Williams, S, Causgrove, T. P, Gilmanshin, R, Fang, K. S, Callender, R. H, Woodruff, W. H, & Dyer, R. B. (1996) Fast Events in Protein Folding: Helix Melting and Formation in a Small Peptide. *Biochemistry* **35**, 691–697.

[213] Walsh, S. T. R, Cheng, R. P, Wright, W. W, Alonso, D. O. V, Daggett, V, Vanderkooi, J. M, & DeGrado, W. F. (2003) The Hydration of Amides in Helices; a Comprehensive Picture from Molecular Dynamics, IR, and NMR. *Protein Science* **12**, 520–531.

[214] Mukherjee, S, Chowdhury, P, & Gai, F. (2007) Infrared Study of the Effect of Hydration on the Amide I Band and Aggregation Properties of Helical Peptides. *Journal of Physical Chemistry B* **111**, 4596–4602.

[215] Wang, T, Zhu, Y. J, Getahun, Z, Du, D. G, Huang, C. Y, DeGrado, W. F, & Gai, F. (2004) Length Dependent Helix-Coil Transition Kinetics of Nine Alanine-Based Peptides. *Journal of Physical Chemistry B* **108**, 15301–15310.

[216] Neri, D, Billeter, M, Wider, G, & Wüthrich, K. (1992) NMR Determination of Residual Structure in a Urea-Denatured Protein, the 434-Repressor. *Science* **257**, 1559–1563.

[217] Wilson, G, Hecht, L, & Barron, L. D. (1996) Residual Structure in Unfolded Proteins Revealed by Raman Optical Activity. *Biochemistry* **35**, 12518–12525.

[218] Mok, Y. K, Kay, C. M, Kay, L. E, & Forman-Kay, J. (1999) NOE Data Demonstrating a Compact Unfolded State for an SH3 Domain under Non-Denaturing Conditions. *Journal of Molecular Biology* **289**, 619–638.

[219] Tang, Y. F, Rigotti, D. J, Fairman, R, & Raleigh, D. P. (2004) Peptide Models Provide Evidence for Significant Structure in the Denatured State of a Rapidly Folding Protein: The Villin Headpiece Subdomain. *Biochemistry* **43**, 3264–3272.

[220] Mok, K. H, Kuhn, L. T, Goez, M, Day, I. J, Lin, J. C, Andersen, N. H, & Hore, P. J. (2007) A Pre-Existing Hydrophobic Collapse in the Unfolded State of an Ultrafast Folding Protein. *Nature* **447**, 106–109.

[221] Zhou, R, Eleftheriou, M, Royyuru, A. K, & Berne, B. J. (2007) Destruction of Long-Range Interactions by a Single Mutation in Lysozyme. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5824–5829.

[222] Das, P, King, J. A, & Zhou, R. (2010) β-Strand Interactions at the Domain Interface Critical for the Stability of Human Lens γ D-Crystallin. *Protein Science* **19**, 131–140.

[223] Zhu, Y. J, Fu, X. R, Wang, T, Tamura, A, Takada, S, Savan, J. G, & Gai, F. (2004) Guiding the Search for a Protein's Maximum Rate of Folding. *Chemical Physics* **307**, 99–109.

[224] Ellis, R. J. (2001) Macromolecular Crowding: Obvious but Underappreciated. *Trends in Biochemical Sciences* **26**, 597–604.

[225] Minton, A. P. (2001) The Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media. *Journal of Biological Chemistry* **276**, 10577–10580.

[226] Zhou, H.-X. (2008) Protein Folding in Confined and Crowded Environments. *Archives of Biochemistry and Biophysics* **469**, 76–82.

[227] Minton, A. P. (2005) Models for Excluded Volume Interaction between an Unfolded Protein and Rigid Macromolecular Cosolutes: Macromolecular Crowding and Protein Stability Revisited. *Biophysical Journal* **88**, 971–985.

[228] van den Berg, B, Wain, R, Dobson, C. M, & Ellis, R. J. (2000) Macromolecular Crowding Perturbs Protein Refolding Kinetics: Implications for Folding inside the Cell. *EMBO Journal* **19**, 3870–3875.

[229] Qu, Y. X & Bolen, D. W. (2002) Efficacy of Macromolecular Crowding in Forcing Proteins to Fold. *Biophysical Chemistry* **101**, 155–165.

[230] Friedel, M, Sheeler, D. J, & Shea, J. E. (2003) Effects of Confinement and Crowding on the Thermodynamics and Kinetics of Folding of a Minimalist β-Barrel Protein. *Journal of Chemical Physics* **118**, 8106–8113.

[231] Ping, G, Yuan, J. M, Sun, Z. F, & Wei, Y. (2004) Studies of Effects of Macromolecular Crowding and Confinement on Protein Folding and Protein Stability. *Journal of Molecular Recognition* **17**, 433–440.

[232] Tokuriki, N, Kinjo, M, Negi, S, Hoshino, M, Goto, Y, Urabe, I, & Yomo, T. (2004) Protein Folding by the Effects of Macromolecular Crowding. *Protein Science* **13**, 125–133.

[233] Ai, X. J, Zhou, Z, Bai, Y. W, & Choy, W. Y. (2006) ^{15}N NMR Spin Relaxation

Dispersion Study of the Molecular Crowding Effects on Protein Folding under Native Conditions. *Journal of the American Chemical Society* **128**, 3916–3917.

[234] Yuan, J.-M, Chyan, C.-L, Zhou, H.-X, Chung, T.-Y, Peng, H, Ping, G, & Yang, G. (2008) The Effects of Macromolecular Crowding on the Mechanical Stability of Protein Molecules. *Protein Science* **17**, 2156–2166.

[235] Charlton, L. M, Barnes, C. O, Li, C, Orans, J, Young, G. B, & Pielak, G. J. (2008) Residue-Level Interrogation of Macromolecular Crowding Effects on Protein Stability. *Journal of the American Chemical Society* **130**, 6826–6830.

[236] Homouz, D, Stagg, L, Wittung-Stafshede, P, & Cheung, M. S. (2009) Macromolecular Crowding Modulates Folding Mechanism of α/β Protein Apoflavodoxin. *Biophysical Journal* **96**, 671–680.

[237] Rivera, E, Straub, J, & Thirumalai, D. (2009) Sequence and Crowding Effects in the Aggregation of a 10-Residue Fragment Derived from Islet Amyloid Polypeptide. *Biophysical Journal* **96**, 4552–4560.

[238] Lavalette, D, Tetreau, C, Tourbez, M, & Blouquit, Y. (1999) Microscopic Viscosity and Rotational Diffusion of Proteins in a Macromolecular Environment. *Biophysical Journal* **76**, 2744–2751.

[239] Kozer, N & Schreiber, G. (2004) Effect of Crowding on Protein - Protein Asso-

ciation Rates: Fundamental Differences between Low and High Mass Crowding Agents. *Journal of Molecular Biology* **336**, 763–774.

[240] Kozer, N, Kuttner, Y. Y, Haran, G, & Schreiber, G. (2007) Protein-Protein Association in Polymer Solutions: From Dilute to Semidilute to Concentrated. *Biophysical Journal* **92**, 2139–2149.

[241] Goins, A. B, Sanabria, H, & Waxham, M. N. (2008) Macromolecular Crowding and Size Effects on Probe Microviscosity. *Biophysical Journal* **95**, 5362–5373.

[242] Neuweiler, H, Lollmann, M, Doose, S, & Sauer, M. (2007) Dynamics of Unfolded Polypeptide Chains in Crowded Environment Studied by Fluorescence Correlation Spectroscopy. *Journal of Molecular Biology* **365**, 856–869.

[243] Mukherjee, S, Chowdhury, P, Bunagan, M. R, & Gai, F. (2008) Folding Kinetics of a Naturally Occurring Helical Peptide: Implication of the Folding Speed Limit of Helical Proteins. *Journal of Physical Chemistry B* **112**, 9146–9150.

[244] Ruhlmann, C, Thieme, M, & Helmstedt, M. (2001) Interaction between Dextran and Human Low Density Lipoproteins (LDL) Observed Using Laser Light Scattering. *Chemistry and Physics of Lipids* **110**, 173–181.

[245] Wenner, J. R & Bloomfield, V. A. (1999) Crowding Effects on EcoRV Kinetics and Binding. *Biophysical Journal* **77**, 3234–3241.

[246] Luby-Phelps, K, Castle, P. E, Taylor, D. L, & Lanni, F. (1987) Hindered Diffusion of Inert Tracer Particles in the Cytoplasm of Mouse 3T3 Cells. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 4910–4913.

[247] Venturoli, D & Rippe, B. (2005) Ficoll and Dextran vs. Globular Proteins as Probes for Testing Glomerular Permselectivity: Effects of Molecular Size, Shape, Charge, and Deformability. *American Journal of Physiology-Renal Physiology* **288**, F605–F613.

[248] Fulton, A. B. (1982) How Crowded is the Cytoplasm. *Cell* **30**, 345–347.

[249] Zimmerman, S. B & Trach, S. O. (1991) Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of Escherichia-Coli. *Journal of Molecular Biology* **222**, 599–620.

[250] Kuhlman, B, Yang, H. Y, Boice, J. A, Fairman, R, & Raleigh, D. P. (1997) An Exceptionally Stable Helix from the Ribosomal Protein L9: Implications for Protein Folding and Stability. *Journal of Molecular Biology* **270**, 640–647.

[251] Phillip, Y, Sherman, E, Haran, G, & Schreiber, G. (2009) Common Crowding Agents Have only a Small Effect on Protein-Protein Interactions. *Biophysical Journal* **97**, 875–885.

[252] Jacob, M & Schmid, F. X. (1999) Protein Folding as a Diffusional Process. *Biochemistry* **38**, 13773–13779.

[253] Bhattacharyya, R. P & Sosnick, T. R. (1999) Viscosity Dependence of the Folding Kinetics of a Dimeric and Monomeric Coiled Coil. *Biochemistry* **38**, 2601–2609.

[254] Hagen, S. J, Qiu, L. L, & Pabit, S. A. (2005) Diffusional Limits to the Speed of Protein Folding: Fact or Friction? *Journal of Physics-Condensed Matter* **17**, S1503–S1514.

[255] Frauenfelder, H, Fenimore, P. W, Chen, G, & McMahon, B. H. (2006) Protein Folding is Slaved to Solvent Motions. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15469–15472.

[256] Barshtein, G, Almagor, A, Yedgar, S, & Gavish, B. (1995) Inhomogeneity of Viscous Aqueous-Solutions. *Physical Review E* **52**, 555–557.

[257] Klimov, D. K & Thirumalai, D. (1997) Viscosity Dependence of the Folding Rates of Proteins. *Physical Review Letters* **79**, 317–320.

[258] Masuda, A, Ushida, K, Koshino, H, Yamashita, K, & Kluge, T. (2001) Novel Distance Dependence of Diffusion Constants in Hyaluronan Aqueous Solution Resulting from Its Characteristic Nano-Microstructure. *Journal of the American Chemical Society* **123**, 11468–11471.

[259] Masuda, A, Ushida, K, Nishimura, G, Kinjo, M, Tamura, M, Koshino, H, Yamashita, K, & Kluge, T. (2004) Experimental Evidence of Distance-Dependent Diffusion Coefficients of a Globular Protein Observed in Polymer Aqueous Solution Forming a Network Structure on Nanometer Scale. *Journal of Chemical Physics* **121**, 10787–10793.

[260] Ogston, A. G, Preston, B. N, Wells, J. D, Ogston, A. G, Preston, B. N, Snowden, J. M, & Wells, J. D. (1973) Transport of Compact Particles through Solutions of Chain-Polymers. *Proceedings of the Royal Society of London Series A - Mathematical Physical and Engineering Sciences* **333**, 297–316.

[261] Rief, M, Fernandez, J. M, & Gaub, H. E. (1998) Elastically Coupled Two-Level Systems as a Model for Biopolymer Extensibility. *Physical Review Letters* **81**, 4764–4767.

[262] Bohrer, M. P, Patterson, G. D, & Carroll, P. J. (1984) Hindered Diffusion of Dextran and Ficoll in Microporous Membranes. *Macromolecules* **17**, 1170–1173.

[263] Reshetnyak, Y. K, Segala, M, Andreev, O. A, & Engelman, D. M. (2007) A Monomeric Membrane Peptide that Lives in Three Worlds: In Solution, Attached to, and Inserted across Lipid Bilayers. *Biophysical Journal* **93**, 2363–2372.

[264] Bunagan, M. R, Gao, J, Kelly, J. W, & Gai, F. (2009) Probing the Folding

Transition State Structure of the Villin Headpiece Subdomain Via Side Chain and Backbone Mutagenesis. *Journal of the American Chemical Society* **131**, 7470–7476.

[265] Guo, L, Chowdhury, P, Fang, J, & Gai, F. (2007) Heterogeneous and Anomalous Diffusion inside Lipid Tubules. *Journal of Physical Chemistry B* **111**, 14244–14249.

[266] Haustein, E & Schwille, P. (2007) Fluorescence Correlation Spectroscopy: Novel Variations of an Established Technique. *Annual Review of Biophysics and Biomolecular Structure* **36**, 151–169.

[267] van den Berg, B, Ellis, R. J, & Dobson, C. M. (1999) Effects of Macromolecular Crowding on Protein Folding and Aggregation. *EMBO Journal* **18**, 6927–6933.

[268] Sasahara, K, Mc Phie, P, & Minton, A. P. (2003) Effect of Dextran on Protein Stability and Conformation Attributed to Macromolecular Crowding. *Journal of Molecular Biology* **326**, 1227–1237.

[269] Ping, G. H, Yang, G. L, & Yuan, H. M. (2006) Depletion Force from Macromolecular Crowding Enhances Mechanical Stability of Protein Molecules. *Polymer* **47**, 2564–2570.

[270] Stagg, L, Zhang, S.-Q, Cheung, M. S, & Wittung-Stafshede, P. (2007) Molecular Crowding Enhances Native Structure and Stability of α/β Protein Flavodoxin.

Proceedings of the National Academy of Sciences of the United States of America **104**, 18976–18981.

[271] Roberts, A & Jackson, S. E. (2007) Destabilised Mutants of Ubiquitin Gain Equal Stability in Crowded Solutions. *Biophysical Chemistry* **128**, 140–149.

[272] Batra, J, Xu, K, & Zhou, H.-X. (2009) Nonadditive Effects of Mixed Crowding on Protein Stability. *Proteins-Structure Function and Bioinformatics* **77**, 133–138.

[273] Batra, J, Xu, K, Qin, S, & Zhou, H.-X. (2009) Effect of Macromolecular Crowding on Protein Binding Stability: Modest Stabilization and Significant Biological Consequences. *Biophysical Journal* **97**, 906–911.

[274] Mukherjee, S, Waegele, M. M, Chowshury, P, Guo, L, & Gai, F. (2009) Effect of Macromolecular Crowding on Protein Folding Dynamics at the Secondary Structure Level. *Journal of Molecular Biology* **393**, 227–236.

[275] Hong, J & Gierasch, L. M. (2010) Macromolecular Crowding Remodels the Energy Landscape of a Protein by Favoring a More Compact Unfolded State. *Journal of the American Chemical Society* **132**, 10445–10452.

[276] Miklos, A. C, Li, C, Sharaf, N. G, & Pielak, G. J. (2010) Volume Exclusion and Soft Interaction Effects on Protein Stability under Crowded Conditions. *Biochemistry* **49**, 6984–6991.

[277] Christiansen, A, Wang, Q, Samiotakis, A, Cheung, M. S, & Wittung-Stafshede, P. (2010) Factors Defining Effects of Macromolecular Crowding on Protein Stability: An *in vitro/in silico* Case Study Using Cytochrome C. *Biochemistry* **49**, 6519–6530.

[278] Kinjo, A. R & Takada, S. (2002) Effects of Macromolecular Crowding on Protein Folding and Aggregation Studied by Density Functional Theory: Statics. *Physical Review E* **66**, 031911–9.

[279] Minh, D. D. L, Chang, C.-e, Trylska, J, Tozzini, V, & McCammon, J. A. (2006) The Influence of Macromolecular Crowding on HIV-1 Protease Internal Dynamics. *Journal of the American Chemical Society* **128**, 6006–6007.

[280] Wieczorek, G & Zielenkiewicz, P. (2008) Influence of Macromolecular Crowding on Protein-Protein Association Rates - A Brownian Dynamics Study. *Biophysical Journal* **95**, 5030–5036.

[281] Pincus, D. L & Thirumalai, D. (2009) Crowding Effects on the Mechanical Stability and Unfolding Pathways of Ubiquitin. *Journal of Physical Chemistry B* **113**, 359–368.

[282] Qin, S & Zhou, H.-X. (2009) Atomistic Modeling of Macromolecular Crowding Predicts Modest Increases in Protein Folding and Binding Stability. *Biophysical Journal* **97**, 12–19.

[283] Tsao, D & Dokholyan, N. V. (2010) Macromolecular Crowding Induces Polypeptide Compaction and Decreases Folding Cooperativity. *Physical Chemistry Chemical Physics* **12**, 3491–3500.

[284] Mittal, J & Best, R. B. (2010) Dependence of Protein Folding Stability and Dynamics on the Density and Composition of Macromolecular Crowders. *Biophysical Journal* **98**, 315–320.

[285] Minton, A. P. (1981) Excluded Volume as a Determinant of Macromolecular Structure and Reactivity. *Biopolymers* **20**, 2093–2120.

[286] Zhou, Y. Q & Hall, C. K. (1996) Solute Excluded-Volume Effects on the Stability of Globular Proteins: A Statistical Thermodynamic Theory. *Biopolymers* **38**, 273–284.

[287] Minton, A. P. (2000) Effect of a Concentrated "Inert" Macromolecular Cosolute on the Stability of a Globular Protein with Respect to Denaturation by Heat and by Chaotropes: A Statistical-Thermodynamical Model. *Biophysical Journal* **78**, 101–109.

[288] Zhou, H.-X. (2008) Effect of Mixed Macromolecular Crowding Agents on Protein Folding. *Proteins-Structure Function and Bioinformatics* **72**, 1109–1113.

[289] Gee, M. B & Smith, P. E. (2009) Kirkwood-Buff Theory of Molecular and

Protein Association, Aggregation, and Cellular Crowding. *Journal of Chemical Physics* **131**, 165101–11.

[290] Qin, S & Zhou, H.-X. (2010) Generalized Fundamental Measure Theory for Atomistic Modeling of Macromolecular Crowding. *Physical Review E* **81**, 031919–5.

[291] Elcock, A. H. (2010) Models of Macromolecular Crowding Effects and the Need for Quantitative Comparisons with Experiment. *Current Opinion in Structural Biology* **20**, 196–206.

[292] De Gennes, P.-G. (1979) *Scaling Concepts in Polymer Physics*. (Cornell University Press, Ithaca, NY).

[293] Sakaue, T & Raphael, E. (2006) Polymer Chains in Confined Spaces and Flow-Injection Problems: Some Remarks. *Macromolecules* **39**, 2621–2628.

[294] Rathore, N, Knotts, T. A, & de Pablo, J. J. (2006) Confinement Effects on the Thermodynamics of Protein Folding: Monte Carlo Simulations. *Biophysical Journal* **90**, 1767–1773.

[295] Mittal, J & Best, R. B. (2008) Thermodynamics and Kinetics of Protein Folding under Confinement. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20233–20238.

[296] Takagi, F, Koga, N, & Takada, S. (2003) How Protein Thermodynamics and Folding Mechanisms Are Altered by the Chaperonin Cage: Molecular Simulations. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11367–11372.

[297] Wang, W, Xu, W.-X, Levy, Y, Trizac, E, & Wolynes, P. G. (2009) Confinement Effects on the Kinetics and Thermodynamics of Protein Dimerization. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5517–5522.

[298] Shaw, M. R & Thirumalai, D. (1991) Free Polymer in a Colloidal Solution. *Physical Review A* **44**, R4797–R4800.

[299] Biben, T & Hansen, J. P. (1991) Phase-Separation of Asymmetric Binary Hard-Sphere Fluids. *Physical Review Letters* **66**, 2215–2218.

[300] Vijay-Kumar, S, Bugg, C. E, & Cook, W. J. (1987) Structure of Ubiquitin Refined at 1.8 Å Resolution. *Journal of Molecular Biology* **194**, 531–544.

[301] Khorasanizadeh, S, Peters, I. D, Butt, T. R, & Roder, H. (1993) Folding and Stability of a Tryptophan-Containing Mutant of Ubiquitin. *Biochemistry* **32**, 7054–7063.

[302] Wintrode, P. l, Makhatadze, G. I, & Privalov, P. L. (1994) Thermodynamics of Ubiquitin Unfolding. *Proteins-Structure Function and Genetics* **18**, 246–253.

[303] Krantz, B. A & Sosnick, T. R. (2000) Distinguishing between Two-State and Three-State Models for Ubiquitin Folding. *Biochemistry* **39**, 11696–11701.

[304] Colley, C. S, Clark, I. P, Griffiths-Jones, S. R, George, M. W, & Searle, M. S. (2000) Steady State and Time-Resolved IR Spectroscopy of the Native and Unfolded States of Bovine Ubiquitin: Protein Stability and Temperature-Jump Kinetic Measurements of Protein Folding at low pH. *Chemical Communications* pp. 1493–1494.

[305] Xu, Y, Oyola, R, & Gai, F. (2003) Infrared Study of the Stability and Folding Kinetics of a 15-Residue β-Hairpin. *Journal of the American Chemical Society* **125**, 15388–15394.

[306] Henry, E. R & Hofrichter, J. (1992) Singular Value Decomposition - Application to Analysis of Experimental Data. *Methods in Enzymology* **210**, 129–192.

[307] Armstrong, J. K, Wenby, R. B, Meiselman, H. J, & Fisher, T. C. (2004) The Hydrodynamic Radii of Macromolecules and Their Effect on Red Blood Cell Aggregation. *Biophysical Journal* **87**, 4259–4270.

[308] Ogston, A. G & Preston, B. N. (1979) Molecular Compression of Dextran. *Biochemical Journal* **183**, 1–9.

[309] Liu, Z, Weng, W, Bookchin, R. M, Lew, V. L, & Ferrone, F. A. (2008) Free

Energy of Sickle Hemoglobin Polymerization: A Scaled-Particle Treatment for Use with Dextran as a Crowding Agent. *Biophysical Journal* **94**, 3629–3634.

[310] Efron, B. (1981) Nonparametric Standard Errors and Confidence Intervals. *Canadian Journal of Statistics* **9**, 139–158.

[311] Laurence, T. A, Kong, X. X, Jager, M, & Weiss, S. (2005) Probing Structural Heterogeneities and Fluctuations of Nucleic Acids and Denatured Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17348–17353.

[312] McCarney, E. R, Werner, J. H, Bernstein, S. L, Ruczinski, I, Makarov, D. E, Goodwin, P. M, & Plaxco, K. W. (2005) Site-Specific Dimensions across a Highly Denatured Protein; A Single Molecule Study. *Journal of Molecular Biology* **352**, 672–682.

[313] Hoffmann, A, Kane, A, Nettels, D, Hertzog, D. E, Baumgaertel, P, Lengefeld, J, Reichardt, G, Horsley, D. A, Seckler, R, Bakajin, O, & Schuler, B. (2007) 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.

[314] Hofmann, H, Golbik, R. P, Ott, M, Huebner, C. G, & Ulbrich-Hofmann, R.

(2008) Coulomb Forces Control the Density of the Collapsed Unfolded State of Barstar. *Journal of Molecular Biology* **376**, 597–605.

[315] Nettels, D, Mueller-Spaeth, S, Kuester, F, Hofmann, H, Haenni, D, Rueegger, S, Reymond, L, Hoffmann, A, Kubelka, J, Heinz, B, Gast, K, Best, R. B, & Schuler, B. (2009) Single-Molecule Spectroscopy of the Temperature-Induced Collapse of Unfolded Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20740–20745.

[316] Müller-Späth, S, Soranno, A, Hirschfeld, V, Hofmann, H, Rueegger, S, Reymond, L, Nettels, D, & Schuler, B. (2010) Charge Interactions Can Dominate the Dimensions of Intrinsically Disordered Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14609–14614.

[317] Bieri, O, Wirz, J, Hellrung, B, Schutkowski, M, Drewello, M, & Kiefhaber, T. (1999) The Speed Limit for Protein Folding Measured by Triplet-Triplet Energy Transfer. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9597–9601.

[318] Hudgins, R. R, Huang, F, Gramlich, G, & Nau, W. M. (2002) A Fluorescence-Based Method for Direct Measurement of Submicrosecond Intramolecular Contact Formation in Biopolymers: An Exploratory Study with Polypeptides. *Journal of the American Chemical Society* **124**, 556–564.

[319] Lapidus, L. J, Steinbach, P. J, Eaton, W. A, Szabo, A, & Hofrichter, J. (2002) Effects of Chain Stiffness on the Dynamics of Loop Formation in Polypeptides. Appendix: Testing a 1-Dimensional Diffusion Model for Peptide Dynamics. *Journal of Physical Chemistry B* **106**, 11628–11640.

[320] Krieger, F, Fierz, B, Bieri, O, Drewello, M, & Kiefhaber, T. (2003) Dynamics of Unfolded Polypeptide Chains as Model for the Earliest Steps in Protein Folding. *Journal of Molecular Biology* **332**, 265–274.

[321] Möglich, A, Joder, K, & Kiefhaber, T. (2006) End-to-End Distance Distributions and Intrachain Diffusion Constants in Unfolded Polypeptide Chains Indicate Intramolecular Hydrogen Bond Formation. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12394–12399.

[322] Buscaglia, M, Lapidus, L. J, Eaton, W. A, & Hofrichter, J. (2006) Effects of Denaturants on the Dynamics of Loop Formation in Polypeptides. *Biophysical Journal* **91**, 276–288.

[323] Fierz, B, Satzger, H, Root, C, Gilch, P, Zinth, W, & Kiefhaber, T. (2007) Loop Formation in Unfolded Polypeptide Chains on the Picoseconds to Microseconds Time Scale. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2163–2168.

[324] Sahoo, H, Roccatano, D, Hennig, A, & Nau, W. M. (2007) A 10-Ångström

Spectroscopic Ruler Applied to Short Polyprolines. *Journal of the American Chemical Society* **129**, 9762–9772.

[325] Singh, V. R & Lapidus, L. J. (2008) The Intrinsic Stiffness of Polyglutamine Peptides. *Journal of Physical Chemistry B* **112**, 13172–13176.

[326] Soranno, A, Longhi, R, Bellini, T, & Buscaglia, M. (2009) Kinetics of Contact Formation and End-to-End Distance Distributions of Swollen Disordered Peptides. *Biophysical Journal* **96**, 1515–1528.

[327] Kohn, J. E, Millett, I. S, Jacob, J, Zagrovic, B, Dillon, T. M, Cingel, N, Dothager, R. S, Seifert, S, Thiyagarajan, P, Sosnick, T. R, Hasan, M. Z, Pande, V. S, Ruczinski, I, Doniach, S, & Plaxco, K. W. (2004) 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.

[328] Lee, J. C, Gray, H. B, & Winkler, J. R. (2005) Tertiary Contact Formation in α-Synuclein Probed by Electron Transfer. *Journal of the American Chemical Society* **127**, 16388–16389.

[329] Tzul, F. O & Bowler, B. E. (2010) Denatured States of Low-Complexity Polypeptide Sequences Differ Dramatically from Those of Foldable Sequences. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11364–11369.

[330] des Cloizeaux, J & Jannik, G. (1989) *Polymers in Solution: Their Modelling and Structure*. (Oxford University Press, Oxford, UK).

[331] Yue, K. Z & Dill, K. a. (1993) Sequence-Structure Relationships in Proteins and Copolymers. *Physical Review E* **48**, 2267–2278.

[332] Taskent-Sezgin, H, Chung, J, Patsalo, V, Miyake-Stoner, S. J, Miller, A. M, Brewer, S. H, Mehl, R. A, Green, D. F, Raleigh, D. P, & Carrico, I. (2009) Interpretation of *p*-Cyanophenylalanine Fluorescence in Proteins in Terms of Solvent Exposure and Contribution of Side-Chain Quenchers: A Combined Fluorescence, IR and Molecular Dynamics Study. *Biochemistry* **48**, 9040–9046.

[333] Kyte, J & Doolittle, R. f. (1982) A Simple Method for Displaying the Hydropathic Character of a Protein. *Journal of Molecular Biology* **157**, 105–132.

[334] Khmelnitsky, Y. L, Mozhaev, V. V, Belova, A. B, Sergeeva, M. V, & Martinek, K. (1991) Denaturation Capacity: A New Quantitative Criterion for Selection of Organic Solvents as Reaction Media in Biocatalysis. *European Journal of Biochemistry* **198**, 31–41.

[335] Haas, E, Wilchek, M, Katchalskikatzir, E, & Steinberg, I. Z. (1975) Distribution of End-to-End Distances of Oligopeptides in Solution as Estimated by Energy-Transfer. *Proceedings of the National Academy of Sciences of the United States of America* **72**, 1807–1811.

[336] Gobush, W, Stockmayer, W. H, Yamakawa, H, & Magee, W. S. (1972) Statistical Mechanisms of Wormlike Chains. I. Asymptotic Behavior. *Journal of Chemical Physics* **57**, 2839–5.

[337] Tanford, C, Kawahara, K, & Lapanje, S. (1966) Proteins in 6M Guanidine Hydrochloride - Demonstration of Random Coil Behavior. *Journal of Biological Chemistry* **241**, 1921–1923.

[338] Wilkins, D. K, Grimshaw, S. B, Receveur, V, Dobson, C. M, Jones, J. A, & Smith, L. J. (1999) Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient Nmr Techniques. *Biochemistry* **38**, 16424–16431.

[339] Tcherkasskaya, O & Uversky, V. N. (2001) Denatured Collapsed States in Protein Folding: Example of Apomyoglobin. *Proteins-Structure Function and Genetics* **44**, 244–254.

[340] Domb, C & Hioe, F. T. (1969) Correlations in a Self-Avoiding Walk. *Journal of Chemical Physics* **51**, 1920–9.

[341] Rosenbluth, M. N & Rosenbluth, A. W. (1955) Monte-Carlo Calculation of the Average Extension of Molecular Chains. *Journal of Chemical Physics* **23**, 356–359.

[342] Redner, S. (1980) Distribution-Functions in the Interior of Polymer-Chains. *Journal of Physics A - Mathematical and General* **13**, 3525–3541.

[343] Brant, D. A & Flory, P. J. (1965) Configuration of Random Polypeptide Chains. II. Theory. *Journal of the American Chemical Society* **87**, 2791–2800.

[344] Brant, D. A & Flory, P. J. (1965) Configuration of Random Polypeptide Chains. I. Experimental Results. *Journal of the American Chemical Society* **87**, 2788–2791.

[345] Miller, W. G, Brant, D. A, & Flory, P. J. (1967) Random Coil Configurations of Polypeptide Copolymers. *Journal of Molecular Biology* **23**, 67–80.

[346] Sahoo, H, Roccatano, D, Zacharias, M, & Nau, W. M. (2006) Distance Distributions of Short Polypeptides Recovered by Fluorescence Resonance Energy Transfer in the 10 Å Domain. *Journal of the American Chemical Society* **128**, 8118–8119.

[347] Zhou, H. X. (2001) Loops in Proteins Can be Modeled as Worm-like Chains. *Journal of Physical Chemistry B* **105**, 6763–6766.

[348] Kohn, J. E, Gillespie, B, & Plaxco, K. W. (2009) Non-Sequence-Specific Interactions Can Account for the Compaction of Proteins Unfolded under "Native" Conditions. *Journal of Molecular Biology* **394**, 343–350.

[349] Wilhelm, J & Frey, E. (1996) Radial Distribution Function of Semiflexible Polymers. *Physical Review Letters* **77**, 2581–2584.

[350] Lakowicz, J. R. (1999) *Principles of Fluorescence Spectroscopy.* (Academic Press Inc, New York, NY).

[351] Weinryb, I & Steiner, R. F. (1970) Luminescence of Tryptophan and Tyrosine Residues of Papain in Solution. *Biochemistry* **9**, 135–146.

[352] Dalosto, S. D, Vanderkooi, J. M, & Sharp, K. A. (2004) Vibrational Stark Effects on Carbonyl, Nitrile, and Nitrosyl Compounds Including Heme Ligands, CO, CN, and NO, Studied with Density Functional Theory. *Journal of Physical Chemistry B* **108**, 6450–6457.

[353] Tucker, M. J, Getahun, Z, Nanda, V, DeGrado, W. F, & Gai, F. (2004) A New Method for Determining the Local Environment and Orientation of Individual Side Chains of Membrane-Binding Peptides. *Journal of the American Chemical Society* **126**, 5078–5079.

[354] Weeks, C. L, Polishchuk, A, Getahun, Z, DeGrado, W. F, & Spiro, T. G. (2008) Investigation of an Unnatural Amino Acid for Use as a Resonance Raman Probe: Detection Limits and Solvent and Temperature Dependence of the νC$\equiv$N Band of 4-Cyanophenylalanine. *Journal of Raman Spectroscopy* **39**, 1606–1613.

[355] Mukherjee, S, Chowdhury, P, DeGrado, W. F, & Gai, F. (2007) Site-Specific Hydration Status of an Amphipathic Peptide in AOT Reverse Micelles. *Langmuir* **23**, 11174–11179.

[356] Zhang, D & Decatur, S. M. (2007) Cyanophenylalanine as a Vibrational Probe for the Aggregation of Aβ16-22 Peptides. *Biophysical Journal* p. 559A.

[357] Zou, H, Liu, J, & Blasie, J. K. (2009) Mechanism of Interaction between the General Anesthetic Halothane and a Model Ion Channel Protein, III: Molecular Dynamics Simulation Incorporating a Cyanophenylalanine Spectroscopic Probe. *Biophysical Journal* **96**, 4188–4199.

[358] Liu, J, Strzalka, J, Tronin, A, Johansson, J. S, & Blasie, J. K. (2009) Mechanism of Interaction between the General Anesthetic Halothane and a Model Ion Channel Protein, II: Fluorescence and Vibrational Spectroscopy Using a Cyanophenylalanine Probe. *Biophysical Journal* **96**, 4176–4187.

[359] Fang, C, Bauman, J. D, Das, K, Remorino, A, Arnold, E, & Hochstrasser, R. M. (2008) Two-Dimensional Infrared Spectra Reveal Relaxation of the Nonnucleoside Inhibitor TMC278 Complexed with HIV-1 Reverse Transcriptase. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1472–1477.

[360] Webb, L. J & Boxer, S. G. (2008) Electrostatic Fields Near the Active Site

of Human Aldose Reductase: 1. New Inhibitors and Vibrational Stark Effect Measurements. *Biochemistry* **47**, 1588–1598.

[361] Krummel, A. T & Zanni, M. T. (2008) Evidence for Coupling between Nitrile Groups Using Dna Templates: A Promising New Method for Monitoring Structures with Infrared Spectroscopy. *Journal of Physical Chemistry B* **112**, 1336–1338.

[362] Watson, M. D, Gai, X. S, Gillies, A. T, Brewer, S. H, & Fenlon, E. E. (2008) A Vibrational Probe for Local Nucleic Acid Environments: 5-Cyano-2 '-Deoxyuridine. *Journal of Physical Chemistry B* **112**, 13188–13192.

[363] Twine, S. M & Szabo, A. G. (2003) Fluorescent Amino Acid Analogs. *Methods in Enzymology* **360**, 104–127.

[364] Ladokhin, A. S, Selsted, M. E, & White, S. H. (1997) Bilayer Interactions of Indolicidin, a Small Antimicrobial Peptide Rich in Tryptophan, Proline, and Basic Amino Acids. *Biophysical Journal* **72**, 794–805.

[365] Rozek, A, Friedrich, C. L, & Hancock, R. E. W. (2000) Structure of the Bovine Antimicrobial Peptide Indolicidin Bound to Dodecylphosphocholine and Sodium Dodecyl Sulfate Micelles. *Biochemistry* **39**, 15765–15774.

[366] Khandelia, H & Kaznessis, Y. N. (2007) Cation-π Interactions Stabilize the

Structure of the Antimicrobial Peptide Indolicidin Near Membranes: Molecular Dynamics Simulations. *Journal of Physical Chemistry B* **111**, 242–250.

[367] Hsu, J. C. Y & Yip, C. M. (2007) Molecular Dynamics Simulations of Indolicidin Association with Model Lipid Bilayers. *Biophysical Journal* **92**, L100–L102.

[368] Li, T, Hassanali, A. A, Kao, Y.-T, Zhong, D, & Singer, S. J. (2007) Hydration Dynamics and Time Scales of Coupled Water-Protein Fluctuations. *Journal of the American Chemical Society* **129**, 3376–3382.

[369] Aprilakis, K. N, Taskent, H, & Raleighj, D. P. (2007) Use of the Novel Fluorescent Amino Acid *p*-Cyanophenylalanine Offers a Direct Probe of Hydrophobic Core Formation during the Folding of the N-Terminal Domain of the Ribosomal Protein L9 and Provides Evidence for Two-State Folding. *Biochemistry* **46**, 12308–12313.

[370] Tang, J, Yin, H, Qiu, J, Tucker, M. J, DeGrado, W. F, & Gai, F. (2009) Using Two Fluorescent Probes to Dissect the Binding, Insertion, and Dimerization Kinetics of a Model Membrane Peptide. *Journal of the American Chemical Society* **131**, 3816–3817.

[371] Eaton, G, Penanunez, A. S, & Symons, M. C. R. (1988) Solvation of Cyanoalkanes [CH_3CN and $(CH_3)_3CCN$] - an Infrared and Nuclear Magnetic-Resonance Study. *Journal of the Chemical Society-Faraday Transactions I* **84**, 2181–2193.

[372] Corcelli, S. A & Skinner, J. L. (2005) Infrared and Raman Line Shapes of Dilute HOD in Liquid H_2O and D_2O from 10 to 90 °C. *Journal of Physical Chemistry A* **109**, 6154–6165.

[373] Hayashi, T, Jansen, T. L, Zhuang, W, & Mukamel, S. (2005) Collective Solvent Coordinates for the Infrared Spectrum of HOD in D_2O Based on an *Ab Initio* Electrostatic Map. *Journal of Physical Chemistry A* **109**, 64–82.

[374] Jansen, T, Hayashi, T, Zhuang, W, & Mukamel, S. (2005) Stochastic Liouville Equations for Hydrogen-Bonding Fluctuations and Their Signatures in Two-Dimensional Vibrational Spectroscopy of Water. *Journal of Chemical Physics* **123**, 114504–11.

[375] Macias, A. T & MacKerell, A. D. (2005) CH/π Interactions Involving Aromatic Amino Acids: Refinement of the Charmm Tryptophan Force Field. *Journal of Computational Chemistry* **26**, 1452–1463.

[376] Frisch, M. J, Trucks, G. W, Schlegel, H. B, Scuseria, G. E, Robb, M. A, Cheeseman, J. R, Montgomery, J A, J, Vreven, T.; Kudin, K. N, Burant, J. C, & et al. (2003) *Gaussian 03* (Gaussian, Inc., Pittsburgh, PA), revision b.05 edition.

[377] Breneman, C. M & Wiberg, K. B. (1990) Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials - the Need for High Sam-

pling Density in Formamide Conformational-Analysis. *Journal of Computational Chemistry* **11**, 361–373.

[378] Bayly, C. I, Cieplak, P, Cornell, W. D, & Kollman, P. A. (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the RESP Model. *Journal of Physical Chemistry* **97**, 10269–10280.

[379] Phillips, J. C, Braun, R, Wang, W, Gumbart, J, Tajkhorshid, E, Villa, E, Chipot, C, Skeel, R. D, Kale, L, & Schulten, K. (2005) Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781–1802.

[380] Zhao, Y & Truhlar, D. G. (2004) Hybrid Meta Density Functional Theory Methods for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions: The MPW1B95 and MPWB1K Models and Comparative Assessments for Hydrogen Bonding and van der Waals Interactions. *Journal of Physical Chemistry A* **108**, 6908–6918.

[381] Zhang, R. B, Somers, K. R. F, Kryachko, E. S, Nguyen, M. T, Zeegers-Huyskens, T. S, & Ceulemans, A. (2005) Hydrogen Bonding to π-Systems of Indole and 1-Methylindole: Is There Any OH–Phenyl Bond? *Journal of Physical Chemistry A* **109**, 8028–8034.

[382] Humphrey, W, Dalke, A, & Schulten, K. (1996) VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **14**, 33–38.

[383] Varsanyi, G & Lang, L. (1974) *Assignments for Vibrational Spectra of Seven Hundered Benzene Derivatives*. (Wiley, New York).

[384] Speranskiy, K & Kurnikova, M. (2004) Accurate Theoretical Prediction of Vibrational Frequencies in an Inhomogeneous Dynamic Environment: A Case Study of a Glutamate Molecule in Water Solution and in a Protein-Bound Form. *Journal of Chemical Physics* **121**, 1516–1524.

[385] Jorgensen, W. L, Chandrasekhar, J, Madura, J. D, Impey, R. W, & Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **79**, 926–935.

[386] Kolano, C, Helbing, J, Kozinski, M, Sander, W, & Hamm, P. (2006) Watching Hydrogen-Bond Dynamics in a β-Turn by Transient Two-Dimensional Infrared Spectroscopy. *Nature* **444**, 469–472.

[387] Luo, P & Baldwin, R. L. (1997) Mechanism of Helix Induction by Trifluoroethanol: A Framework for Extrapolating the Helix-Forming Properties of Peptides from Trifluoroethanol/Water Mixtures Back to Water. *Biochemistry* **36**, 8413–8421.

[388] Prompers, J. J, Hilbers, C. W, & Pepermans, H. A. M. (1999) Tryptophan Mediated Photoreduction of Disulfide Bond Causes Unusual Fluorescence Behaviour of *Fusarium Solani Pisi* Cutinase. *FEBS Letters* **456**, 409 – 416.

[389] Vanhooren, A, Devreese, B, Vanhee, K, Van Beeumen, J, & Hanssens, I. (2002) Photoexcitation of Tryptophan Groups Induces Reduction of Two Disulfide Bonds in Goat α-Lactalbumin. *Biochemistry* **41**, 11035–11043.

[390] Neves-Petersen, M. T, Gryczynski, Z, Lakowicz, J, Fojan, P, Pedersen, S, Petersen, E, & Bjørn Petersen, S. (2002) High Probability of Disrupting a Disulphide Bridge Mediated by an Endogenous Excited Tryptophan Residue. *Protein Science* **11**, 588–600.

[391] Vanhooren, A, De Vriendt, K, Devreese, B, Chedad, A, Sterling, A, Van Dael, H, Van Beeumen, J, & Hanssens, I. (2006) Selectivity of Tryptophan Residues in Mediating Photolysis of Disulfide Bridges in Goat α-Lactalbumin. *Biochemistry* **45**, 2085–2093.

[392] Kehoe, J. J, Remondetto, G. E, Subirade, M, Morris, E. R, & Brodkorb, A. (2008) Tryptophan-Mediated Denaturation of β-Lactoglobulin A by UV Irradiation. *Journal of Agricultural and Food Chemistry* **56**, 4720–4725.

[393] Wu, L.-Z, Sheng, Y.-B, Xie, J.-B, & Wang, W. (2008) Photoexcitation of

Tryptophan Groups Induced Reduction of Disulfide Bonds in Hen Egg White Lysozyme. *Journal of Molecular Structure* **882**, 101 – 106.

[394] Neves-Petersen, M. T, Klitgaard, S, Pascher, T, Skovsen, E, Polivka, T, Yartsev, A, Sundstrom, V, & Petersen, S. B. (2009) Flash Photolysis of Cutinase: Identification and Decay Kinetics of Transient Intermediates Formed upon UV Excitation of Aromatic Residues. *Biophysical Journal* **97**, 211–226.

[395] Bent, D. V & Hayon, E. (1975) Excited State Chemistry of Aromatic Amino Acids and Related Peptides. III. Tryptophan. *Journal of the American Chemical Society* **97**, 2612–2619.

[396] Hoffman, M. Z & Hayon, E. (1972) One-Electron Reduction of the Disulfide Linkage in Aqueous Solution. Formation, Protonation, and Decay Kinetics of the RSSR-Radical. *Journal of the American Chemical Society* **94**, 7950–7957.

[397] Pirie, A. (1971) Formation of *N*'-Formylkynurenine in Proteins from Lens and Other Sources by Exposure to Sunlight. *Biochem. J.* **125**, 203–208.

[398] Huang, C.-Y, Klemke, J. W, Getahun, Z, DeGrado, W. F, & Gai, F. (2001) Temperature-Dependent Helix-Coil Transition of an Alanine Based Peptide. *Journal of the American Chemical Society* **123**, 9235–9238.

[399] Milanesi, L, Tomas, S, Hunter, C. A, Weinstein, J. A, Edge, R, Navaratnam, S, Waltho, J. P, & Best, J. (2008) A Pulse-Radiolysis Approach to Fast Reductive

Cleavage of a Disulfide Bond to Uncage Enzyme Activity. *Free Radical Biology and Medicine* **45**, 1271 – 1278.