2017

Evaluating Protein Structure And Dynamics Using Co-Solvents, Photochemical Triggers, And Site-Specific Spectroscopic Probes

Rachel Marie Abaskharon

University of Pennsylvania, rroesch@sas.upenn.edu

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Physical Chemistry Commons

Recommended Citation
https://repository.upenn.edu/edissertations/2154

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/2154
For more information, please contact repository@pobox.upenn.edu.
Evaluating Protein Structure And Dynamics Using Co-Solvents, Photochemical Triggers, And Site-Specific Spectroscopic Probes

Abstract
As ubiquitous and diverse biopolymers, proteins are dynamic molecules that are constantly engaging in inter- and intramolecular interactions responsible for their structure, fold, and function. Because of this, gaining a comprehensive understanding of the factors that control protein conformation and dynamics remains elusive as current experimental techniques often lack the ability to initiate and probe a specific interaction or conformational transition. For this reason, this thesis aims to develop methods to control and monitor protein conformations, conformational transitions, and dynamics in a site-specific manner, as well as to understand how specific and non-specific interactions affect the protein folding energy landscape. First, by using the co-solvent, trifluoroethanol (TFE), we show that the rate at which a peptide folds can be greatly impacted and thus controlled by the excluded volume effect. Secondly, we demonstrate the utility of several light-responsive molecules and reactions as methods to manipulate and investigate protein-folding processes. Using an azobenzene linker as a photo-initiator, we are able to increase the folding rate of a protein system by an order of magnitude by channeling a sub-population through a parallel, faster folding pathway. Additionally, we utilize a tryptophan-mediated electron transfer process to a nearby disulfide bond to strategically unfold a protein molecule with ultraviolet light. We also demonstrate the potential of two ruthenium polypyridyl complexes as ultrafast phototriggers of protein reactions. Finally, we develop several site-specific spectroscopic probes of protein structure and environment. Specifically, we demonstrate that a 13C-labeled aspartic acid residue constitutes a useful site-specific infrared probe for investigating salt-bridges and hydration dynamics of proteins, particularly in proteins containing several acidic amino acids. We also show that a proline-derivative, 4-oxoproline, possesses novel infrared properties that can be exploited to monitor the cis-trans isomerization process of individual proline residues in proteins.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Chemistry

First Advisor
Feng Gai

Keywords
Co-Solvents, Infrared Spectroscopy, Photochemical Triggers, Protein Folding, Site-Specific Spectroscopic Probes

Subject Categories
Physical Chemistry

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/2154
EVALUATING PROTEIN STRUCTURE AND DYNAMICS USING CO-SOLVENTS, PHOTOCHEMICAL TRIGGERS, AND SITE-SPECIFIC SPECTROSCOPIC PROBES

Rachel M. Abaskharon

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

2017

Supervisor of Dissertation

______________________________

Feng Gai

Edmund J. and Louise W. Kahn Endowed Term Professor of Chemistry

Graduate Group Chairperson

______________________________

Gary A. Molander

Hirschmann-Makineni Professor of Chemistry

Dissertation Committee

Marsha I. Lester (chair), Edmund J. Kahn Distinguished Professor
Tobias Baumgart, Associate Professor of Chemistry, Graduate Chair
Sergei Vinogradov, Associate Professor of Biochemistry and Biophysics
EVALUATING PROTEIN STRUCTURE AND DYNAMICS USING CO-SOLVENTS,
PHOTOCHEMICAL TRIGGERS, AND SITE-SPECIFIC SPECTROSCOPIC PROBES

COPYRIGHT

2017

Rachel M. Abaskharon

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License

To view a copy of this license, visit

https://creativecommons.org/licenses/by-nc-sa/3.0/us/
DEDICATION

To my Nanny:

You always believed in me even when I did not believe in myself
ACKNOWLEDGEMENTS

The past five years at Penn have proved to be challenging and rewarding, educational and inspiring, confidence building and fulfilling. I came into graduate school as an excited yet naïve chemist, ready to see where my research would take me. I am leaving, still excited, but much wiser, knowing that I now possess the skills, knowledge, and experience to follow my dreams to my full capability. Professor Feng Gai has been instrumental in helping me develop into the chemist I am today. He took a chance on an eager, new college graduate and saw the potential in me, early on, urging me to stretch myself academically and professionally, while showing his support in my endeavors. His trust in my abilities provided me with the confidence I needed to tackle any obstacle and push myself beyond the limits I thought I had. I am very grateful for his time, dedication, insight, and guidance throughout my time in his lab, which has been invaluable in allowing me to accomplish what I have today.

I would also like to thank my committee members, collaborators and labmates for all of their guidance and support. Professors Marsha Lester, Tobias Baumgart, and Sergei Vinogradov have not only showed interest in my research but have also encouraged me in my scientific accomplishments and goals. I value the discussions I have had with each of them which have allowed to expand my own perspective of my work. Additionally, I appreciate the efforts of Dr. Jeffery Saven in helping me secure external funding throughout my time at Penn. My collaborators and labmates deserve a great amount in credit all of the work here. The collaborations of Professor G. Andrew Woolley, Dr. Stephen Brown, Professor Amos B. Smith III, Teresa Rapp, and Professor Ivan Dmochowski have helped tremendously with the synthesis of compounds and proteins,
protein bioconjugation, and scientific insights for which I am very grateful. Additionally, I would like to thank all the former and current Gai lab group members and the UOPL staff members that I have had the pleasure of working with including: Dr. Robert Culik, Dr. Ileana Pazos, Dr. Beatrice Markiewicz, Dr. Chun-Wei Lin, Dr. Wenkai Zhang, Dr. Bei Ding, Dr. Kwang-Im Oh, Dr. Natalia Rubtsova, Dr. Patrick Walsh, Dr. Thomas Troxler, Dr. Jian-Xin Chen, Mary Rose Hilaire, Ismail Ahmed, Jeffrey Rodgers, Debopreeti Mukherjee, and Arusha Acharyya. I have been blessed with working with such a talented, kind and inspirational group of people, and I am sincerely thankful for all of your help and friendships. In particular, I want to thank Drs. Robert Culik, Ileana Pazos, and Beatrice Markiewicz for training, mentoring and guiding me in lab and life.

Finally, I would never have accomplished what I have today if not for my friends and family supporting me. I am truly thankful to my friends, Melanie Hutnick and Leigh Matano, who were always there through the ups and downs of grad school despite the distance that separated us. To my sister-from-a-different-mister, Dr. Michelle Bunagan, I would not be here without you sharing your love of thermodynamics with me and encouraging me to pursue my Ph.D. I am very thankful to my parents, William and Carol Roesch, who taught me at a young age that in all I do, I should do it with all of my heart and ability to the glory of God. I also want to thank my “little” sister, Jillian Roesch, for making sure I grew up okay. To my husband, Ray Abaskharon, thank you for helping me chase my dreams and believing that “it’s you and me against the world, and the world has nothing on us.” Lastly and most importantly, I want to thank God for guiding me throughout my whole time here and blessing me with everything I have.
ABSTRACT

EVALUATING PROTEIN STRUCTURE AND DYNAMICS USING CO-SOLVENTS, PHOTOCHEMICAL TRIGGERS, AND SITE-SPECIFIC SPECTROSCOPIC PROBES

Rachel M. Abaskharon

Feng Gai

As ubiquitous and diverse biopolymers, proteins are dynamic molecules that are constantly engaging in inter- and intramolecular interactions responsible for their structure, fold, and function. Because of this, gaining a comprehensive understanding of the factors that control protein conformation and dynamics remains elusive as current experimental techniques often lack the ability to initiate and probe a specific interaction or conformational transition. For this reason, this thesis aims to develop methods to control and monitor protein conformations, conformational transitions, and dynamics in a site-specific manner, as well as to understand how specific and non-specific interactions affect the protein folding energy landscape. First, by using the co-solvent, trifluoroethanol (TFE), we show that the rate at which a peptide folds can be greatly impacted and thus controlled by the excluded volume effect. Secondly, we demonstrate the utility of several light-responsive molecules and reactions as methods to manipulate and investigate protein-folding processes. Using an azobenzene linker as a photo-initiator, we are able to increase the folding rate of a protein system by an order of magnitude by channeling a sub-population through a parallel, faster folding pathway.
Additionally, we utilize a tryptophan-mediated electron transfer process to a nearby disulfide bond to strategically unfold a protein molecule with ultraviolet light. We also demonstrate the potential of two ruthenium polypyridyl complexes as ultrafast phototriggers of protein reactions. Finally, we develop several site-specific spectroscopic probes of protein structure and environment. Specifically, we demonstrate that a $^{13}$C-labeled aspartic acid residue constitutes a useful site-specific infrared probe for investigating salt-bridges and hydration dynamics of proteins, particularly in proteins containing several acidic amino acids. We also show that a proline-derivative, 4-oxoproline, possesses novel infrared properties that can be exploited to monitor the cis-trans isomerization process of individual proline residues in proteins.
# TABLE OF CONTENTS

1 Introduction .................................................................................................................... 1
   1.1 The Protein Folding Problem .................................................................................. 1
   1.2 Recent Progress in the Field .................................................................................. 3
      1.2.1 Single-molecule fluorescence studies .......................................................... 3
      1.2.2 Single-molecule force studies ......................................................................... 7
      1.2.3 Ensemble spectroscopic studies ....................................................................... 9
      1.2.4 Computational studies .................................................................................. 13
   1.3 Thesis Overview ..................................................................................................... 14

2 Theory .............................................................................................................................. 19
   2.1 Protein Folding Thermodynamics .......................................................................... 19
   2.2 Protein Folding Kinetics ......................................................................................... 22
      2.2.1 Rate Expressions and Derivations ................................................................. 22
      2.2.2 Reaction Rate Theory .................................................................................... 24
   2.3 Vibrational Spectroscopy Theory .......................................................................... 26
      2.3.1 Vibrational Lineshapes .................................................................................. 26
      2.3.2 Normal Modes and Transition Dipole Couplings .......................................... 30

3 Methods .......................................................................................................................... 35
   3.1 Circular Dichroism Spectroscopy .......................................................................... 35
   3.2 Infrared Spectroscopy .............................................................................................. 37
      3.2.1 Fourier Transform Infrared Spectroscopy ..................................................... 39
3.2.2 Laser-Induced Temperature-Jump Infrared Spectroscopy .................. 40

3.2.3 Flash Photolysis Infrared Spectroscopy........................................ 45

3.2.4 Two-Dimensional Infrared Spectroscopy........................................ 45

3.3 Cosolvents.................................................................................. 47

3.4 Photochemical Triggering Systems............................................... 50

3.5 Site-Specific Spectroscopic Probes............................................... 52

4 Experimental Validation of the Role of Trifluoroethanol as a Nanocrowder ..... 60

4.1 Original Publication..................................................................... 60

4.2 Introduction............................................................................. 60

4.3 Experimental Section............................................................... 63

4.3.1 Sample Preparation and Spectroscopy ..................................... 63

4.3.2 Fractional Helicity Calculation ................................................ 64

4.4 Results..................................................................................... 64

4.4.1 Helicity and Relaxation Rates of pKID in TFE Solutions ............... 64

4.4.2 Viscosity and Crowding Effects.............................................. 66

4.4.3 Control Experiments............................................................. 67

4.5 Discussion............................................................................... 69

4.6 Conclusions.......................................................................... 70

4.7 Acknowledgements.................................................................. 71

5 Tuning the Attempt Frequency of Protein Folding Dynamics via Transition- State Rigidification: Application to Trp-Cage ........................................ 83

5.1 Original Publication................................................................. 83
6.3.3 Steady-State Measurements ................................................................. 111
6.3.4 Time-Resolved Measurements ............................................................ 112
6.3.5 Estimating the Triplet State Quenching Rate by Disulfide ...................... 112
6.3.6 Removal of Solvent Contribution to the Kinetic Offset ......................... 112
6.3.7 Electron Scavenger Experiment ........................................................... 113
6.3.8 Free Thiol Determination ................................................................... 113
6.4 Results and Discussion .......................................................................... 114
6.4.1 Disulfide Cleavage and Unfolding ....................................................... 114
6.4.2 Kinetics of Disulfide Cleavage ............................................................ 114
6.4.3 Mechanism of Action ......................................................................... 116
6.4.4 Reaction Products ............................................................................. 119
6.5 Conclusions ......................................................................................... 120
6.6 Acknowledgements ............................................................................. 121

7 Light-Initiated Ultrafast Kinetics and Two-Ligand Dissociation of Ruthenium Polypyridyl Compounds ................................................................. 133
7.1 Introduction ............................................................................................ 133
7.2 Experimental .......................................................................................... 135
7.2.1 Sample Preparation ........................................................................... 135
7.2.2 Irradiation ......................................................................................... 135
7.2.3 Steady-State Measurements ............................................................... 136
7.2.4 Time-Resolved Measurements ............................................................ 136
7.3 Results and Discussion ........................................................................................................ 137
  7.3.1 Photodissociation of Nitrile Ligands ........................................................................... 137
  7.3.2 Kinetics of Ligand Photodissociation ....................................................................... 140
7.4 Conclusions .......................................................................................................................... 142
7.5 Acknowledgements ............................................................................................................. 143

8 Isotope-Labeled Aspartate Sidechain as a Non-perturbing Infrared Probe:
Application to Investigate the Dynamics of a Carboxylate Buried inside a
Protein ...................................................................................................................................... 151
  8.1 Original Publication ........................................................................................................... 151
  8.2 Introduction ....................................................................................................................... 151
  8.3 Experimental ..................................................................................................................... 154
    8.3.1 Preparation of Cap-Asp and Cap-Asp* .................................................................. 154
    8.3.2 Peptide Synthesis and Purification ........................................................................ 156
    8.3.3 Circular Dichroism (CD) Measurements ............................................................... 156
    8.3.4 Fourier Transform Infrared (FTIR) Measurements ................................................ 156
    8.3.5 2DIR Measurements ............................................................................................. 157
  8.4 Results and Discussion ....................................................................................................... 157
    8.4.1 Synthesis of Fmoc-Asp* ....................................................................................... 157
    8.4.2 Vibrational Frequencies of Isotope Label ............................................................. 158
    8.4.3 Spectral Diffusion Analysis ..................................................................................... 160
  8.5 Conclusions ....................................................................................................................... 163
  8.6 Acknowledgements ........................................................................................................... 164
9 The C=O Stretching Vibration of 4-Oxoproline as an Infrared Probe of Protein Structure and Local Electric Field ........................................................................ 178

9.1 Introduction ........................................................................................................ 178

9.2 Experimental ...................................................................................................... 180

9.2.1 Materials ........................................................................................................ 180

9.2.2 Peptide Synthesis, Purification, and Preparation ......................................... 181

9.2.3 Fourier Transform Infrared (FTIR) Measurements ..................................... 181

9.2.4 MD Simulations .............................................................................................. 182

9.2.5 Gaussian Calculations .................................................................................... 182

9.3 Results and Discussion ...................................................................................... 183

9.3.1 Model Compound Vibrational Frequency and Electric Field ................. 183

9.3.2 Tripeptides and Backbone Effects ............................................................... 185

9.3.3 Investigation of TTR Aggregates ................................................................. 188

9.4 Conclusions ....................................................................................................... 190

9.5 Acknowledgements .......................................................................................... 190

10 Summary and Future Directions ....................................................................... 207

Bibliography ........................................................................................................... 212
LIST OF TABLES

4.1 Estimated fractional helicity ($f_H$) for pKID and LEA at 1°C ..........................72
4.2 Relaxation time constants ($\tau_R$) of pKID and LEA at the indicated final
temperature ($T_f$) ..................................................................................................73
5.1 Unfolding thermodynamic parameters of Trp-cage 10b in a 20/80
TFE/water solution ..................................................................................................97
5.2 Fast and slow time constants of the conformational kinetics of 10b-azob
and 10b-h-azob in 20/80 TFE/D$_2$O solutions at the temperatures ($T$) and
frequencies ($\omega_{\text{probe}}$) indicated.................................................................98
6.1 Amplitudes (in mOD) of the four kinetics components at different
probing frequencies ...............................................................................................122
8.1 Unfolding thermodynamic parameters of Asp$^*$34-psbd41 at pH 8 ...........165
9.1 Center frequencies and electric fields of 1M3P in various solvents ............191
9.2 Full width at half maximum and standard deviations of the experimental
and calculated vibrational frequencies of 1M3P .............................................192
9.3 Gaussian calculated vibrational frequencies and Mulliken partial charges
of P$_{\text{oxt}}$ tripeptide systems ..............................................................................193
9.4 Thermodynamic Parameters for Proline Trans$\rightarrow$Cis Isomerization in
KP$_{\text{oxt}}$G..............................................................................................................194
LIST OF FIGURES

1.1 Plot of the number of computational and experimental protein folding articles per year .................................................................18

2.1 Schematic representations of free energy diagrams .........................33

2.2 Representative normalized lorentzian and gaussian functions ..........34

3.1 Diagram of the laser-induced temperature-jump (T-jump) infrared spectroscopy setup ..........................................................56

3.2 Diagram of the flash photolysis infrared spectroscopy setup .............57

3.3 Cartoon representation of the 2D IR pulse sequence .......................58

3.4 Schematic illustration of spectral diffusion in FTIR and 2D IR spectra ...59

4.1 CD Spectra of pKID in different TFE percentages ..........................74

4.2 CD T-melts of pKID in different TFE percentages ..........................75

4.3 Representative trace of the relaxation kinetics of the pKID peptide ....76

4.4 Temperature dependence of the relaxation rate constants of pKID in different TFE solutions ....................................................77

4.5 Cartoon illustration of the effect of TFE on the folding and unfolding free energy barriers of pKID ....................................................78

4.6 CD spectra of LEA in different TFE percentages ............................79

4.7 CD T-melts of LEA in different TFE percentages ............................80

4.8 Representative trace of the relaxation kinetics of the LEA peptide ....81

4.9 Temperature dependence of the relaxation rate constants of LEA in different TFE solutions ....................................................82
5.1 Schematic representation of how the azobenzene cross-linker alters the curvature of the folding free energy surface of Trp-cage .........................99
5.2 Absorption spectra of dark-equilibrated and light-irradiated 10b-azob ..........100
5.3 CD spectra of dark-equilibrated and light-irradiated 10b-azob ..................101
5.4 CD T-melt of light-irradiated 10b-azob ........................................102
5.5 CD T-melt of Trp-cage 10b in a 20/80 TFE/water solution ..................... 103
5.6 Conformational kinetics of 10b-azob probed at different frequencies ........104
5.7 Conformational kinetics of 10b-h-azob probed at different frequencies .......105
5.8 Temperature dependence of the fast and slow rate constants of 10b-azob ... 106
6.1 NMR structure of Z34C .................................................................123
6.2 CD spectra of Trp-Z34C before and after irradiation ............................124
6.3 CD spectra of wild-type Z34C and Trp-Z34C before and after irradiation ....125
6.4 FTIR difference spectrum of Trp-Z34C in the amide I’ band region ..........126
6.5 IR transient kinetic traces of Trp-Z34C at 1630 cm\(^{-1}\) and 1664 cm\(^{-1}\) ..........127
6.6 IR transient kinetic traces of Trp-Z34C obtained at different probing frequencies .................................................................128
6.7 Transient IR spectrum of Trp-Z34C generated using the offsets of the decay kinetics at the corresponding probing frequencies .........................129
6.8 Normalized CD spectra of two Trp-Z34C samples with and without N\(_2\)O .... 130
6.9 UV-Vis spectra of an Ellman’s reagent sample with and without light-irradiated Trp-Z34C .................................................................131
6.10 Absorption spectra of Trp-Z34C before and after irradiation ..................132
7.1 Reaction scheme for the photochemical ligand exchange process of the ruthenium polypyridyl complexes ................................................. 144
7.2 Representative UV-Vis spectrum of the ruthenium polypyridyl complexes ..... 145
7.3 FTIR spectra of [Ru(bpy)_2(4PN)_2]^{2+} before and after irradiation ............... 146
7.4 FTIR spectra of [Ru(bpy)_2(5HN)_2]^{2+} before and after irradiation .............. 147
7.5 Representative IR transient kinetic traces of [Ru(bpy)_2(4PN)_2]^{2+} ................. 148
7.6 Representative IR transient kinetic trace of [Ru(bpy)_2(5HN)_2]^{2+} ............... 149
7.7 FTIR difference spectrum and transient IR difference spectrum of
[Ru(bpy)_2(4PN)_2]^{2+} ........................................................................ 150
8.1 NMR structure of pdsd41 .................................................................. 166
8.2 Scheme of the synthesis of Fmoc-Asp* ............................................. 167
8.3 FTIR spectra of Cap-Asp and Cap-Asp* at pH 8 ................................. 168
8.4 FTIR spectra of Cap-Asp and Cap-Asp* at pH 1.5 ............................... 169
8.5 CD spectrum of Asp*34-psbd41 at pH 8 ............................................ 170
8.6 CD T-melt of Asp*34-psbd41 at pH 8 ................................................. 171
8.7 FTIR spectrum of Asp*34-psbd41 at pH 8 .......................................... 172
8.8 FTIR spectra of deprotonated Cap-Asp* in dimethyl sulfoxide (DMSO) ...... 173
8.9 Representative 2D IR spectra of Asp*34-psbd41 at pH 8 ...................... 174
8.10 2D IR spectra of Cap-Asp* at pH 8 and $T = 150$ fs .............................. 175
8.11 CLS$^{-1}$ versus $T$ plot of the 2D IR spectra of Asp*34-psbd41 at pH 8 ......... 176
8.12 CLS$^{-1}$ versus $T$ plot of the 2D IR spectra of Cap-Asp* at pH 8 .............. 177
9.1 Normalized FTIR spectra of 1M3P in various solvents .......................... 195
1 Introduction


1.1 The Protein Folding Problem

Proteins are perhaps the most versatile and diverse group of biological molecules in the cell, yet all of these biopolymers are built from a set of twenty standard amino acids. From the stabilization of cell structure to the signaling of physiological responses to the transport of small molecules, protein function requires each molecule to fold into a very specific structure to accomplish its task. Ever since Anfinsen’s discovery in the early 1960s that a denatured protein can spontaneously and rapidly refold to its native conformation upon removal of denaturant, the question of how proteins fold has fascinated many people and inspired many studies. Part of this captivation stems from the notion that even the folding of small proteins (i.e., ~100 amino acids in length) involves a vast number of conformational degrees of freedom and, hence, cannot be achieved through a random search process but rather through a well-defined folding pathway(s). In other words, the linear amino acid sequence of a protein encodes not only its native structure but also the mechanism by which this structure is attained. Thus, the pursuit of this code, or more precisely the underlying principles that govern the thermodynamics and dynamics of protein folding, which are at the core of the protein folding problem, has become a major undertaking in many laboratories around the world. As shown (Figure
1.1), the development of the protein folding field has been quite dynamic in the past 50 years, as judged by the total number of relevant articles published per year. Based on this indicator, it is clear that the protein-folding field as a whole had enjoyed an enormous and exponential-like growth in the early 1990s, followed by a decade-long period of high activity and productivity. However, the trend in the past decade seems to indicate that the field is going through a declining phase, with a decreasing rate of \(~110\) articles per year. If this declining trend were to continue with this rate, the field faces a dim future.

Although we have made great strides in deepening our understanding of many of the aspects of protein folding, we still have not developed a mechanism general enough to predict the “movie” of any protein’s pathway down the energy landscape as accurately or quickly as nature. Advancing a research field requires new ideas, new findings, new theoretical models, and new techniques. The rapid take off of the protein folding field in the early 1990s was fueled by such driving forces, for example, the development of the folding energy landscape theory,\(^4,5\) introduction of the \(\Phi\)-value analysis method,\(^6\) discovery of two-state folders,\(^7\) and advancements in various biological, chemical, physical, and computational techniques that made it possible to study previously inaccessible questions and protein systems. For example, the advent of ultrafast triggering methods, such as the laser-induced temperature-jump (T-jump) technique,\(^8\) has greatly enhanced the time resolution of kinetic studies, enabling investigation of protein folding dynamics on the nanosecond and microsecond timescale. Similarly, application of single-molecule-based techniques allowed for elucidation of conformational transitions and dynamics that were not accessible by traditional ensemble measurements.\(^9\) Although
the prevalence of protein folding publications in the literature has been decreasing over the last several years, this may not be because the problem is solved. Rather, those questions that still need answering may be currently unattainable. The next leap forward in the protein-folding field will most likely require the development of new techniques that would allow us to explore those questions that still remain.

Solving the protein-folding problem has become a multi-disciplinary endeavor drawing the expertise of biologists, chemists, and physicists alike. Experimental studies in the past 5 years have made significant progress in characterizing and understanding the dynamics and mechanism of protein folding. In the first three sections below, we describe some of the key findings that have emerged from those studies, which are organized based on the techniques they used, namely single-molecule fluorescence spectroscopy, single-molecule force spectroscopy, and ensemble spectroscopic methods. In the last section, we provide a brief summary of the major developments in computer simulations of protein folding dynamics.

1.2 Recent Progress in the Field

1.2.1 Single-molecule fluorescence studies

Single-molecule fluorescence-based techniques, such as single-molecule fluorescence resonance energy transfer (smFRET), have become increasingly important in elucidating the fine features of the protein folding free energy landscape and mechanism. One distinct advantage of single-molecule-based spectroscopic techniques is that they can extract information that otherwise would be difficult to attain from ensemble
measurements. In particular, the past 5 years have witnessed the rapid growth of using these techniques to assess the role of native and nonnative interactions, internal friction, unfolded state structure and dynamics, and transition-path time on protein folding pathways and kinetics. A few representative examples are given below.

Several recent studies have focused on the role of internal friction and frustration on the dynamics of protein folding.10–15 For example, Schuler and co-workers12,13 used smFRET, fluorescence correlation spectroscopy (FCS), and microfluidic mixing to study the effect of internal friction on the folding energy landscape of the spectrin domains. They found that the internal friction localized at the early transition state plays a major role in determining the overall folding times of these proteins, which led them to suggest that the curvature of the transition state barrier is larger than that of the unfolded state potential well.12 Using a small cold shock protein (Csp) as a model, they further showed that internal friction, as measured by the conformational reconfiguration or relaxation time, is also a key determinant of the conformational dynamics in the unfolded potential well.13 Under native-like conditions where the polypeptide chain is more compact, the reconfiguration time of the unfolded ensemble of Csp is 100 ns but the reconfiguration time extrapolated to zero viscosity is nearly zero when the polypeptide chain becomes more extended (i.e., under high denaturant concentrations). Based on these finding, they suggested that internal friction is particularly important in determining the barrier crossing dynamics of microsecond folders. On the other hand, Sherman and Haran10 found, based on the mean first-passage times extracted from their FCS experiments using the theory of Szabo, Schulten, and Schulten,16 that the intrachain diffusion coefficient of
protein L remained approximately constant from a denaturant concentration of 3 to 7 M guanidinium chloride (GdmCl). Similarly, by using FCS to measure the intrachain motions of a set of unstructured peptides with and without side chains, Teufel et al. concluded that side-chain interactions slow down loop formation while backbone-backbone hydrogen bonds accelerate intrachain interactions. Voelz et al. also studied the protein unfolded state and concluded, based on smFRET measurements and molecular dynamics (MD) simulations, that there is a large network of metastable states in the unfolded ensemble of the ACBP protein, which interconvert on a timescale of ~100 μs. Although previous studies had identified this kinetic phase as a folding intermediate, they attributed this time constant to the protein’s slow achievement of the unfolded state structure that leads to productive folding. Most recently, Chung et al. showed that interresidue contacts, particularly nonnative salt bridges, were able to definitively decrease the folding rate for a designed α-helical protein, α3D. Although these protein interactions are able to create local minima along the folding coordinate thus increasing the roughness of the energy landscape and the time required to traverse the free energy barrier, they do not, however, significantly alter the folding free energy barrier height.

The power of smFRET in revealing the detailed features underlying the folding dynamics of large proteins was recently demonstrated by Haran and co-workers. By analyzing smFRET trajectories using a hidden Markov model, they showed that the folding landscape of adenylate kinase, a 214-residue, multidomain protein, encompasses six metastable states, with their connectivity depending on denaturant concentration. Although many intersecting folding pathways were observed at low denaturant
concentrations (~0.5 M urea), one sequential mechanism became dominant when the
denaturant concentration was increased to 1 M. Liu et al.\textsuperscript{19} also showed that smFRET can be applied to study the dynamics of a downhill protein folding process. Additionally, Clarke and co-workers\textsuperscript{20} used a single-molecule total internal reflection fluorescence microscopy technique as well as other computational and biophysical methods to study the SasG tandem repeat protein, which is known to form extended fibrils on the surface of \textit{Staphylococcus aureus} bacteria. They found that the center domain, although intrinsically disordered, can mediate long-range folding cooperativity of the terminal domains as a result of the high stability of the interfaces.

Perhaps one of the most significant advancements in the assessment of protein folding dynamics is the ability to determine the protein folding transition-path time, namely the time required for the successful crossing of the free energy barrier by a protein molecule.\textsuperscript{21} Assuming an attempt frequency value of $10^5$ to $10^6$ s$^{-1}$, the transition-path time $\langle t_{TP} \rangle$ is estimated to be between 0.6 and 6 $\mu$s. The very short nature of $\langle t_{TP} \rangle$, as well as the fact that this quantity manifests itself in the barrier-crossing process of individual molecules, have made it very difficult to measure experimentally. By fully characterizing the smFRET time traces obtained with the GB1 protein, including the number of photons emitted, their polarizations, their relative and absolute arrival times, and their wavelengths, Eaton and co-workers\textsuperscript{22} were able to isolate single-molecule barrier crossing events and used them to determine an upper bound of 200 $\mu$s for the transition-path time for this protein. Fundamentally, $\langle t_{TP} \rangle$ is relatively insensitive to the barrier height.\textsuperscript{22} Indeed, in a later study,\textsuperscript{21} they were not only able to more accurately
determine $\langle t_{TP} \rangle$, but also showed that the transition path time (i.e., 10 $\mu$s) of a slow-folding protein (folding time = 1 s) is similar to that (i.e., 2 $\mu$s) of a fast-folding protein (folding time = 100 $\mu$s). This is an exciting finding because it experimentally illustrates that correct folding takes approximately the same time even for proteins that are drastically different in structure and size.

1.2.2 Single-molecule force studies

Like single-molecule fluorescence studies, those based on single-molecule force measurements have also made outstanding contributions toward our understanding of how proteins fold. A distinct advantage of using force spectroscopy to study protein-folding dynamics is that the reaction coordinate can be defined as the extension of the molecule, thus allowing for characterization of the folding energy landscape, identification of parallel folding pathways, and observation of anisotropic folding behaviors. Below, we highlight several recent studies, showcasing the advancement in this area.

In one study, Zhang and co-workers$^{23}$ used optical tweezers to investigate the heterogeneity in the folding-unfolding process of the coiled coil GCN4 and found that the force-induced transition rate was highly anisotropic when the folded and unfolded states were equally populated, depending greatly on the pulling direction. This finding was corroborated by the study of Marquesee and co-workers$^{24}$ who found that a two-state folding protein, the src SH3 domain, was more resistant to a low force applied in the longitudinal direction than the perpendicular direction. Moreover, they observed biphasic behavior along the longitudinal pulling axis, indicating that the protein is capable of
accessing parallel unfolding pathways. In a more recent study,\textsuperscript{25} they showed that for the same protein, it is possible to use mutations and denaturant to modulate the flux among the different pathways. Similarly, using atomic force microscopy, Li and co-workers\textsuperscript{26} demonstrated that the mechanical unfolding and untwisting process of a slipknot protein, AFV3-109, also occurs via multiple pathways that are either two-state or three-state in nature. Furthermore, the high-resolution force measurements of Rief and co-workers\textsuperscript{27} revealed that the folding and unfolding transitions of single calmodulin molecules involve two on-pathway and two off-pathway intermediates and identified cooperative and anticooperative interactions between the domains.

In the past 5 years, single-molecule force spectroscopy has also been actively applied to characterize other aspects of the folding free energy landscape of interest.\textsuperscript{28–30} For example, by using a high-resolution optical trap to apply tension to the prion protein, Yu et al.\textsuperscript{28–30} were able to determine the free energy barrier height and position along the reaction coordinate and further used this information to determine the transition-path time and the conformational diffusion coefficient. Similarly, Fernández and co-workers\textsuperscript{29} employed force clamp spectroscopy to examine the energy landscape of an engineered I27 protein undergoing two separate reactions, namely unfolding and disulfide bond reduction. In the framework of static disorder theory, they showed that the disulfide-containing mutants had a high degree of heterogeneity in their unfolding pathways; however, the disulfide cleavage event itself followed a rather homogeneous reactive pathway. By performing a force-quench experiment,\textsuperscript{28} they also showed that the
heterogeneous collapse trajectories observed for ubiquitin and I27 arise from a force-
dependent free energy barrier.

1.2.3 Ensemble spectroscopic studies

Ensemble measurements based on various spectroscopic techniques continue to play a
major role in the investigation of the protein-folding problem. In particular, new insights
into the dynamics and mechanism of protein folding have been generated from studies
that have used new triggering and probing methods, as well as new strategies to enhance
the time and structural resolution of the experiments, manipulate the folding energy
landscape of interest, and extract detailed mechanistic information from conventional
kinetic measurements. Below we highlight a few examples, with a focus on studies that
employ light-based triggering and detection methods.

Unlike single-molecule fluorescence studies, which rely on spontaneous
conformational fluctuations of individual protein molecules, measurements of protein
folding kinetics at the ensemble level require a well-defined triggering event to define the
zero time. Although flow- and mixing-based triggering methods continue to play an
important role in this regard,\textsuperscript{31–33} recent years have seen an increased interest in applying
various photo-induced processes to initiate and control protein folding reactions. In
comparison to commonly used chemical and thermal triggering methods, those based on
photo-induced isomerization or bond cleavage have the advantage of being site-specific.
In addition, they allow interrogation of ultrafast protein folding events. Of course, the
disadvantage of using a photoliable group or photoswitchable trigger is the possibility
that it may introduce an undesirable structural perturbation to the protein/peptide system in question.

Currently, the most commonly used phototriggers are azobenzene and its derivatives,\(^{34}\) whose cis- to trans- (and vice versa) isomerization can be controlled with ultraviolet (UV) or visible light. Because this isomerization process occurs on a picoseconds timescale, it is suitable for interrogation of ultrafast protein folding events. This has been nicely demonstrated by the works of Zinth and coworkers,\(^{35}\) Hamm and co-workers,\(^ {36,37}\) and Kliger and co-workers.\(^ {38}\) A relatively less used phototriggering method due to the very fast geminate recombination rate of the underlying photocleavage reaction is based on disulfide bond cleavage via UV excitation.\(^ {39}\) However, Volk and co-workers\(^ {40}\) took advantage of this geminate recombination and used it to monitor the conformational dynamics of the N-terminal domain of phosphoglycerate kinase, initiated by UV excitation of an aromatic disulfide bond.

Besides those well-established phototriggering strategies, the past 5 years have also seen the development of new photochemical methods. For example, Zinth and co-workers\(^ {41}\) demonstrated the feasibility of using hemithioindigohemistilbene derivatives as ultrafast protein folding triggers. These stilbene-based chromophores were shown to isomerize in tens of picoseconds and to induce strong structural changes in a model \(\beta\)-hairpin where the phototrigger was covalently linked to the termini of the peptide. Another very promising and useful phototriggering strategy is based on photodissociation of an \(S,S\)-tetrazine moiety. Brown and Smith\(^ {42}\) showed that this moiety can be easily incorporated into unprotected peptides and proteins via two cysteine residues, and Tucker
et al. demonstrated that the underlying photodissociation process, in response to an excitation with 330–400 nm light, occurs on the picoseCONDS timescale. They further illustrated the use of this phototigger by using it to trigger a local conformational relaxation event in an α-helix and found, via two-dimensional infrared (2D IR) spectroscopy, that the dynamics of this reorganization process, which involves only a single helix turn, occurs in ~100 ps. Finally, we note that Zewail and co-workers have expanded the time resolution of the T-jump technique to the picosecond regime, allowing for the study of ultrafast protein folding processes, including helix nucleation.

Cross-linking is not only the basis for photo-induced structural transitions, but it can also be exploited to perform other novel applications in protein folding studies. For example, in one such application, Markiewicz et al. used a strategically placed m-xylene cross-linker to assess how local friction or frustration affects protein folding dynamics, a topic that has also been explored in a recent study by Matthews, Brooks, and co-workers. By placing this cross-linker in a congested region of Trp-cage, Markiewicz et al. were able to decrease both the folding and unfolding rates of this miniprotein without significantly affecting its stability. They attributed this phenomenon to an increase in local internal friction, due to the presence of the m-xylene cross-linker, which acts as a local mass crowding agent. Using their kinetic results and a theoretical model developed by Thirumalai, Straub, and co-workers, they were able to further show that this local crowder increases the roughness of the folding free energy surface of Trp-cage by 0.4–1.0 k_B T. In yet another study, Markiewicz et al. hypothesized that it is possible to use cross-linking strategies to create structural analogs of protein-folding transition
states. To verify this hypothesis, they employed a disulfide bond to enforce the $\beta$-turn of the Trpzip4 $\beta$-hairpin to be in a native-like configuration, which had been shown to be formed in the transition state. They found that this disulfide-bonded version of Trpzip4 folds 10 times faster than the uncross-linked peptide, supporting the notion that it is possible to engineer a mimic of the transition state structure of interest via cross-linking.

In a different application, Sosnick and co-workers used cross-linking and $\psi$-analysis to probe the transition state heterogeneity of ubiquitin and found that despite the incorporation of a cross-linker, the structural content of the transition state was not substantially altered.

Another distinct effort of the field in the past 5 years has been to enhance the structural resolution and site-specificity in protein folding kinetic studies using various spectroscopic probes. In one example, Culik et al. demonstrated that it is possible to monitor the formation of individual secondary structural elements in protein folding via isotopic labeling of selective amide carbonyls. This isotope editing method was also used by Keiderling, Kubelka, and co-workers to study the hydrogen-bonding pattern in the aggregation kinetics of polyglutamic acid peptides. In other examples, Tokmakoff and co-workers, Zanni and co-workers, and Hochstrasser and co-workers showed that more detailed structural content can be obtained on the folding reactions of interest by monitoring couplings or interactions between two or more isotopic amide labels using 2D IR spectroscopy. This is because distinct interactions between two specific groups in a protein reveal distance and thus structural information. Although similar ideas, for instance those based on FRET, have long been used in fluorescence based protein folding.
studies, the past 5 years have seen a renewed interest in developing amino acid-based FRET or fluorophore-quencher pairs for this purpose. Examples include the \( p \)-cyanophenylalanine-tryptophan FRET pair,\(^{59}\) amino acid fluorophore-thioamide quencher pair,\(^{60}\) and \( p \)-cyanophenylalanine-selenomethionine fluorophore- quencher pair.\(^{61,62}\) A recent example highlighting the use of such structure-sensitive probes is nicely illustrated by Gruebele and co-workers,\(^{63}\) who used three tryptophan-tyrosine quencher pairs to measure contact formation between three helices during the folding of a fast folding protein, \( \lambda_{6-85} \). Acquiring site-specific folding kinetic information is also actively pursued in recent studies.\(^{64}\) For example, Dyer and co-workers\(^{65}\) used the aspartic acid side chain as a local probe to show that the formation of the first hairpin in the WW domain is tightly linked to the protonation state of this charged residue and used an azide IR probe to specifically monitor side-chain reordering events in the folding of the N-terminal domain of the L9 protein.\(^{66}\) Similarly, Kiefhaber and co-workers\(^{67}\) showed that by replacing a native oxoamide with a thioamide, it is possible to probe backbone-backbone H-bond formation in a site-specific manner, a notion also verified by the study of Culik et al..\(^{68}\)

### 1.2.4 Computational studies

In parallel, computer simulation of protein folding has also made significant progress in the past 5 years.\(^{69-73}\) First, the ability to perform long-time (e.g., millisecond timescale) MD simulations with atomic-level resolution has made it possible to directly fold a protein in silico, allowing visualization of its folding process\(^{74,75}\) and, perhaps more importantly, a direct comparison with experiment.\(^{76}\) Second, the development of Markov
state models, which use a statistical approach to group structures observed in a simulation into microstates and connect them by a transition matrix, has allowed a more physical-based and mechanistic interpretation of MD trajectories.\textsuperscript{77,78} Third, continued efforts in the improvement of the molecular force fields, such as those on protein secondary structures and hydration,\textsuperscript{79,80} have led to more accurate characterization of various interactions underlying protein folding.

1.3 Thesis Overview

The overarching motif of this Thesis is developing and utilizing new chemical methods to study protein structure and dynamics. These methods range from co-solvents to photoresponsive small molecules to site-specific spectroscopic reporters. In each case, the goal is to expand the toolbox currently available to study these processes in an effort to gain new insights into protein conformation, folding, and environment. However, prior to describing these new methods, in Chapter 2, we first discuss protein thermodynamics and kinetics as well as the theory behind vibrational spectroscopy. Then, in Chapter 3, a summary of the various techniques utilized is provided.

In Chapter 4, we investigate the ability of trifluoroethanol (TFE), a commonly used protein secondary structure inducer, to act as a nanocrowder at certain concentrations. Although TFE is best known as a co-solvent that can stabilize $\alpha$-helical structure in proteins, molecular dynamics simulations have shown that it can also self-associate to form small, nanometer-sized clusters due to its amphiphilic structure. Such clustering can affect the dynamics of protein folding due to the excluded volume effect.
Thus, we measure the conformational relaxation kinetics of an intrinsically disordered protein, the phosphorylated kinase inducible domain (pKID), using an infrared temperature-jump technique. This protein forms a helix–turn–helix structure in TFE solutions; however, we demonstrate that the rate at which this happens is intricately dependent on TFE percentage. In fact, the fastest conformational relaxation kinetics are observed between 15−30% TFE, in agreement with maximum cluster formation. The notion that this cluster-formation creates a crowding effect was further supported by an experiment performed on a monomeric α-helix, which shows no such kinetic trend due to the predominately local interactions needed for its folding process.

Chapter 5 explores the use of an azobenzene photoisomer as a way to tune the attempt frequency of a protein folding reaction. The attempt frequency or prefactor ($k_0$) of the transition-state rate equation of protein folding kinetics has been estimated to be several orders of magnitude smaller than that of small-molecule chemical reactions. This is due, in part, to the flexible nature of the protein structure. In this investigation, we demonstrate that by strategically placing an azobenzene phototrigger into the α-helix involved in the major folding transition state of the mini-protein Trp-cage, it is possible to significantly increase the attempt frequency and thus the folding rate of this protein by rigidifying the transition state structure upon photoisomerization. Moreover, this method exposes parallel folding pathways, allowing us to estimate the curvature of the transition-state free-energy surface of a protein for the first time.

In Chapter 6, we investigate the kinetics and mechanism of a protein photodamage process involving electron transfer from a tryptophan (Trp) residue to a
nearby disulfide bond. Although there have been a wide-variety of previous studies investigating this process in small molecules or observing its effects in peptide/protein systems, many details about the mechanism involved in protein systems are unknown, including the rate at which this happens in a protein environment. Therefore, we devise a method to directly assess the kinetics of photo-induced disulfide cleavage, a key pathway leading to protein damage, using transient infrared spectroscopy. We find that this event occurs in ~2 μs via a mechanism involving electron transfer from an excited state of a Trp residue to the disulfide bond. Additionally, our study indicates that a Trp-SR adduct is formed on one side of the peptide, which prevents the protein from re-cross-linking after this radical process. Therefore, a Trp-disulfide pair could be used as a phototrigger to initiate protein-folding reactions and control the dynamics and activities of disulfide-containing peptides in applications slower than 2 μs.

In Chapter 7, infrared investigations of photosensitive ruthenium polypyridyl complexes are conducted, which we show are able to undergo a ligand exchange reaction on the picosecond timescale with light. By implementing nitrile groups into the ligands of this complex, the photoreaction is monitored by ultrafast time-resolved infrared spectroscopy. We study the photophysics of two ruthenium polypyridyl complexes, which differ by one methylene group in each photocleavable ligand. Our studies show that these two complexes appear to replace both of their nitrile-containing ligands with solvent molecules by 1 ns. Therefore, these systems could be useful photochemical triggers of various protein folding and binding events by conjugating them to biomolecules using azide-alkyne cycloaddition.
Chapter 8 investigates the utility of an isotope-labeled aspartate (Asp) residue as a site-specific infrared probe of protein structure and environment. Asp residues are often involved in the active or binding site of proteins and can also be critical to the stabilization of protein structure. Although these carboxylate moieties absorb outside the congested amide I’ region of the infrared (IR) spectrum, it is still difficult to study a specific Asp residue in proteins that contain multiple acidic amino acids with IR spectroscopy. However, by isotopic-labeling of the Asp sidechain, we show that individual Asp residues can be monitored by shifting the vibrational frequency of the $^{13}$COO$^-$ asymmetric stretch. We use this site-specific vibrational probe in conjunction with two-dimensional IR spectroscopy to study the dynamics of a structurally important Asp ion buried inside a small protein.

In Chapter 9, we show that a proline analog, 4-oxoproline (P$_{ox}$), has a carbonyl stretching frequency that is sensitive to both local electrostatic field and backbone conformation. Proline residues are important motifs in many protein-binding and control mechanisms because of their distinct structure and isomerization abilities. However, there are currently very few analogs of proline, which are used as spectroscopic probes of this amino acid’s structure and environment. Since the carbonyl stretching frequency of P$_{ox}$ resides in an uncongested region of the infrared spectrum, we use it to assess the local environment of the aggregating peptide, transthyretin. Additionally, we find that due to the integration of this side-chain into the peptide backbone, the carbonyl stretch is integrally tied to the partial charge of the amide nitrogen. Therefore, we are also able to use this probe to monitor the cis-trans isomerization process of a tripeptide system.
Figure 1.1 Plot of the number of computational and experimental protein folding articles found on the ISI Web of Science database per year. The number of computational articles per year was produced by using the search, “protein folding” and (comput* or simulat* or theor*), whereas the number of experimental articles per year was obtained by subtracting this value from that of the search using “protein folding.”
2 Theory

2.1 Protein Folding Thermodynamics

Protein folding is a very complex process that is governed by a variety of thermodynamic interactions ranging from ionic and hydrogen bonding (H-bonding) to hydrophobic effects and van der Waal’s forces.\textsuperscript{81} Not only can these interactions take place within the protein molecule itself, but the solvent environment can also greatly affect the magnitude and number of these interactions by competing for them as well. Despite Anfinsen’s discovery that a protein molecule will spontaneously fold to its native state after removal of a denaturant,\textsuperscript{2} the difference in energy between the folded and unfolded state of the protein is actually not very large as a result of these competing forces. In order to understand this quantitatively, we must look at the thermodynamic theory underlying protein folding reactions. Although there has been extensive study in this field,\textsuperscript{4,82–85} we will only discuss the thermodynamics of a highly cooperative system, which can be represented as a two-state folding protein, although there are more complex cases such as folding with intermediates or downhill folding (Figure 2.1).

A two-state folding protein follows the general model that there are two ensembles that can be populated, the folded ($F$) and unfolded ($U$) states,

\[
F \overset{k_f}{\rightleftharpoons} U \tag{2.1}
\]

which are able to interconvert with rates $k_f$ for the folding process and $k_u$ for the unfolding process. Therefore, the Gibbs free energy of this protein folding reaction ($\Delta_r G^r(T)$) is defined to be

\[
\Delta_r G^r(T) = \Delta_r H^r(T) - T \Delta_r S^r(T) \tag{2.2}
\]
where \( T \) is the temperature, \((\Delta_r H^\circ(T))\) is the change in enthalpy, and \((\Delta_r S^\circ(T))\) is the change in entropy of the reaction.\(^{86}\) Specifically, these energetic and entropic changes are between the denatured state and native state such that,

\[
\Delta_r G^\circ(T) = G^\circ_U(T) - G^\circ_F(T) \quad (2.3)
\]

\[
\Delta_r H^\circ(T) = H^\circ_U(T) - H^\circ_F(T) \quad (2.4)
\]

\[
\Delta_r S^\circ(T) = S^\circ_U(T) - S^\circ_F(T) \quad (2.5)
\]

where the subscripts \( U \) and \( F \) refer to the unfolded and folded states, respectively. Therefore, for a protein to unfold spontaneously, \( \Delta_r G^\circ(T) < 0 \). However, there are competing interactions that can affect the energetics of this process. For example, although the chain conformational entropy of the protein becomes more disordered and thus increases in entropy, the disruption of the hydrophobic effect partially counteracts this by creating more organization in the solvent molecules around non-polar residues.\(^{81,87}\) Additionally, the destruction of H-bonding, van der Waals and electrostatic interactions increases the enthalpy of the system.

As can be seen above, all of these thermodynamic functions are temperature dependent. Specifically, \((\Delta_r H^\circ(T))\) and \((\Delta_r S^\circ(T))\) can be defined as follows,

\[
\Delta_r H^\circ(T) = \Delta_r H^\circ(T_s) + \int_{T_s}^{T} \frac{\Delta_r C_p}{T} dT \quad (2.6)
\]

\[
\Delta_r S^\circ(T) = \Delta_r S^\circ(T_s) + \int_{T_s}^{T} \frac{\Delta_r C_p}{T} dT \quad (2.7)
\]

where \( T_s \) is a reference temperature and \( \Delta_r C_p \) is the heat capacity of the reaction and is generally assumed to be constant over the biologically relevant temperature range.\(^{86}\)
Typically, the melting temperature ($T_m$) is used as this reference. $T_m$ is defined as the temperature at which 50% of the protein is unfolded and thus, $\Delta_r G^\circ (T) = 0$, i.e.

$$T_m = \frac{\Delta_r H^\circ(T_m)}{\Delta_r S^\circ(T_m)}.$$  \hspace{1cm} (2.8)

Therefore, Eqs 2.6 and 2.7 can be rewritten as,

$$\Delta_r H^\circ(T) = \Delta_r H^\circ(T_m) + \Delta_r C_p (T - T_m)$$ \hspace{1cm} (2.9)

$$\Delta_r S^\circ(T) = \Delta_r S^\circ(T_m) + \Delta_r C_p \ln \left(\frac{T}{T_m}\right)$$ \hspace{1cm} (2.10)

Substituting Eqs 2.9 and 2.10 into Eq 2.2 provides the following relation,

$$\Delta_r G^\circ(T) = \Delta_r H^\circ(T_m) - T \Delta_r S^\circ(T_m) + \Delta_r C_p \left[ T - T_m - T \ln \left(\frac{T}{T_m}\right) \right].$$  \hspace{1cm} (2.11)

Eq 2.11 is often used to fit $\Delta_r G^\circ(T)$ versus $T$ data to obtain values for the three thermodynamic quantities and $T_m$. In this way, we are able to see how a specific mutation or alteration of the protein system, affects the protein stability by determining the value of $T_m$. Previous studies have shown that $\Delta_r H^\circ(T)$, $\Delta_r S^\circ(T)$, and $\Delta_r C_p$ vary linearly with the number of amino acids in a protein. Therefore, for a 100 residue protein, their correlation at $T = 60$ °C would yield $\Delta_r H^\circ(T) = 292$ kJ mol$^{-1}$, $\Delta_r S^\circ(T) = 0.88$ kJ mol$^{-1}$ K$^{-1}$, and $\Delta_r C_p = 5.8$ kJ mol$^{-1}$ K$^{-1}$. By using these values in Eq 2.11 (assuming the $T_m$ is 60 °C), it is easy to see that $\Delta_r G^\circ(T) = 0$ at two temperatures, indicating that the protein can be denatured by both heating and cooling processes. This leads to both a melting temperature ($T_m$) for the heat denaturation process and a cold denaturation temperature ($T_c$); however, it is often hard to experimentally measure $T_c$.  

21
2.2 Protein Folding Kinetics

2.2.1 Rate Expressions and Derivations

In order to gain kinetic information in a protein folding experiment, it is necessary to perturb the protein from its equilibrium by altering a thermodynamic variable. If we consider the two-state folding protein again, as shown in Eq 2.1, this necessity is obvious since at equilibrium the change of the concentration of the folded ([F]_eq) or unfolded ([U]_eq) state as a function of time is zero \( \frac{d[F]_\text{eq}}{dt} = -\frac{d[U]_\text{eq}}{dt} = 0 \). Therefore, to gain kinetic information about the protein folding process on the ensemble level, the thermodynamic conditions of the system must be rapidly changed to form a non-equilibrium state, which is then able to relax to a new equilibrium. From this type of experiment, \( k_f \) and \( k_u \) values can be learned.

Given the reaction in Eq 2.1, the differential rate expressions are

\[
\frac{d[F](t)}{dt} = k_f[U](t) - k_u[F](t) \quad (2.12)
\]

\[
\frac{d[U](t)}{dt} = k_u[F](t) - k_f[U](t) \quad (2.13)
\]

for the folding and unfolding reactions, respectively.\(^{89,90}\) Thus, at equilibrium, where Eqs 2.12 and 2.13 are zero, \( k_f[U]_\text{eq} = k_u[F]_\text{eq} \) which allows the equilibrium constant \( (K) \) to be written as

\[
K = \frac{[U]_\text{eq}}{[F]_\text{eq}} = \frac{k_u}{k_f}. \quad (2.14)
\]

However, for a system out of equilibrium, \([U](t)\) can be represented as
\[ [U](t) = [F]_0 + [U]_0 - [F](t) \]  \hspace{1cm} (2.15)

where \([F]_0\) and \([U]_0\) are the concentrations of the folded and unfolded state respectively at time \(t=0\). Combining Eqs 2.12 and 2.15 yields

\[
\frac{d[F](t)}{dt} = k_f ([F]_0 + [U]_0 - [F](t)) - k_u ([F](t)). \hspace{1cm} (2.16)
\]

Solving this differential equation from \(t = 0\) to \(t\) leads to the following:

\[
-(k_f + k_u)[F](t) + k_f ([F]_0 + [U]_0) \bigg| _0^t = t. \hspace{1cm} (2.17)
\]

Further rearrangement of Eq 2.17 allows \([F](t)\) to be expressed as

\[
[F](t) = [F]_0 e^{-(k_f + k_u)t} + \frac{k_f ([F]_0 + [U]_0)}{k_f + k_u} \left[1 - e^{-(k_f + k_u)t}\right]. \hspace{1cm} (2.18)
\]

illustrating that when a protein is perturbed from equilibrium, it relaxes exponentially as a function of time. Moreover, the rate extracted by fitting the exponential signal resulting from a protein folding reaction is the sum of the folding and unfolding rates and is often called the relaxation rate, \(k_r\). In order to determine both \(k_f\) and \(k_u\), the equilibrium constant, \(K\) must be determined in addition to \(k_r\). Lastly, Eq 2.18 indicates that \([F]_{eq}\) can be expressed as

\[
[F]_{eq} = \frac{k_f ([F]_0 + [U]_0)}{k_f + k_u} \hspace{1cm} (2.19)
\]

since \(\frac{d[F]_{eq}}{dt} = 0\) and \(k_f [U]_{eq} = k_u [F]_{eq}\) at equilibrium.
2.2.2 Reaction Rate Theory

In order to correlate protein folding thermodynamics with kinetics, it is necessary to have a model that explains how the reaction rate is dictated by thermodynamic parameters. Given the two-state folding protein in Eq 2.1, the folded state of the protein is more stable than the unfolded state by some amount of energy, $\Delta_f G^r(T)$. At equilibrium, $\Delta_f G^r(T)$ can be related to the equilibrium constant, $K$, by the equation:

$$\Delta_f G^r(T) = -RT \ln(K)$$

(2.20)

where $R$ is the gas constant. Additionally, in most cases, the folded and unfolded states are separated by a free energy barrier, which must be overcome to interconvert from one ensemble to the other. Therefore, the free energy of this barrier can be expressed as

$$\Delta G^b_{fu}(T) = \Delta G^b_f(T) + \Delta_f G^r(T)$$

(2.21)

where $\Delta G^b_{fu}(T)$ is the free energy difference between the folded state and the top of the free energy barrier and $\Delta G^b_f(T)$ is the free energy difference between the unfolded state and the top of the free energy barrier. By combining Eqs 2.14, 2.20, and 2.21, the following relations can be attained,

$$k_u \propto e^{\frac{\Delta G^b_{fu}}{RT}}$$

(2.22)

$$k_f \propto e^{\frac{\Delta G^b_f}{RT}}.$$  

(2.23)

Despite these proportions, getting the exact model for relating the reaction rate to the energy and temperature has been the subject of a number of studies. The first person to describe the reaction rate of such a process was Arrhenius in 1889 using the following equation:
where $E_a$ is the activation energy or the energy needed to undergo a productive reaction and $A$ is a pre-exponential factor.\textsuperscript{91} This equation was developed empirically by combining the Boltzmann distribution property with activation energy. Moreover, $E_a$ and $A$ were assumed to be temperature independent. In this view of kinetic rates, the energy barrier separating the folded and unfolded potential wells was thought to be purely enthalpic and thus $E_a$ is related to reaction enthalpy. However, by the 1930s, after introduction of Gibbs free energy, another equation was developed which more strictly quantified $A$ and $E_a$. This equation, known as the Eyring-Polanyi equation, came as the result of Transition-State Theory and was shown to be

$$k = A e^\left(\frac{E_a}{RT}\right)$$

(2.24)

where $k_B$ is the Boltzmann constant and $h$ is Planck’s constant.\textsuperscript{92,93} This equation was derived using statistical mechanical methods, which assumed the reaction was bimolecular in the gas phase where the reactants were in quasi-equilibrium with the activated complexes at all times. Moreover, it did not account for instances when a molecule has enough energy to undergo a chemical reaction but does not do so successfully. These underlying assumptions cause the pre-exponential factor, or attempt frequency, in this equation to be $6.2 \times 10^{12}$ s$^{-1}$ for $T = 298$ K, which is orders of magnitude too large for reactions in solution that are limited by diffusion. Therefore, in 1940, Kramers derived an alternate reaction rate equation using Langevin dynamics that explained barrier crossing in both the inertial and diffusive regimes.\textsuperscript{94,95} By including a
damping force, $\gamma$, to account for the coupling between the solvent system and reactants, the rate of Brownian motion of barrier crossing in the high friction limit (i.e. in solution) was described as

$$k = \frac{\omega_B \omega_R}{2\pi\gamma} e^{\frac{\Delta G^\circ}{RT}}$$

(2.26)

where $\omega_B$ is the frequency of the free energy barrier, $\omega_R$ is the frequency of the reactant potential well, and $\gamma$ is the friction coefficient. Despite the accuracy of this rate equation, it is actually quite difficult to determine all of the pre-exponential terms experimentally. Thus, the following equation,

$$k = k_0 e^{\frac{\Delta G^\circ}{RT}}$$

(2.27)

which is directly related to Eqs 2.22 and 2.23, is often used to describe protein-folding reactions. The attempt frequency, $k_0$, has been proposed to be in the range of $10^3$ to $10^6 \text{s}^{-1}$ from both experiments$^9$ and calculations.$^{96}$

2.3 Vibrational Spectroscopy Theory

2.3.1 Vibrational Lineshapes

In order to understand the origin of vibrational lineshapes, it is helpful to first understand the concept of the density matrix.$^{97-99}$ For a pure quantum mechanical state of a single molecule, which can be described by a single wavefunction, let us consider the Hamiltonian to be

$$\hat{H}(t) = \hat{H}_0(t)$$

(2.28)
where $\hat{H}_0(t)$ is a time-dependent molecular Hamiltonian. The time dependent Schrödinger equation,

$$\frac{\partial}{\partial t} |\psi(t)\rangle = -i\frac{\hbar}{\hbar} \hat{H}(t) |\psi(t)\rangle$$  \hspace{1cm} (2.29)

thus governs the evolution of this molecular wavefunction. Moreover, the molecular wavefunction can be described as

$$|\psi(t)\rangle = \sum_a c_a(t) |a\rangle$$  \hspace{1cm} (2.30)

where $c_a$ are the amplitudes of the probability of being in different states, $a$. Therefore, the expectation value of any operator, such as the dipole operator $\hat{\mu}$ is

$$\langle \hat{\mu} \rangle = \langle \psi(t) |\hat{\mu}|\psi(t)\rangle = \sum_{ab} c_b^* c_a \mu_{ba}$$  \hspace{1cm} (2.31)

where $\rho_{ab} = c_a c_b^*$ is the density matrix and thus gives the probability of finding the particle in the state to which $\rho_{ab}$ is applied. Moreover, by taking the time dependence of the density matrix, it can be shown that

$$\frac{\partial}{\partial t} \rho(t) = -i\frac{\hbar}{\hbar} [\hat{H}(t), \rho(t)]$$  \hspace{1cm} (2.32)

in what is known as the Liouville-von Neumann equation.\textsuperscript{99} It can be shown that for a density matrix, $\rho_{01}(t)$, of a single molecule which is expanded in the eigenbasis of $\hat{H}(t)$,

$$\frac{\partial}{\partial t} \rho_{01}(t) = -i\omega_{01}(t) \rho_{01}(t)$$  \hspace{1cm} (2.33)

where $\omega_{01}(t)$ is the frequency of the transition from state 0 to 1. Upon integration with respect to time,
\[ \rho_{01}(t) = \rho_{01}(0)e^{-i\int_{0}^{t} \omega_{01}(\tau)d\tau} \]  
\[ \text{(2.34)} \]

indicating that the off-diagonal density matrix oscillates as a function of time as the frequency fluctuates around some average value. However, upon taking the ensemble average, these oscillations become out of phase leading to a damping of the density matrix in time leading to homogeneous and inhomogeneous dynamics. If the time-dependent transition frequency is defined to be

\[ \omega_{01}(t) = \omega_{01} + \Delta \omega_{01}(t) \]  
\[ \text{(2.35)} \]

where the time-independent average, \( \omega_{01} \), and time-dependent fluctuations, \( \Delta \omega_{01}(t) \), are separated, then the ensemble average density matrix can be rewritten as

\[ \rho_{01}(t) \propto e^{-i\int_{0}^{t} \omega_{01}(\tau)d\tau} \left\langle e^{-i\int_{0}^{t} \Delta \omega_{01}(\tau)d\tau} \right\rangle. \]  
\[ \text{(2.36)} \]

Using the cumulant expansion method,\textsuperscript{97,99} this function can be expanded in powers of \( \Delta \omega_{01}(t) \) leading to

\[ \left\langle e^{-i\int_{0}^{t} \Delta \omega_{01}(\tau)d\tau} \right\rangle = 1 - i \int_{0}^{t} d\tau \left\langle \Delta \omega_{01}(\tau) \right\rangle - \frac{1}{2} \int_{0}^{t} d\tau_{1} \int_{0}^{t} d\tau_{2} \left\langle \Delta \omega_{01}(\tau_{1}) \Delta \omega_{01}(\tau_{2}) \right\rangle + \ldots. \]  
\[ \text{(2.37)} \]

Since \( \left\langle \Delta \omega_{01}(t) \right\rangle = 0 \), the lineshape function can be written as the second term,

\[ g(t) = \frac{1}{2} \int_{0}^{t} d\tau_{1} d\tau_{2} \left\langle \Delta \omega_{01}(\tau_{1}) \Delta \omega_{01}(\tau_{2}) \right\rangle \]  
\[ \text{(2.38)} \]
as this is the only term that contributes significantly for a system with Gaussian fluctuations. Since $\langle \delta \omega_{01}(\tau_1) \delta \omega_{01}(\tau_2) \rangle$ is both an even function and stationary, Eq. 2.38 can be rewritten as

$$g(t) = \frac{1}{2} \int_0^t \int_0^t d\tau_1 d\tau_2 \langle \delta \omega_{01}(\tau_2) \delta \omega_{01}(0) \rangle$$ (2.39)

where $\langle \delta \omega_{01}(\tau_2) \delta \omega_{01}(0) \rangle$ is called the frequency frequency correlation function (FFCF).

Ryogo Kubo proposed that this FFCF could be written as a function of both correlation time, $\tau_c$, and frequency fluctuation amplitude, $\Delta$, such that the FFCF decays to zero over time as follows:

$$\langle \delta \omega_{01}(\tau_2) \delta \omega_{01}(0) \rangle = \Delta^2 e^{-\frac{\tau_2}{\tau_c}}.$$ (2.40)

Thus, integrating this equation twice reveals the lineshape function,

$$g(t) = \Delta^2 \tau_c^2 \left[ e^{-\frac{t}{\tau_c}} + \frac{t}{\tau_c} - 1 \right].$$ (2.41)

The different lineshape functions (Figure 2.2) can be generated for the different limits of the frequency fluctuations. In the homogeneous limit, where the frequency fluctuations are very fast, that is when $\tau_c \ll 1$, $g(t)$ can be reduced to just the second term. This leads to a linear absorption spectrum of

$$S(\omega) \propto \frac{1/T_2}{(\omega - \omega_0)^2 + 1/T_2^2}$$ (2.42)

where $T_2 = (\Delta \tau_c)^{-1}$, thus yielding a band with a Lorentzian lineshape. Conversely, in the inhomogeneous limit, where the frequency fluctuations are slow, the linear absorption spectrum can be written as
since \( \langle \partial \omega_{01}(\tau_2) \partial \omega_{01}(0) \rangle \approx \Delta^2 \), simplifying the lineshape function to \( g(t) = \frac{\Delta^2}{2} t^2 \). Thus, the absorption band in this limit is a Gaussian line and reveals all the heterogeneity in the system.

### 2.3.2 Normal Modes and Transition Dipole Couplings

The main vibrational band used in the assessment of protein structure and dynamics is the amide I band which arises from the C=O stretching vibration and the N-H bend, both of which are present in the backbone units of the protein.\(^{101,102}\) It has found great utility because of its large extinction coefficient, spectral separation from other protein vibrations and sensitivity to protein secondary structure.\(^ {103}\) This sensitivity arises from the fact that the amide oscillators throughout the protein are able to interact with each other both through-bond and through-space thus delocalizing the amide I mode over the whole protein backbone. Therefore, distance and orientation changes of these dipole moments influence the overall features of the amide I band. Several models have been developed to describe and reproduce this vibrational band,\(^ {104–106}\) but this thesis will only discuss one in which each amide group is considered a single oscillator.\(^ {107}\)

Given this consideration, the Hamiltonian for the vibrational excitation can be written as the summation of the Hamiltonian for each, individual peptide group (\( \hat{H}_0 \)) and the interpeptide potential, or coupling term, (\( \hat{W} \)):\(^ {107}\)
\[ \hat{H} = \hat{H}_0 + \hat{W} = \begin{bmatrix} \omega_{11} & \beta_{12} & \beta_{13} & \ldots & \beta_{1n} \\ \beta_{21} & \omega_{22} & \beta_{23} & \ldots & \beta_{2n} \\ \beta_{31} & \beta_{32} & \omega_{33} & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \beta_{n,n-1} \\ \beta_{n1} & \beta_{n2} & \ldots & \beta_{n,n-1} & \omega_{nn} \end{bmatrix} \]  

(2.44)

where the diagonal matrix elements, \( \omega_{ii} \), are the frequencies of each oscillator and the cross-terms, \( \beta_{ij} \), are the couplings between the \( i \)th and \( j \)th oscillator.\(^{108}\) Additionally, the matrix is \( n \)-dimensional where \( n \) is the number of residues and thus amide bonds in the protein. Since the vibrational frequencies can be influenced by interactions with the solvent bath, such as electrostatic and hydrogen bonding, they can be modeled into the diagonal terms by

\[ \omega_{ii} = \omega_{gas} + \sum_{m=1}^{\infty} \left[ \lambda_m \phi_m + \sum_{\alpha} \kappa_{m,\alpha} E_{m,\alpha} + \sum_{\alpha > \beta} \gamma_{m,\alpha,\beta} E_{m,\alpha,\beta} \right] \]  

(2.45)

where \( \phi_m \) is the electrostatic potential imparted by the solvent at site or atom \( m \), \( E_{m,\alpha} \) is the vector electric field and \( E_{m,\alpha,\beta} \) is the gradient of the electric field tensor through three-dimensional space (\( \alpha, \beta \in \{x, y, z\} \)), \( \lambda_m, \kappa_m, \) and \( \gamma_m \) are coefficients and \( \omega_{gas} \) is the frequency of the oscillator in the gas phase.\(^{109}\) Additionally, the couplings in Eq 2.44 can be defined as

\[ \beta_{ij} = \frac{\bar{\mu}_i \cdot \bar{\mu}_j - 3(\bar{a}_i \cdot \bar{\mu}_j)(\bar{a}_i \cdot \bar{\mu}_j)}{r_{ij}^3} \]  

(2.46)

where \( \bar{\mu}_i \) is the unit vector of the transition dipole moment of the \( i \)th amide group, \( \bar{a}_i \) is the unit vector connecting the \( i \)th and \( j \)th amide group, and \( r_{ij} \) is the distance between
these two points.\textsuperscript{108} For each amide group, the magnitude of the transition dipoles are assumed to be the same.

The frequencies of the normal modes of the protein are the eigenvalues of Eq 2.44 and thus can be attained by diagonalizing the Hamiltonian. Moreover, the associated eigenvectors are the different vibrational modes and thus the transition dipole moment for a particular vibrational mode can be calculated by projecting (or taking the dot product) a dipole matrix, $\mathbf{d}$, for the $n$ different amide bonds onto a particular vibrational mode vector. The entire IR spectra can then be developed by taking the sum of these projections squared for all the vibrational eigenstates. Since this results in a stick spectrum, convolution of each vibrational mode with a Lorentzian line shape is performed to account for lifetime broadening as follows:

$$S(\omega) = \text{Im} \sum_n \left| \langle n | \mathbf{d} | 0 \rangle \right|^2 \frac{1}{(\omega - \omega_{e0} + i\Gamma)^2}.$$  \textsuperscript{109,110} (2.47)
Figure 2.1 Schematic representations of free energy diagrams: a) a two-state folding scenario where the folded (F) and unfolded (U) state are separated by a free energy barrier where the transition state (TS) complex is formed, b) a folding diagram involving the formation of an intermediate (I) complex such that F and U are separated by two free energy barriers, and c) a downhill folding scenario where there is no free energy barrier separating F and U.
Figure 2.2 Representative normalized lorentzian and gaussian functions with full width at half maximum (FWHM) of 20 cm$^{-1}$ and center frequency ($\omega_0$) of 1650 cm$^{-1}$.
3 Methods

3.1 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy is a spectroscopic method that uses circularly polarized light to gain insight into the chirality of molecules and thus has found extensive application in biochemistry.\textsuperscript{111–113} Due to the differential absorption of right and left circularly polarized light by the chiral centers in any biomolecule, the secondary and tertiary structure of protein and peptide systems can be assessed. Additionally, since the structure of a protein changes in response to environmental conditions, CD spectroscopy can be used with temperature or pH denaturation to dissect the thermodynamics of the folding process. Since one of the circularly polarized components is absorbed more than the other as it passes through the sample, the transmitted electric field vector oscillates in an elliptical pattern and thus CD spectroscopy measures the ellipticity angle (the inverse tangent of the semiminor axis of the ellipse over the semimajor axis) of this process as a function of wavelength.

The amide group of proteins is known to absorb in the far-UV region of the electromagnetic spectrum.\textsuperscript{114} In particular, there are three different electronic transitions in this region, namely an n→π* transition at 220 nm, a π→π* transition at 190 nm (NV\textsubscript{1}) and a π→π* transition at 140 nm (NV\textsubscript{2}). The n→π* transition is only weakly allowed; however, due to mixing of the n→π* transition with the NV\textsubscript{1} transition of either the same molecule (one-electron effect) or that of another peptide (μ→m coupling) this signal is enhanced in protein and peptide systems. Additionally, the π→π* transitions are able to couple among identical peptide groups in a protein and thus the N-fold degenerate excited state is split into N levels where N is the number of amide bonds in the protein. The
symmetry of the molecule then dictates which of these transitions are allowed or forbidden.

The different secondary structural elements of protein systems thus affect these coupling interactions and give rise to CD signatures, which are particular to the structural motif. For \( \alpha \)-helices, the \( n\rightarrow\pi^* \) transition produces a negative band at 222 nm whereas the \( \pi\rightarrow\pi^* \) transition undergoes exciton splitting to produce both a positive band at 192 nm and a negative band at 208 nm.\(^{115}\) The 222 nm band can be used to determine the helical fraction \((f_H)\) of a peptide, that is, the ratio of the average number of helical hydrogen bonds (H-bonds) over the total number possible in the peptide, using the following equation:

\[
f_H = \frac{[\theta]_{222} - [\theta]_C}{([\theta]_H - [\theta]_C)}
\]  

(3.1)

where \([\theta]_{222}\) is the mean residue ellipticity at 222 nm, \([\theta]_H\) is the mean residue ellipticity of a 100% helix and \([\theta]_C\) is the mean residue ellipticity of a 100% random coil.\(^{116}\) Additionally, these last two parameters can be calculated by

\[
[\theta]_H = (-44000 + 250 \cdot T) \cdot \frac{\left(n_H - a\right)}{n_T} 
\]  

(3.2)

\[
[\theta]_C = 640 - 45 \cdot T
\]  

(3.3)

where \(n_H\) is the number of residues in a helical conformation in the folded peptide, \(a\) is the number of helical residues with carbonyls that are not in intramolecular helical hydrogen bonds, \(n_T\) is the total number of residues in the peptide, and \(T\) is the temperature in Celsius. On the other hand, for \(\beta\)-sheets, these same electronic transitions are seen except the \(n\rightarrow\pi^* \) transition is located at 216 nm and the exciton splitting
produces a positive band at 195 nm and negative band at 175 nm.\textsuperscript{115} Moreover, CD spectroscopy can also report on the tertiary structure of proteins as interactions between nearby aromatic residues contribute to exciton couplets (one positive and one negative band) arising between 215 – 230 nm.\textsuperscript{117}

In order to assess the thermodynamics of a given protein system, thermal denaturation is often used in conjunction with CD spectroscopy. Depending on the secondary structure of the protein, a single wavelength (either 222 nm for $\alpha$-helix or 216 nm for $\beta$-sheet) is monitored over a wide temperature range. The sigmoidal curve that results can then be fit with the following thermodynamic equation provided the two-state approximation can be made:

$$[\theta]_f = \frac{[\theta]_u + K[\theta]_f}{1 + K}$$

(3.4)

where $K$ is the equilibrium constant, $[\theta]_u$ is the mean residue ellipticity of the unfolded state baseline, and $[\theta]_f$ is the mean residue ellipticity of the folded state baseline.\textsuperscript{118} These baselines are assumed to be linear functions of temperature. Additionally, $K$ is defined as in Eq 2.20 and thus Eq 2.11 can be used to determine $\Delta_r G^\circ(T)$ with all the thermodynamic functions being treated as fitting variables under the constraint of Eq 2.8. In this way, the $T_m$ of a given protein can be determined.

### 3.2 Infrared Spectroscopy

Infrared (IR) spectroscopy has emerged as a key spectroscopic technique to gain insight into biological systems, particularly those involving protein molecules. Not only is this
method sensitive to the chemical composition and structure of these molecules, but it is also able to access a wide range of timescales and thus provide valuable information on protein dynamics.\textsuperscript{119} Moreover, IR spectroscopy can be used with both soluble and membrane protein systems as well as cell and tissue samples. When studying proteins, the mid-IR region (4000 – 200 cm\(^{-1}\)) of the IR spectrum is often utilized as it excites fundamental vibrations which are specific to the molecule being studied.\textsuperscript{120} However, due to the vast number of different vibrational moieties in a protein, there are a large number of oscillators with overlapping absorbance bands and vibrational couplings leading to very complex spectra with unresolved features. Still, global protein structural information, such as secondary structure, can be elucidated. Additionally, the use of non-natural amino acid vibrational probes can be used to gain more site-specific information of protein structure, dynamics, and environments.\textsuperscript{64}

Due to its large extinction coefficient and structural sensitivity, the amide I absorption band of the protein backbone is often used as a probe of protein conformation and conformational changes. This mode is mainly composed of the carbonyl stretch, which is repeated along the whole protein backbone, but does not have considerable contribution from amino acid side-chains. Hydrogen-bonding strength and transition dipole coupling affect the frequency and lineshape of the amide I band causing it to be sensitive to secondary structural changes.\textsuperscript{103} For example, \(\alpha\)-helices absorb around 1652 cm\(^{-1}\) in D\(_2\)O, which shifts closer to 1630 cm\(^{-1}\) upon more solvent exposure. This redshift is caused by water hydrogen bonding to the C=O oscillator thus reducing the covalent bond vibrational energy due to the redistribution of electrons. On the other hand,
antiparallel β-sheets give rise to two bands: a high intensity band near 1630 cm$^{-1}$ and a weaker band around 1685 cm$^{-1}$. This exciton splitting is the result of interstrand and intrastrand coupling between in-phase and out-of-phase (relative to each other) amide oscillators, respectively. However, parallel β-sheets often only exhibit a single band around 1620 cm$^{-1}$. Lastly, random coil or disordered protein structure exhibits a very broad band centered around 1650 cm$^{-1}$. The width of this band is dictated by inhomogeneous broadening as each amide oscillator is experiencing a different microscopic environment. Below, we discuss specific IR-based techniques that allow us to dissect protein structure and dynamics.

3.2.1 Fourier Transform Infrared Spectroscopy

Fourier Transform Infrared (FTIR) spectroscopy is a commonly used method to attain steady-state information on protein systems. This technique works by using broadband radiation that undergoes wave interference to produce different combinations of frequencies that are focused onto the sample. Specifically, light from a broadband, incoherent infrared lamp is directed onto a beam splitter, which then creates the two arms of the interferometer with 50% of the light traversing each. One of these arms contains a stationary mirror from which light is reflected back to the beam splitter while the other arm has a moving mirror that does the same. The recombined beam is then focused onto the sample and then onto a detector. Due to the difference in length of the two arms, also known as the optical path difference (OPD), the two beams undergo constructive and destructive interference, thus blocking and transmitting different combinations of frequencies onto the sample as the moving mirror changes position. Therefore, an FTIR
interferogram is created as a function of the moving mirror position with the largest intensity being at zero OPD. The interference intensity follows:

\[ I(x) = \int_0^\infty A(\omega) \cos(2\pi \omega x) d\omega \]  

(3.5)

where \( A(\omega) \) is the intensity of the light, \( \omega \) is the frequency, and \( x \) is the OPD. By taking the inverse Fourier transform,

\[ A(\omega) = \int_0^\infty I(x) \cos(2\pi \omega x) dx \]  

(3.6)

the spectrum as a function of \( \omega \) can be recovered.

In this thesis, for each measurement, a buffer solution is collected along with the sample for background subtraction. Teflon spacers are used to create the optical pathlength, and in most cases, two-compartment cells are used which contain buffer on one side of the cell and sample on the other. Moreover, for temperature dependent studies, a heating stage is used to control the temperature of the cell and solution. Upon subtracting lower temperature spectra from those of higher temperature, FTIR difference spectra can be obtained which provide further insight into the conformational changes occurring in the protein system.

### 3.2.2 Laser-Induced Temperature-Jump Infrared Spectroscopy

In order to determine kinetic quantities of protein systems such as rate constants and relaxations times, it is necessary to quickly perturb the sample’s equilibrium state. After its discovery by Eigen and coworkers,\(^{122}\) laser-induced temperature jump infrared spectroscopy has become a primary way to achieve this perturbation. The disturbance induced by the temperature jump on a pico- to nanosecond timescale displaces the
chemical equilibrium of the system and the relaxation back to equilibrium can be monitored.

The schematic (Figure 3.1) illustrates the spectrometer setup for laser-induced temperature jump experiments performed in this thesis. As is shown, typically a Raman-shifted Q-switched Nd:YAG laser is used to initiate the temperature jump in the sample via a pulsed beam. This pulse is absorbed by the solvent and produce radiationless decay in the form of heat, thus providing the temperature jump in a few nanoseconds. Although modelocked lasers are able to provide shorter pulses on the picoseconds timescale, they need to be pulse-selected and amplified to intensify the magnitude of the induced temperature jump. Therefore, Q-switched lasers are often used as they provide enough energy to make a sufficient temperature change on the order to 10 °C. However, as is shown in the diagram, the Q-switched Nd:YAG laser produces an emission wavelength around 1064nm. Water does not strongly absorb at this wavelength, so early studies used absorbing dyes for the heating. One disadvantage to this method is that the dye may interfere with the chemistry of the system. Therefore, a focal lens is used to focus the beam into a hydrogen Raman cell, which shifts the laser wavelength to a strong absorbing region of water.

The hydrogen Raman cell is able to change the frequency of the Q-switched Nd:YAG laser by means of Raman shifts. For Stokes and Anti-Stokes shifts, the resulting frequencies are given by Eqs 3.7 and 3.8 respectively where \( v_{\text{photon}} \) is the frequency of the incoming laser, \( v_{\text{vibration}} \) is the vibrational frequency of the molecules (hydrogen in this case), and \( n \) is the order of the overtone,
\[ v_{\text{Stokes}} = v_\text{photon} - n v_\text{vibration} \quad (3.7) \]

\[ v_{\text{Anti-Stokes}} = v_\text{photon} + n v_\text{vibration} \quad (3.8) \]

For the hydrogen Raman cell, the Stokes frequency for \( n = 1 \) is 5,238 cm\(^{-1}\) or 1,909 nm, which falls directly in the overtone stretch region for H\(_2\)O and D\(_2\)O. However, H\(_2\)O has a larger absorption coefficient at this wavelength than does D\(_2\)O, so often D\(_2\)O is used as the solvent medium. Methane Raman cells are more appropriate to use with H\(_2\)O as the solvent.\(^{124}\)

It is worth noting, however, that by collimating the Nd:YAG laser beam into the Raman converter, the input beam has sufficiently high power to produce a stimulated Raman Effect. As a stimulated beam, the output pulse from the Raman cell is coherent and can have up to 30% conversion efficiency. Moreover, instead of scattering as in spontaneous Raman scattering, this stimulated pulse continues in the direction of the incident beam. Unfortunately, above a certain threshold Stokes beam intensity, this stimulated beam can act as a pump to produce higher order Stokes shifts. Additionally, mixing with the incident pulse produces Anti-Stokes frequency radiation. This development reduces the energy of the beam corresponding to the Stokes frequency for \( n = 1 \) and thus is unfavorable. Since the Anti-Stokes generation is dependent on the phase vector matching, a weakly focused beam is necessary to optimize Stokes generation. This often requires Raman cells to be 1-2 m long to prevent window damage.\(^{125}\) The 1,909nm wavelength is the only frequency of interest in these studies. Therefore, after focusing the
output pulse with a focal lens, a prism is used to disperse the various Raman shifted frequencies and then the Stokes frequency for n=1 is chosen and focused onto the sample.

Lastly, an infrared laser, such as a quantum cascade laser or lead salt laser, is used to supply a probe beam into the sample, overlapped with the pump spot. This source is tunable and continuous wave. The frequency can be tuned by varying the temperature and the current via temperature controllers and current sources respectively. Another alternative is to use a monochromator to pick the frequency of interest. This beam is focused into the sample by a focal lens and then recollimated after exiting the sample. At this point, a Mercury Cadmium Telluride (MCT) detector detects the IR intensity change of the probe laser with respect to the temperature jumps. Furthermore, an oscilloscope is used to digitize this information, and process the data to extract the transient signal.

Without considering thermal diffusion, the temperature jump \( \delta T \) generated by the laser pulse in the sample is given by\(^{123}\)

\[
\delta T(r, z, t) = \frac{k}{\rho c_v} \left[ I(r, z, t) - I_0 \right] dt.
\]  

(3.9)

In this equation, \( I(r, z, t) \) is the instantaneous laser fluence (W/cm\(^2\)) or flux integrated over time at time \( t \) and axial position \( z \). Moreover, \( k \) is an absorption coefficient, \( \rho \) is the solution density, and \( c_v \) is the heat capacity. The Beer-Lambert law governs the intensity dependence of an absorbing medium on the coordinate \( z \) by\(^{125}\)

\[
I(r, z) = I_0(r, t) e^{-kz}.
\]

(3.10)

However, this pulse is temporary and the thermal diffusivity of the system determines its duration. If the sample is considered a uniformly heated infinite cylinder, which is an
approximation for the case where the pathlength is much longer than the beam diameter, then the decay of the temperature jump is given by,

$$\delta T(t) = \delta T_{\text{max}} \left(1 - e^{-\frac{t^2}{4a^2}}\right)$$ (3.11)

where $\kappa$ as the diffusivity. However if the opposite case is true, the pathlength is shorter than the beam diameter, the decay of the temperature jump can be represented as

$$\delta T(z,t) = \delta T_{\text{max}} \left(\text{erf} \left(\frac{a-z}{2\sqrt{\kappa a}}\right) + \text{erf} \left(\frac{a+z}{2\sqrt{\kappa a}}\right)\right)$$ (3.12)

which approximates this scenario as an infinite uniformly heated slab in a homogeneous medium. In this equation, $\text{erf}$ is the error function, $2a$ is the path length, and $z$ is the displacement from the center. This would be the duration equation used most often for an IR pump whereas Eq 3.11 is most likely used for UV pumps.

Lastly, in performing these experiments, a few difficulties must be overcome. First, often non-uniform heating along the beam path results due to the extinction of the beam intensity as it propagates, placing restrictions on the pathlength and the probe beam diameter. Additionally, the rapid temperature jump can also cause photo-acoustic effects that can interfere with the signal. Cavitation or bubble formation can also be produced by the large pressure change. In both of these cases, precautions such as degassing the sample, reducing the sample path length, and repositioning the beam help reduce their probability. Finally, the expansion that takes place in the system can cause “thermal
lensing” or a refractive index gradient, which can affect the probe beam direction. Once again, careful alignment of the pump and probe beams can avoid this.

### 3.2.3 Flash Photolysis Infrared Spectroscopy

Flash photolysis infrared spectroscopy (Figure 3.2) utilizes UV or visible light to trigger a protein-folding event by exciting a chromophore in or attached to the protein system. The setup used in this thesis is very similar to that of the $T$-jump setup with a few key differences. An Nd:YAG laser with harmonic generation is used to generate the UV or visible pulses. Harmonic generation using a beta barium borate (BBO) crystal can be used to double, triple or quadruple the frequency of the input beam. In the case of Nd:YAG laser, the 2$^{\text{nd}}$, 3$^{\text{rd}}$, and 4$^{\text{th}}$ harmonic wavelengths generated are 532nm, 355nm, and 266nm respectively. This beam is then focused into the sample compartment with UV optics and attenuated using pinholes and neutral density filters as needed and overlapped with the IR probe beam similar to the $T$-jump experiment described above.

### 3.2.4 Two-Dimensional Infrared Spectroscopy

Two-dimensional infrared (2D IR) spectroscopy is a time-resolved spectroscopic technique that elucidates vibrational coupling and energy redistribution details of vibrational transitions by spreading the spectroscopic information over two-dimensions. This method is conducted in the time-domain using a three-pulse sequence derived from a femtosecond Ti:Sapphire amplifier (Figure 3.3). The 800nm light originating from the laser is converted into the near-IR region (~1400 nm and ~1900 nm) by a home-built optical parametric amplifier (OPA). A difference frequency generator (DFG) is then used to generate the necessary IR frequency ($5–6 \mu$m), which is then split into three pulses that
are focused onto the sample in a boxcar geometry. The first pulse creates a coherence between the ground vibrational state and the first excited state, which the second pulse then converts to a population. The time between these two pulses is thus known as the coherence time. Another time delay, known as the waiting time, follows this second pulse after which a third pulse generates a coherence once again. An echo signal that contains information about what frequencies survived the waiting time results from this last interaction. This radiated signal is then heterodyned with a local oscillator pulse to dissect the frequency and phase of the signal. The signal is then dispersed onto a monochromator and detected by an MCT array detector. One Fourier transform is then needed to convert this mixed frequency-time domain information into the frequency-frequency domain of a 2D IR spectrum.

Spectral diffusion is the process by which the instantaneous frequency of each oscillator evolves in time to explore a larger distribution of frequencies (Figure 3.4). These frequency fluctuations can be the result of hydrogen bond making and breaking, conformational changes, as well as a number of other different processes. Therefore, although the system of interest begins in an inhomogeneously broadened distribution, during the waiting time, it undergoes homogenization. By varying this waiting time, the frequency correlation decays can be measured. In particular, for polar solvents, the frequency-frequency correlation function (FFCF) is often modeled using,

\[
C(t) = \frac{\delta(t)}{T_2} + \Delta^2 e^{-\frac{t}{\tau}} + \Delta_s^2
\]  

(3.15)
where $T_2^*$ is the pure dephasing time, $\Delta$ is the frequency fluctuation amplitude, $\tau$ is the correlation time, and $\Delta_s$ is the static offset. Previous studies have shown that the inverse of the center line slope (CLS$^{-1}$), which is the slope of the line connecting the maxima at each pump frequency, as a function of the waiting time, $T$, is directly related to the FFCF and thus reports on spectral diffusion dynamics. By fitting the CLS$^{-1}$ as a function of waiting time, $T$, with an exponential and an offset, the value of $\tau$ can be determined.

3.3 Cosolvents

Cosolvents and cosolutes play very important roles in protein folding as they can affect both the structure and dynamics of this folding process. In fact, the cellular environment consists of a large number of different molecules, both large and small, to help maintain homeostasis in the cellular environment and provide the machinery for cell function. Some of these molecules, such as urea, are known to denature proteins whereas others, such as trimethylamine $N$-oxide (TMAO), help stabilize protein structure. In general, these small organic molecules are known as osmolytes as they help maintain the osmotic pressure in the cell while also aiding in protein folding and stability. $^{127}$ Moreover, due to the vast number and types of molecule in the cell, proteins have to fold in very crowded environments, which is quite different from the in vitro conditions typically used. Thus, it is important to understand the effects that these cosolvents and cosolutes can have on the protein folding process as well as the mechanisms involved. Additionally, these same small molecules, along with others, are often used as tools in in vitro studies to alter or direct protein folding structure and dynamics.
Perhaps the most commonly used cosolvents and cosolutes in protein folding studies are small molecule denaturants. In this regard, urea and guanidinium chloride (GdmCl) can be used to unfold a protein for a variety of purposes including labeling strategies,\textsuperscript{128} folding kinetic studies,\textsuperscript{129,130} and stability measurements.\textsuperscript{131} Additionally, protein stabilizers, such as TMAO and alcohols at certain concentrations, are used to help stabilize the folded state of the protein molecule, both in the cell and in vitro. For example, low concentrations of 2,2,2-trifluoroethanol (TFE) are often used to induce secondary structure, particularly \(\alpha\)-helical structure, into intrinsically disordered peptides and proteins.\textsuperscript{132–134} Despite their general uses, the mechanisms by which these small molecules affect protein structure is still not fully understood.

In general, there are two proposed mechanisms by which stabilizers and denaturants affect protein structure.\textsuperscript{135} First, several studies have supported the notion that these small molecules actually do not directly interact with the protein but instead indirectly affect the stability of the protein by changing the structure of the water.\textsuperscript{136} In the case of denaturants, the solute disrupts the hydrogen-bonding network of the water, decreasing its structure and thereby weakening the hydrophobic effect. Conversely, stabilizers are thought to increase the number of hydrogen bonds within the solvents thus discouraging interactions with the protein.\textsuperscript{137} The second theory is a direct mechanism whereby the cosolutes interact with the protein molecule itself. Therefore, denaturants form hydrogen bonds with polar moieties of the protein thus reducing intramolecular hydrogen bonding.\textsuperscript{138,139} However, many studies have shown that stabilizing osmolytes are actually depleted from the protein surface thus allowing hydration of the protein.
Despite the mechanism involved, it is obvious that these small molecules can greatly change the interactions of protein molecules and thus alter their structural and dynamic properties.

It is well known that the cell consists of many large molecules such as proteins, lipids, and carbohydrates, at high concentrations, which all create a very crowded environment for protein folding and function. This process, known as macromolecular crowding, can affect both the thermodynamics and kinetics of protein folding by the excluded volume effect, that is, a reduction in the solvent or space available to the protein molecule. In fact, all the biomolecules in the cell take up between 5% to 40% of the available solvent volume.\textsuperscript{140} Since the unfolded state of the protein has less space to extend, it is often more compact than it would be in pure water, and thus has a reduced configurational entropy. In several studies it has been shown that the excluded volume effect can increase the rate of protein folding,\textsuperscript{141,142} however, macromolecular crowding can also increase the viscosity which could alternately affect the folding rate.\textsuperscript{143} Typically, this macromolecular crowding affect is simulated in vitro using large inert polymers, such as dextran and ficoll, or proteins, such as lysozyme. However, lesser known is that small molecules, such as osmolytes and alcohols, can also cause protein crowding by clustering around protein molecules in a process termed, “nano-crowding”.

Thirumalai, Straub and coworkers first used the word “nano-crowding” in 2011 in their computational study of TMAO induced protein folding,\textsuperscript{144} and since then, many recent studies have found that this term is an accurate description of the observed cosolvent interactions in various biological systems. Using molecular dynamics (MD)
simulations of several peptide systems, their results indicated that TMAO can act as a crowding agent by entropically destabilizing a peptide’s unfolded state via TMAO-peptide H-bonds while stabilizing the folded state by the excluded volume effect. Trehalose and glycine betain, which are other organic osmolytes, have also been shown to counteract denaturation of proteins by similar entropic mechanisms.\textsuperscript{145,146} Additionally, despite its denaturation abilities, urea has actually been shown to produce a local crowding effect on proteins in the presence of GdmCl as it is expelled from the protein surface by the more favorable GdmCl.\textsuperscript{147} Therefore, these crowding affects should be taken into account when designing experiments.

3.4 Photochemical Triggering Systems

Another method that has found wider use over the past decade in protein folding studies is the use of photochemical triggering systems. In ensemble measurements, a perturbation of equilibrium must take place in order to observe protein kinetics. Oftentimes, this is implemented by changing a thermodynamic variable such as temperature\textsuperscript{148} or pressure.\textsuperscript{149} However, these methods lack control over the conformation of the protein at any given point in the experiment and thus only general structural information can be gathered. Therefore, phototriggering systems are helpful in this regard, as they reduce the distribution of conformations that are accessible to the protein molecule at the start of the experiment thereby allowing for a better structural understanding of the process. They also allow the user control over a biochemical process as they can be strategically and site-specifically manipulated with light. Additionally, phototriggers can come in several
varieties including photocages, photoswitches, and photocrosslinkers, which can be used for different experimental needs from secondary structure analysis to assembly formation. Below we discuss some of the key considerations for designing photochemical triggers as well as a brief summary of different types of phototriggers.

For a phototriggering system to be useful in protein folding studies, there are key properties that must be considered.\textsuperscript{34} First, the molecule of interest must have a large photochemical yield, such that enough of the reactant is converted to product so that an experiment can produce a noticeable signal. Secondly, the rate of the photochemical process must be faster than the timescale being monitored in the experiment. For example, for protein folding processes that occur on a nanosecond or microsecond timescales, photochemical triggers that complete their photoreaction on the picosecond timescale would be most useful. These phototriggers must also be able to be easily incorporated in the protein system. Click chemistry\textsuperscript{150} or cysteine and lysine bioconjugation methods\textsuperscript{151} are typical means to do just this. Additionally, the phototrigger of interest must be relatively small or leave behind small, if any, perturbations to the protein after photolysis. Moreover, the products of the reaction must be inert and there should be minimal side reactions. It is also preferable, in some cases, that visible or near UV light is used in the photolysis, particularly for applications in vivo or for proteins containing aromatic amino acids, which might otherwise be excited in the process.

Photocages are a variety of phototriggers, which are localized to a single amino acid residue. In general, these are photocages are bulky, ring structures, typically
nitrobenzyl derivatives, which can be attached to cysteine, lysine or serine residues. Upon photocleavage, the natural amino acid is regained. This method allows for site-specific control over protein interactions as a single amino acids that are important to protein structure or function can essentially be irreversibly turned on or off with light. Some examples of the use of such photocages include manipulation of peptide hydrogel disassembly, control over protein phosphorylation and triggering of protein splicing. Alternatively, photoswitches offer the ability to study the interconversion between two protein structures as the reversible phototrigger changes its end-to-end distance upon isomerization. These moieties are typically azobenzene or stilbene derivatives and have found applications in protein aggregation studies. Lastly, photochemical cross-linkers which completely break apart with light are perhaps the most useful type of phototriggers in studying protein folding reactions. These phototriggers link a protein of peptide between two points at the start of a protein folding reaction and then release it to allow the protein to fold or unfold naturally. Tetrazine as well as some disulfide cleavage methods have found use in this area.

3.5 Site-Specific Spectroscopic Probes

Site-specific spectroscopic probes are useful tools in the investigation of protein environment, structure and function. Since the amide I region of the infrared spectrum provides a global picture of the whole protein backbone, site-specific information on protein environment or conformation is often lacking. However, by localizing a vibrational moiety onto a specific region of the protein sequence, this information can be
garnered. For infrared spectroscopy, these probes must absorb in an uncongested region of the infrared spectrum with a relatively high extinction coefficient. Moreover, they should be sensitive to environmental changes such as electric field or solvent conditions. There are several nitrile (e.g. 5-cyanotryptophan\textsuperscript{159,160} and 4-cyanophenylalanine\textsuperscript{161}) and carbonyl-derivatized (e.g. l-aspartic acid 4-methyl ester and l-glutamic acid 5-methyl ester)\textsuperscript{162,163} unnatural amino acids that have these properties with the added advantages that they can be easily incorporated into protein or peptide systems and are small in size so they do not alter the protein structure. The nitrile stretching vibration absorbs between 2200 cm\(^{-1}\) and 2300 cm\(^{-1}\) and the carbonyl stretching vibration of esters, ketones, and aldehydes takes place between 1700 cm\(^{-1}\) and 1800 cm\(^{-1}\) and are thus outside backbone absorbing regions. Additionally, besides extrinsic vibrational reporters, isotope-labeling strategies\textsuperscript{164} are also useful to alter the vibrational frequency of the native amino acid. For example, a backbone \(^{13}\)C=O label redshifts the carbonyl stretching frequency by \(\sim 40\) cm\(^{-1}\). This provides a non-perturbative way to assess local interactions in the protein.

Hydrogen bonding within the protein and between the protein and water molecules plays a very critical role protein stability, structure, and function. Not only is the hydrogen bonding pattern the key determinant of protein secondary structure, but hydrogen bonds between sidechains aid in the packing of proteins\textsuperscript{165} and water-protein hydrogen bonds can help mediate protein-ligand interactions.\textsuperscript{166} The hydrogen bonding structure of bulk water\textsuperscript{167–169} is very different from water surrounding a protein molecule.\textsuperscript{170} In bulk, water forms a very large extended network with some local areas of clustering; however, dissolved biomolecules disrupt this pattern as some favorable
hydrogen bonding interaction can no longer be satisfied, specifically at a hydrophobic interface. The reorganization of the protein molecule (protein folding) and the water molecules then is crucial to the correct folding of the protein. However, because hydrogen bonding is a relatively weak interaction that can be disrupted easily by small thermal fluctuation energies, during protein folding or function, the hydrogen bonding network is constantly changing. These hydrogen bonding dynamics often take place on the picosecond timescale and thus infrared spectroscopy is helpful in the analysis of them.\textsuperscript{171} Moreover, using site-specific infrared probes allows for investigation of local hydrogen bonding interactions and motions.

Site-specific infrared probes have also found use in the investigation of local electrostatic fields. Electric fields are ubiquitous in nature as biology is made up of a variety of non-covalent interactions, which arise as a result of charge separation.\textsuperscript{172,173} It is this separation of charges and the fluctuation thereof that actually helps dictate protein function and folding. Therefore, insight into the modulation and properties of these local electric fields can aid in our understanding of and even control over protein interactions. Changes in the electric field present in protein systems can actually produce a distinguishable signal in the infrared spectrum as a result of the vibrational Stark effect. Essentially the presence of an electric field produces a shift in the vibrational frequency of an oscillator according to

\[ E = h c \nu = -\vec{F} \cdot \vec{\mu} \quad (3.16) \]

where $E$ is the energy, $h$ is Planck’s constant, $c$ is the speed of light, $\nu$ is the observed vibrational frequency, $\vec{F}$ is the electric field vector, and $\vec{\mu}$ is the Stark tuning rate.\textsuperscript{174} The
Stark tuning rate is highly dependent on the vibrational oscillator with higher values leading to more dramatic Stark shifts. Moreover, some vibrational probes have a more complex dependence on electric field rather than just the first order linear term. Thus, there is a need to find vibrational reporters, which are site-specific with a high Stark tuning rate and a vibrational frequency that is linearly dependent on local electric field. The carbonyl and nitrile stretching frequency have found utility in the regard and several unnatural amino acids containing these moieties have been studied and utilized to measure the electric field in protein systems.\textsuperscript{163,175,176}
**Figure 3.1** Diagram of the laser-induced temperature-jump (T-jump) infrared spectroscopy setup. The Q-switched Nd:YAG laser is used to produce a 1064 nm nanosecond pulse which is Raman shifted to the ~1.9 \( \mu \text{m} \) T-jump pulse. The CW tunable quantum cascade laser is used to probe the 1600 – 1700 cm\(^{-1}\) region of the infrared spectrum. The MCT detector is then able to monitor changes in the absorbance of the sample as a result of the temperature-jump which are then displayed by the oscilloscope.
Figure 3.2 Diagram of the flash photolysis infrared spectroscopy setup. The Minilite II Nd:YAG laser can be used to produce either a 1064 nm, 532 nm, 355 nm or 266 nm nanosecond pulse which is directed into the sample. The CW tunable quantum cascade laser is used to probe the 1600 – 1700 cm\(^{-1}\) region of the infrared spectrum. The MCT detector is then able to monitor changes in the absorbance of the sample as a result of the flash photolysis which are then displayed by the oscilloscope.
Figure 3.3 Cartoon representation of the 2D IR pulse sequence. The first two pump pulses are separated by a coherence time which is followed by a probe pulse after some waiting time. The emitted field is heterodyned with a local oscillator for detection.
Figure 3.4 Schematic illustration of the appearance of spectral diffusion in a 2D IR spectrum. At early waiting time, each excited frequency remains static over the waiting time and the spectra are inhomogeneously broadened. At longer waiting time, the excited frequency fluctuates and thus creates a more homogeneously broadened spectrum.
4 Experimental Validation of the Role of Trifluoroethanol as a Nanocrowder

4.1 Original Publication


4.2 Introduction

While there are many ways to experimentally perturb a protein’s stability, perhaps one of the most common is through the use of cosolvents. For example, guanidinium chloride (GdmCl) and urea are frequently used to denature proteins, whereas several alcohols, such as hexafluoroisopropanol (HFIP) and trifluoroethanol (TFE), are known to induce secondary structure formation in polypeptides. Although there have been numerous efforts to understand how cosolvents act to change a protein’s conformational preference, in each case, unanswered questions still remain. Herein, we study the conformational relaxation kinetics of two intrinsically disordered proteins (IDP) in different water/TFE mixtures, aiming to gain a better understanding of the mechanism with which this cosolvent influences the dynamics of protein folding.

The protein-stabilizing effect of TFE has been studied extensively both experimentally and computationally since its discovery by Goodman and Listowsky.178–180 One view on TFE’s mechanism of action is that it more favorably surrounds the
protein than water, effectively leading to dehydration of the protein backbone, which, consequently, leads to backbone-backbone hydrogen bond formation and hence promotes secondary structure stabilization.\textsuperscript{181–189} Conversely, other studies suggest that rather than stabilize the folded state, TFE acts to destabilize the unfolded state by structuring the solvent and, as a result, increasing the folded population.\textsuperscript{190–192} Not surprisingly, some proposed mechanisms fall somewhere in-between.\textsuperscript{193–197} In addition, it has been shown that the amphiphilic TFE molecule is capable, at large volume percentages, of exposing and interacting with hydrophobic side chains, thereby leading to disruption of hydrophobic tertiary interactions.\textsuperscript{198} Due to the complexity of protein-TFE interactions, one expects that TFE will affect protein folding kinetics in a nonlinear manner. Indeed, Hamada et al.\textsuperscript{199} found that the folding rates of a set of globular proteins follow a chevron-like trend with increasing TFE concentration. The interpretation for these results was that at low TFE percentages, folding rates are increased due to stabilization of native hydrogen-bonding groups, whereas at higher percentages, folding rates are decreased in a similar manner as is found with denaturants, due to TFE’s interaction with buried residues, as determined by a high correlation between the $m$-values of TFE and GdmCl.\textsuperscript{199}

One factor that is potentially important to TFE’s effect on protein folding, but not considered by previous studies, is the ability of TFE to self-associate. For example, dynamic light scattering (DLS) and nuclear magnetic resonance (NMR) measurements, as well as molecular dynamics (MD) simulations, found that TFE molecules can form clusters.\textsuperscript{200–205} This clustering is thought to be the result of the cosolvent’s hydrophobic
CF$_3$ groups shielding themselves from water in micellelike structures that have Stokes’ radii of 0.55 nm. Furthermore, TFE clustering does not show a monotonic dependence on its percentage; it reaches a maximum at about 30% (v/v), above which the clusters disassemble and the solution becomes more homogeneous. Taken together, these findings suggest that TFE could act as a molecular crowder, thus increasing folding rates at certain percentages via the excluded volume effect. In addition, the viscosity of TFE/ water mixtures doubles from 0% to 60% TFE. Such a drastic increase in solvent viscosity could also have notable impacts on the folding rates of proteins in these solutions.

In order to gain insight into the effect of viscosity and cosolvent aggregation on protein folding kinetics, we have examined the conformational relaxation rates of two IDPs in different water/TFE solutions. IDPs are ideal candidates for this study, because they lack appreciable tertiary structure when isolated in buffer, simplifying our interpretations. Specifically, we studied the phosphorylated kinase inducible domain (pKID) peptide and the late embryogenesis abundant (LEA) peptide. We chose these two systems because both have folded states that are rich in $\alpha$-helical content; however, pKID forms a helix-turn-helix (HTH) structure, whereas LEA folds into a monomeric $\alpha$-helix. Previous experiments have shown that macromolecular crowding only has a small effect on the folding rate of monomeric $\alpha$-helices, whereas proteins with appreciable nonlocal contacts experience more of a change. Our hypothesis is that if TFE indeed behaves as a nanocrowder, it will affect the folding rate of pKID differently than that of LEA. Our results indeed reveal that the relaxation rate of pKID shows a complex
dependence on the TFE percentage (in the range of 0−50%), with a maximum occurring between 15 and 30%, whereas that of LEA does not show such a dependence.

4.3 Experimental Section

4.3.1 Sample Preparation and Spectroscopy

Deuterated TFE was purchased from Cambridge Isotope Laboratories and stored in a drybox upon opening. Peptides were synthesized on a PS3 automated peptide synthesizer (Protein Technologies, MA) using Fmoc-protocols, purified by reverse-phase chromatography, and identified by matrix-assisted laser desorption ionization (MALDI) mass spectroscopy. Phosphorylated serine was incorporated into pKID (sequence DSVTDSQKRREILSRRPS*YRKILNDLSSDAPG−CONH$_2$, with S* representing phosphoserine) via the modified amino acid Fmoc-Ser(HPO$_3$Bzl)−OH. The sequence of the LEA peptide is AADGAKEKAGEAADGAKEKAGE−CONH$_2$. CD measurements were carried out on an Aviv 62A DS spectropolarimeter (Aviv Associates, NJ) with a 1 mm sample holder. The peptide concentration was in the range of 50−60 μM in H$_2$O and various concentrations of TFE (pH 7). Fourier transform infrared (FTIR) spectra were collected with 1 cm$^{-1}$ resolution on a Magna-IR 860 spectrometer (Nicolet, WI) using a two-compartment CaF$_2$ sample cell of 56 μm path length. The details of the laser-induced temperature jump (T-jump) IR setup have been described elsewhere.$^{52}$ The amide hydrogen of peptides used in IR measurements has been exchanged to deuterium; the samples were prepared by directly dissolving lyophilized solids in D$_2$O solutions.
containing desired percentages of deuterated TFE (pH* 7). The final peptide concentration was between 1–2 mM.

4.3.2 Fractional Helicity Calculation

The fractional helicity of the peptide, \( f_H \), was estimated based on its mean residue ellipticity at 222 nm, [\( \theta \)]\(_{222} \), using the following relationship:

\[
f_H = \frac{([\theta]_{222} - [\theta]_C)}{([\theta]_H - [\theta]_C)}
\]

(4.1)

where [\( \theta \)]\(_{H} \) is defined as:

\[
[\theta]_H = (-44000 + 250T)(n_H - a)/n_T
\]

(4.2)

and [\( \theta \)]\(_{C} \) is defined as:

\[
[\theta]_C = 640 - 45T
\]

(4.3)

where \( n_H \) is the number of helical residues in the peptide folded state (\( n_H \) was defined as 21 for pKID and 22 for LEA), \( n_T \) is the total number of residues in the peptide, \( a \) is the number of carbonyls in helical structure not involved in intramolecular helical hydrogen bonding (\( a = 6 \) for pKID and 3 for LEA), and \( T \) is the temperature in Celsius.

4.4 Results

4.4.1 Helicity and Relaxation Rates of pKID in TFE Solutions

We chose pKID as our model system because a previous study has shown that TFE (10–40%) can significantly increase its helical content.\(^{210} \) Consistent with this finding, our CD measurements indicate that the helicity of pKID increases with increasing TFE percentage from 0 to 30%, above which this increase levels off (Figure 4.1). In addition, the thermal melt (\( T \)-melt) of this peptide, probed at 222 nm, indicates that TFE has an
effect on the nature of the thermal unfolding transition (Figure 4.2). Specifically, it appears that the unfolding transition becomes most cooperative when the percentage of TFE is approximately 15%, whereas at higher TFE concentrations (e.g., 50%) the $T$-melt is essentially linear. This type of transition has also been seen in other studies where TFE was used to induce helical structure formation,\textsuperscript{116,211} however, a microscopic interpretation of this phenomenon is lacking. Due to the lack of baselines in these CD $T$-melts, as well as the changing nature of the $T$-melts themselves as a function of TFE percentage, no quantitative analysis was performed to extract additional information from this data. We did, however, use the mean residue ellipticity at 222 nm and the method developed by Baldwin and co-workers (Experimental Section) to estimate the fractional helicity ($f_H$) formed for each case. As shown (Table 4.1), the $f_H$ values obtained for pKID in 0% and 30% TFE, 21% and 54%, respectively, are in good agreement with those obtained in previous studies.\textsuperscript{210,212}

To determine the effect of TFE on the folding-unfolding kinetics of pKID, we measured its conformational relaxation rates in various concentrations of TFE using a laser-induced $T$-jump IR technique.\textsuperscript{213} As shown (Figure 4.3), the $T$-jump-induced relaxation kinetics, probed at 1630 cm\textsuperscript{-1}, can be described by a single-exponential function. In addition, the relaxation rate does not show any measurable dependence on the initial temperature, suggesting that the folding-unfolding process of pKID involves a significant ($\geq 1.5k_B T$) free energy barrier.\textsuperscript{214}
4.4.2 Viscosity and Crowding Effects

As indicated (Figure 4.4 and Table 4.2), in comparison to those measured in the absence of TFE, the relaxation rates obtained at low TFE percentages (up to 5% TFE) show a small but measurable decrease (~24%), which disappears completely upon increasing the TFE percentage to 15%. This non-monotonic dependence is interesting, since such a trend has not been reported before. One possible explanation for the initial decrease in the relaxation rate is that it arises from a TFE induced increase in the solution viscosity (\( \eta \)), since a previous study\(^{207}\) has shown that a 5% TFE solution (in H\(_2\)O) has a viscosity of 1.00 cP, compared to 0.89 cP for pure water. Interestingly, this viscosity increase cannot completely account for the observed decrease in the relaxation rate (\( k_R \)). This is because, assuming \( k_R \propto (\eta)^{-\alpha} \), where \( \alpha \) ranges from 0.6 to 1.0,\(^{215-217}\) an increase of \( \eta \) from 0.89 to 1.00 cP only leads to a decrease of \( k_R \) by ~12%, less than observed. This finding is entirely expected since, besides viscosity, the protein stability, which in this case is a function of TFE concentration, can also affect \( k_R \). As discussed above, the helicity of pKID increases in the presence of TFE, suggesting that under these conditions the folded state is stabilized. As shown (Figure 4.5), an increased stability can result from either an increase in the folding rate (\( k_F \)), a decrease in the unfolding rate (\( k_U \)), or both. However, an increase in \( k_F \) would lead to an increase in \( k_R \), as \( k_R = k_F + k_U \). Thus, the decreased relaxation rate at 5% TFE is consistent with the notion that this alcohol, at relatively low percentages, can selectively stabilize the folded state, which kinetically manifests as an increase in the unfolding free energy barrier.
What is more surprising, however, is that upon further increasing the TFE percentage from 5 to 15%, the relaxation rate of pKID becomes larger (Figure 4.4). This faster relaxation rate remains unchanged, within experimental error, up to 30% TFE. Since both the helicity of pKID and the solution viscosity increase with increasing TFE concentration in this range of TFE percentages (i.e., 5–30%), both of which, as discussed above, would lead to a decrease in $k_R$, this kinetic trend is not anticipated and hence suggests that one needs to consider additional factors. One possible explanation, according to our hypothesis, is that this rate increase results from a crowding effect of TFE, which is known to form clusters in this concentration range. Such clusters, typically consisting of nine TFE molecules, can occupy approximately 30% of the volume at 40% TFE, based on MD simulations. Crowding, which preferentially destabilizes the more extended unfolded state through the excluded volume effect, increases protein folding rates. However, unlike other commonly used macromolecular crowders, such as ficoll and dextran, which typically are assumed to be repulsive toward proteins, TFE interacts specifically with pKID. Thus, the observed concave upward dependence of the relaxation rate on TFE percentage, in the range of 0–30%, is a manifestation of the interplay of three factors, i.e., viscosity, stability and crowding.

4.4.3 Control Experiments

While the results discussed above are consistent with the notion that TFE can act as a crowding agent at certain volume percentages, further validation of this claim is needed. Fortunately, TFE self-association or aggregation is not a monotonic function of its concentration, which peaks at around 30% and effectively vanishes at 70%.
characteristic property of TFE clustering provides a simple means to test the validity of our hypothesis. Should the increased relaxation rate of pKID observed at 15–30% TFE solutions arise from crowding due to nearby TFE clusters, we would expect at higher concentrations of TFE, where these aggregates are less prevalent, the relaxation rates to, once again, decrease. Indeed, at 50% TFE the conformational relaxation rates of pKID become appreciably slower than those at 30% TFE, by a factor of approximately 1.7 (Figure 4.4 and Table 4.2). Thus, these results provide additional evidence in support of the notion that TFE clusters can act as nanocrowders. Furthermore, measurements of the conformational relaxation kinetics of another IDP, i.e. the LEA peptide, in water/TFE solutions also help support this claim. In nature, LEA proteins fold upon desiccation and are responsible for reducing aggregation of proteins in water deficient conditions in both plants and animals.\textsuperscript{219,220} Therefore, TFE, which causes the peptide backbone to be dehydrated, should be very effective in promoting LEA’s folding to a monomeric $\alpha$-helical structure, as observed (Figure 4.6 and Table 4.1). In addition, unlike that of pKID, the CD T-melt of LEA is more cooperative (Figure 4.7), which may reflect its intrinsic ability to fold upon the removal of water. Perhaps more importantly, the folding of monomeric $\alpha$-helices involves predominantly local interactions and thus diffusive motions over a relatively small length scale. As such, a previous study\textsuperscript{209} has shown that their folding kinetics are much less affected by macromolecular crowding in comparison to folding processes that involve formation of substantial nonlocal interactions, such as the folding of $\beta$-sheet structures. In other words, we expect, unlike pKID, that LEA’s relaxation rate will be less dependent on TFE clustering. Indeed, as shown (Figures 4.8
and 4.9 and Table 4.2), the T-jump-induced conformational relaxation rates of LEA are, within experimental uncertainties, practically the same in the range of 30–50% TFE. Taken together, we believe that this difference in the relaxation kinetics of pKID and LEA supports the conclusion that TFE can act as a nanocrowder at certain concentrations. Additionally, it is worth noting that the relaxation rate of LEA is similar to that of a monomeric helical peptide derived from the ribosomal protein L9, providing further evidence that the folding kinetics of naturally occurring helices are at, or near, the folding speed limit.

4.5 Discussion

One alternative theory that has been proposed in the literature concerning TFE’s interactions with proteins is that the clusters that are formed at certain volume percentages directly bind to hydrophobic residues in proteins; however, both pKID and LEA are composed mainly of hydrophilic residues, with LEA having a slightly larger nonpolar residue composition. Although direct binding of these clusters to pKID could result in a change in the relaxation rates, such an event seems unlikely, since the kinetics of LEA are relatively unchanged throughout the TFE percentages examined.

In protein conformational studies, it is common to use high concentrations of cosolvents, such as urea, alcohol, or TMAO, to experimentally control protein stability. Since many of these cosolvents have the tendency to self-associate, the crowding effect observed for TFE may also occur in other systems, an important aspect that has been largely overlooked. For example, in their MD simulations Cho et al. found that a high
concentration of TMAO leads to a reduction in the radius of gyration of several peptides, which led them to propose that TMAO can act as a molecular crowder. Using two-dimensional infrared (2D IR) spectroscopy, Ma et al.\textsuperscript{224} also showed that TMAO can reduce the conformational entropy of proteins, thus further validating the crowding effect of TMAO aggregates. In this context, we expect that our observations in this study may be common for other cosolvents and thus should be taken into consideration in future studies when these molecules are used to tune the folding thermodynamics of proteins.

4.6 Conclusions
Alcohols are frequently used as cosolvents to enhance structure formation in peptides and proteins. In particular, TFE is remarkably effective in this regard and thus has found broad application. While previous studies have provided many insights into how TFE acts to achieve its structure-enhancing role, the potential effect of TFE clustering, which is maximized at approximately 30% TFE (v/v), has often been overlooked. To investigate whether TFE clusters affect the folding kinetics of proteins, herein we study the conformational relaxation kinetics of two intrinsically disorder proteins: one (i.e., pKID) forms a HTH conformation and the other (i.e., LEA) folds into an $\alpha$-helix when TFE is present. Our results show that the relaxation rate of pKID has a complex dependence on TFE percentage in the range of 0–50%, whereas that of LEA is insensitive to TFE concentration. In particular, the maximum relaxation rate of pKID occurs at a TFE percentage (15–30%) where TFE clustering is also prevalent. Thus, based on these results, we propose that TFE can act as a nanocrowder and, through the excluded volume
effect, increase the folding rate of proteins containing a substantial amount of nonlocal contacts.

4.7 Acknowledgements

We gratefully acknowledge financial support from the National Institutes of Health (GM-065978). R.M.C. is an NIH Ruth Kirschstein Predoctoral Fellow (GM-008275). R.M.A. is an NIH Structural Biology Training Grant Fellow (T32-GM-008275).
<table>
<thead>
<tr>
<th>TFE%</th>
<th>$f_{Ht}$ (%)</th>
<th>TFE%</th>
<th>$f_{Ht}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>21</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>28</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>15</td>
<td>36</td>
<td>40</td>
<td>38</td>
</tr>
<tr>
<td>20</td>
<td>47</td>
<td>50</td>
<td>39</td>
</tr>
<tr>
<td>30</td>
<td>54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>57</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 4.1* Estimated fractional helicity ($f_{Ht}$) for pKID and LEA at 1°C.
<table>
<thead>
<tr>
<th>TFE%</th>
<th>$T_i$ ($^\circ$C)</th>
<th>$\tau_R$ (μs)</th>
<th>TFE%</th>
<th>$T_i$ ($^\circ$C)</th>
<th>$\tau_R$ (μs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>15.5</td>
<td>1.07 ± 0.09</td>
<td>30</td>
<td>24.2</td>
<td>0.21 ± 0.03</td>
</tr>
<tr>
<td>2.5</td>
<td>15.7</td>
<td>1.40 ± 0.10</td>
<td>40</td>
<td>24.7</td>
<td>0.36 ± 0.09</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>1.43 ± 0.08</td>
<td>50</td>
<td>23.1</td>
<td>0.18 ± 0.05</td>
</tr>
<tr>
<td>7.5</td>
<td>16.2</td>
<td>1.22 ± 0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>16.3</td>
<td>1.00 ± 0.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>16.1</td>
<td>1.40 ± 0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>16.1</td>
<td>2.40 ± 0.50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.2** Relaxation time constants ($\tau_R$) of pKID and LEA at the indicated final temperature ($T_i$).
**Figure 4.1** CD spectra of pKID collected at 1 °C and in aqueous solutions of different TFE percentages, as indicated.
Figure 4.2 CD 7-melt spectra of pKID in aqueous solutions of different TFE percentages collected at 222 nm.
Figure 4.3 Representative trace of the relaxation kinetics of the pKID peptide in a 30% TFE solution in response to a $T$-jump from 5.7 to 11 °C, probed at 1630 cm$^{-1}$. The smooth line represents the best fit of this curve to a single-exponential function with a time constant of $1.8 \pm 0.1 \mu$s.
Figure 4.4 Temperature dependence of the relaxation rate constant of pKID measured for different TFE solutions, as indicated. For easy comparison, the results are presented in two panels: (A) 0–15% TFE and (B) 15–50% TFE. The solid lines shown are to guide the eye.
Figure 4.5 Cartoon illustration of the effect of TFE on the folding and unfolding free energy barriers of pKID. In scenario (A), $\Delta G_{U,W}^{f} > \Delta G_{U,TFE}^{f}$ and $\Delta G_{F,W}^{f} = \Delta G_{F,TFE}^{f}$, whereas in scenario (B) $\Delta G_{U,W}^{f} = \Delta G_{U,TFE}^{f}$ and $\Delta G_{F,W}^{f} < \Delta G_{F,TFE}^{f}$. 
Figure 4.6 CD spectra of LEA collected at 1 °C and in aqueous solutions of different TFE percentages, as indicated.
Figure 4.7 CD T-melt spectra of LEA in aqueous solutions of different TFE percentages collected at 222 nm.
Figure 4.8 A representative trace of the relaxation kinetics of the LEA peptide in a 40% TFE solution in response to a $T$-jump from 3.8 to 8.4 °C, probed at 1664 cm$^{-1}$. The smooth line represents the best fit of this curve to a single-exponential function with a time constant of $0.9 \pm 0.1 \mu s$. 
Figure 4.9 Relaxation rate constants of LEA versus temperature for different TFE solutions, as indicated.
5 Tuning the Attempt Frequency of Protein Folding Dynamics via Transition-State Rigidification: Application to Trp-Cage

5.1 Original Publication

5.2 Introduction
Conformational diffusion on an energy landscape that is biased toward the native state ensures that protein folding is a thermodynamically robust and productive event. However, during folding, the free energy of the system does not always show a monotonic decrease; instead, it can increase over a relatively small region of the landscape, leading to the formation of folding free energy barriers. Because these barriers contain key information for achieving a comprehensive understanding of the mechanisms of protein folding, significant efforts have been made to investigate how and why such kinetic bottlenecks are generated as well as the structural characteristics of the associated transition states. More recently, several studies have focused on elucidating the dynamic aspects of the folding free energy barrier, such as the roughness of the underlying free energy surface and the transition path time. According to Kramers’ theory

\[ k = \frac{\omega_r \omega_B}{2\pi\gamma} \exp \left( -\frac{\Delta G^\ddagger}{RT} \right), \]  

(5.1)
where R is the gas constant and T is the absolute temperature, the rate of a barrier-crossing process is determined not only by the height of the barrier (\(\Delta G^\ddagger\)) but also by the curvatures of the reactant (\(\omega_R^2\)) and transition-state (\(\omega_B^2\)) potential wells as well as the friction coefficient (\(\gamma\)). The latter manifests as the roughness of the potential energy surface. While direct experimental assessments of \(\omega_R\) and \(\omega_B\) are currently not possible, many previous studies\(^{96,232}\) have been performed to determine the pre-exponential factor, often referred to as the attempt frequency (\(k_0\)). Interestingly, the value of \(k_0\) for protein folding is estimated to be in the range of \(10^3\) to \(10^6\) s\(^{-1}\),\(^9\) which is several orders of magnitude smaller than that observed for chemical reactions and thus suggests that the curvature of the protein folding transition-state potential well (i.e., \(\omega_B^2\)) is intrinsically small. This is consistent with the well-recognized notion that the folding transition state consists of an ensemble of structures that contain only a fraction of the native contacts and hence is inherently flexible. Gas-phase chemical reactions between small molecules often encounter a transition state that contains a single, distinct species in a highly constrained geometric configuration. In this regard, we hypothesize that by rigidifying the folding transition state one could significantly increase \(\omega_B\) and hence \(k_0\) (eq 5.1). In this proof-of-concept study, we chose a mini-protein, Trp-cage,\(^{233,234}\) as our model system and employed an azobenzene cross-linker to modify the curvature of its free-energy barrier.

Trp-cage is one of the most extensively studied model peptide systems in protein folding,\(^{53,235–273}\) which has led to a fairly detailed understanding of its folding mechanism. For example, both experimental\(^{53,242,260,267–269}\) and computational studies\(^{236–}\)
have shown that the $\alpha$-helix is either partially or completely formed in the major folding transition state, often without the presence of many native tertiary stabilizing interactions. Thus, this feature provides a unique opportunity to modify the characteristics of the folding transition state of Trp-cage via a photoactivatable cross-linker. As shown (Figure 5.1), our working hypothesis is that upon imposing a geometric constraint on the Trp-cage $\alpha$-helix via photoinduced isomerization of an azobenzene cross-linker, we will be able to not only initiate folding but also force the conformational search to pass through a more rigidified transition state, thus making the attempt frequency (i.e., $k_0$) of this folding “reaction” larger.

We chose an amidoazobenzene derivative as the photoactivatable cross-linker based on the fact that (1) its $cis$ isomeric form supports or stabilizes $\alpha$-helical conformations when attached between the i and i+7 positions of a peptide,\textsuperscript{274} whereas its $trans$ form does not, (2) its $trans$ form is thermodynamically more favorable (>95%) in the dark at room temperature,\textsuperscript{275} (3) upon irradiation with 355 nm light, the $trans$ to $cis$ isomerization occurs on the picosecond timescale,\textsuperscript{276} which is significantly faster than the folding time of Trp-cage, and (4) the spontaneous back-reaction, that is, the $cis$ to $trans$ isomerization, takes place on the time scale of minutes at room temperature.\textsuperscript{277}

Specifically, we introduced the azobenzene moiety into a mutant of the Trp-cage 10b variant containing cysteine substitutions at residues 1 and 8 (sequence: CAYAQWLCDDGGPSSGRPPPS), using standard cysteine alkylation methods.\textsuperscript{274} As previously indicated, the resultant Trp-cage peptide (hereafter referred to as 10b-azob) should fold only when the azobenzene cross-linker is in its $cis$ isomeric form (Figure 5.1).
5.3 Experimental Section

5.3.1 Peptide Synthesis and Sample Preparation

Di-iodoacetamide azobenzene was synthesized as previously described.\textsuperscript{274} The Trp-cage peptides were synthesized on a PS3 automated peptide synthesizer (Protein Technologies, MA) using Fmoc-protocols, purified by reverse-phase chromatography, and verified by matrix-assisted laser desorption ionization (MALDI) mass spectroscopy. Incorporation of the di-iodoacetamide azobenzene into Trp-cage was accomplished by following established methods.\textsuperscript{274} Briefly, peptide (1.13 mM) was dissolved in Tris·Cl buffer (pH 8) containing Tris-carboxyethylphospine (TCEP, 1.13 mM) and incubated at room temperature for two hours. Afterwards, di-iodoacetamide azobenzene was dissolved in DMSO and added to the peptide solution to a final concentration of 0.66 mM. The mixture was placed under foil in a hood with the lights out and allowed to stir for 10 minutes. Afterwards, 56 µl of a 10 mM solution of di-iodoacetamide azobenzene was added to the mixture, and this was allowed to stir in the dark for 10 minutes. Following this, another 56 µl of 10 mM di-iodoacetamide azobenzene was added to the mixture and allowed to stir in the dark for 10 minutes. Then, the mixture was exposed to light and stirred for another 10 minutes. The mixture was purified using reverse-phase chromatography, and a second round of MALDI spectroscopy was performed to identify the pure labelled compound (i.e., 10b-azob or 10b-h-azob). Trifluoroacetic acid (TFA) removal and H-D exchange were achieved by multiple rounds of lyophilization. Peptide samples used in the experiments were prepared by directly dissolving lyophilized peptide solids in a 20/80 trifluoroethanol (TFE)/D\textsubscript{2}O solution.
5.3.2 **Irradiation**

Irradiation was performed at room temperature by placing the corresponding dark-equilibrated peptide sample (in a standard 1 cm quartz cuvette) in a Fluorolog 3.10 spectrofluorometer (Jobin Yvon Horiba, NJ), and exposing it to 355 ± 15 nm light for 5-10 minutes. The intensity of the irradiation light was approximately 8.8 mW cm\(^{-2}\).

5.3.3 **UV-Vis Measurements**

UV-Vis spectra were collected on a Lambda 25 UV-Vis spectrometer (Perkin Elmer, MA) using a 1 cm quartz cuvette and an integration time of 0.5 s/nm. Sample concentrations were ~10 µM.

5.3.4 **Circular Dichroism Measurements**

CD spectra were obtained on an Aviv 62A DS spectropolarimeter (Aviv Associates, NJ) with a 1 mm sample holder. The peptide concentration was ~30 µM for the 10b-azob measurements. For the T-melt of Trp-cage 10b in a 20/80 TFE/water solution, the peptide concentration was ~40 µM.

5.3.5 **Time-Resolved Measurements**

The transient IR setup used is similar to the T-jump IR setup described previously.\(^{278}\) Briefly, a 3 ns 355 nm pulse derived from a Minilite II Nd:YAG laser (Continuum, CA) was used as the pump, while a continuous wave (CW) quantum cascade (QC) mid-IR laser (Daylight Solutions, CA) was used to probe the pump-induced conformational kinetics. The pump beam was gently focused, yielding a spot size of ~30 µm at the sample with an energy around 8 mJ. Peptide solution was placed between two CaF\(_2\)
windows in a home-made sample cell with an optical pathlength of 400 μm. The peptide concentration was 80-220 μM, which gave rise to an absorbance of 0.1-0.25 at 355 nm. During the experiment, the sample cell was moving constantly with an appropriate speed to ensure that each laser pulse hit a different spot than the previous one.

5.4 Results and Discussion

5.4.1 Steady-State Measurements of 10b-azob

As shown (Figure 5.2), the π−π* transition of trans-amidoazobenzene at ~367 nm has a significant decrease in intensity upon irradiation of 10b-azob with 355 nm light, whereas there is a gain in absorbance at ~258 nm, which corresponds to the π−π* transition of cis-amidoazobenzene. Furthermore, as expected (Figure 5.3), the circular dichroism (CD) spectrum of the dark-equilibrated 10b-azob sample (in a 20/80 TFE/water mixture) indicates that the peptide adopts mostly disordered conformations, whereas the CD spectrum of the light-irradiated sample indicates that light absorption indeed prompts α-helix formation. The reason that we added TFE, which is known to promote α-helix formation, is that in pure water the light-irradiated peptide exhibits relatively low helicity. This is most likely due to the fact that addition of the azobenzene cross-linker eliminates the favorable N-terminal helical cap, which has been shown to be detrimental to the stability of Trp-cage. More importantly, in the presence of 20% TFE the CD signal of the light-irradiated 10b-azob sample at 222 nm shows a similar sigmoidal dependence on temperature as that of the wild-type peptide (Figure 5.4), suggesting that the peptide’s cage structure is formed when the azobenzene moiety is in its cis form and
that the addition of TFE compensates for the loss of helix stability upon cross-linking. The latter is supported by the fact that addition of 20% TFE only leads to a small increase (~7 °C) in the thermal melting temperature of Trp-cage 10b (Figure 5.5 and Table 5.1).

5.4.2 Conformational Dynamics of 10b-azob

The light-induced folding kinetics of 10b-azob were probed using a time-resolved infrared (IR) apparatus. Briefly, the 355 nm pump pulse (3–5 ns) was derived from a Minilite II Nd:YAG laser (Continuum, CA), and a tunable 1001-TLC quantum cascade (QC) laser (Daylight Solutions, CA) was used as the continuous-wave (CW) IR probe. As indicated (Figure 5.6), the light-induced conformational dynamics of 10b-azob at 25 °C, probed at 1630 cm⁻¹ where helical content is known to absorb, show an increase in absorbance as a function of time, consistent with the CD results (Figure 5.3). What is more interesting, however, is that this kinetic trace is best fit to a double-exponential with time constants that differ by an order of magnitude (i.e., 90 ns versus 1.1 μs) (Table 5.2). Further measurements at 1680 cm⁻¹, where disordered conformations have a larger absorbance, show identical results (Figure 5.6). Previously, we have shown that the folding time of Trp-cage 10b is ~1.6 μs at 25 °C. Thus, the slower kinetic phase in the current case is similar to the folding kinetics of the wild type peptide, whereas the faster kinetic phase represents a previously unobserved folding event. Taken together, these results indicate that by photoinitiating isomerization of an azobenzene crosslinker added to the α-helical segment of the Trp-cage sequence, we are creating either two parallel pathways that have distinctly different folding rates or a sequential pathway that involves a folding intermediate.
5.4.3 Parallel Folding Pathways

It has recently been shown that the $3_{10}$-helix of Trp-cage 10b folds on the order of hundreds of nanoseconds and the formation of this structure is considered to be the last step in the folding process.\textsuperscript{53} However, it is unlikely that the fast component seen in these experiments comes from $3_{10}$-helix formation for two reasons. First, $3_{10}$-helices typically absorb in the 1660 cm\(^{-1}\) region,\textsuperscript{280} yet the ~100 ns component observed for 10b-azob is detected at both 1630 and 1680 cm\(^{-1}\). Also, the $3_{10}$-helix of Trp-cage 10b is relatively unstable and, as a result, the previous study\textsuperscript{53} was only able to detect its folding-unfolding kinetics at temperatures below ~20 °C. Another possibility is that the fast phase reports on the formation of an intermediate state that contains a native or native-like $\alpha$-helix, which goes on to form the folded Trp-cage structure with a slower folding rate. To test this possibility, we studied the photoinduced conformational dynamics of another azobenzene cross-linked peptide that corresponds to the Trp-cage 10b $\alpha$-helix (sequence: CAYAQWLC, hereafter referred to as 10b-h-azob). As indicated (Figure 5.7), the light-induced kinetics of 10b-h-azob in the presence of 20% TFE, probed at 1630 and 1680 cm\(^{-1}\), can be described by a single-exponential function, with a time constant of ~1.0 $\mu$s for both cases (Table 5.2). This result is consistent with the study of Serrano et al.,\textsuperscript{281} which showed that the folding time of a helical peptide with a side chain–side chain cross-linker is on the order of 1 $\mu$s. Perhaps most importantly, our 10b-h-azob results are in line with those of Hamm and coworkers,\textsuperscript{37} who observed that the presence of an azobenzene cross-linker in a short $\alpha$-helical peptide acts as a thermodynamic constraint rather than a dynamic one. In this regard, they observed that rather than initiating a fast
downhill folding process, the azobenzene photoswitch allowed for the stabilization of metastable, non-native free-energy traps. Therefore, these results prompt us to conclude that the fast (i.e., \(~100\text{ ns}\)) component seen in the case of 10b-azob does not arise from an early, partially folded, on-pathway Trp-cage intermediate wherein only the \(\alpha\)-helix is formed; instead, it corresponds to an alternative but much faster folding pathway. Similarly, a sequential scenario in which the \(\alpha\)-helix is formed in \(~1\mu\text{s}\) followed by a 100 ns folding event can also be ruled out, as the current experimental strategy is unable to detect a fast kinetic event following a slower one. Moreover, these kinetic results also argue against the idea that the faster folding component of 10b-azob results from a decrease in the folding free energy barrier, as we would expect similar double exponential behavior for 10b-h-azob in this case. This is because, as previously discussed, the rate-limiting step in Trp-cage folding corresponds to helix formation. Thus, tertiary interactions with the rest of the 10b-azob peptide seem to play an influential role in creating this alternate protein-folding pathway. Indeed, kinetic measurements carried out on 10b-azob at different temperatures reveal that both rates have very similar dependences on temperature (Figure 5.8), further supporting the idea that the azobenzene cross-linker is not affecting the free-energy barrier height but rather altering the frequency with which the system leaves the transition state region.

In summary, because an additional parallel pathway originating from the same reactant can only lead to an increase in the overall reaction rate, our interpretation implies, as shown in the following kinetic scheme, that upon photoisomerization of the
azobenzene cross-linker two distinguishable conformational ensembles (\(U_{1,\text{cis}}\) and \(U_{2,\text{cis}}\)) in the unfolded potential well of 10b-azob are rapidly formed

\[
\begin{align*}
U_{\text{trans}} & \quad \xrightarrow{355 \text{ nm}} \quad U_{1,\text{cis}} \quad k_1 = A_1 \exp(-\Delta G_1^e/RT) \\
 & \quad \xrightarrow{355 \text{ nm}} \quad U_{2,\text{cis}} \quad k_2 = A_2 \exp(-\Delta G_2^e/RT) \\
& \quad \xrightarrow{\text{ps timescale}} \quad F_{\text{cis}}
\end{align*}
\]

where \(A_1 \approx 10A_2\), \(\Delta G_1^e \approx \Delta G_2^e\), and the exchange rate between \(U_{1,\text{cis}}\) and \(U_{2,\text{cis}}\) is significantly slower than their folding rates to form \(F_{\text{cis}}\). Also of note, both \(k_1\) and \(k_2\) are significantly faster than the single-exponential folding rate (\(\tau_F\) is in the range of 5−7 \(\mu\)s) of another cross-linked Trp-cage peptide\(^{46}\). It was shown previously that when a helix cross-linker, m-xylene, was placed between positions 4 and 8 of the Trp-cage 10b sequence, both the folding and unfolding rates of the resultant peptide were significantly decreased in comparison with those of the wild-type Trp-cage 10b\(^{34,46}\). This was attributed to a frictional effect of m-xylene, as it was located at the most sterically congested region of the peptide. Because the azobenzene cross-linker is not only longer but also more flexible than m-xylene, it is expected to cause a much smaller perturbation due to internal friction. In addition, in keeping with the present hypothesis, the findings obtained with the m-xylene cross-linker suggest that crossing-linking a single \(\alpha\)-helical turn is insufficient to significantly increase the rigidity of the folding transition state.
5.4.4 Attempt Frequency Calculation

The notion that the two kinetic phases of 10b-azob arise from parallel folding pathways that have identical or comparable free energy barriers suggests that we could further estimate the value of $\omega_B$, which, to the best of our knowledge, has never been done before. On the basis of Eq 5.1, it is easy to show that

$$\frac{k_S}{k_F} = \frac{\omega_{BS}}{\omega_{BF}},$$

(5.2)

assuming that $\omega_r$ for both the fast and slow pathways is the same, where $k_S$ and $k_F$ are the rate constants of the slow and fast components, respectively, while $\omega_{BS}$ and $\omega_{BF}$ are the frequencies of the respective transition-state harmonic potential wells. In turn, these frequencies determine the free energy ($G_B$) of motion along the folding coordinate ($q$) near the transition state

$$G_B = \Delta G^* - \frac{1}{2} m \omega^2 q^2$$

(5.3)

where $m$ is the effective mass of the particle. Following Eq 5.3, one can easily show that for the same displacement along the folding coordinate, that is, $\Delta q$, the free-energy difference between the two aforementioned harmonic wells would be

$$\Delta G_B = \frac{1}{2} m (\omega_{BF}^2 - \omega_{BS}^2) (\Delta q)^2$$

(5.4)

Thus, by combining Eqs 5.2 and 5.4 and using the experimentally determined values of $k_S$ and $k_F$, one could solve for $\omega_{BS}$ and $\omega_{BF}$ if $\Delta G_B$ and $\Delta q$ are known. Whereas both are difficult, if not impossible, to be determined, we can make reasonable estimates in the current case. As previously concluded, the fast folding phase arises from a more rigid
transition state. In other words, it is the entropic effect of the azobenzene crosslinker that makes $\omega_{BF}$ larger. Using the values for change in conformational entropy upon helix formation determined by Hofrichter et al.,$^{282}$ we estimated the maximum entropic stabilization of helical structure arising from the azobenzene cross-linker to be ~6 kcal/mol. By further assuming that the peptide, which has a molecular weight of 2049.2 g/mol, needs to move a distance that is one-fourth of the radius of gyration of Trp-cage to cross the transition state, or $\Delta q = 3$ Å,$^{243}$ we found that $\omega_{hs} = 5.24 \times 10^{10}$ rad/s. This estimate provides what, to the best of our knowledge, is the first experimental assessment of the frequency of the protein folding transition state. By further assuming that $\omega_R$ is on the same order of magnitude as $\omega_B$, an assumption commonly used in the literature,$^{9,96}$ and $D = 10^{-6}$ cm$^2$/s as an upper limit,$^{283}$ we estimated $k_0$ to be $6.9 \times 10^6$ s$^{-1}$ for the folding kinetics of the unconstrained Trp-cage peptide.

5.5 Discussion

Despite its approximate nature, the previous calculation yields a $k_0$ value that is in good agreement with previously estimated values based on measurements of the folding rate of ultrafast folders$^{232}$ and the rate of contact formation in unfolded protein ensembles$^{284}$ as well as those based on simulations$^{96}$ and theoretical predictions.$^{285}$ In particular, this value compares well with that ($10^{7\pm1}$ s$^{-1}$) determined by Yu et al.,$^{30}$ who used single molecule force spectroscopy to characterize the folding free energy landscape and rate of a prion protein. Therefore, these agreements provide further support, albeit indirectly, of our interpretation and analysis of the kinetics results obtained with 10b-azob.
Although all of the evidence supports the aforementioned folding mechanism of 10b-azob, it is worth mentioning that an alternative interpretation for the observed nonexponential behavior is due to projection of the protein onto an incipient downhill folding landscape upon azobenzene isomerization. Gruebele and coworkers\textsuperscript{286,287} have found that under certain conditions proteins can be engineered to fold in a complex manner, in which there is both a slow phase due to some molecules diffusing on a landscape containing a barrier (activated folding) and a fast phase resulting from other molecules navigating a barrierless landscape (downhill folding). We have tentatively ruled out this possibility based on the kinetic results of 10b-h-azob (Figure 5.7) and the similar temperature dependence of the fast and slow rate constants of 10b-azob (Figure 5.8).

5.6 Conclusions

While extensive effort has gone into identifying the structures of folding transition states of peptides and proteins, aside from the ability to further stabilize these proteins and to obtain generic protein design strategies, there have not been many examples of using this knowledge to actively change the nature of a protein’s folding, for example, altering the shape of the protein folding free-energy barrier. Here we show, using Trp-cage as a testbed, that it is possible to tune the attempt frequency of protein folding dynamics via rigidification of the transition state. Specifically, we exploit the trans to cis isomerization of an azobenzene cross-linker via phototriggering to not only initiate folding but also provide a certain degree of constraint on the conformational flexibility of the α-helix of
Trp-cage, which has previously been shown to be formed in the transition state. Transient IR measurements reveal that this strategy produces biphasic kinetics of folding, with time constants that differ by an order of magnitude (i.e., 100 ns versus 1 \( \mu \)s). Further control experiments on a truncate of Trp-cage containing just the \( \alpha \)-helix segment provide strong evidence indicating that the fast kinetic phase does not arise from an intermediate; instead, it is confirmation of a parallel folding pathway whose transition-state potential well has a larger curvature in comparison with that of the wild-type Trp-cage. Moreover, from these experimental results, we are able to estimate the frequency of the transition state of Trp-cage to be on the order of \( 10^{10} \) rad/s.

5.7 Acknowledgements

We gratefully acknowledge financial support from the National Institutes of Health (P41GM-104605). R.M.A. is a NSF Graduate Research Fellow (DGE-1321851). R.M.C. is an NIH Ruth Kirschstein Predoctoral Fellow (GM-008275).
Table 5.1 Unfolding thermodynamic parameters of Trp-cage 10b in a 20/80 TFE/water solution.
<table>
<thead>
<tr>
<th>$\omega_{\text{probe}}$ (cm$^{-1}$)</th>
<th>T (°C)</th>
<th>$\tau_{\text{fast}}$ (ns)</th>
<th>$\tau_{\text{slow}}$ (ns)</th>
<th>$\omega_{\text{probe}}$ (cm$^{-1}$)</th>
<th>T (°C)</th>
<th>$\tau_{\text{slow}}$ (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1630</td>
<td>7.2</td>
<td>240 ± 30</td>
<td>1950 ± 160</td>
<td>1630</td>
<td>24.4</td>
<td>960 ± 60</td>
</tr>
<tr>
<td>13.5</td>
<td>250 ± 30</td>
<td>2050 ± 280</td>
<td>1100 ± 100</td>
<td>1680</td>
<td>24.4</td>
<td>860 ± 40</td>
</tr>
<tr>
<td>25</td>
<td>90 ± 20</td>
<td>1100 ± 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1680</td>
<td>25</td>
<td>120 ± 20</td>
<td>1000 ± 90</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.2** Fast and slow time constants of the conformational kinetics of 10b-azob and 10b-h-azob in 20/80 TFE/D$_2$O solutions at the temperatures (T) and frequencies ($\omega_{\text{probe}}$) indicated.
Figure 5.1 Schematic representation of how the azobenzene cross-linker alters the curvature of the folding free energy surface of Trp-cage.
**Figure 5.2** Absorption spectra of dark-equilibrated and light-irradiated 10b-azob peptides (~ 10μM), as indicated. The light-irradiated sample was prepared by irradiating the dark-equilibrated sample with 355 nm light (~8.8 mW cm⁻²) for 5 min.
Figure 5.3 CD spectra of dark-equilibrated and light-irradiated 10b-azob samples (~33 μM in a 20/80 TFE/water solution), as indicated. The light-irradiated sample was prepared as described in the caption of Figure 5.2.
Figure 5.4 CD T-melt of the light-irradiated 10b-azob sample, monitored at 222 nm. The solid line is a fit of the data to a two-state model using the same thermodynamic parameters determined for the wild-type Trp-cage 10b.53
Figure 5.5 CD thermal melting curve (open circles) of Trp-cage 10b in a 20/80 TFE/water solution. The smooth line here represents the best fit of the data to a two-state model. The slopes of the folded and unfolded CD baselines were treated as global fitting parameters. The resultant unfolding thermodynamic parameters are given in Table 5.1.
Figure 5.6 Conformational kinetics of 10b-azob (in a 20/80 TFE/water solution) induced by a nanosecond 355 nm laser pulse and probed at different frequencies, as indicated. These kinetic traces were collected at 25 °C and in each case, a linear and instrument-limited signal arising from the solvent due to the pump-induced temperature jump (~1 °C) has been subtracted for clarity. The smooth lines are fits of these traces to a double-exponential function with the following time constants provided in Table 5.2.
Figure 5.7 Conformational kinetics of 10b-h-azob (in a 20/80 TFE/water solution) induced by a nanosecond 355 nm laser pulse and probed at different frequencies, as indicated. These kinetic traces were collected at 24.4 °C, and in each case a linear background signal arising from the solvent has been subtracted for clarity. The smooth lines are fits of these traces to a single-exponential function with the time constants provided in Table 5.2.
Figure 5.8 Temperature dependence of the fast and slow rate constants of 10b-azob obtained at 1630 cm$^{-1}$. 
6 Direct Measurement of the Tryptophan-Mediated Photocleavage
Kinetics of a Protein Disulfide Bond

6.1 Original Publication

Reproduced from [Abaskharon, R. M.; Gai, F. Phys. Chem. Chem. Phys. 2016, 18, 9602-
9607.] DOI: 10.1039/C6CP00865H. http://pubs.rsc.org/en/content/articlehtml/2016/cp/
c6cp00865h with permission from the PCCP Owner Societies.

6.2 Introduction

The effect of ultraviolet (UV) radiation on protein structure and function has been an area
of active research for over a century. Although UV light is necessary for vitamin D
synthesis in cells and offers other therapeutic benefits, it can also cause DNA and protein damage, and hence, diseases. For proteins, such photodamage is
linked to the strong molar absorptivity of three aromatic amino acids, i.e., tryptophan (Trp), tyrosine (Tyr) and phenylalanine (Phe), in the UVB region. Among the known
protein photodegradation pathways, one arises from disulfide bond cleavage or reduction.

For example, the study of Correia et al. showed that the UV-induced damage of human
insulin, characterized by a decrease in the β-sheet and α-helical content and also a
reduction in the antibody binding ability and hormonal function, is correlated with an
increase in the number of free thiols present in the protein. Additionally, Wang and
coworkers have utilized UV-induced disulfide bond cleavage as a means to form
nanoparticles with free thiol groups that can be modified to load small molecules for
drug-delivery. While a disulfide bond can be broken by directly absorbing a UV photon,
this mechanism is unlikely the main cause of photoinduced disulfide cleavage in proteins due to the weak absorbance of a disulfide bond around 280 nm ($\varepsilon = 125 \text{ M}^{-1} \text{ cm}^{-1}$). Instead, disulfide bond cleavage has been connected to UV excitation of aromatic residues, especially Trp. It is well known that upon photoexcitation, the indole ring of Trp is capable of donating an electron to either the solvent or a nearby acceptor. Thus, it has been proposed that the photoinduced disulfide bridge cleavage in proteins is mediated by Trp, which acts as an electron donor. The transient absorption experiments of Hayon and coworkers on disulfide model systems suggested that the Trp-mediated disulfide cleavage could occur via two mechanisms. The first one, as indicated in Schemes (6.1) – (6.4), involves transfer of an electron from the triplet state of Trp ($^3\text{Trp}$), generated from its excited singlet state ($^1\text{Trp}^*$), to the disulfide (RSSR), leading to the formation of a disulfide radical anion ($\text{RSSR}^-$), which further reacts to cause the disulfide linkage to break apart:

\[
\text{Trp} + h\nu \rightarrow ^1\text{Trp}^* \quad (6.1)
\]

\[
^1\text{Trp}^* \rightarrow ^3\text{Trp} \quad (6.2)
\]

\[
^3\text{Trp} + \text{RSSR} \rightarrow \text{Trp}^{++} + \text{RSSR}^- \quad (6.3)
\]

\[
\text{RSSR}^- \rightarrow \text{RS}^- + \text{RS}^* . \quad (6.4)
\]

The second one (Schemes (6.5) and (6.6)) involves production of a solvated electron ($e^-_{\text{aq}}$) upon photoexcitation of Trp, which then reacts with the disulfide to generate the precursor RSSR$^-$, as shown below:

\[
^1\text{Trp}^* \rightarrow \text{Trp}^{++} + e^-_{\text{aq}} \quad (6.5)
\]

\[
e^-_{\text{aq}} + \text{RSSR} \rightarrow \text{RSSR}^- . \quad (6.6)
\]
While studies on small molecules suggest that both mechanisms are plausible, mechanistic studies on proteins are few and have not provided a complete picture of the individual processes involved in disulfide bond cleavage. For example, to the best of our knowledge, the only kinetics study of a Trp-mediated disulfide bond cleavage process in a protein environment is that of Petersen and coworkers. By performing a flash photolysis experiment on the protein cutinase, they were able to monitor the formation kinetics and lifetimes of several transient species involved in this process, including those of $e_{aq}$ and RSS$^\cdot$. However, due to spectral congestion, they were neither able to directly determine every key kinetic event leading to disulfide cleavage nor able to confirm the aforementioned mechanisms.

To provide further insights into the kinetics and mechanism of this important biological process, herein we designed and carried out an experiment to directly assess the Trp-mediated cleaving kinetics of a disulfide bond in a protein. The strategy is to induce the protein in question to unfold via photocleavage of a key disulfide cross-linker, allowing the disulfide cleavage kinetics to be assessed by measuring the protein unfolding kinetics. Specifically, we used a 34-residue mini-protein (referred to as Z34C) designed by Wells and coworkers. Z34C forms a helix-turn-helix (HTH) structure (Figure 6.1) that is stabilized by a disulfide bond near the termini. Thus, removal of this disulfide cross-linker leads to unfolding of the HTH structure, allowing for direct assessment of its cleavage by monitoring the kinetics of unfolding. Since Z34C does not contain Trp, we inserted this electron donor next to the disulfide bridge through a double
mutation with the resulting sequence: FNMQCWRRFY-EALHDPNLNE-EQRNAIKSI-RDWC (hereafter referred to as Trp-Z34C).

Photoinduced release of a conformational constraint is a very useful strategy to trigger a protein folding or unfolding event.\textsuperscript{34} In comparison to other commonly used triggering methods, such as laser-induced temperature-jump (T-jump) techniques,\textsuperscript{148} the advantage of this strategy is that it offers site-specific control. Thus, an additional goal of this study is to show that the Trp-mediated disulfide cleavage can be used to trigger protein folding/unfolding kinetics,\textsuperscript{314} while preventing geminate recombination.\textsuperscript{39,40,315,316}

6.3 Experimental

6.3.1 Peptide Synthesis and Sample Preparation

The Z34C and Trp-Z34C peptides were synthesized on a Liberty Blue microwave peptide synthesizer (CEM Corporation, NC) and a PS3 automated peptide synthesizer (Protein Technologies, MA), respectively, using Fmoc-protocols. They were then purified by reverse-phase chromatography, and verified by matrix assisted laser desorption ionization (MALDI) mass spectroscopy. For each peptide, the purified sample was dissolved in a 20% DMSO/water solution and stirred overnight, allowing oxidation of the thiol groups to form a disulfide bond. The oxidized sample was subject to a second round of purification and the peptide product was further verified by MALDI. Trifluoroacetic acid (TFA) removal and H–D exchange for IR measurements were achieved by multiple rounds of lyophilization in acidic D$_2$O solution. Peptide samples used in the experiments
were prepared by dissolving an appropriate amount of the lyophilized peptide in a 20 mM sodium phosphate buffer in D₂O and the final pH of the solution was 7.

### 6.3.2 Irradiation

Irradiation was performed at 25 °C by placing the peptide sample in either a 1mm quartz cuvette (for CD and UV-Vis measurements) or an IR cell with a 56 μm spacer (for FTIR measurements) in a Fluorolog 3.10 spectrofluorometer (Jobin Yvon Horiba, NJ) and illuminating it with 266 ± 15 nm light derived from the excitation source of the fluorometer for the amount of time reported. The intensity of the irradiation light was ~2.5–5.0 mW cm⁻².

### 6.3.3 Steady-State Measurements

CD spectra were recorded on an Aviv 62A DS spectropolarimeter (Aviv Associates, NJ) using a 1 mm sample holder. A 67 μM sample of Trp-Z34C was used and the sample was irradiated for 23 h. For the CD comparison to the wild-type peptide, the optical densities of both samples at 266 nm were kept equivalent. To do so, the concentration (83 μM) of the wild-type Z34C sample was higher than that (13 μM) of the Trp-Z34C sample. Both samples were irradiated for 2 h. UV-Vis spectra were recorded on a Lambda 25 UV-Vis spectrometer (Perkin Elmer, MA) using a 1 mm quartz cuvette at room temperature with a peptide concentration of ~100 μM and an irradiation time of 2 h. Fourier transform infrared (FTIR) spectra were recorded on a Magna-IR 860 spectrometer (Nicolet, WI) using a two-compartment CaF₂ sample cell of 56 μm pathlength with 1 cm⁻¹ resolution and at room temperature.
6.3.4 Time-Resolved Measurements

The transient IR setup is similar to the T-jump IR setup described previously.\textsuperscript{278} In brief, a Minilite II Nd:YAG laser (Continuum, CA) was used to produce the pump pulse of 0.05 mJ, 3 ns, and 266 nm pulse which was focused to a spot size of \( \sim 60 \mu m \) at the sample. A continuous wave (CW) quantum cascade (QC) mid-IR laser (Daylight Solutions, CA) was used as the probe. The peptide sample was held in a sample cell consisting of two CaF\(_2\) windows with an optical pathlength of 120 \( \mu \)m. The concentration of the peptide samples used in the time-resolved IR measurements was \( \sim 500 \mu M \), which led to an absorbance of \( \sim 0.1 \) at 266 nm.

6.3.5 Estimating the Triplet State Quenching Rate by Disulfide

Using the software of VMD, the distance between the tryptophan (Trp) residue and the disulfide bond in Trp-Z34C was estimated to be between 5 – 10 Å. Using this distance as the diameter of a sphere wherein a single pair of Trp and disulfide resides, the effective quencher concentration was estimated to be 3.17 – 25 M. Lapidus \textit{et al.}\textsuperscript{317} have shown that the bimolecular quenching rate constant of the triplet state of Trp by a disulfide is \( 1.9 \times 10^8 \) M\(^{-1}\) s\(^{-1}\). Using this value and the determined effective quencher concentration, we estimated the triplet state quenching rate by the disulfide in Trp-Z34C to be (0.2 – 2 ns)\(^{-1}\).

6.3.6 Removal of Solvent Contribution to the Kinetic Offset

The absorption spectrum of D\(_2\)O in the amide I’ region depends on temperature. As discussed in the main text, photoexcitation of Trp with a 266 nm ns pulse leads to an increase in the sample temperature and hence a long-lived D\(_2\)O signal due to this dependence. To subtract the D\(_2\)O signal, we collected the absorption spectra of D\(_2\)O in
the amide I' region at different temperatures and, based on these data, we can calculate
the D$_2$O signal for a given set of known parameters, i.e., frequency, pathlength, and
temperature change (i.e., $\Delta T$). In the current case, $\Delta T$ was estimated to be ~0.3 °C, using
the transient kinetic curve obtained at 1652 cm$^{-1}$, which is dominated by the temperature
jump signal. Using this information and also the optical pathlength of the sample cell
used in UV-pump and IR-probe measurements, we can determine the D$_2$O contribution to
the kinetic offset at each probing frequency.

6.3.7 Electron Scavenger Experiment

In this experiment, two Trp-Z34C samples of identical concentrations (~40 $\mu$M), with and
without the presence of N$_2$O, were used. The N$_2$O was introduced by bubbling pure N$_2$O
gas through the sample for at least 20 minutes, which was shown to accrue a sufficient
amount of N$_2$O as determined by FTIR. These samples were then subject to UV light
irradiation in 1 h intervals with shaking of the sample holder every 30 min. The CD
spectra of these samples were then taken at different time points.

6.3.8 Free Thiol Determination

The thiol determination UV-Vis spectra were collected on a V-650 UV-Vis
spectrophotometer (Jasco, MD) using a 1 cm quartz cuvette. First, the UV-Vis spectrum
of a 1.6 mM solution of an Ellman’s reagent, 5,5'-dithio-bis-[2-nitrobenzoic acid], in a 20
mM sodium phosphate buffer (pH 7) was recorded. Second, a Trp-Z34C sample solution
that had been irradiated with 266 nm light for ~15 h was added to this solution to yield a
final protein concentration of 8 $\mu$M. Third, the UV-Vis spectrum of this mixture was
taken. The change in absorbance at 412 nm was ~0.06, which indicates that the
concentration of 2-nitro-5-thiobenzoate dianion (TNB$^2-$), a product of the reaction between the Ellman’s reagent and a free thiol, is \(~4\ \mu\text{M}\), determined using the molar extinction coefficient of TNB$^2-$ at 412 nm ($\varepsilon = 14,150\ \text{M}^{-1}\ \text{cm}^{-1}$).

6.4 Results and Discussion

6.4.1 Disulfide Cleavage and Unfolding

As shown (Figure 6.2), upon irradiation with a low power ($\sim2.5 - 5.0\ \text{mW\ cm}^2$), continuous wave (CW), 266 nm light source (derived from a Xenon arc lamp) for a long period of time, the CD spectrum of Trp-Z34C lost intensity at 208 and 222 nm, both of which are characteristic of $\alpha$-helical conformations. However, under the same irradiation conditions, wild-type Z34C does not show the same degree of spectral changes (Figure 6.3). Measurements of the amide I’ (amide I in D$_2$O) band of Trp-Z34C also show that UV irradiation results in a decrease in its $\alpha$-helical content (Figure 6.4). As the reduced form (or uncross-linked) of Z34C (and Trp-Z34C) has significantly less $\alpha$-helical character than its oxidized (or cross-linked) counterpart, these results, taken together, not only confirm the notion that UV radiation can break the disulfide cross-linker in Trp-Z34C, but also indicate that the underlying process is mediated by photoexcitation of Trp.

6.4.2 Kinetics of Disulfide Cleavage

To directly assess the timescale of this disulfide cleavage process, we measured the relaxation kinetics of Trp-Z34C in response to photoexcitation of the Trp residue by a 3 ns, 266 nm laser pulse using transient infrared (IR) spectroscopy.$^{52}$ As indicated (Figure 6.4), photocleavage of the disulfide in Trp-Z34C leads to a change in its amide I’ band.
that is characterized by a loss/gain of absorbance at 1630 cm\(^{-1}\)/1664 cm\(^{-1}\). In addition, the folding–unfolding process of Z34C has been shown to follow two-state or single-exponential kinetics on the ms timescale.\(^{313}\) Thus, if the UV pump-induced transient signal contained contribution only from the unfolding process of the HTH motif, we would expect to observe a single, negative-going kinetic component on the ms timescale at 1630 cm\(^{-1}\) due to decreased helical content upon unfolding.\(^{279}\) However, the transient kinetics obtained at this frequency show very different behaviors (Figure 6.5 – blue curve), exhibiting a negative-going spike followed by a positive-going signal evolving on a slower (ns and \(\mu s\)) timescale. In particular, the negative-going component is fully developed within the time resolution of the instrument, and hence, cannot arise from the unfolding kinetics of Trp-Z34C. Since the indole ring of Trp (i.e., the chromophore absorbing the 266 nm light in the current case) does not have any appreciable absorbance between 1600 and 1700 cm\(^{-1}\) (except a weak C=C band of indole at 1618 cm\(^{-1}\)),\(^{318}\) this initial kinetic phase must manifest as a spectral change in the amide I’ band of Trp-Z34C induced by photoexcitation of the Trp residue. This assessment is corroborated by the kinetics obtained at 1664 cm\(^{-1}\) (Figure 6.5 – red curve), the initial kinetic phase of which is positive-going, as expected for a spectral change that results in a decrease/increase of the absorbance at 1630 cm\(^{-1}\)/1664 cm\(^{-1}\). A previous study by Huang et al.\(^{319}\) showed that photoexcitation of a Ru-complex covalently attached to the N-terminus of an \(\alpha\)-helical peptide produces a transient IR signal that arises instantaneously in the amide I’ band region of the peptide. They attributed this signal to a Stark effect,\(^{320,321}\) arising from a large difference (~8 D) between the dipole moments of the ground and excited states of
the chromophore. We believe that the transient IR signals in the current case are also caused by a Stark effect, induced by either a dipole moment change or generation of a charged species upon excitation of the Trp chromophore (see below).

6.4.3 Mechanism of Action

To further substantiate this assignment, we globally analyzed the kinetic traces obtained at 1630 cm\(^{-1}\) and 1664 cm\(^{-1}\) and found that both curves can be satisfactorily fit by convoluting the instrument response function (IRF) with a three-exponential function (plus an offset) with the following time constants: 7 ns, 150 ns, and 2.0 \(\mu\)s. Because our IRF is about 30 ns, the 7 ns component is unresolved. This component most likely arises from the lowest excited singlet state of Trp \((^1L_a)\), which has a lifetime between 2 and 4 ns and hence cannot be time-resolved with our current time resolution. More importantly, the dipole moment (me) of the \(^1L_a\) state is much larger than that (\(\mu_g\)) of the ground state (i.e., \(\mu_e - \mu_g = 3.5\) D),\(^{308}\) making it detectable via the aforementioned Stark effect. Because the other two decay components do not exhibit a resolvable rise time, they must originate from two longer-lived species that are not only quickly generated (i.e., within the instrument response time), but also carry a charge or a large dipole moment (thus becoming detectable). The only species that meet these requirements are Trp\(^{+}\) and RSSR\(^-\), produced via Scheme (6.3). This assessment is supported by the following reasons. First, \(^3\)Trp is formed during the lifetime of \(^1\)Trp\(^*\) or within 2–4 ns.\(^{322}\) Second, the subsequent \(^3\)Trp quenching by the nearby disulfide bond is expected to occur between 0.2 ns and 2 ns, estimated based on the reported bimolecular rate constant (i.e., \(1.9 \times 10^8\) M\(^{-1}\) s\(^{-1}\)) for this reaction\(^{317}\) and the effective disulfide concentration. Thus, combined, these
rates lead to rapid (and unresolved) formation of \( \text{Trp}^{++} \) (and \( \text{RSSR}^{-} \)). It is worth noting, based on previous studies,\(^{323,324}\) that \( \text{Trp}^{++} \) could also be formed from \( ^{1}\text{Trp}^{*} \) via an electron transfer process to the disulfide on a picosecond timescale. However, the current study cannot distinguish between these two mechanisms. Third, the time constant (150 ns) of the second kinetic component is comparable to the lifetime of \( \text{Trp}^{++} \) observed in other studies. For example, the study of Aubert et al.\(^{325}\) on the photoinduced electron transfer kinetics in DNA photolyase indicated that a \( \text{Trp}^{++} \) species can live as long as 300 ns. Similarly, a study\(^{311}\) on the photophysics of free Trp in aqueous solution using a UV pump and UV/visible probe spectroscopy showed that the spectral signature associated with \( \text{Trp}^{++} \) decays in a multi-exponential manner, and the fastest component has a time constant of 340 ns. Fourth, the same study also indicated that while the spectroscopic signal of \( \text{RSSR}^{-} \) for two different proteins exhibits multi-exponential decay kinetics, one of the decay time constants is 1.5–4 \( \mu s \), which matches that (2.0 \( \mu s \)) of the third kinetic component in the present case. Thus, based on these similarities, we attribute the 150 ns and 2.0 \( \mu s \) components to the decay kinetics of \( \text{Trp}^{++} \) and \( \text{RSSR}^{-} \), respectively. Although Tyr has also been shown to be able to promote disulfide bond cleavage,\(^{326,327}\) it is unlikely that in the current case the single Tyr residue in Trp-Z34C is responsible for the observed kinetics because it is several residues away from the disulfide bond and its absorbance at the excitation wavelength is approximately an order of magnitude lower than that of the two Trp residues.

According to Scheme (6.6), one of the reaction outcomes of \( \text{RSSR}^{-} \) is to break up the disulfide bond, which, in the current case, would result in unfolding of the HTH
structure. As discussed above, this unfolding process will result in a \( \mu s \) kinetic component that shows a decrease (increase) in absorbance at 1630 cm\(^{-1}\) (1664 cm\(^{-1}\)) as a function of time. However, such a kinetic component was not observed. We believe that this is caused by the similarity in the rates of unfolding and disulfide cleavage, and the larger amplitude of the latter makes the former not directly observable. In support of this idea, a previous study\(^{313}\) revealed that the unfolding time of the un-cross-linked variant of Z34C is \( 3.0 \pm 0.5 \mu s \) at 25 \( {\circ}C \), which is indeed very similar to the time constant (\( \sim 2.0 \mu s \)) of the third kinetic component. To provide more direct proof of this claim, we turned our attention to the offset of the transient kinetics. The basic idea is that if the disulfide bond indeed breaks up in 2.0 \( \mu s \), leading to the formation of an un-cross-linked and also unfolded Trp-Z34C, the transient spectrum obtained at a longer delay time (e.g., the offset) should reflect the underlying structural change thus induced. To make this comparison, we collected additional transient kinetic traces at frequencies across the whole amide I’ region (Figure 6.6). As expected, those traces can also be globally fit by the function used to describe the kinetics obtained at 1630 cm\(^{-1}\) and 1664 cm\(^{-1}\) (Table 6.1). More importantly, as shown (Figure 6.7), upon subtracting out the signal from water, due to an increase in temperature caused by part of the photoexcitation energy that is dissipated to the solvent, which is negative and has a linear dependence on frequency in the amide I’ band region, the offsets of these data produce a transient spectrum that resembles the difference spectrum (Figure 6.4) generated by the IR spectra of Trp-Z34C obtained before and after irradiation with 266 nm light. Thus, this analysis provides direct
support for the notion that the 2.0 μs kinetic component originates from the disulfide bond cleavage.

Taken together, our results support a mechanism wherein the precursor that leads to cleavage of a disulfide, RSSR⁻, is produced via ³Trp (Scheme (6.3)) but not via a solvated electron (Scheme (6.5)). To further validate this point, we compared the UV-induced disulfide cleaving efficiencies of two Trp-Z34C samples, with and without the presence of N₂O, an efficient electron scavenger, using CD spectroscopy as a probe. As shown (Figure 6.8) upon irradiation with 266 nm light for 2 h, the CD spectra of both samples show the same changes, thus providing strong evidence indicating that RSSR⁻ is not formed via the solvated electron pathway, i.e., Schemes (6.5) and (6.6).

6.4.4 Reaction Products

In addition, we carried out an experiment to determine the number of free thiols generated upon UV excitation of Trp-Z34C using Ellman’s reagent, which reacts with a free thiol group in a 1:1 ratio to produce 2-nitro-5-thiobenzoate dianion (TNB²⁻). As shown (Figure 6.9), the absorbance of the TNB²⁻ anion at 412 nm, obtained by reacting Ellman’s reagent with a Trp-Z34C sample that had been illuminated with 266 nm light overnight, indicates the presence of at least one thiol in the photoproduct of the disulfide cleavage reaction. This implies that the other thiol radical, RS⁺, produced in Scheme (6.6), must undergo a different reaction. As indicated (Figure 6.10), the UV-Vis spectra of Trp-Z34C obtained before and after irradiation with 266 nm CW light for 2 hours show an increase in the absorbance at both 320 nm and 280 nm, similar to the changes observed for the formation of a Trp-SR adduct in goat α-lactalbumin upon UV
excitation. Thus, this result, in conjunction with the fact that photoexcitation of the Trp residue induces Trp-Z34C to unfold, suggests that this adduct is formed on one side of the peptide. In other words, upon RSSR\(^-\) dissociation, the sulfur atom closer to Trp\(^{++}\) is converted to a radical (i.e., RS\(^-\) in Scheme (6.3)). Further reaction between this radical and Trp\(^-\), formed via Trp\(^{++}\) deprotonation, results in uncross-linking and hence unfolding of Trp-Z34C. However, it is worth noting that over an extended period of UV irradiation (i.e., overnight), the peptide sample showed an increase in light scattering, suggesting the existence of a slower process that leads to peptide aggregation.

### 6.5 Conclusions

In summary, we have designed and carried out a UV-pump and IR probe experiment on a mini-protein (Trp-Z34C), aiming to provide new insights into the mechanism of Trp-mediated disulfide cleavage in proteins. Our strategy exploits the key stabilizing role of a terminal disulfide cross-linker in this protein and uses the UV-triggered unfolding process to assess the underlying disulfide cleavage kinetics. Our results are consistent with a mechanism wherein the cleavage is initiated by an electron transfer event from the triplet state of the photoexcited Trp residue to the disulfide in question, leading to the formation of a reactive disulfide species (i.e., RSSR\(^-\)), which further dissociates to break up the disulfide bond on a timescale of 2–3 \(\mu s\). Furthermore, we find that one of the photoproducts of this reaction, a Trp-SR adduct, is formed locally, thus does not yield a new cross-linker. Because a disulfide can be easily incorporated into a polypeptide sequence, our finding suggests that it is possible to use a Trp and disulfide pair to site-
specifically control the starting point of a folding or unfolding process. Another potential application is to use UV light to control biological functions as disulfides are commonly found in naturally occurring and designed peptides, such as antimicrobial peptides,\textsuperscript{330} which perform a wide range of biological activities.

### 6.6 Acknowledgements

We gratefully acknowledge financial support from the National Institutes of Health (GM-065978). R.M.A. is an NSF Graduate Research Fellow (DGE-1321851). The transient IR experiments were performed on instruments that were developed under an NIH Research Resource Grant (P41-GM104605).
<table>
<thead>
<tr>
<th>Frequency (cm(^{-1}))</th>
<th>Offset</th>
<th>2.0 ± 0.2 μs</th>
<th>150 ± 50 ns</th>
<th>7 ± 3 ns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1605</td>
<td>-1.00 ± 0.002</td>
<td>-0.79 ± 0.07</td>
<td>-1.6 ± 0.3</td>
<td>-5.7 ± 0.4</td>
</tr>
<tr>
<td>1615</td>
<td>-1.06 ± 0.01</td>
<td>-1.1 ± 0.3</td>
<td>-2.8 ± 0.5</td>
<td>-9.2 ± 0.5</td>
</tr>
<tr>
<td>1625</td>
<td>-1.15 ± 0.01</td>
<td>-1.2 ± 0.1</td>
<td>-3.3 ± 0.5</td>
<td>-12.0 ± 0.8</td>
</tr>
<tr>
<td>1630</td>
<td>-1.11 ± 0.04</td>
<td>-1.0 ± 0.7</td>
<td>-3.0 ± 0.4</td>
<td>-11.9 ± 0.8</td>
</tr>
<tr>
<td>1635</td>
<td>-1.05 ± 0.01</td>
<td>-0.9 ± 0.1</td>
<td>-2.9 ± 0.4</td>
<td>-11.4 ± 0.7</td>
</tr>
<tr>
<td>1645</td>
<td>-0.96 ± 0.01</td>
<td>-0.35 ± 0.04</td>
<td>-1.3 ± 0.3</td>
<td>-7.1 ± 0.6</td>
</tr>
<tr>
<td>1664</td>
<td>-0.51 ± 0.02</td>
<td>0.26 ± 0.06</td>
<td>0.71 ± 0.07</td>
<td>2.2 ± 0.2</td>
</tr>
<tr>
<td>1675</td>
<td>-0.46 ± 0.02</td>
<td>0.30 ± 0.07</td>
<td>0.9 ± 0.2</td>
<td>3.0 ± 0.3</td>
</tr>
<tr>
<td>1685</td>
<td>-0.417 ± 0.003</td>
<td>0.25 ± 0.03</td>
<td>0.7 ± 0.1</td>
<td>2.37 ± 0.09</td>
</tr>
<tr>
<td>1695</td>
<td>-0.445 ± 0.004</td>
<td>0.26 ± 0.02</td>
<td>0.42 ± 0.06</td>
<td>1.6 ± 0.2</td>
</tr>
</tbody>
</table>

**Table 6.1** Amplitudes (in mOD) of the four kinetics components at different probing frequencies
Figure 6.1 NMR structure of Z34C (PDB code: 1ZDC), showing the location of the disulfide cross-linker (green) and the tryptophan mutations (pink).
Figure 6.2 CD spectra of a 67 μM Trp-Z34C sample taken at 25 °C before and after irradiation with 266 nm light for 23 h, as indicated.
**Figure 6.3** CD spectra of wild-type Z34C (a) and Trp-Z34C (b) taken before and after irradiation with 266 nm light for 2 hrs, as indicated.
Figure 6.4 FTIR difference spectrum of Trp-Z34C in the amide I’ band region. This spectrum was generated by subtracting the FTIR spectrum of a non-irradiated Trp-Z34C sample (2.8 mM in D$_2$O buffer at pH 7) from that of the same sample irradiated for 2.5 days.
Figure 6.5 IR transient kinetic traces of Trp-Z34C obtained at two representative probing frequencies, 1630 cm\(^{-1}\) (blue) and 1664 cm\(^{-1}\) (red). The smooth lines are global fits of these traces to a three-exponential function plus an offset with the following time constants: 7 ns, 150 ns, and 2.0 \(\mu\)s. Shown in the inset are the same data with a logarithm time axis.
Figure 6.6 IR transient kinetic traces of Trp-Z34C obtained at different probing frequencies, as indicated. For clarity, these traces have been offset.
Figure 6.7 Transient IR spectrum of Trp-Z34C generated using the offsets of the decay kinetics at the corresponding probing frequencies. A contribution from the solvent at each frequency, due to a pump-induced temperature jump, has been removed.
**Figure 6.8** Effect of an electron scavenger (N$_2$O) on the photoinduced disulfide cleavage in Trp-Z34C, assessed via CD spectroscopy. These normalized CD spectra of two Trp-Z34C samples, one containing N$_2$O and the other with no electron scavenger, show that the presence of N$_2$O has no significant effect on the disulfide cleaving efficiency.
Figure 6.9 UV-Vis spectra of an Ellman’s reagent sample with and without the presence of light-irradiated Trp-Z34C peptide, as indicated.
Figure 6.10 Absorption spectra of Trp-Z34C before and after irradiation with 266 nm light for 2 h, as indicated.
7 Light-Initiated Ultrafast Kinetics and Two-Ligand Dissociation of Ruthenium Polypyridyl Compounds

7.1 Introduction

The use of light to manipulate and site-specifically modulate the activities and properties of molecular systems via photochemical means has seen an extensive growth over the past three decades. In biology, this approach has been used to not only control neurological processes by genetically incorporating light-sensitive ion channels into the neuronal network, but has also found relevance in external initiation of nanocarrier payload release. Moreover, light-sensitive molecules have been used in materials applications in organic electronics to tune the functionalities of these devices. Since the timescales and properties of the applications of these photochemical moieties are so diverse, the need for tunable, ultrafast photochemical triggering systems is growing. One such application is the use of these photochemical triggers as a means to control and initiate protein folding reactions to explore protein dynamics over a range of timescales from picoseconds to seconds. Ruthenium polypyridyl complexes might be useful in this regard as they are easily varied by interchanging the associated ligands and have been shown to dissociate on an ultrafast timescale.

Since it was first discovered in 1959 that excitation of the metal ligand charge-transfer (MLCT) band of tris-2,2'-bipyridine ruthenium (II) ([Ru(bpy)₃]²⁺) would lead to luminescence, this compound and its derivatives have been extensively explored and utilized for their interesting photochemical properties. It has since been shown that the lowest exited state of [Ru(bpy)₃]²⁺ is actually a triplet state and thus accounts for the long
lifetime of this process. Interestingly, substitution of one of the bipyridine ligands with two monodentate ligands (i.e. \([\text{Ru(bpy)}_2(L_1)(L_2)]^{2+}\) where \(L_1\) and \(L_2\) are monodentate ligands) has been shown to not only phosphoresce when this MLCT band is excited, but an additional photochemical pathway involving loss of the monodentate ligands through thermal population of antibonding orbitals opens up. The quantum yield of this last pathway is dependent on both the excitation maxima and redox potential of the compound. A few studies have shown that the dissociation of a single ligand occurs on the picosecond timescale, however, there has been no kinetic investigation into the photodissociation of two ligands. In addition, because the ligands are directly attached to the ruthenium core, these reactions are generally devoid of side reactions. Since the MLCT absorption bands typically resides between 400 – 500 nm, visible light, as well as IR light in cases which allow for two-photon absorption, can be used to initiate this reaction. Changing the ligands allows for “tuning” of this frequency and photochemical reaction properties and thus these compounds have found a variety of applications from biological imaging to therapeutic agents. For example, Ru(bipyridine)_2(3-ethynylpyridine)_2 (RuBEP) coordinated to antisense morpholinos was used to initiate early developmental gene knockdown in zebrafish upon irradiation and uncaging of this morpholino structure. However, despite the interest in these compounds, the photochemistry of these processes, particularly as they relate to biological coordination, has still not been extensively studied.

Infrared spectroscopy has proven to be an invaluable means to assess the ultrafast dynamics of various compounds due to its high temporal resolution. When coupled with
an ultraviolet or visible pump source, it can also be useful in the investigation of
photochemical processes as well. In this regard, we designed a system which contained
infrared absorbing chemical groups, specifically nitrile groups, on the monodentate
ligands of [Ru(bpy)₂(L)₂]²⁺ in collaboration with the Dmochowski lab. These 4-
pentynenitrile (4PN) and 5-hexynenitrile (5HN) ligands (Figure 7.1) attach to the metal
center via the lone pair on the nitrogen of the nitrile moiety. Moreover, the alkyne ends
leave available the possibility of using azide-alkyne cycloaddition to attach this
photosystem to a protein containing azido amino acids. However, as a first step, the
photophysics and design of this reaction are investigated.

7.2 Experimental

7.2.1 Sample Preparation

[Ru(bpy)₂(L)₂]²⁺ compounds were synthesized and purified by Teresa Rapp from the
Dmochowski lab and used as received. Lyophilized samples were stored in a -30 °C
freezer prior to use and were directly dissolved in either water or methanol to the
appropriate concentration before each experiment.

7.2.2 Irradiation

Irradiation was performed at room temperature using either a Fluorolog 3.10
spectrofluorometer (Jobin Yvon Horiba, NJ) for 420 nm irradiation or an LED for 450
nm irradiation. All irradiation was performed with the sample in a homebuilt IR cell with
a 56 μm pathlength for the amount of time reported.
7.2.3 Steady-State Measurements

UV-Vis spectra were collected on a Lambda 25 UV-Vis spectrometer (Perkin Elmer, MA) in a 1 cm cuvette with a sample concentration of ~25 μm in water at room temperature. Fourier Transform Infrared (FTIR) spectra were obtained on a Magna-IR 860 spectrometer (Nicolet, WI) at room temperature with 1 cm\(^{-1}\) resolution. A two-compartment cell with a 56 cm\(^{-1}\) Teflon spacer was used with one side containing the solvent and the other the sample solution such that an accurate background subtraction could be conducted by modulating between sides during collection. Sample concentrations of ~100 mM for \([\text{Ru(bpy)}_2(4\text{PN})_2]^{2+}, [\text{Ru(bpy)}_2(5\text{HN})_2]^{2+}, 4\text{PN}, \text{and } 5\text{HN}\) were used. FTIR spectra before and after the transient measurements were collected on a Nicolet 6700 FT-IR spectrometer (Nicolet, WI) with 1 cm\(^{-1}\) resolution at room temperature using a cell with a 400 μm pathlength and 1 mM concentration.

7.2.4 Time-Resolved Measurements

The setup of the femtosecond IR transient absorption setup have been described in detail elsewhere.\(^{356}\) Briefly, a 355 nm excitation pulse was overlapped with Fourier-transform limited 70 fs IR pulses with center frequencies of 2150 cm\(^{-1}\) such that the beam radius of the IR probe pulse was 75% of that of the UV pump pulses at the sample cell. The timing between these two pulses was modulated by using two automated translation stages (Melles Griot, Sigma Koki Co., LTD) such that fs to ns resolution was achieved. The transmitted IR pulse was focused onto a monochromator with a 64-element MCT array detector (InfraRed Associates, Stuart, FL). A sample cell with a 400 μm spacer and a sample concentration between 1 – 2.5 mM in methanol was assembled into a rotation
stage such that the sample was continually rotated during collection. The absorbance of each sample was between 0.2 and 0.4 at 355 nm. Kinetic measurements were collected up to 1 ns with a 100 fs interval.

7.3 Results and Discussion

7.3.1 Photodissociation of Nitrile Ligands

Previous experiments have suggested that photoinitiation of ligand loss in ruthenium polypyridyl complexes does not lead to a change in the oxidation state of the ruthenium but rather leads to exchange with nearby polar molecules in order to fill the octet that ruthenium prefers.\textsuperscript{358–360} Moreover, if no other potential ligand is present, water or solvent molecules can coordinate to the metal in place of the lost moiety. As an example, del Campo and coworkers\textsuperscript{361} showed that a surface bound [Ru(bpy),\textsubscript{2}(PMe\textsubscript{3})(APTS)] (where PMe\textsubscript{3} is trimethylphosphine and APTS is (3-aminopropyl)triethoxysilane) is able to act as a photocleavable cage in surface applications by exchanging one of the APTS ligand with a single water molecule. Therefore, for the present case, we expect water to coordinate to the central ruthenium metal to form [Ru(bpy),\textsubscript{2}H\textsubscript{2}O\textsubscript{2}]\textsuperscript{2+} as a result of this reaction. Previous studies have shown that nitriles are particularly good leaving groups in this reaction because they are weak donor ligands. Moreover, for system in which the monodentate ligands are the same, photoexcitation should lead to loss of both ligands if the σ-donating ability of the second ligand is lower than that of the solvent. Since [Ru(bpy),\textsubscript{2}(ACN)\textsubscript{2}]\textsuperscript{2+} (where ACN is acetonitrile) is known to dissociate both ligands\textsuperscript{345} and [Ru(bpy),\textsubscript{4}(ACN)\textsubscript{4}]\textsuperscript{2+} has been shown to lose two ACN ligands upon photolysis,\textsuperscript{362} we
hypothesize that 4PN and 5HN behave similarly. Therefore, the proposed scheme of the reaction under investigation (Figure 7.1) whereby the ligands are replaced with water molecules should lead to the same product in both cases.

The UV-Vis absorption spectra of both [Ru(bpy)$_2$(4PN)$_2$]$^{2+}$ and [Ru(bpy)$_2$(5HN)$_2$]$^{2+}$ are essentially the same (Figure 7.2) and are characterized by several features. First two absorption bands, one below 190 nm and one at 285 nm, have been previously identified to be LC transitions from the bipyridine ligands.$^{363}$ Moreover, the bands around 240 nm and 420 nm are proposed to be the MLCT transitions with the 420 nm band populating the singlet $^1$MLCT state.$^{364}$ Previous studies$^{345}$ have shown that the more blue-shifted the $^1$MLCT band lies in the 400 – 500 nm range, the greater the yield of photosubstitution. Thus, since the $^1$MLCT absorbance of these compounds lies at the higher frequency side of this region, it is likely that the reaction will occur with appreciable yield. In fact, upon photolysis of these compounds for 5 min with 450 nm light from an LED, this lower frequency MLCT band undergoes a substantial redshift to about 500 nm as a result of the ligand loss (data not shown) as has been shown by the Dmochowski lab. The reaction seems to be complete within this timeframe; however, it is not known whether one or two ligands come off in this process.

In order to investigate the dissociation of this ruthenium complex further, infrared spectroscopy was utilized to monitor the absorbance of the nitrile moiety before and after photolysis. The nitrile stretching mode is known to absorb between 2100 – 2400 cm$^{-1}$, is generally sensitive to its environment.$^{161,365,366}$ As shown (Figure 7.3), the FTIR spectrum of [Ru(bpy)$_2$(4PN)$_2$]$^{2+}$ in water exhibits a relatively sharp band at 2277 cm$^{-1}$ which upon
irradiation with 450 nm light for ~ 10 minutes shifts to 2258 cm\(^{-1}\) and grows in intensity. This redshift makes sense as dissociation of the nitrile moiety from the ruthenium metal increases the electron density in an antibonding orbital of the nitrile thus decreasing its bond order and vibrational frequency.\(^{367}\) Additionally, it appears that the initial peak disappears almost completely indicating that the reactant is fully converted to product. It is thus likely that both ligands are dissociated in this process as a single ligand dissociation would mostly likely lead to two peaks, one from the ligand free in solution and the other from [Ru(bpy)\(_2\)(4PN)(H\(_2\)O)]\(^{2+}\). However, in order to verify this conclusion, the FTIR spectrum of the ligand alone in solution was collected. As expected, the ligand absorption nearly overlays that of the product, indicating that the absorption is indeed from the ligand free in solution and the photolysis of [Ru(bpy)\(_2\)(4PN)]\(^{2+}\) does produce [Ru(bpy)\(_2\)(H\(_2\)O)\(_2\)]\(^{2+}\). It is worth noting that several studies\(^{368,369}\) have shown that Cl\(^-\) ions, which are present in this solution, can also replace dissociated ligands in these complexes and thus it is possible that the final product in this reaction is [Ru(bpy)\(_2\)(Cl)]\(^{2+}\) or [Ru(bpy)\(_2\)(H\(_2\)O)(Cl)]\(^{2+}\). However, due to their dilute nature with respect to the abundant water and the need to maintain ionic balance in the system, it is more likely that water acts as the new ligand.

These same experiments were conducted on the [Ru(bpy)\(_2\)(5HN)]\(^{2+}\) compound in methanol using 420 nm light (Figure 7.4) and 355 nm light (data not shown), since transient measurements would be later conducted in methanol with more blue-shifted frequencies. Once again, similar observations are made as there is a 21 cm\(^{-1}\) redshift after photolysis, which correlates with the absorption of the ligand alone in solution.
Interestingly, there is about a 2 cm$^{-1}$ shift of the higher frequency band in the spectrum of a partially photolyzed state, which might be due to an intermediate with one 5HN ligand attached and one dissociated. FTIR of the monosubstituted complex, [Ru(bpy)$_2$(5HN)(H$_2$O)]$^{2+}$, would be useful in determining if this assignment is indeed correct.

7.3.2 Kinetics of Ligand Photodissociation

Since both of these complexes exhibited strong IR absorption bands, time-resolved measurements were carried out using a femtosecond IR transient absorption setup. In both cases, ultrafast kinetics were observed by monitoring the nitrile-stretching region of the infrared spectrum. For the [Ru(bpy)$_2$(4PN)$_2$]$^{2+}$ photosystem, the resulting trace probed at 2244 cm$^{-1}$ was fit with a biexponential function with time constants of 3.3 ps and 43.7 ps for the fast and slow phase, respectively (Figure 7.5). Since the free ligand in solution absorbs in this region, there should be an increase in the kinetic trace obtained if the observed dynamics resulted from an increase in the abundance of this moiety as a result of the photocleavage reaction. However, the kinetic trace actually has a decreasing absorbance with respect to time. Additionally, essentially the same biexponential kinetics are observed across the entire region monitored (1800 cm$^{-1}$ – 2500 cm$^{-1}$), as for instance, the kinetic trace obtained at 2050 cm$^{-1}$ can be fit to a biexponential function with 8.3 ps and 50.0 ps time constants (Figure 7.5 − inset). This supports the notion that the kinetics observed are not arising from the dissociation but instead from some other source, most likely a heating effect, as some of the absorbed radiation most likely goes through a non-radiative process leading to heating of the surrounding solvent molecules. Moreover, the
methanol solvent alone did not show such kinetics, supporting this assignment. The transient measurements of the \([\text{Ru(bpy)}_2(5\text{HN})_2]^{2+}\) complex displayed similar results as negatively-going kinetics traces were observed across the entire region monitored. However, the traces from this compound decayed with slightly faster rates as the kinetics probed at 2231 cm\(^{-1}\) were fit best to a biexponential with time constants of 0.3 ps and 5.5 ps and those at 2071 cm\(^{-1}\) with 0.4 ps and 7.5 ps (Figure 7.6).

In order to determine if the reaction still took places during the timescale of the experiment, a time slices of the transient experiment were compared to the FTIR difference spectrum before and after the experiment. As shown (Figure 7.7), the 430 ps time slice matches the FTIR difference spectrum reasonably well for \([\text{Ru(bpy)}_2(4\text{PN})_2]^{2+}\). The same spectral features are observed in both cases, including the small difference in the nitrile stretch in the 2200 – 2300 cm\(^{-1}\) region. This would indicate that the reaction is complete within a few hundred picoseconds. However, for the \([\text{Ru(bpy)}_2(5\text{HN})_2]^{2+}\), the two spectra do not match until ~ 1ns. These findings were also supported by the fact that nanosecond UV transient absorption measurements indicated that the reaction of \([\text{Ru(bpy)}_2(4\text{PN})_2]^{2+}\) is already complete by the earliest timepoint. Therefore, since both of these reactions seem to take place within hundreds of picoseconds, which is much faster than the rates of protein folding, these two systems hold promise for being suitable phototriggers for initiating protein folding reactions.
7.4 Conclusions

Taken together, these experiments suggest that the ruthenium polypyridyl complexes studied exchange both nitrile containing ligands with solvent molecules upon irradiation with 350 – 450 nm light. Although the reaction appears to be complete in hundreds of picoseconds, the exact kinetics of these processes were not able to be directly measured using infrared spectroscopy as a result of solvent heating. Ultrafast UV/Vis spectroscopy would be useful in this regard and it would allow for direct monitoring of the bleach of the 420 nm peak and an increase in the 500 nm band as the reactant is converted to product. This method would also be advantageous in monitoring different ruthenium polypyridyl, which might not have strong IR absorbers. Nonetheless, the experimental results from the IR transient absorption measurements indicate that both of these complexes convert to their final product on the ultrafast timescale. Therefore, utilizing “click” chemical reactions between the alkyne groups on the labile ligands of these compounds and azide-containing peptides and proteins, these compounds should be able to be attached to strategic places in the protein structure based on the location of azido unnatural amino acids. Upon photolysis, the circularized structure would open up as the 4PN or 5HN ligands are exchanged on the metal center but remain attached to the protein system. Infrared spectroscopy could then be used to monitor protein secondary structural changes in response to this triggering event.
7.5 Acknowledgements

We are thankful for the financial support from the National Institutes of Health (P41-GM104605 and GM-065978). R.M.A. is also an NSF Graduate Research Fellow (DGE-1321851). This project is a collaboration with the Dmochowski lab and has contributions from Teresa Rapp, Ivan Dmochowski, as well as Jianxin Chen from the Ultrafast Optical Processes Laboratory.
Figure 7.1 Reaction scheme for the photochemical ligand exchange process of the ruthenium polypyridyl complexes of the form \( [\text{Ru(bpy)}_2(L)_2]^{2+} \) with \( L \) being either A) 4-pentynenitrile (4PN) or B) 5-hexynenitrile (5HN).
**Figure 7.2** Representative UV-Vis spectrum of the complexes in Figure 1. The spectrum is that of $\text{[Ru(bpy)$_2$(5HN)$_2$]}^{2+}$ in water at room temperature prior to photolysis.
Figure 7.3 FTIR spectra of \([\text{Ru(bpy)}_2(4\text{PN})_2]^{2+}\) in water at room temperature before and after photolysis with 455 nm light, as indicated. Dotted line is the FTIR spectrum of the 4PN ligand in water with its absorbance normalized to the absorbance of the photolyzed product.
**Figure 7.4** FTIR spectra of \([\text{Ru(bpy)}_2(5\text{HN})_2]^{2+}\) in methanol at room temperature before and after photolysis with 420 nm light for the amount of time indicated. Dotted line is the FTIR spectrum of the 5HN ligand in methanol with its absorbance normalized to the absorbance of the photolyzed product.
Figure 7.5 IR transient kinetic trace of [Ru(bpy)$_2$(4PN)$_2$]$^{2+}$ in methanol at room temperature probed at 2244 cm$^{-1}$. The smooth black line is a biexponential fit of this trace with 3.3 ps and 43.7 ps time constants. Inset: Kinetic trace of the same complex probed at 2050 cm$^{-1}$. The biexponential fit of this trace has time constants 8.3 ps and 50.0 ps.
Figure 7.6 IR transient kinetic trace of \([\text{Ru(bpy)}_2(5\text{HN})_2]^{2+}\) in methanol at room temperature probed at 2231 cm\(^{-1}\). The smooth black line is a biexponential fit of this trace with 0.3 ps and 4.9 ps time constants. Inset: Kinetic trace of the same complex probed at 2071 cm\(^{-1}\). The biexponential fit of this trace has time constants 0.4 ps and 7.5 ps.
Figure 7.7 FTIR difference spectrum (green) and transient IR difference spectrum at a 430 ps time delay (pink) of [Ru(bpy)$_2$(4PN)$_2$]$^{2+}$. The amplitudes of the FTIR difference spectrum are on the left axis and the amplitudes of the 430 ps trace are on the right axis.
8. Isotope-Labeled Aspartate Sidechain as a Non-perturbing Infrared Probe: Application to Investigate the Dynamics of a Carboxylate Buried inside a Protein

8.1 Original Publication


8.2 Introduction

Various linear and nonlinear spectroscopic techniques based on probing the spectral and/or dynamic properties of various vibrational modes present in biological molecules have found a wide range of applications in biochemistry and biophysics. For example, vibrational transitions arising from protein backbone units, such as the amide I vibration, have long been exploited to study the structure, dynamics, interactions, folding and aggregation of proteins and peptides. Despite their broad utility, the commonly used intrinsic protein vibrational modes are often unable to yield site-specific structural or environmental information, due to vibrational coupling or spectral overlapping. An effective strategy to overcome this limitation is to introduce an ‘isolated’ and distinct vibrational mode via either isotopic labeling of an intrinsic vibrator or incorporation of an extrinsic vibrational probe. Examples of the former include replacing a backbone $^{12}$C=O unit with $^{13}$C=O or $^{13}$C=O or substituting a –CH$_3$ group with –CD$_3$, while the latter uses unnatural amino acids that exhibit a spectrally-distinguishable and
environmentally-sensitive vibrational transition. In the current study, we expand the toolbox of protein vibrational probes by showing that a $^{13}$C sidechain-isotope-labeled aspartate can be used as a non-perturbing IR reporter to study, for example, the dynamics of a charged carboxylate buried in a solvent-inaccessible position inside a protein.

The sidechains of aspartic acid (Asp) and glutamic acid (Glu) can exist in either a neutral or an anionic form (i.e., aspartate and glutamate), thus making them versatile and, in many cases, allowing them to play a vital role in protein folding and function. For example, salt-bridges formed between the carboxylate groups of Asp or Glu and basic amino acid sidechains are important to protein structure and stability as well as protein-protein interactions. Additionally, Asp and Glu residues often play important functional roles in proton pumps, catalysis, and ligand or protein binding. These carboxylate groups of deprotonated Asp and Glu give rise to a strong vibrational band (mainly the COO$^-$ asymmetric stretching vibration) around 1585 cm$^{-1}$ and 1568 cm$^{-1}$ in D$_2$O, respectively. However, only a few studies have utilized this naturally-occurring vibrational band as a site-specific IR probe because oftentimes multiple Asp and/or Glu residues are present in any one protein molecule. For a C=O stretching vibration, changing $^{12}$C to $^{13}$C typically red-shifts the vibrational frequency by $\sim$40 cm$^{-1}$. Previous studies have used a biological approach to incorporate a $^{13}$C label into the sidechain of Asp for analysis of the protonation states changes of Asp residues during the photocycle of bacteriorhodopsin. This approach was limited, however, as it also isotopically-labeled the Thr (threonine) and Glu residues in the protein of interest and the labeling efficiency of the Asp residue was low ($\sim$40%). Herein, we present a new method
to introduce $^{13}$C site-specifically into Asp residues (hereafter this isotopically-labeled Asp is referred to as Asp*), and demonstrate that the ionic form of an Asp residue with a $^{13}$COO$^-$ group is capable of acting as a reporter of Asp sidechain interactions and dynamics, particularly in proteins which have multiple acidic residues.

In order to show the utility of this site-specific IR probe, we employ it to probe the dynamics of a charged residue located in the interior of a small protein, psbd41, which corresponds to a truncated version of the peripheral subunit-binding domain of dihydrolipoamide acetyltransferase (E2) from the pyruvate dehydrogenase multienzyme complex from *Bacillus stearothermophilus*. This small protein consists of 41 amino acids, including three Asp and two Glu residues. In particular, a charged Asp residue (i.e., Asp34), although buried in the interior of the protein, has been shown to be critical to the folding and stability of psbd41. Although bringing a charged moiety into a protein’s interior, which is generally hydrophobic, is energetically unfavorable, it is not uncommon to find buried charges in proteins due to structural and/or functional purposes. In such scenarios, the mechanism by which nature oftentimes chooses to overcome this energetic penalty is by forming various favorable inter-residue electrostatic interactions, such as hydrogen-bonds (H-bonds) and salt-bridges. In addition, the presence of water molecules near a buried charge can also help significantly reduce the associated energetic penalty. According to an NMR structure of the peripheral subunit-binding domain of the E2 chain (Figure 8.1), the otherwise unfavorable burial of the charged sidechain of Asp34 is alleviated by multiple hydrogen-bonding (H-bonding) interactions formed between its carboxylate ion and the backbone amides of Gly23, Thr24, Gly25 and Leu31.
as well as the sidechain of Thr24.\textsuperscript{389} However, it is not clear whether water is also present near Asp34 because the NMR structure suggests that this site is inaccessible by solvent.\textsuperscript{389} Thus, in order to verify the spectroscopic utility of Asp*, as well as provide insight into the dynamics of the interactions that stabilize a charged group inside a protein, a topic important for protein electrostatics and energetic,\textsuperscript{394} we chose to replace Asp34 with Asp* in psbd41 and, in turn, use two-dimensional infrared (2D IR) spectroscopy\textsuperscript{97,99} to probe, site-specifically, the spectral diffusion dynamics of the \textsuperscript{13}COO\textsuperscript{−} asymmetric stretching vibration arising from Asp*34. The spectral diffusion dynamics of an inhomogeneously broadened molecular vibration would report on the time evolution of microscopic states contributing to the vibrational bandwidth and thus reveal information about the environmental fluctuations of the infrared (IR) reporter in question.\textsuperscript{395} For example, spectral diffusion dynamics measured in aqueous solution typically contain a component on the 1-2 ps timescale, due to water dynamics.\textsuperscript{396}

Interestingly, our results not only show that Asp* can be used as a site-specific IR probe in the presence of other carboxylates, but also provide evidence suggesting that water may exist near the Asp34 site in psbd41.

8.3 Experimental

8.3.1 Preparation of Cap-Asp and Cap-Asp*

In order to synthesize N-terminal acetylated and C-terminal N-methyl amidated, or capped, Asp and Asp* (hereafter referred to as Cap-Asp and Cap-Asp*), Fmoc-N-methyl indole resin (0.12 mmol) was first placed in a peptide synthesis vessel and swelled in
dichloromethane (CH$_2$Cl$_2$) (10 mL) for 1 h. The solvent was drained and the resin was washed with dimethylformamide (DMF) (3 × 6 mL) and then treated with 20% piperidine/DMF (2 × 6 mL), allowing the solution to contact the resin for 10 minutes. The resin was washed with DMF (5 × 6 mL) and a pre-mixed solution of Fmoc-amino acid (0.1 mmol), 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) (40 mg, 0.105 mmol, 1.05 equiv), oxyma (15 mg, 0.105 mmol, 1.05 equiv) and N,N-diisopropylethylamine (DIPEA) (35 µL, 0.2 mmol, 2 equiv) dissolved in DMF (1 mL) was added to the resin. The contents were rocked gently for 1 h, then drained and the resin was washed with DMF (3 × 6 mL).

The resin-bound Fmoc-amino acid (0.1 mmol) was washed with DMF (3 × 5 mL) and then treated with a solution of 20% piperidine/DMF (2 × 6 mL) allowing each treatment to contact the resin for 5 minutes. The resin was washed with DMF (5 × 6 mL), and then a pre-mixed solution of acetic anhydride (95 µL, 1.0 mmol, 10 equiv) and DIPEA (348 µL, 2.0 mmol, 20 equiv) dissolved in DMF (2 mL) was added to the resin. The contents were rocked gently for 1 h, then drained and the resin was washed with DMF (3 × 6 mL) and CH$_2$Cl$_2$ (3 × 6 mL).

The resin-bound peptide (~0.1 mmol) was pre-swelled in CH$_2$Cl$_2$ for 30 minutes and then treated with a cleavage cocktail of CH$_2$Cl$_2$, trifluoroacetic acid (TFA), triisopropylsilane (TIPS), and water (70:25: 2.5: 2.5; 5 mL) and stirred for 1 hour. The filtrate was collected and an additional cleavage cocktail (3 × 1 mL) was used to wash the resin. The pooled filtrates were evaporated to dryness using a stream of air. The residue was dissolved in water/acetonitrile (MeCN) (4:1, 1 mL) and purified by reverse-phase
high-pressure liquid chromatography (5 – 30% organic over 10 minutes) to give the desired product.

8.3.2 Peptide Synthesis and Purification

The Asp*34-psbd41 peptide (sequence: AMPSVRKY AREKGVDIRL VQGTGKNGRV LKE-Asp*-IDAFLA GGA) was synthesized on a Liberty Blue microwave peptide synthesizer (CEM, NC) using the synthesized Fmoc-Asp* described below and standard Fmoc-protocols. The peptide was cleaved from the rink amide resin using a TFA cleavage cocktail and purified by reverse-phase high performance liquid chromatography using a C18 column. This peptide was then identified by matrix assisted laser desorption ionization (MALDI) mass spectrometry. Peptide samples were exchanged with 0.01M DCl in D$_2$O two times to remove residual TFA from purification and titrated to either pH 1.5 or pH 8 with DCl or NaOD prior to dissolving in the respective buffer for analysis.

8.3.3 Circular Dichroism (CD) Measurements

The CD spectrum and thermal-melt (T-melt) of Asp*34-psbd41 were measured on an Aviv 62A DS spectropolarimeter (Aviv Associates, NJ) using a 1 mm sample holder. Concentrations of the protein sample (in 50 mM sodium phosphate buffer, pH 8) were 40 and 65 μM for the wavelength scan and T-melt measurement, respectively.

8.3.4 Fourier Transform Infrared (FTIR) Measurements

FTIR spectra were collected on an iS50 FT-IR spectrometer (Nicolet, WI) with 1 cm$^{-1}$ resolution. A two-compartment CaF$_2$ sample cell with a 56 μm Teflon spacer was used to allow back-to-back acquisition of the single-beam spectra of the protein sample and the
D$_2$O buffer via a home-made device.\textsuperscript{213} All samples used in the FTIR measurements had concentrations of \( \sim 5 \) mM in either a 50 mM deuterated sodium phosphate buffer (pH 8) or a 50 mM DCl-KCl D$_2$O solution (pH 1.5).

### 8.3.5 2DIR Measurements

The details of the 2D IR instrument have been described elsewhere.\textsuperscript{397} Briefly, three ultrafast mid-IR pulses were used to produce a third-order response (i.e., a photon echo signal) in the sample of interest, which was collected in the phase-matched direction following heterodyning with a local oscillator pulse. The resultant signal was dispersed by a monochromator and detected by a 64-element MCT array detector (InfraRed Associates). The final 2D spectra were obtained from three Fourier transform operations performed on the spectra collected in the time domain. The sample was held in a CaF$_2$ cell with either a 56 \( \mu \)m spacer for Asp*34-psbd41 (\( \sim 11 \) mM) or a 25 \( \mu \)m spacer for Cap-Asp* (\( \sim 23 \) mM), both of which were prepared in 50 mM deuterated sodium phosphate buffer (pH 8.0).

### 8.4 Results and Discussion

#### 8.4.1 Synthesis of Fmoc-Asp*

The synthesis scheme of Fmoc-\( L \)-Aspartic-4-$^{13}$C-Acid (\( \beta \)- tert-butyl ester) is shown (Figure 8.2). Specifically, the reaction was initiated from O’Donnell’s benzophenone imine of glycine 1,\textsuperscript{398–400} which, upon a trimethylaluminum mediated acylation with Oppozler sultam auxilliary,\textsuperscript{401} was converted to 2.\textsuperscript{402} Treatment of 2 with \( n \)-butyllithium formed the enolate which was combined with 3.\textsuperscript{403} After further reaction at room
temperature for three days, two molar equivalents of hydrochloric acid were used to hydrolyze the imine, yielding 4 with good yield after purification. Removal of the sultam from 4 was accomplished by the classical lithium hydroxide saponification. While a side reaction involving cleavage of the sultam ring occurred, as previously encountered in similar reactions,\textsuperscript{402,404} the desired reaction was achieved. The sultam was extracted from the solution containing the amino acid, and the solution pH was adjusted to 8. Lastly, the final protection was conducted with Fmoc-N-hydroxysuccinimide ester in acetone to produce the desired substrate, 5, after chromatography.

8.4.2 Vibrational Frequencies of Isotope Label

In order to determine the effect of isotopic-labeling on the COO\textsuperscript{−} asymmetric stretching vibrational frequency of the carboxylate group of Asp, we measured the FTIR spectra of the two model compounds, Cap-Asp and Cap-Asp*. As expected (Figure 8.3), the FTIR spectrum of Cap-Asp at pH 8, which has a deprotonated and hence a charged sidechain, consists of two peaks, one at \(\sim 1637\) cm\(^{-1}\) arising from the amide carbonyls and one at \(\sim 1581\) cm\(^{-1}\) arising from the carboxylate.\textsuperscript{318} In comparison, the lower-frequency band of Cap-Asp* at pH 8 is shifted to \(\sim 1538\) cm\(^{-1}\) (Figure 8.3), indicating that isotopic labeling of the C atom in the carboxylate group of Asp results in a 44 cm\(^{-1}\) red-shift in the corresponding COO\textsuperscript{−} asymmetric stretching vibrational frequency. However, at pH 1.5 where the sidechain of Asp* becomes protonated, the IR band arising from the sidechain is shifted to a spectral region that is overlapped with the amide I’ band (Figure 8.4). Thus, taken together these results indicate that the depotonated form of Asp* can be used as a site-specific IR probe of proteins, as none of the naturally-occurring protein vibrations
reside in the frequency region of 1520–1550 cm\(^{-1}\). For example, it can be used to probe the protonation status of a specific Asp residue as well as its H-bonding dynamics. As the pK\(_a\) of an acidic amino acid buried inside a protein can sample a large range, \(4.5 \text{ -- } 9.4\), the ability of Asp\(^*\) to be useful as an IR probe in a buried position then depends on the system of interest. However, the pK\(_a\) of a solvent-exposed Asp residue is ~4. Therefore, Asp\(^*\) can be used to study proteins with solvent-exposed Asp residues in solutions with a pH of 4 or larger.

To demonstrate the potential utility of Asp\(^*\), we employed it to probe protein H-bonding dynamics in a site-specific manner. Specifically, we replaced Asp34 in psbd41 with Asp\(^*\) (hereafter the isotopically labeled psbd41 is referred to as Asp\(^*\)34-psbd41), as the sidechain of this amino acid is not only engaged in multiple H-bonding interactions, but is also buried in the interior of the protein.\(^{389}\) As expected, this mutation is non-perturbing as circular dichroism (CD) measurements (Figures 8.5 and 8.6 and Table 8.1) revealed that the melting temperature of Asp\(^*\)34-psbd41 is within uncertainty the same as that reported for the wild-type protein by Raleigh and coworkers.\(^{390}\) In addition, their study showed that the folded structure was maintained and stable between pH 5.3 and pH 10. As shown (Figure 8.7), the FTIR spectrum of Asp\(^*\)34-psbd41 at pH 8 consists of four resolvable spectral features between 1500–1700 cm\(^{-1}\), with peak frequencies at approximately 1516, 1540, 1584, and 1643 cm\(^{-1}\), respectively. Besides the apparent amide I’ band at 1643 cm\(^{-1}\), the 1516 cm\(^{-1}\) band can be assigned to a C-C ring mode of the single tyrosine (Tyr) in psbd41, whereas the relatively broad feature between 1550–1600 cm\(^{-1}\) can be attributed to the other four non-labeled carboxylates (i.e., two from Asp and
two from Glu) in the protein. Thus, the 1540 cm$^{-1}$ band must arise from Asp*34. This assignment is consistent not only with the result obtained with Cap-Asp* (Figure 8.3) but also with previous studies demonstrating that the Asp34 is deprotonated at neutral pH.$^{389}$ Moreover, in dimethyl sulfoxide (DMSO), a polar but aprotic solvent, the carboxylate IR band of Cap-Asp* is blue-shifted to 1545 cm$^{-1}$ (Figure 8.8). Thus, the above assignment is also consistent with the notion that the carboxylate group of Asp34 is involved in H-bonding interactions. Furthermore, and perhaps most importantly, the FTIR spectrum of Asp*34-psbd41 provides convincing evidence that Asp* can be used to provide site-specific spectroscopic information, even for proteins that contain multiple Asp and Glu residues.

### 8.4.3 Spectral Diffusion Analysis

To demonstrate further the utility of Asp*, next we employed 2D IR spectroscopy to measure dynamics underlying the spectral bandwidth of the carboxylate asymmetric stretching transition of Asp*34 in Asp*34-psbd41. For an inhomogeneously broadened vibrational transition, the corresponding 2D IR spectra are often elongated along the diagonal at early waiting times and subsequently become more circular as time proceeds due to a redistribution of vibrational frequencies. Interestingly, the 2D IR spectrum of Asp*34-psbd41 obtained at an early waiting time (i.e., $T$), as shown (Figure 8.9), only exhibits a modest elongation along the diagonal. That is, the CLS$^{-1}$ at waiting time $T = 0$ fs is $\sim$0.2, which is relatively small compared to that of other inhomogeneously broadened molecular vibrations. For example, Hochstrasser and coworkers found that 2D IR experiments on an isotopically-labeled backbone carbonyl in the transmembrane
region of the M2 proton channel produced a CLS$^{-1}$ at waiting time $T = 0$ fs of 0.74. Additionally, Chung et al. showed that the CLS$^{-1}$ for the CN stretching vibration of cyanophenylalanine mutants of HP35 were all ~0.6 at time $T = 0$ fs. Therefore, there must be a very fast dynamic event, which contributes to the Lorentzian bandwidth (~17 cm$^{-1}$) of Asp*34 but evades detection due to our experimental time resolution. Since the 2D IR spectrum of Cap-Asp* displays a similar behavior (Figure 8.10), this result suggests that this phenomenon is intrinsic to the vibrator and is not caused by an environmental factor. One potential interpretation is that this phenomenon is caused by the rapid electronic equilibration dynamics between the two resonant structures of this carboxylate ion, leading to an ultrafast dephasing time of the vibration of interest; however, any rapid electronic equilibration between two states would lead to this ultrafast dephasing. Further analysis of the 2D IR spectra of Asp*34-psbd41 collected at different $T$ values using the commonly used center line slope (CLS) method permitted us to determine the spectral diffusion dynamics occurring on the picoseconds (ps) timescale that contribute partially to the inhomogeneous bandwidth of the $^{13}$COO$^{-}$ asymmetric stretching vibration of Asp*34. As indicated (Figure 8.11), the value of CLS$^{-1}$ decays exponentially with $T$ with a time constant of $1.1 \pm 0.2$ ps. Since the carboxylate of Asp*34 is H-bonded with multiple sites, the straightforward interpretation of this result is that it manifests the underlying dynamics of these H-bonds.

However, most existing examples in the literature have attributed such ultrafast spectral diffusion dynamics to mobile water molecules. For instance, the study of Hochstrasser and coworkers found that the spectral diffusion dynamics of the amide
I vibration of certain isotopically-labeled amide units in an amyloid β fibril occur on a 1-2 ps timescale, which was interpreted as the motion of nearby water molecules. Similar studies on other biological systems revealed that only when mobile or bulk-like water molecules are present the spectral diffusion of the vibrational probe occurs on such an ultrafast timescale. On the other hand, spectral diffusion of an H-bonded vibrator arising from protein backbone motions typically occurs on a slower time scale, as observed by Hamm and coworkers.

Based on these previous studies, we tentatively assign the ~1.1 ps spectral diffusion component of Asp*34 to mobile water near the carboxylate group rather than from intra-molecular H-bond interactions. Given the small size of psbd41, it is not unreasonable to assume that water can transiently penetrate into its interior. In order to provide further evidence in support of this notion, an ideal control experiment would be to measure the spectral diffusion dynamics of the IR probe in the unfolded state of Asp*34-psbd41, wherein Asp*34 is expected to be solvent exposed. However, both urea and guanidinium chloride, which can denature psbd41 at high concentrations, have a strong absorbance near the IR band of Asp*34, making it difficult, if not impossible, to perform the required 2D IR measurements when these denaturants are present. While psbd41 can also be denatured by lowering the pH to approximately 2.5, we cannot use this strategy because Asp*34 will become protonated under such acidic pH condition. For these reasons, we simply measured the spectral diffusion dynamics of the $^{13}\text{COO}^{-}$ asymmetric stretching vibration of Cap-Asp* in water at pH 8, where the sidechain of Asp* is expected to be fully hydrated. As indicated (Figure 8.12), the corresponding CLS$^{-1}$ vs. $T$ curve can be described by a single-exponential function.
with a time constant of ~1.4 ps. This control experiment therefore corroborates the above conclusion that Asp34 in psbd41 is solvated by water.

8.5 Conclusions

Achieving broad site-specificity in the IR study of protein structure, dynamics and function necessitates the development of a wide array of not only environmentally-sensitive but also structurally-diverse vibrational probes that can be incorporated into proteins. The availability of a large set of such probes is important, as different applications may require substitution of different amino acids in order to avoid or minimize structural perturbation and to obtain certain structural or dynamic information. Despite many previous efforts, until this study we have lacked the ability to target and examine the underlying vibrational properties of a specific carboxylate ion in a protein system. Given the fact that the carboxylates of Asp and Glu are often involved in electrostatic interactions that are crucial for protein folding and function, the development of a method that allows selective interrogation of the structural and environmental properties of any individual acidic residues in a protein would be highly valuable. In this study, we have demonstrated that a $^{13}$COO$^-$ isotope-labeled aspartate now provides a convenient, non-perturbing means to achieve this goal. Additionally, using this site-specific IR probe and 2D IR spectroscopy, we are able to provide evidence that water may exist in the interior of the small protein, psbd41.
8.6 Acknowledgements

We gratefully acknowledge financial support from the National Institutes of Health (P41-GM104605). RMA is supported by a National Science Foundation Graduate Research Fellowship (DGE-1321851).
### Table 8.1 Unfolding thermodynamic parameters of Asp*34-psbd41 in a 50 mM sodium phosphate buffer (pH 8).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta H_m$ (kcal mol$^{-1}$)</td>
<td>24.5</td>
</tr>
<tr>
<td>$\Delta S_m$ (cal K$^{-1}$ mol$^{-1}$)</td>
<td>73.5</td>
</tr>
<tr>
<td>$\Delta C_p$ (cal K$^{-1}$ mol$^{-1}$)</td>
<td>569.5</td>
</tr>
<tr>
<td>$T_m$ ($^\circ$C)</td>
<td>59.0</td>
</tr>
</tbody>
</table>
Figure 8.1 NMR structure of pdsd41 (PDB code: 2PDD), showing the sidechain of Asp34 as well as the various hydrogen-bonding interactions.
Figure 8.2 Scheme of the synthesis of Fmoc-Asp*.
Figure 8.3 FTIR spectra of Cap-Asp and Cap-Asp* in a 50 mM sodium phosphate D$_2$O buffer (pH 8), as indicated. In both cases, the bands at ~1637 cm$^{-1}$ corresponds to the amide I’ transition of the backbone carbonyl, whereas the bands centered at 1581 and 1537 cm$^{-1}$ arise from the sidechains.
Figure 8.4 FTIR spectra of Cap-Asp and Cap-Asp* in a 50mM deuterium chloride-potassium chloride D$_2$O buffer (pH 1.5), as indicated. In both cases, the bands at ~1643 cm$^{-1}$ correspond to the amide I' transition of the backbone carbonyl, whereas the band centered at 1707 cm$^{-1}$ and the shoulder around 1672 cm$^{-1}$ arise from the sidechains.
Figure 8.5 CD spectrum of Asp*34-psbd41 in a 50 mM sodium phosphate D$_2$O buffer (pH 8) at 25°C.
**Figure 8.6** CD thermal melting curve (open circles) of Asp*34-psbd41 in a 50 mM sodium phosphate D_2O buffer (pH 8). The smooth line represents the best fit of this data to a two-state model. The melting temperature (T_m) was determined to be 59.0 °C and the values from the fit are listed in Table 8.1.
Figure 8.7 FTIR spectrum of Asp*34-psbd41 in a 50 mM sodium phosphate D_{2}O buffer (pH 8). The arrows indicate the $^{13}$COO$^{-}$ asymmetric stretching band of the Asp*34 sidechain at 1540 cm$^{-1}$. The inset is a zoom of the spectrum in the region of 1500–1600 cm$^{-1}$. 
Figure 8.8 FTIR spectra of deprotonated Cap-Asp* in dimethyl sulfoxide (DMSO). The band at 1665 cm$^{-1}$ arises from the amide I’ transition of the backbone carbonyl, and the band centered at 1545 cm$^{-1}$ corresponds to the deprotonated sidechain.
Figure 8.9 Representative 2D IR spectra of Asp*34-psbd41 in a 50 mM sodium phosphate D₂O buffer (pH 8).
Figure 8.10 2D IR spectra of Cap-Asp* in a 50 mM sodium phosphate D$_2$O buffer (pH 8) at $T = 150$ fs.
Figure 8.11 CLS$^{-1}$ versus $T$ plot of the 2D IR spectra of Asp*34-psbd41. The smooth line is the fit of these data to a single-exponential function with a time constant of 1.1 ps.
Figure 8.12 CLS$^{-1}$ versus $T$ plot of the 2D IR spectra of Cap-Asp* in a 50 mM sodium phosphate D$_2$O buffer, pH 8. The smooth line is the fit of these data to a single-exponential function with a time constant of 1.4 ps.
9 The C=O Stretching Vibration of 4-Oxoproline as an Infrared Probe of Protein Structure and Local Electric Field

9.1 Introduction

Proline is an important amino acid in protein conformation and function due to its unique cyclized structure. As the only amino acid with its side-chain integrated into the backbone units, proline has a very rigid structure with a backbone φ angle of $\sim -65^\circ$. Moreover, this inflexibility also affects the proximal N-terminal amino acid as the dihedral angle is restrained. Because of this angle constraint, chains containing several proline residues are often restricted to form a polyproline II (PP II) helix with backbone dihedral angles ($\phi, \psi$) of ($-75^\circ, 150^\circ$). In this structure, the prolines form a hydrophobic strip around the helix with the backbone carbonyls open for hydrogen-bonding (H-bonding) and the prolines in their trans conformation. These proline-rich-motifs are often easily recognizable sites for binding and since proline is a rigid structure, it loses very little conformational entropy upon binding. Additionally, the cyclized proline structure also allows for cis/trans isomerization along the peptide bond preceding it. This is due to the steric repulsion in both the cis and trans forms of Xaa-Pro (where Xaa is any amino acid) are approximately equal whereas favorable electrostatic interactions in non-proline peptide bonds help stabilize the trans conformation. This isomerization ability thus often acts as a rate-determining step of protein folding reactions taking tens to hundreds of seconds to occur in large globular proteins. Nonetheless, nature also uses this isomerization mechanism in protein function. As examples, 5-Hydroxytryptamin type 3 receptor ion channels use proline
isomerization as a gating mechanism to transfer messages across cell membranes, whereas other biomolecular systems, such as Oxytocin peptides and Cry adaptor proteins use proline isomerization as a means to control or regulate ligand binding. Therefore, not only does proline play a critical role in maintaining protein structure, but also acts as a switchable moiety to modulate protein activities.

In many of these proline interactions, there are large changes in local electrostatic fields. For example, biomolecular interactions are often governed by electrostatic effects and protein folding, which involves the formation of hydrogen bonds (H-bonds) and van der Waals interactions, also produces changes in the local electric field throughout the folding pathway. Therefore, probing and quantifying these changes in electrostatics would be an effective way to monitor and understand the interactions taking place. Vibrational stark spectroscopy is useful in this regard as it relies on the notion that certain vibrational modes are sensitive to electric field in a quantitative, predictive manner and thus their vibrational frequencies can be used to assess the local electrostatic field. The amide I’ band of the protein is often used to assess these protein interactions as the carbonyl stretching frequency of the backbone amide unit is sensitive to electrostatics. However, due to the delocalization of this mode over the whole protein backbone and the contribution of other vibrational modes to this band, the amide I’ vibration does not allow for a site-specific assessment of the local electric field. Thus, localized vibrational modes are necessary in order to achieve a rigorous evaluation of the local electric field. For this reason, in protein systems, we rely on unnatural amino acid reporters to site-specifically probe these interactions.
Currently, there are only a few analogs of proline that have been utilized as site-specific IR probes\textsuperscript{437} with none being sensitive to local electric field. In order to be used in this way, several criteria must be met by the amino acid substitute.\textsuperscript{64} First, it should be minimally perturbative to the protein structure and environment. Additionally, it should be easily incorporated either chemically or biologically. The vibration of interest should be a local mode with a large cross-section that resides in an uncongested region of the infrared spectrum. Lastly, the frequency reporter should be sensitive to a particular environmental factor such as hydration status or pH. Previous studies have shown that the carbonyl stretch in various chemical groups including ketones,\textsuperscript{438,439} amides,\textsuperscript{321} and esters\textsuperscript{162} is a sensitive IR probe to local electrostatic field. Moreover, ketones reside between 1665 cm\textsuperscript{-1} and 1850 cm\textsuperscript{-1} in the infrared spectrum with cyclic ketones tending towards the higher frequency end of this range.\textsuperscript{440} Herein, we show that the uncommon amino acid 4-oxoproline (P\textsubscript{ox}) can be used as a sensitive probe of local electric field in proline-containing peptides as well as a reporter of proline \textit{cis-trans} isomerization. We then use this probe to assess the structure and environment of a protein aggregate.

9.2 Experimental

9.2.1 Materials

1-Methyl-3-pyrrolidinone was purchased from Sigma-Aldrich with 97% purity and used as received. FMOC-Pro(4-keto)-OH was purchased from Chem-Impex International Inc. for the peptide syntheses. Methanol, acetonitrile, and dimethyl sulfoxide were purchased
from Acros Organics and hexanes, diethyl ether, and dimethylformamide were purchased from Fisher Scientific and used as received.

9.2.2 Peptide Synthesis, Purification, and Preparation

All peptides were synthesized using standard FMOC protocols on a Liberty Blue microwave peptide synthesizer (CEM, NC) and then cleaved from the rink amide resin using a TFA cleavage cocktail. Peptides were then purified using reverse-phase high performance liquid chromatography (HPLC; Agilent Technologies, CA) and identified using either matrix-assisted laser desorption ionization mass spectrometry (MALDI-TOF; Bruker, MA) for TTR-P\textsubscript{ox} or liquid chromatography-mass spectrometry (LC-MS; Water, MA) for KP\textsubscript{ox}G and GP\textsubscript{ox}G. Peptides were then exchanged with multiple rounds of 0.01M DCl in D\textsubscript{2}O to remove residual TFA. The KP\textsubscript{ox}G and GP\textsubscript{ox}G peptides were titrated to the respective pH using 0.1M NaOD in D\textsubscript{2}O or 0.1M DCl in D\textsubscript{2}O prior to analysis. The TTR-P\textsubscript{ox} peptide was monomerized with HFIP and lyophilized prior to being dissolved in 10% ACN in D\textsubscript{2}O, pH 2 for aggregation at 4.4mM. Aggregation was prompted by seeding with a 1:10 mol ratio of aggregated TTR wild-type to the TTR- P\textsubscript{ox} and sonicated for 5 minutes. The sample was sonicated for another 10 min after one day of aggregation to ensure proper fibril formation.

9.2.3 Fourier Transform Infrared (FTIR) Measurements

FTIR measurements were collected on a Nicolet iS50 FT-IR spectrometer (Nicolet, WI) for the 1M3P, KP\textsubscript{ox}G, and GP\textsubscript{ox}G samples with 1 cm\textsuperscript{-1} resolution. A two-compartment home-built sample cell comprised of two CaF\textsubscript{2} windows separated by a Teflon spacer of 56 μm pathlength was used for the measurements allowing for back-to-back monitoring.
of both the sample and reference/solution spectra. Sample concentrations for these measurements were 10 mM for 1M3P, between 10 – 12 mM for KP_oxG, and between 5 – 7 mM for GP_oxG. A Nicolet 6700 FT-IR spectrometer (Nicolet, WI) with 1 cm\(^{-1}\) resolution was used for the TTR-P_ox measurements using a single-compartment cell with a 56 μm spacer. All spectra were collected in solution at ~4.4mM except for the spectrum of the dry film, which was dried onto the CaF\(_2\) window under vacuum.

### 9.2.4 MD Simulations

PSF, and PDB files for the 1M3P molecule was created using the Molefacture built-in module in VMD. The Force Field Toolkit Plugin was then used to create the parameter file for this molecule. MD simulations were run using the molecular dynamics program NAMD2.7 and force field parameters for the solvent molecules from CHARMM22 by first centering the 1M3P molecule in an equilibrated solvent box for each of the solvents used. A 1 ns equilibration run at 298 K and 1 atm in the NPT ensemble was then run followed by two production runs for 10 ns each at 298 K in the NVT ensemble saving every 1ps for a total of 20000 frames. EF calculations were then performed using VMD. In order to separate the MD frames of 1M3P in protic solvents into different hydrogen bonded states, the built-in module, Hydrogen Bonds, was used with the geometric criteria that carbonyl hydrogen bonds occurred with distances < 3.2 Å and angles of < 40°.

### 9.2.5 Gaussian Calculations

The Gaussian 09 software package was used to optimize the geometry of the GP_oxG and KP_oxG peptide systems as well as calculate their vibrational frequencies in vacuo. The
B3LYP level of theory was used in the 6-31+G(d,p) orbital basis set. The resulting vibrational frequencies were scaled by a factor of 0.9632.441

9.3 Results and Discussion

9.3.1 Model Compound Vibrational Frequency and Electric Field

In order to assess the environmental dependence of the ketone stretching frequency of the Pox amino acid, we performed vibrational solvatochromic measurements on a small molecule mimic, 1-methyl-3-pyrrolidinone (1M3P). The reasons for choosing this small molecule were two-fold: 1) it was miscible/soluble with various solvents whereas the amino acid itself was not, thus we could study a wide range of dielectric constants and 2) by removing the possibility for H-bond donating at the nitrogen site, the methyl capped nitrogen is a better mimic of Pox in a peptide than a small molecule with a secondary amine would be (such as Pox alone). As can be seen (Figure 9.1), the stretching frequency and peak width of 1M3P are highly dependent on the local electric field of the solvent. For example, by moving from water to hexanes, the peak frequency of 1M3P blue shifts by 20 cm$^{-1}$ and the FWHM narrows significantly by $\sim$10 cm$^{-1}$. Moreover, at least two resolvable peaks can be observed in methanol suggesting that the vibrational frequency of this probe is sensitive to the H-bonding state of the carbonyl as well as its local electric field. Similar trends have been seen by Pazos et al.162 in which case their ester carbonyl was able to quantitatively distinguish both the H-bonding status as well as the local electric field of the carbonyl moiety.
Given the complex behavior of the carbonyl stretching frequency of this model compound, molecular dynamics (MD) simulations of 1M3P in the various solvents were performed similarly to the method by Boxer and coworkers. The local electric field contribution in the direction of the carbonyl dipole (C to O) was calculated at both the carbonyl C and O atoms in 1M3P for each of 20000 frames in a series of solvents. For aprotic solvents, the average electric field value at each atom was calculated whereas for protic solvents, frames were first divided up based on geometric criteria for H-bond formation. In each of these cases, the lower frequency peaks in the FTIR were attributed to the more H-bonded vibrators as H-bonds pull electron density from the chemical bond thus weakening it. However, the correlation between the calculated average electric field (taken as the sum of the electric field on the carbonyl carbon \( E_C \) and the carbonyl oxygen \( E_O \)) and the center frequency was not as linear as in previously analyzed compounds (Figure 9.2). Specifically, the \( R^2 \) value for the best-fit line was only 0.84. Therefore, unlike linear compounds, the stretching frequency of 1M3P cannot be accurately determined by solely using the calculated electric field on only the two carbonyl atoms as other atoms in the ring structure must contribute.

In order to consider contributions from other atoms in the 1M3P ring, the electric field was calculated in the same way for each non-hydrogen atom in the molecule. A frequency-frequency map was then generated using a linear combination of the electric field values for the carbonyl C, O, and one other atom in the ring. Although all the atoms in the ring produced a linear correlation with \( R^2 \) values of greater than 0.9, it was found that the carbon atom (C2) between to the carbonyl and the other ring carbon
led to the best linear fit with an $R^2$ value of 0.98 (Table 9.1). Moreover, as shown (Figures 9.3 and 9.4), the frequency distributions generated as a result of this fit are Gaussian in nature with the relative percentages of each H-bonded state in the MD simulations of the protic solvents in agreement with the ratio of the integrated areas of the IR peak fits. For the case of methanol, the FTIR fits and MD simulations suggest the presence of three hydrogen-bonded states whereas only two are discernible for water. Additionally, the standard deviation of the calculated vibrational frequencies as a result of the fit shows a similar correlation to the FWHM of the FTIR peaks, as has been previously seen162 (Figure 9.5 and Table 9.2). Together, these indicate that the MD simulations were able to accurately represent the H-bonding patterns of 1M3P as well as encompass the heterogeneity of the system. Therefore, the frequency-frequency correlation map (Figure 9.6) generated using the electric field at the C, O, and C2 atoms provides a quantitative method to correlate the experimentally measured frequency of this ketone to electric field.

9.3.2 Tripeptides and Backbone Effects

In order to see whether this electric field dependence still held in a peptide system, $P_{ox}$ was implemented into a tripeptide system, $GP_{ox}G$. As shown (Figure 9.7), the $P_{ox}$ carbonyl in the tripeptide system in D$_2$O absorbs at a frequency that is blue-shifted by ~13 cm$^{-1}$ from the 1M3P carbonyl. This further indicates that the backbone units or atoms in/near the pyrrolidinone ring also influence the stretching vibration of this carbonyl, as had been seen in the quantitative determination of the frequency-frequency map. However, the electric field dependence is maintained, as there is a substantial increase in
the vibrational frequency upon moving this peptide system to the more hydrophobic solvent, tetrahydrofuran (THF). Interestingly, this shift is 4.3 cm\(^{-1}\) larger than that observed for the model compound. Thus, these results indicate that there might be some other effect, in addition to solvent electrostatics, that governs the frequency shifts of this probe.

Moreover, the sensitivity of this stretching frequency to different pH conditions was tested using the tripeptide, KP\(_{\text{ox}}\)G which has two pKa values due to the presence of a charged sidechain (the C-terminus is amidated and thus not ionizable). Previous studies utilizing an ester sidechain showed that it was able to sense the charged groups on nearby amino acids. Therefore, we would expect to see large shifts upon deprotonation of the lysine sidechain. The FTIR spectra collected at pH 2, 7, 11, and 12 show noticeable changes in the vibrational frequency of the P\(_{\text{ox}}\) carbonyl (Figure 9.8). The band shifts slightly from 1760 cm\(^{-1}\) to 1758 cm\(^{-1}\) upon increasing the pH from 2 to 7. However, after deprotonation of the N-terminus and lysine sidechain, the band starts to blue shift again to 1768 cm\(^{-1}\). This large shift makes sense as the positively charged sidechain becomes neutral thus decreasing the magnitude of local electric field. Thus, the C=O stretch of P\(_{\text{ox}}\) is dependent on and sensitive to local electrostatics and can sense the protonation/deprotonation events on nearby basic residues.

If in fact local backbone atoms of the protein influence the stretching frequency of this carbonyl vibration in addition to the local electric field, then this probe should be able to report on local backbone conformations. To test this hypothesis, we utilized the two tripeptide systems, GP\(_{\text{ox}}\)G and KP\(_{\text{ox}}\)G, and monitored the conformational changes of
the backbone upon increasing the temperature. Since proline is able to occupy both its cis and trans isomeric forms at room temperature, increasing the temperature should change the ratio of molecules in the cis:trans conformation. Therefore, temperature-dependent IR spectra of KPoxG were collected and, as the temperature is increased, there is a large red shift (~ 10 cm⁻¹) in the frequency of the Pox carbonyl (Figure 9.9). Moreover, at each temperature, the C=O band can be fit to two pseudo-Voigt profiles. If these two bands are from the cis and trans peptide populations, then the lower frequency band must be the result of the cis conformer as it grows in at higher temperatures. Interestingly, the temperature-dependent FTIR of the GPoxG does not show such a large shift (Figure 9.10). This may be due to the lack of sidechains on the amino acids adjacent to the proline moiety, which do not significantly change the charge distribution on the molecule from one conformation to the other.

In order to see if the assignment of cis and trans conformers to these two populations is correct, Gaussian frequency calculations in vacuo were conducted on both tripeptide systems. As shown (Table 9.3), in both cases, the carbonyl stretching frequency in the cis isomeric form is more red shifted than that of the trans form. Moreover, whereas the KPoxG peptide has a 5 cm⁻¹ shift between the two conformers, the GPoxG peptide shifts by less than 1 cm⁻¹. These calculations are in agreement with the observed experimental results as the KPoxG shows a significant shift upon isomerization, but the GPoxG peptide does not. This phenomenon must be due to a rearrangement of partial charges within the peptide. For these two peptides, the N atom shows the greatest change in its partial charge for the KPoxG peptide. This makes sense given the lone pair of
electrons on this atom that are easily polarizable. Moreover, the partial charge of this atom does not change significantly between the two isomeric forms for GP\textsubscript{ox}G.

Since the two FTIR bands for the KP\textsubscript{ox}G peptide are due to two isomeric forms, the ratio between the areas under these bands would report on the equilibrium constant for this process. Thus, as shown in the van’t Hoff plot (Figure 9.11), there is a linear dependence of this ratio on 1/T. Using this information, the $\Delta S$, $\Delta H$, and $\Delta G$ values can be calculated. As can be seen (Table 9.4), there is a rather small free energy difference between these two isomeric forms of only 0.9 kcal/mol. A previous study\textsuperscript{422} on a pentapeptide system showed that for a Lys-Pro peptide bond, the $\Delta G_{\text{trans} \rightarrow \text{cis}}$ is 1.4 kcal/mol. Additionally, NMR studies\textsuperscript{447,448} on Staphylococcal Nuclease A showed that this protein exists in two different folded states which differ in the isomerization state of the Lys\textsubscript{116}−Pro\textsubscript{117} peptide bond. A thermodynamic assessment of these two states showed that $\Delta G_{\text{trans} \rightarrow \text{cis}}$ is $-1.2$ kcal/mol, as the cis isomeric form is dominant due to its enthalpic favorability. One possible reason for the small free energy difference in KP\textsubscript{ox}G is due to the fact that both the cis and trans isomeric forms are stabilized by intramolecular H-bonding interactions. As revealed by Gaussian structural optimization calculations in vacuo, in its cis isomeric form, KP\textsubscript{ox}G is able to form an H-bond between the N- and C-termini whereas an H-bond can form between the Lys sidechain and the C-terminus in the trans isomer.

### 9.3.3 Investigation of TTR Aggregates

In order to test the utility of this unnatural amino acid as a probe of protein conformational and environmental changes, we implemented it into the 113 position of
the transthyretin (TTR) peptide segment (105-115) studied by Dobson and coworkers (sequence: YTIAALLSPYS, hereafter referred to as TTR-\(P_{\text{ox}}\)) in place of the proline residue.\textsuperscript{449,450} All of the residues of this short peptide are involved in backbone-backbone H-bonds except for the proline residues, which adopts a different structure than that observed in the full-length protein. By mutating this residue, we can site-specifically monitor the changes that take place at this location during aggregation. Early in the aggregation process (Figure 9.12), the \(P_{\text{ox}}\) band resides around 1762 cm\(^{-1}\), as has been noticed with all solvent exposed \(P_{\text{ox}}\) carbonyl groups in peptides. As aggregation progresses, however, there is a shoulder at higher frequencies that forms and a slight red shift in the main vibrational band. Upon drying and dessication of the aggregates, this vibrational band broadens and needs multiple Gaussian functions to be fit appropriately, indicating the presence of \(P_{\text{ox}}\) in at least two different environments. One of these bands is \(\sim 1760\) cm\(^{-1}\), indicating the presence of water in the system despite the removal of the outside solvent. Therefore, there must be trapped water in the aggregates that are formed. Interestingly, this peak is at slightly lower frequencies than the original band indicating that there is either an increase in the local electric field or the proline might occupy the \textit{cis} isomer to some extent. Additionally, the more blue shifted peak at \(\sim 1771\) cm\(^{-1}\) indicates that a subset of the carbonyl groups are facing a more apolar environment, such as the interior of the aggregate. Thus, the \(P_{\text{ox}}\) probe is able to provide insight into the heterogeneity of local environments in these aggregates as well as monitor changes along the course of the aggregation process.
9.4 Conclusions

Herein, we introduced the utilization of a carbonyl-derivative of proline, $P_{ox}$, as a site-specific IR probe of local electric field and protein backbone conformation. Due to the complex relationship of this C=O stretch to both nearby solvent charges and intramolecular partial charges, we were able to show that it can be useful in assessing the cis-trans isomerization process of proline residues as well as the solvent environment surrounding the moiety. In this way, we used the frequency changes of this moiety as a function of temperature to determine the change in free energy for the cis isomerization process in a small tripeptide, KP$_{ox}$G. Additionally, by implementing $P_{ox}$ into the TTR$_{105-115}$ peptide, changes in the local environment of the proline residue were able to be assessed during the aggregation process.

9.5 Acknowledgements

We would like to thank the National Institutes of Health (GM-065978) for their financial support. R.M.A. is supported by an NSF Graduate Research Fellowship (DGE-1321851).
Table 9.1 Experimentally determined carbonyl stretching frequencies ($\omega_{\text{expt}}$), calculated electric fields at the carbonyl C ($E_C$), carbonyl O ($E_O$), and proximal C2 ($E_{C2}$) positions and vibrational frequency fit values ($\omega_{\text{fit}}$) of 1M3P in various solvents.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>$\omega_{\text{expt}}$ (cm$^{-1}$)</th>
<th>$E_C$ (MV cm$^{-1}$)</th>
<th>$E_O$ (MV cm$^{-1}$)</th>
<th>$E_{C2}$ (MV cm$^{-1}$)</th>
<th>$\omega_{\text{fit}}$ (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEXA</td>
<td>1768.6</td>
<td>-0.41</td>
<td>-0.44</td>
<td>0.01</td>
<td>1770.2</td>
</tr>
<tr>
<td>DETE</td>
<td>1765.0</td>
<td>-11.77</td>
<td>-16.08</td>
<td>-1.26</td>
<td>1765.0</td>
</tr>
<tr>
<td>ACN</td>
<td>1758.4</td>
<td>-21.62</td>
<td>-32.44</td>
<td>-1.33</td>
<td>1758.9</td>
</tr>
<tr>
<td>DMF</td>
<td>1757.8</td>
<td>-22.86</td>
<td>-33.35</td>
<td>-0.80</td>
<td>1757.1</td>
</tr>
<tr>
<td>DMSO</td>
<td>1754.2</td>
<td>-25.70</td>
<td>-35.48</td>
<td>-0.50</td>
<td>1754.5</td>
</tr>
<tr>
<td>MeOH 0 H-bonds</td>
<td>1765.0</td>
<td>-19.09</td>
<td>-32.38</td>
<td>-2.41</td>
<td>1762.8</td>
</tr>
<tr>
<td>MeOH 1 H-bond</td>
<td>1757.8</td>
<td>-37.08</td>
<td>-84.83</td>
<td>-4.75</td>
<td>1757.5</td>
</tr>
<tr>
<td>MeOH 2 H-bonds</td>
<td>1748.1</td>
<td>-57.29</td>
<td>-110.89</td>
<td>-7.22</td>
<td>1748.6</td>
</tr>
<tr>
<td>Water 1 H-bond</td>
<td>1751.5</td>
<td>-39.37</td>
<td>-98.47</td>
<td>-1.30</td>
<td>1750.6</td>
</tr>
<tr>
<td>Water 2 H-bonds</td>
<td>1744.4</td>
<td>-55.16</td>
<td>-122.38</td>
<td>-4.03</td>
<td>1745.4</td>
</tr>
<tr>
<td>Solvent</td>
<td>FWHM of $\omega_{\text{expt}}$ (cm$^{-1}$)</td>
<td>Std. Dev. of $\omega_{\text{fit}}$ (cm$^{-1}$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------------------</td>
<td>---------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEXA</td>
<td>9.6 ± 0.8</td>
<td>4.1 ± 0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETE</td>
<td>9.6 ± 0.8</td>
<td>15.7 ± 1.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACN</td>
<td>10.0 ± 0.8</td>
<td>21.4 ± 1.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMF</td>
<td>9.6 ± 0.8</td>
<td>22.7 ± 1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMSO</td>
<td>12.2 ± 1.0</td>
<td>25.1 ± 2.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MeOH 0 H-bonds</td>
<td>9.7 ± 0.8</td>
<td>26.6 ± 2.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MeOH 1 H-bond</td>
<td>16.9 ± 1.4</td>
<td>27.5 ± 2.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MeOH 2 H-bonds</td>
<td>19.7 ± 1.6</td>
<td>27.2 ± 2.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water 1 H-bond</td>
<td>15.2 ± 1.2</td>
<td>34.5 ± 2.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water 2 H-bonds</td>
<td>18.1 ± 1.4</td>
<td>34.6 ± 2.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 9.2** Full width at half maximum (FWHM) of the carbonyl stretching frequencies of 1M3P and the standard deviation (Std. Dev) of the respective vibrational frequency fit in various solvents. The error bars are the 92% confidence interval of the data.
<table>
<thead>
<tr>
<th></th>
<th>Frequency (cm⁻¹)</th>
<th>N Charge (au)</th>
<th>O Charge (au)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KoxoPG - Charged Lysine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cis</td>
<td>1795.058048</td>
<td>0.105512</td>
<td>-0.376498</td>
</tr>
<tr>
<td>trans</td>
<td>1800.095584</td>
<td>0.419078</td>
<td>-0.363728</td>
</tr>
<tr>
<td>trans-cis</td>
<td>5.037536</td>
<td>0.313566</td>
<td>0.01277</td>
</tr>
<tr>
<td><strong>KoxoPG - no charged Lysine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cis</td>
<td>1787.98816</td>
<td>0.126303</td>
<td>-0.393855</td>
</tr>
<tr>
<td>trans</td>
<td>1789.307744</td>
<td>0.241906</td>
<td>-0.38155</td>
</tr>
<tr>
<td>trans-cis</td>
<td>1.319584</td>
<td>0.115603</td>
<td>0.012305</td>
</tr>
<tr>
<td><strong>GoxoPG</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cis</td>
<td>1790.126464</td>
<td>0.073486</td>
<td>-0.381391</td>
</tr>
<tr>
<td>trans</td>
<td>1790.935552</td>
<td>-0.002854</td>
<td>-0.375806</td>
</tr>
<tr>
<td>trans-cis</td>
<td>0.809088</td>
<td>-0.07634</td>
<td>0.005585</td>
</tr>
</tbody>
</table>

Table 9.3 Gaussian calculated vibrational frequencies and Mulliken partial charges for the Pox tripeptides indicated
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value (units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta H_{\text{trans} \rightarrow \text{cis}}$ (kcal mol$^{-1}$)</td>
<td>7.8</td>
</tr>
<tr>
<td>$\Delta S_{\text{trans} \rightarrow \text{cis}}$ (cal K$^{-1}$ mol$^{-1}$)</td>
<td>25.1</td>
</tr>
<tr>
<td>$\Delta G_{\text{trans} \rightarrow \text{cis}}$ (kcal mol$^{-1}$)</td>
<td>0.9</td>
</tr>
</tbody>
</table>

**Table 9.4** Thermodynamic Parameters for Proline $\text{Trans} \rightarrow \text{Cis}$ Isomerization in KP$_{\text{ox}}$G
Figure 9.1 Normalized FTIR spectra of 1M3P in various solvents, as indicated. Inset: Structure of the 1M3P molecule.
Figure 9.2 Center frequencies of 1M3P versus the calculated local electric field from MD simulations in various solvents, as indicated. The electric field values have been scaled by a factor of 2.5 in the same way as Boxer and coworkers.⁴³⁹
Figure 9.3 Histogram of the vibrational frequency fit ($\omega_{\text{fit}}$) distributions of the C=O stretching frequency of 1M3P in aprotic solvents, as indicated.
**Figure 9.4** Histogram of the vibrational frequency fit ($\omega_{\text{fit}}$) distributions of the C=O stretching frequency of 1M3P in protic solvents, a) MeOH and b) D$_2$O. The differently H-bonded states have been calculated separately, as indicated.
Figure 9.5 Correlation between the full width at half maximum (FWHM) of the experimentally determined carbonyl stretching frequencies ($\omega_{\text{expt}}$) of 1M3P and the standard deviation (Std. Dev.) of the vibrational frequency fit ($\omega_{\text{fit}}$) for each solvent.
Figure 9.6 Vibrational frequency fit ($\omega_{\text{fit}}$) of 1M3P versus the experimental center frequencies ($\omega_{\text{expt}}$) in various solvents. The $\omega_{\text{fit}}$ values were generated by using a linear combination of the electric field values in the C=O bond direction at three atoms in the molecule: the carbonyl carbon (C), the carbonyl oxygen (O), and the carbon between the carbonyl and another carbon (C2). The resulting fit is given by the following equation:

$$\omega_{\text{fit}} = 0.77E_C - 0.08E_O - 1.81E_{C2} + 1770.5.$$
**Figure 9.7** FTIR spectra of GP$_{ox}$G in D$_2$O and THF, as indicated. The amide I’ contributions have been subtracted out for clarity.
Figure 9.8 FTIR spectra of KP$_{\alpha}$G in D$_2$O at various pH conditions, as indicated. The amide I’ contributions have been subtracted out for clarity.
Figure 9.9 Temperature-dependent FTIR spectra of KP_{ox}G in D_{2}O, pH 7.
Figure 9.10 Temperature-dependent FTIR spectra of $\text{GP}_x\text{G}$ in $\text{D}_2\text{O}$ at pH 7.
Figure 9.11 Van’t Hoff plot of the Cis-Trans reaction of KP$_{ox}$G in D$_2$O, pH 7. The equation of the line of best fit is provided. The thermodynamic parameters corresponding to this fit are given in Table 9.4.
**Figure 9.12** Normalized FTIR spectra of the carbonyl stretch of $P_{ox}$ in TTR-$P_{ox}$ at different stages of aggregation, as indicated. The amide I’ band has been subtracted for clarity.
10 Summary and Future Directions

Protein molecules are diverse polypeptides that are able to undergo a wide range of interactions and motions across many different timescales. Although current spectroscopic methods have been useful in elucidating many of the fundamental aspects of protein structure and dynamics, a comprehensive understanding of protein mechanisms and interactions is still lacking. As such, there is still a need to develop experimental methods and strategies that can control and site-specifically monitor these biomolecular processes. To this end, this thesis has aimed to use spectroscopic techniques along with chemical strategies to gain insight into the factors, which affect protein structure, dynamics, and function. A second aim has been to develop new methods that can be used to answer some open questions in the field. There are many potential extensions to the work presented herein and below we summarize the major finding as well as future directions.

In Chapter 4, we demonstrated that the commonly used co-solvent, trifluoroethanol (TFE), at certain concentrations, can act as a nanocrowder to increase the rate of protein folding reactions by the excluded volume effect. Although computational studies had previously been shown in that TFE can form small nanometer-sized clusters in water, this study was the first experimental validation of this claim. By monitoring the conformational kinetics of two intrinsically disordered peptides using temperature-jump infrared spectroscopy, we observed that these clusters increase the rate of tertiary structure formation, while local secondary structure formation is not affected. Not only does this study provide evidence in support of the potential for TFE to form small
clusters, but also provides a caution for experimental measurements taking place in binary mixtures. A potential future direction could be to use 2D IR spectroscopy to gain a better molecular level understanding of how TFE forms clusters, the mechanisms involved, and how TFE clusters interact with protein systems. Additionally, this effect could easily be used to tune the rate at which a protein reaction occurs, for instance, to potentially increase the rate of protein binding events without the need for large macromolecular crowders.

In Chapter 5, we developed a new method to increase the rate of a protein folding reaction by tuning the attempt frequency of this process. Specifically, we utilized an azobenzene cross-linker strategically placed in the α-helix involved in the major folding transition state of the Trp-cage mini-protein to not only initiate the folding of this protein but also to increase the attempt frequency of this protein folding reaction upon photoisomerization. By increasing the curvature of the free energy barrier to folding, we increased the rate of this protein folding reaction by an order of magnitude with respect to the normal folding time of this protein. Moreover, by opening up parallel folding pathways using this method, the curvature of the transition-state barrier was estimated for the first time. As a next step, it would be interesting to see if azobenzene or similar photoisomers could also be used to tune the attempt frequency of other protein systems, particularly larger proteins with more tertiary interactions. Moreover, learning what physical factors affect the presence of a parallel folding pathway versus a single faster pathway by using these other systems would also help our understanding of the free energy landscape and its underlying physical characteristics.
Chapter 6 focused on monitoring the photochemical reaction involving electron transfer from tryptophan (Trp) residues to a nearby disulfide bond because of its importance in protein damage and degradation. The reaction was determined to occur in \(~2 \mu s\) using nanosecond transient IR spectroscopy. Moreover, this reaction was shown to proceed through an excited state quenching mechanism with end products involving formation of a localized Trp-S adduct between a Trp and cysteine (Cys) residue as a result of the radical reaction. This study indicates that tryptophan-mediated electron transfer to a structure stabilizing disulfide bond could be used as a phototigger to initiate protein-folding reactions for protein systems folding slower than 2 \(\mu s\). Therefore, a natural extension of this work would be to implement this same Trp disulfide system into a slower folding protein to initiate a folding or unfolding reaction. For example, ribonuclease A\(^{451,452}\) might be a good candidate for this as it contains four interweaving disulfide bonds, two of which are shown to be important to the stability of the protein. Additionally, this process can also be used to control the biological activity of a protein system, for instance that of an antimicrobial peptide. The activity of several antimicrobial peptides\(^{330}\) has been shown to be dependent on the presence of disulfide bonds in their structure thus cleaving this bond could eliminate this function.

Chapter 7 explored the photophysics of the ligand dissociation process in two ruthenium polypyridyl complexes. Specifically, it was found that both nitrile-containing ligands are exchanged upon excitation using near UV or blue light within a few hundred picoseconds. Since this timescale is faster than that of protein folding reactions, this system would be suitable for the study of protein folding processes, even those occurring
on the nanosecond or microsecond timescale. Click chemistry\textsuperscript{372} can be used to attach the labile ligands in the ruthenium complex to azido amino acid sidechains in a protein and folding can be triggered using visible light. Another possibility would be to use nitrile derivatized amino acids, such as 4-cyanophenylalanine or 5-cyanotryptophan, as the detachable ligands in this complex such that photodissociation would yield that natural peptide, free on any non-native chemical moieties.

In Chapter 8, we demonstrated the use of a $^{13}$C label on the sidechain of aspartate (Asp) as a site-specific reporter of protein structure and environment. By shifting the vibrational frequency of this carboxylate to an uncongested region of the infrared spectrum, local protein interactions can be probed. In this regard, we were able to use this probe in a protein, which contains several acidic residues, to monitor the dynamics of an aspartate residue buried in the interior of this protein. We found that water is able to penetrate inside this protein structure and thus most likely help stabilize this interior charge. Future work will aim to incorporate this probe into the transmembrane region of the M2 channel,\textsuperscript{453} as there is an aspartic acid residue near the C-terminal end of this peptide, which is thought to form hydrogen bonds with the Trp residues in the closed state to prohibit the passage of water. Since no studies have been able to site-specifically probe the water dynamics of the four-helix bundle below the Trp residues, implementing sidechain-labeled $^{13}$C Asp into this structure would provide insight in the regard. Moreover, rimantadine, a drug which inhibits the functionality of this channel, is known to bind near this Asp residue and thus this probe will be used to interrogate these interactions as well.\textsuperscript{454}
Finally, in Chapter 9, we showed that a proline analog, 4-oxoproline, has a carbonyl stretching frequency that is sensitive to both local electrostatic field and backbone conformation. As such, this probe was used to monitor a cis-trans isomerization process in a small peptide system as well as assess the local environmental changes during the aggregation of the peptide, transthyretin (TTR). Since there are very few infrared probes which are proline analogs, this amino acid shows potential as being a reporter of local interactions or dynamics affecting proline-containing amino acids. For example, proline rich motifs are common sequences recognized by protein-protein interaction domains, and thus implementing 4-oxoproline into one of these domains could be helpful in determining the changing electric field associated with this process.
Bibliography


216


219


(142) Zhou, B.-R.; Liang, Y.; Du, F.; Zhou, Z.; Chen, J. Mixed Macromolecular Crowding Accelerates the Oxidative Refolding of Reduced, Denatured Lysozyme:


(306) Xie, J.; Li, Y.; Cao, Y.; Xu, C.; Xia, M.; Qin, M.; Wei, J.; Wang, W. Photo Synthesis of Protein-Based Drug-Delivery Nanoparticles for Active Tumor Targeting. *Biomater. Sci.* **2013**, *1*, 1216.


250