2D IR Spectroscopy and Computational Modeling: Application to Protein Folding and Binding

by

Ziad Ganim

B. S., University of California (2003)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUN 0 2 2010
LIBRARIES

Submitted to the Department of Chemistry
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physical Chemistry

ARCHIVES

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2010

© Massachusetts Institute of Technology 2010. All rights reserved.

Author ....
Department of Chemistry
May 21, 2010

Certified by ....
Andrei Tokmakoff
Professor
Thesis Supervisor

Accepted by ....
Robert W. Field
Chairman, Department Committee on Graduate Theses

This doctoral thesis has been examined by a Committee of the Department of Chemistry that included,

Professor Robert G. Griffin
Chair

....

Professor Keith A. Nelson

....

Professor Andrei Tokmakoff
Thesis Supervisor

....

2D IR Spectroscopy and Computational Modeling: Application to Protein Folding and Binding

by

Ziad Ganim

Submitted to the Department of Chemistry
on May 21, 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physical Chemistry

Abstract

In this thesis, dynamics experiments are developed that can be used to study protein conformational changes such as folding and binding. Every functional motion of a protein is inextricably linked to conformational dynamics. However, most of our insight into protein folding and binding is indirectly obtained through kinetics experiments that measure reaction rates and reveal how fast populations of stable states interconvert. Two-dimensional infrared spectroscopy (2D IR) is the central tool developed in this thesis for protein dynamics experiments due to its combination of time and structural resolution. As a vibrational spectroscopy, 2D IR potentially offers femtosecond time resolution. Its advantages over linear, absorption spectroscopy come through correlating excitation and emission frequencies to allow for a separation of homogenous and inhomogeneous line shape components, and to give rise to structurally sensitive cross-peaks. One general problem was repeatedly addressed in this thesis: how can 2D IR spectra best be modeled to reveal atomistic structural information? The key feature that now sets 2D IR apart from other fast protein probes is that the data can readily be calculated from an atomistic structure or molecular dynamics simulation using the methods developed in this thesis work. Demonstrative applications are presented for the amide I-II spectroscopy of NMA, the amide I'-II' spectroscopy of poly-L-lysine, isotope-edited 2D IR spectroscopy of trpzip2, and transient 2D IR spectroscopy of ubiquitin unfolding after a temperature jump. The emerging paradigm is to interpret 2D IR spectra with the aid of an atomistic, molecular dynamics simulation. The applications to protein binding use the monomer-dimer transition of insulin as a model system. Using a combination of experiments and simulations, this equilibrium was characterized as a function of protein concentration, temperature, and solvent. Finally, as a complement to the structural information provided by 2D IR, dye-labeling and intrinsic tyrosine fluorescence experiments on insulin are described.

Thesis Supervisor: Andrei Tokmakoff
Title: Professor

To my first math and science teachers-
my mother and father

Acknowledgments

Graduate school has been an incredible ride for me. I would like to start my acknowledgments by reflecting on what brought me here. For as long as I can remember, I've known my parents to be incredibly hard workers and I've tried to emulate their example. They instilled in me a love for school and wound my spring tightly enough to keep me excited about learning all this time. I knew that I had to be good at math like my mom and dad before I even knew what math was. They made sure that I did know before my first day of kindergarten, and continued to take my education seriously and personally from then on. They gave me great confidence in myself, enrolled me in the best public schools, and always made sure I had all the tools necessary to learn. As the culmination of the schooling they began, I dedicate this thesis to my mother and father.

My parents are only the beginning of the tremendous encouragement I've received from my family. I am very grateful for the unwavering support of my brother, Sam. I am fortunate enough to have many aunts, uncles, and cousins, which guarantees a constant flux of their well-wishes. Thank you for everything.

I first knew that I wanted to go to graduate school when I was an undergraduate in Rich Mathies' group. I started as a biochemistry major, barely out of freshman chemistry, and interested in proteins. Rich gave me the responsibility for purifying rhodopsin and it was through him that I was able to work closely with three graduate students- Judy Kim, David McCamant, Mike Tauber- and two post-docs- Elsa Yan and Duohai Pan. Discussions with them about quantum mechanics, vibrational spectroscopy, and optics captured my imagination; I got to see the life of a graduate student up close and personal, and I learned that graduate school was *the place* for curious people to learn, teach other, and discuss all day. Rich gave me the freedom to work on my first computational project, and by then I was hooked. I am so grateful that I had such a wonderful group of mentors to help me find my path.

When I came to MIT, I was not expecting to stay into the next decade. I credit Andrei for advising me through a set of projects that leave me excited enough at the end of the day that I'm always ready to pick up the next morning. I believe the scariest feeling in graduate school is being trapped; to be unable to leave a dark, locked room until you've slain a beast, which you've already failed to slay and is not guaranteed to be mortal. These character-building moments are indeed necessary to appreciate success. However, these moments have been rare for me, and for that, I am thankful for Andrei's creativity, open-mindedness, and trust. With his encouragement, I've been able to grow protein crystals, learn gene expression, work on ion-caging complexes and amyloid fibrils. While these projects have at times been tangent to my top priorities, I would never trade the experiences. As a direct result, when I look back at graduate school my vivid memories are of thoughts such as, "Maybe this will work, and wouldn't it be great if it did?" and, "I can't wait for the next step. This is going to be amazing." In addition to helping me self-actualize, such a work environment has given me strength during some challenging personal times.

As a young student, I learned the most about lab from Hoi Sung Chung. I remember watching him and building an expectation that whenever I would touch any optic, the signal would go up. His talent was matched by patience and generosity; this was the perfect safety net for someone new to the world of corralling invisible beams around the table. I also benefitted from working with Adam Smith and Benjamin Dietzek in lab. Through my early interactions with senior students, Chris Fecko and Joel Eaves, I quickly learned that the Tokmakoff Group set high standards for the quality of work, and I have been compelled to maintain that reputation.

I learned a great deal about optics and spectroscopy from Joe Loparo, Matt DeCamp, Poul Petersen, and Lauren DeFlores. For a while, our desk arrangement was most conducive to continuing the flowing conversations from lunch that would meander from academic life, interferometry, and photon echoes into politics and policy. As I began to discover my niche as the computational student for protein-related projects, I was thankful that other people were willing to share their projects with me. I first worked with Lauren on NMA because I thought it was fun; the fact that it ended up being both fun and productive helped shape these types of collaborations into a pattern for me.

I feel a kindred bond to students who entered my year and went through all the same trials with me. As the last person in my incoming class to graduate, I've seen them all come and go, and I look forward to keeping in touch. Sean Roberts is a talented researcher who shares my vision of combining experiments and simulations, and has always been a wonderful resource and a sympathetic ear whose opinion I trust and respect greatly.

A very special mention needs to go out to the "Muddy Crew" who are the people that deal with my anguish and share in my joys on nearly a daily basis, and certainly on Wednesday nights- Marvin Bayro, Marc Caporini, Xander Barnes, Galia Debelouchina, Ashley Ball, Leo Gomez, Björn Corzilius, and Albert Smith are a tremendous support group that have surely helped me more than they realize. A dubious honor goes to Becky Nicodemus for her dual membership in the Muddy Crew and Tokmakoff group. As a close friend and a careful scientist, I have depended a lot on her and will always value her direct opinion. I've been reminded that life doesn't end at MIT through friendships with John Higby, Joe Osborne, and people from CAMS and Berkeley. They've all helped me maintain perspective and allowed me to live vicariously in the world outside academia.

As someone who enjoys both teaching and learning, I'm thankful for the time I've spent working with younger students. I've enjoyed plenty of scientific and lighthearted discussions with Joshua Lessing and Krupa Ramasesha. In lab, I quickly realized that Kevin Jones, and more recently, Chunte Sam Peng are both far more than just extra hands. I've learned from their expertise and they have helped me become a better teacher by challenging my perspective. They have helped me find and plug knowledge gaps, and I look forward to seeing their successes in the Big Lab.

Contents

1 Introduction **25**

1.1 Dynamics Experiments 27

1.2 Thesis Outline 31

1.3 Introduction to Two-Dimensional Infrared Spectroscopy 33

1.3.1 Challenges in Interpreting Protein Vibrational Spectra 34

1.3.2 Amide I Vibrations 35

1.3.3 Amide I 2D IR of Secondary Structure 35

1.3.4 Protein 2D IR Spectra 38

1.4 Afterword 40

2 Third-Order Infrared Experimental Methods **41**

2.1 Implementations of 2D IR Spectroscopy 41

2.2 Acquiring 2D IR Spectra 43

2.2.1 Generating 6 μm Light 43

2.2.2 2D IR Spectrometer 44

2.2.3 2D IR Experiments 46

2.3 2D IR Data Workup 50

2.4 Representations of 2D IR Spectra 52

2.5 Basic Features of 2D IR Spectra 54

2.5.1 Diagonal Peaks 54

2.5.2 Cross Peaks 55

2.6 Line Shapes and Intensity 57

2.6.1 Polarization-Selectivity 59

2.6.2 Phase Twist in 2D IR Line Shapes 59
2.6.3 2D IR Peak Positions 61
2.7 Slices and Projections: One-Dimensional Representations of 2D IR Spectra 62

3 Numerical and Analytical Methods for Evaluating Third-Order Response Functions 65
3.1 Response Function Formalism 67
3.2 Amide I' - Amide II' 2D IR Spectroscopy of *N*-methylacetamide . . . 77
3.3 Modeling the Amide I'-II' 2D IR Spectra of Idealized Secondary Structures 81
3.3.1 Experimental Results 81
3.3.2 Amide I'-II' Model 84
3.4 Afterword and Acknowledgements 88

4 Simulating 2D IR Spectra Using MD Simulations: Spectral Signatures of Heterogeneous Protein Ensembles 91
4.1 Introduction 91
4.2 Methods 94
4.2.1 The Amide I Subspace 94
4.2.2 MD Simulations 98
4.2.3 IR Spectroscopy 99
4.2.4 Experimental 102
4.3 Results and Discussion 102
4.3.1 Information Content of FTIR vs. 2DIR 102
4.3.2 Static Averaging for 2DIR 104
4.3.3 The Origin of Frequency Shifts in Site Energy Models 108
4.3.4 Relative Importance of Coupling and Site Energies 109
4.4 Conclusions 112
4.5 Acknowledgements 114
4.6 Afterword 114

5 Melting of a β-hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations 119

5.1 Isotope Labeling Experiments 120

5.2 Modeling Trpzip2 Spectroscopy using Hidden Markov Model Macrostates 122

5.2.1 Time-Correlation Methods vs. Static Averaging Methods for Calculating 2D IR Spectra 123

5.3 Results for Trpzip2 124

5.4 Conclusions 132

5.5 Acknowledgements and Afterword 132

6 Transient Two-Dimensional Spectroscopy of Ubiquitin Unfolding: A Comparison of Dynamics from Experiment and MD Simulation 135

6.1 Introduction 135

6.2 Materials and Methods 139

6.2.1 2D IR Spectroscopy 139

6.2.2 Transient 2D IR spectroscopy 141

6.2.3 Temperature Jump Synchronization 141

6.2.4 Sample Preparation 142

6.2.5 Simulating 2D IR Spectra 143

6.3 Experimental Results and Discussion 145

6.3.1 Equilibrium Measurements 145

6.3.2 Transient Thermal Unfolding of Ubiquitin 147

6.3.3 Line Broadening During Unfolding 151

6.3.4 Comparison to T-jump 2D IR Spectrum of Ubiquitin at $T_i =$ 25 °C 152

6.3.5 Comparison with DVE Results 153

6.3.6 Temperature Dependence of Homogeneous Broadening 155

6.3.7 Unfolding Mechanism of Ubiquitin 155

6.3.8 Comparison with Simulation 157

6.4 Conclusions 162

6.5 Acknowledgements 162
6.6 Afterword 162

7 The Biophysics of Coupled Protein Folding and Binding 163
7.1 Diffusive Encounter 163
7.1.1 Single Particle Diffusion 163
7.1.2 Association Kinetics from Two Particle Diffusion 165
7.2 Recognition and Binding 166
7.2.1 Methods for Studying Protein Association Biophysics 167
7.2.2 Computational Methods for Predicting Protein Association . . 169
7.2.3 Conclusions about the Current State of Protein Association Models 171
7.3 Free Energy Surfaces for Coupled Protein Folding/Binding Processes 174
7.3.1 Binding Mechanisms Characterized by Free Energy Surfaces . 175
7.4 Insulin Background 177

8 Insulin Dimer Dissociation and Unfolding Revealed by Amide I Two-Dimensional Infrared Spectroscopy 181
8.1 Introduction 181
8.2 Methods 182
8.2.1 Sample Preparation 182
8.2.2 Window Treatment 182
8.2.3 Simulations 183
8.3 Results 185
8.3.1 Concentration-Dependent 2D IR Spectral Features 185
8.3.2 K_D Measurements 187
8.3.3 Temperature-Dependent 2D IR Spectral Features 189
8.3.4 Solvent-Dependent Thermal Dissociation and Unfolding 190
8.3.5 Thermodynamic Modeling 194
8.3.6 Structural Modeling of Insulin 2D IR Spectra 197
8.3.7 Bright State Analysis of the Calculated Spectra 198

8.4 Discussion 200
8.4.1 Monomer and Dimer Spectral Signatures Across the Various IR Spectra 200
8.4.2 Solvent and Temperature Effects 202
8.5 Conclusions 207
8.6 Acknowledgements 209
8.7 Afterword 209

9 Insulin Monomer and Dimer Equilibrium Fluorescence Experiments 211
9.1 Motivation 211
9.2 Dye-Labeled Insulin Fluorescence 212
9.2.1 Synthesis, Purification, and Spectroscopic Methods 213
9.2.2 Characterization of Labeled Insulin 216
9.2.3 Measuring the Dissociation Constant, K_D 219
9.2.4 Thermal Dimer Dissociation 223
9.2.5 Conclusions 224
9.3 Intrinsic Insulin Fluorescence 227
9.3.1 Background 227
9.3.2 Experimental Setup 228
9.3.3 Results and Discussion for Insulin Fluorescence 230
9.3.4 Conclusions 234
9.4 Acknowledgement 234

A MATLAB Code for Calculating 2D IR Spectra of Six-Level Systems 235

B Texas Red-Insulin Mass Spectra 243

List of Figures

1-1 Illustration of considerations for a dynamics experiment. (a) A folding reaction with two structurally distinct equilibrium states that is characterized by the rates k_f and k_r. (b) A spectroscopic coordinate that distinguishes between the reactant (R), product (P), and transition state (TS) with a dividing line on the free energy surface. (c) A free energy surface where the product well and transition state are not separable from the reactant without inclusion of an unobservable, dark coordinate. (d) A schematic of a *dynamics experiment* that directly reveals time-resolved structural changes and the barrier-free rate. (c) Time-resolved populations observed in a *kinetics experiment* that measure the overall rate, k_f, and require additional assumptions to reveal dynamical information. 28

1-2 Overview of time scales for protein dynamics and kinetics and directly accessible time scales by experiments. 29

1-3 The Amide I Mode in Proteins. (Top) Empirical protein structure-frequency relationships in the amide I region. (Bottom) A color-coded visualization of the IR active β-sheet and α-helix states where the shading intensity is proportional to the participation ratio of a unit oscillator and the color denotes its phase; blue and red are 180° out-of-phase. Figure excerpted from Ganim et al.[1] 36

1-4 Amide I FTIR and 2D IR spectra demonstrating secondary structure sensitivity. (Top) FTIR and 2D IR spectra of the three forms of poly-L-lysine compared to calculated FTIR and 2D IR spectra of idealized structures (Bottom). Gaussian random site energies are sampled around a mean of 1650 cm^{-1} for the α helix, and random coil ($\sigma = 10$ cm^{-1} and 12 cm^{-1}) with a homogeneous line width of γ=12 cm^{-1} for all. Figure adapted from Ganim et al.[1] 37

1-5 FTIR and 2D IR spectra for proteins with a varying amounts of β sheet and α helix secondary structure. The data was compiled from Demirdöven et al.,[2] Chung et al.,[3] Smith et al.,[4] and Ganim et al.[1] . . 39

2-1 2D IR Inteferometer, Sample, and Detection Schematic. Figure adapted from Khalil et al.[5] t^*_{LO}- The local oscillator (LO) and tracer (d) are delayed using the same translation stage, and are thus alternately used. 45

2-2 Pulse Sequence and Wavevector Matching for Rephasing and Non-Rephasing Experiments. Figure adapted from Khalil et al.[5] 47

2-3 Calibrating the ω_3 axis using spectral interferometry. 49

2-4 Rephasing and non-rephasing t_1-ω_3 surfaces for the insulin dimer 2D IR surface shown in Figure 8-2A. 50

2-5 Slices of correctly phased data from Figure 1-4, and comparison to the pump-probe. 53

2-6 One- and two-quantum dipole allowed transitions (in red and blue, respectively) and energy levels for a two-dimensional anharmonic oscillator. 55

2-7 Sample 2D IR spectrum for the six-level system. 56

2-8 ZZZZ (Top) and ZZYY (Bottom) Real, Imaginary, and Power Spectra of Rephasing, Non-Rephasing, and Summed 2D IR spectra 58

2-9 Slices of the ZZYY 2D IR surfaces in Figure 1-8 at ω_1=1635 cm^{-1}. . 60

2-10 FT IR Spectrum, HDVE Spectra, Pump-Probe Spectra, and 2D IR Diagonal Slices for the Six-Level System in Figure 1-8. FT IR spectrum for the six-level system shown in Figure 1-8, projections (HDVE and pump-probe), and slices (Correlation Spectrum and NRPS Diagonal). 64

3-1 Layout of the approximations made to derive the third-order response functions presented in this chapter. 66

3-2 Feynman Diagrams corresponding to third-order signals for a multilevel system. 71

3-3 Simulated and experimental data for NMA-d_7 in D_2O. Adapted from DeFlores et al.[6]. 78

3-4 Six level system for NMA-d7 in D_2O 79

3-5 Experimental amide I'-II' spectra of poly-L-lysine in its three forms. 2D IR spectra in (A) ZZZZ and (B) ZZYY polarization are plotted along with (C) ω_1 projections. Contours are plotted from ±80% of the amide I' normalized peak at a 5% spacing. For the β-sheet conformation, contours are plotted at 1.25% between ±20% to enhance subtle features of the cross peak region. 82

3-6 Calculated amide I'-II' spectra of three idealized secondary structures. 2D IR spectra in (A) ZZZZ and (B) ZZYY polarization are plotted along with (C) ω_1 projections. Contours are plotted from ±100% of the amide I' normalized peak at 7% spacing and 1.25% spacing for ±15%. 83

3-7 The Amide I-II Model Hamiltonian and Transition Dipoles. (Top) A color-coded visualization of the amide I-II Hamiltonian. Each square on the diagonal represents a site and each off-diagonal square represents a coupling value. The top left quadrant is the amide I manifold and the bottom right quadrant is the amide II manifold; the coupling between the amide I and amide II blocks is diagonal, which indicates that amide I-II couplings are only non-zero if the oscillators are at the same backbone site. (Bottom) A visual representation of the transition dipoles. 86

3-8 Comparison of Amide II' Coupling Mechanisms- Nearest Neighbor Coupling and Mutual Coupling to Amide I'. 87

4-1 Comparison of Homogeneous and Inhomogeneous Broadening Mechanisms. As the Lorentzian HWHM parameter γ is increased, the predicted FTIR linewidth begins to match experiment, but the predicted tilt in the 2DIR peaks is qualitatively incorrect. The spectra are for bovine pancreatic trypsin inhibitor (BPTI) calculated using the heuristic site energy model from the PDB structure 1BPI[7]. 21 evenly spaced contours are plotted from ±50% of the maximum for each spectrum. . 103

4-2 Averaging over structural heterogeneity in 2D IR lineshapes. Two sample spectra and structures are shown for slightly different hydrogen bonding environments calculated using the Skinner 6P site energy model. Note that the fine structure disappears and the diagonal elongation of the 2DIR peaks arises from the static average over the equilibrium distribution of hydrogen bonding environments characterized by N, the number of internal hydrogen bonds, and R_{O-N}, the oxygen-nitrogen hydrogen bond distance. 21 evenly spaced contours are plotted from ±50% of the maximum for each spectrum. 105

4-3 Comparison of Calculated and Experimental FTIR and 2DIR for Trpzip2, D-Arg, and Ubiquitin. 21 evenly spaced contours are plotted from ±50% of the maximum for each spectrum. 106

4-4 Trpzip2 site energy correlation functions. The normalized site energy fluctuation correlation function (solid line), calculated with the Keiderling 5P (dark) and Jansen 4F (light) models (a representative potential and field site energy model, respectively) are shown, in addition to the contribution from water (dashed line) for both models. 110

4-5 Site energy fluctuations give rise to diagonal broadening. Six exploratory calculations are presented in an attempt to identify the physical origins of spectral features in trpzip2, (A) All elements of the Hamiltonian set to their average value, (B) Site energies set to their average value, couplings sampled from MD, (C) Couplings set to their average value, site energies sampled, (D) All elements sampled, (E) Couplings set to their average value, site energies sampled from a Gaussian distribution with center and width corresponding to MD site energy statistics. 21 evenly spaced contours are plotted from ±50% of the maximum for each spectrum. 111

4-6 Comparison of Calculated and Experimental FTIR and 2DIR for Trpzip2, D-Arg, and Ubiquitin for all electrostatic potential-based site energy models. Calculated spectra from the manuscript of Ganim et al.[8] 21 evenly spaced contours are plotted from ±50% of the maximum for each spectrum. 116

4-7 Comparison of Calculated and Experimental FTIR and 2DIR for Trpzip2, D-Arg, and Ubiquitin for all non-electrostatic and electrostatic field-based site energy models. Calculated spectra from the manuscript of Ganim et al.[8] The heuristic model uses PDB structures 1LE1 for trpzip2, 1UBQ for ubiquitin, and a random equilibrated MD structure for D-Arg. 21 evenly spaced contours are plotted from ±50% of the maximum for each spectrum. 117

5-1 Sequence, backbone structure, and isotope labeling sites for trpzip2 experiments. 120

5-2 2D IR spectra of trpzip2 isotopologues focusing on the isotope-shifted peaks. 121

5-3 Comparison between numerical integration of the Schrödinger equation and static average methods for calculating 2D IR spectra. 2D IR spectra are plotted for the trpzip2 bulged Markov state (BT) using the static averaging method with different values of the Lorentzian linewidth parameter, γ, and the NISE method. 124

5-4 Histograms of the site energies for each trpzip2 site separated by the five different Markov states. 125

5-5 Histograms of the site energies for each trpzip2 site separated by the number of hydrogen bonds. 127

5-6 Calculated 2D IR spectra of trpzip2 and isotopologues for five different Markov states. 128

5-7 FT IR and FT IR difference spectra of trpzip2 isotopologues. Equilibrium FTIR spectra are shown for each TZ2 isotopologue at 25°C and pH = 2.5. Each spectrum is baseline corrected with a linear subtraction, and area normalized for comparison. Difference data (bottom) are obtained by subtracting the unlabeled spectrum (TZ2-UL) from each of the spectra. The S1 difference spectra is presented at 2x magnification for clarity. Arrows have been drawn to emphasize the isotope labelled peaks. 129

5-8 Calculated trpzip2 FT IR and FT IR Difference Spectra. The difference spectra are relative to the unlabeled peptide for each Markov state. . 130

6-1 Downhill unfolding dynamics. A T-jump induces a barrier shift toward the folded state. A subensemble is trapped at the shifted transition state and unfolds in a downhill manner on the nanosecond-microsecond time scale (A). The downhill unfolding appears as a burst phase in experiments with millisecond time resolution. Subsequently, the excess population in the folded well unfolds across the barrier, which results in a millisecond exponential relaxation (B). The dotted vertical lines indicate barrier positions before and after the T-jump. 138

6-2 Structure of ubiquitin. (a) Crystal structure of ubiquitin[9] rendered with MOLMOL.[10] (b) Projection of the β-sheet of ubiquitin. A square box with a digit n represents a peptide group formed by residues n and $n+1$. Red and purple lines indicate native contacts used in calculating the interstrand distance coordinates R_1 and R_2, respectively, in Fig. ??. 139

6-3 Beam alignment and data processing for transient 2D IR spectroscopy. Alignment of beams in the sample region (a). The LO is vertically displaced by 100 μm from the other three beams (a, b, and c) that generate the third-order signal. 140

6-4 Equilibrium thermal unfolding of ubiquitin monitored by 2D IR spectroscopy. Parallel (ZZZZ) (a) and perpendicular (ZZYY) (b) polarization geometries. Spectra are normalized to the maximum of the 63 ° C spectrum. Twenty-one contours are plotted for ±60% of the spectra at 63 and 72 ° C and for 15% of the difference spectra (Right). Positive and negative peaks are indicated by red and blue. Green and purple arrows represent cross-peaks. In the difference spectra, red and blue arrows indicate the diagonal peaks on the red and blue sides of the ω_3 axis, respectively. $\nu_\perp$ and $\nu_\parallel$ transitions are marked with lines. 146

6-5 Transient 2D IR difference spectra (ZZZZ) after a T-jump from 63 to 72°C. Transient difference spectra are plotted as a function of delay τ. Twenty-one contours are plotted at 1.5% of the maximum of the reference spectrum at T_i = 63 °C. The spectra from 1 to 7 ms are obtained from the same data set as the 100-ns spectrum. The red ellipse in Upper Left indicates depletion on the red side of the diagonal region. 148

6-6 Semilog plot of transient changes in 2D IR spectra. (a) Temporal profile of unfolding and refolding of ubiquitin constructed from the first SVD component of the transient difference spectra shown in Fig. ??. (b) Transient changes of slices at $\omega_1 = 1642$ cm^{-1} for representative delays between $\tau = 100$ ns (blue solid line) and 7 ms (red dashed line). (c) Relaxation profiles from ω_3 slices in b at $\omega_3 = 1639$ cm^{-1} (light blue) and 1663 cm^{-1} (magenta). (d and e) Transient slices at $\omega_1 =$ 1620 cm^{-1} and the corresponding frequency shift of the positive peak ω_3*. (f) Relative changes in the antidiagonal width of the $\nu_{||}$ (light blue) and random coil (magenta) components marked with arrows in the Inset. The normalized transient temperature relaxation profile (black dashed line) is also shown. 150

6-7 Transient 2D IR spectra of ubiquitin (25→35°C). (Left) Reference 2D IR spectrum of ubiquitin at $T_i = 25$ ° C in the ZZZZ polarization geometry. Twenty-one contours are plotted in ±60% of the maximum. (Right) Transient 2D IR difference spectra [$\Delta S(t)$]. Twenty-one contours are plotted from ±1.5% of the maximum of the reference spectra. Depletion of the signal on the red side of the diagonal regions at 200 ns is marked with a red ellipse. The guidelines parallel to the two frequency axes (ω_1 and ω_3) mark the two resonances of the β-sheet in ubiquitin. 153

6-8 Comparison of DVE spectra measured and reconstructed from 2D IR spectra. (a) DVE difference spectra obtained from the T-jump at $T_i =$ 63 ° C. Each spectrum is divided by the maximum of the reference DVE spectrum. (b) DVE difference spectra constructed from the absolute value square of the projection of complex 2D IR spectra onto the ω_3 axis. (c) Comparison of relaxation profiles constructed from the first singular value decomposition component of the $\nu_\perp$ region (1577-1651 cm^{-1}). 154

6-9 Relative changes in the antidiagonal width for a T-jump from $T_i =$ 25 °C, plotted with a temperature relaxation profile normalized to the initial change (dashed line). 156

6-10 Thermal unfolding trajectory of ubiquitin. The unfolding simulation[11] is plotted along the coordinates R_1 and R_1 defined in the text. Protein snapshots correspond to five persistent structural regions (shown in green, red, blue, magenta, and cyan, respectively). The trajectory after a 17-ps window average is plotted in black. 159

6-11 Calculated transient 2D IR spectra of ubiquitin unfolding. Ubiquitin 2D IR spectra calculated at equilibrium (25 °C) (adapted from Ganim et al.[8]), for the transient states A, B, C, D, and E, and the difference spectra relative to A (B-A, C-A, D-A, and E-A). The trends in $\nu_\perp$ frequency and intensity and in ν_R intensity were obtained from $\omega_1 =$ 1640 cm^{-1} and $\omega_1 = 1662$ cm^{-1} slices, respectively. 160

6-12 Doorway modes of the transient ubiquitin unfolding states. Doorway modes (calculated as described previously[12]) for selected regions of the $\nu_\perp$ and $\nu_\parallel$ bands. 161

7-1 Examples of protein oligomer complexes displaying shape complementary and symmetry. Reproduced from Goodsell and Olson.[13] 168

7-2 Schematic of a thermodynamic cycle used for calculating free energy of association.[14] 171

7-3 Prediction accuracy results of the CAPRI challenge. 167

7-4 Example free energy surfaces for protein dimerization. Insulin monomer and dimer structures are shown as illustrative examples rather than constraints or specific predictions. 169

7-5 Structural visualizations of insulin dimer and monomer models. Strucutral models from PDB IDs 4INS (a,b), 2JV1(c), and 1JCO (d). Structures visualized using VMD[15] and POV-Ray. 173

8-1 Structural visualizations of insulin dimer and monomer models. Strucutral models from PDB IDs 4INS (a,b), 2JV1(c), and 1JCO (d). Structures visualized using VMD[15] and POV-Ray. 177

8-2 Concentration-dependent Insulin ZZYY 2D IR spectra and difference spectra. Concentration-dependent ZZYY 2D IR spectra and difference spectra acquired in 0.27 M DCl 20% (v/v) EtOD-D2O. Top: absorptive spectra; bottom: NRPS. All contours are plotted in 8.3% intervals. Slices along the diagonal ($\omega_1 = \omega_3$) are plotted above each spectrum, except for panel A and B, where the slices chosen to coincide with the peak maximum ($\omega_1 + \Delta = \omega_3$). 180

8-3 Measurement of the insulin dimer dissociation constant, K_D. Extraction of dimer fraction from FTIR and 2D IR rephasing power spectra (RPS). Representative RPS spectra, and the FTIR series and 2nd SVD component are shown on top. Concentration-normalized off-diagonal integration of 2D IR spectra and second component SVD amplitudes of FTIR spectra were fit to the dimer fraction to extract K_D (bottom). 182

8-4 Temperature-dependent Insulin ZZYY 2D IR Spectra. ZZYY spectra were acquired for 1.7 mM insulin concentration in 0.27 M DCl, 0.1 M NaCl 20% EtOD-D2O. All spectra show contours plotted in 8.3% intervals. 183

8-5 Temperature-dependent HDVE spectra were acquired as a function of ethanol co-solvent concentration. Spectra were acquired in 2.5 °C intervals from 10-55 °C. Representative spectra and 2nd SVD component are shown in panel A for 30% EtOD. Extracted melting temperatures and amplitudes of the second SVD are shown in B and C. 185

8-6 Extracted parameters for the two-state insulin dimer dissociation model. 186

8-7 Temperature-dependent HDVE series showing the aggregate peak used to exclude spectra from two-state fitting. 187

8-8 Solvent-dependent HDVE spectra. The HDVE spectra are shown for 25°C with reconstructed spectra using a two-state basis derived from 0.27 M DCl, D2O:20% EtOD + 100 mM NaCl. The fit residuals are indicative of experimental irreproducibility, such as shifts in the spectrum of the laser or variations in scattering. 189

8-9 Simulated 2D IR spectra for insulin monomers and dimers. Simulated 2D IR spectra for dimers (4INS), compact monomers (1JCO) and extended B-chain monomers (2JV1). The intensity at 1620 cm^{-1} includes an exaggerated contribution from proline. Structures visualized using VMD[15] and POV-Ray. 190

8-10 Bright state analysis and calculated FT IR spectra. Top: plot of the mode composition from bright state analysis decomposed by secondary structure, and the calculated IR absorption spectrum for the (a) dimer, (b) compact monomer, and (c) extended monomer ensembles. For the mode decomposition, the independent axis gives the beginning of the 5 cm^{-1} window. Below: Visualizations of the bright states for representative spectral regions, prepared using PyMol.(Delano Scientific) The color (red or blue) represents the vibrational phase of backbone amide I oscillators, while the intensity of color reflects the amplitude of vibration. For each bright state, the vibrational amplitudes were normalized to the large magnitude value. 198

8-11 Insulin melting and dissociation curves. Melting and dissociation curves in all tested solvents along with two-state model fits, shown in Figure 8-6. The total insulin concentration was 1.7 mM and all solvents included 0.27 M DCl. HDVE experiments were repeated 3-7 times in each solvent, and the sets of spectra chosen for analysis had the least aggregation. For solvents that caused persistent aggregatation, aggregate spectra (points indicated with an X) were excluded from the fit (see Figure 8-7). 205

9-1 *N*-hydroxysuccinimide ester Texas Red reacts with primary amines to form a very stable amide bond. 213

9-2 A digital camera photograph of the Texas Red-labeled insulin (dark, purple) and free dye bands (light, pink). 214

9-3 Diagram of the primary amines in the insulin dimer crystal structure,[16] with Texas Red conjugated to one of the N-termini, drawn to scale. . 215

9-4 Insulin, Free Dye, and Dye-Labeled Insulin Absorption and Fluorescence Spectra. Normalized absorption and fluorescence spectra for all NHS dyes used for labeling. Malachite green is a nonfluorescent dye. All solutes were optically dilute (OD < 0.5) in water. Fluorescence spectra were collected with excitation at 532 nm. 218

9-5 Characterizing aliquots from the column-separated TR-ins synthesis using UV and visible absorption. The absorption at 593 nm arises from Texas Red chromophores, which may be covalently bound to insulin. The absorption at 280 nm arises from insulin, Texas Red, and possibly also the NHS leaving group. 219

9-6 Concentration-dependent fluorescence decays of Texas Red-Labeled Insulin. 220

9-7 Extracting the Dimer Fraction from Fluorescence Decay Amplitudes. 222

9-8 Temperature-dependent fluorescence spectra of TR-insulin. Dimer dissociation is observed in acidic conditions as an increase in fluorescence due to the loss of quenching. At pH 8.1, where dissociation is strongly inhibited, a much smaller increase is seen. 224

9-9 Sensitivity to the dimer fraction as a function of concentration and labeling efficiency for $K_D = 100\ \mu M$. 226

9-10 Diagram of the six pheynalanine and eight tyrosine residues in the insulin dimer crystal structure.[16] 228

9-11 Absorption and emission spectra of phenylalanine and tyrosine. Adapted from Lakowicz et al.[17] 229

9-12 Experimental scheme for tyrosine fluorescence experiments. L1 (BK7, f=15 cm), BBO 1-3 (β-BaB_2O_4), WP (quartz waveplate 14 λ at 400 nm, 6.5 λ at 800 nm, 0.584 mm), M (silver mirror), CM (curved aluminum mirror, r=40 cm), PR (fused silica, 1 cm), PIN (Thor Labs DET210 High Speed Silicon Detector), L2 (fused silica f=7.5 cm), L3 (fused silica, f=7.5 cm), S (brass sample cell), FIL (UG 11 Schott Glass Filter), ASD (Electro-Optics Technology ET 2030A 1.2 GHz amplified silicon photodetector) 230

9-13 Insulin and tyrosine fluorescence. 232

9-14 Temperature-dependent intrinsic insulin fluorescence. 233

B-1 Texas Red-Insulin HPLC Log 244

B-2 UV-Vis assay of the LC/MS separation, and demarcation of the fractions. 245

B-3 Mass Spec for Fraction 8 246

B-4 Mass Spec for Fraction 8, expanded 247

B-5 Mass Spec for Fraction 9 248

B-6 Mass Spec for Fraction 10 249

B-7 Mass Spec for Fraction 13 250

B-8 Mass Spec for Fraction 14 251

B-9 Mass Spec for Fraction 15 252

B-10 Mass Spec for Fraction 19 253

List of Tables

2.1 Six level system parameters. Parameters for the six-level system are explained in Chapter 2, Section 2.2. The code to calculate the 2D IR spectra appears in Appendix A 57

2.2 Peak positions for the 1635 cm^{-1} fundamental mode in the ZZZZ spectra of Fig. 1-8. Spectra are defined in Equations 1.18. 61

3.1 Tensor components of the orientational response. 73

3.2 Numerical parameters for six-level system model of NMA-d7 in D_2O. Values in parentheses indicate those expected from harmonic scaling. 80

3.3 Amide I'-II' Modeling Parameters. 88

4.1 Site Energy Models and Fluctuation Statistics for Trpzip2. Listing of site energy models tested, comparison of electrostatic property correlated, site correlated, nomenclature, and fluctuation statistics for Trpzip2. $\langle\omega\rangle$ is the average site frequency, which is broken down to contributions from water, backbone, and sidechain respectively in $\langle\delta\omega_{Water}\rangle$, $\langle\delta\omega_{Backbone}\rangle$, and $\langle\delta\omega_{Sidechain}\rangle$. $\langle\sigma\rangle$ is the average site standard deviation. All quantities are in cm^{-1}. 115

8.1 Measures of Insulin Dimerization K_D 206

9.1 Availability of Primary Amines in Insulin as a Function of pH, and half lives for NHS-ester hydrolysis.[18-20]. N terminus pKa was taken to be 8.2 and Lys pKa was taken to be 10.5. 216

Chapter 1

Introduction

Proteins are the nanoscopic automatons that direct every process that we associate with life. Proteins control our growth and reproduction by carefully unraveling our genome, faithfully duplicating every base in the two meter length of our DNA, and then repackaging it.[1] Proteins break down our food into the tiniest usable blocks, extract its energy, and use the pieces to construct sophisticated devices like lungs and eyeballs and more proteins. Proteins drive the thunderstorm of electrical currents responsible for memory and thought, they patiently capture photons one at a time to allow for vision, and synthesize the hormones and other chemicals that guide our personalities; they synthesize the adrenaline, oxytocin, and seratonin that help us rise to action, bond with people, and relax. It's not much of an exaggeration to say that every biochemical bond in the body that is formed or broken has a protein's fingerprints on it.

Proteins perform a vast array of functions, and we know that each can be decomposed into a nested application of chemical forces acting on short distances and fast time scales. Thus, proteins act as gears in a transmission device that couples the world of pairwise molecular interactions acting on Ångstrom distances and femtosecond time scales into a universe of emergent synthetic and regulatory changes appearing on μm-m distances and ms-min time scales. What is the best way to understand the functions of proteins? Because atomistic structure elucidation methods such as x-ray crystallography and NMR have been the traditional workhorses for molecular biophysics, it is unsurprising that these methods have driven the major advances in our understanding of protein function. However, the increasing availability of atomistic models has not been accompanied by growing confidence in the ability to intuit biochemical mechanisms from the inspection of static protein structures. Universal application of the biological dogma that "form determines function" has not been feasible because protein functions often arise as a result of a competition between weak, non-covalent interactions that are difficult to predict based on the structure and are not amenable to localized descriptions.

---

[1] DNA fully stretched is $\approx 10^7$ the length of DNA polymerase. A similar scaling of length scales is the ratio of the height of the average human to the diameter of the earth.

Every functional motion of a protein is inextricably linked to conformational dynamics. Thus, an alternative route towards understanding protein function is to complement structural studies with dynamics experiments aimed at elucidating a time-resolved view of the order of events. The ultimate goal of these dynamics experiments is to characterize the free energy surface, which is a model that encapsulates dynamical information, thermodynamics, and probable reaction pathways. In this thesis, there is an emphasis on protein folding because the associated molecular forces and search problems are ubiquitous elements; the same physical principles underlie phenomena such as allostery, cooperativity, fibrilization, and assembly. The tools required to probe these events require considerations of disorder, collective coordinates, and multidimensional reaction surfaces. One outstanding problem that highlights many of the challenges in understanding protein functions is the biophysics of protein-protein interactions. For flexible proteins that undergo a conformational change upon binding, this process is intimately related to the folding problem. A wide variety of length and time scales are juxtaposed in the process of diffusing several intermolecular lengths, recognizing a binding partner, and refolding. Chapter 7 provides a detailed introduction to the questions that arise in protein-protein binding and the framework for designing dynamics experiments aimed at protein association.

1.1 Dynamics Experiments

Ultimately, this thesis is concerned with developing *dynamics experiments* that can be used to study the biophysics of protein folding and binding. Most of our insight into protein folding and binding is indirectly obtained through *kinetics* experiments that measure reaction rates and reveal how fast populations of stable states interconvert. In contrast, a *dynamics* experiment reveals evolution along a reaction coordinate; this type of experiment is challenging because most biophysical processes are encumbered by activation barriers that obscure the fleeting motion along reaction coordinates relative to the many failed attempts to cross reaction barriers (non-reactive dynamics). Furthermore, in equilibrium bulk experiments these brief barrier-crossing events are unsynchronized, and cannot be observed in a time-dependent fashion.

These considerations detail the needs for a dynamics experiment, which are illustrated using the reaction shown in Figure 1-1a. Figure 1-1b and c show two associated free energy surface with obligatory, two-dimensional reaction coordinates. While no one-dimensional reaction coordinate fully describes the pathway from reactant to product, the spectroscopic coordinate in Fig. 1-1b allows for a continuous description of the changes from reactant, through the transition state, to the product; this is not so in Fig. 1-1c because the reactant well overshadows the product well when only the spectroscopic coordinate is observed.

Figure 1-1d shows a schematic of the spectral changes that would be observed if a barrier-free reaction was initiated on the surface in Fig. 1-1b and probed with sufficiently fast time-resolution. Because no barriers are encountered to desynchronize the reaction, the barrier-free timescale, A^{-1} is revealed. Further, because the spectroscopic coordinate projects favorably upon the pathway from reactant to product, the time-dependent frequency characterizes the structural changes. If the spectroscopic coordinate does not distinguish continuous changes along the reaction pathway, or if thermally activated barrier crossing is initiated a kinetics experiment such as shown in Fig. 1-1e is still possible that measures the overall conversation rate, k_f^{-1} without structural information.

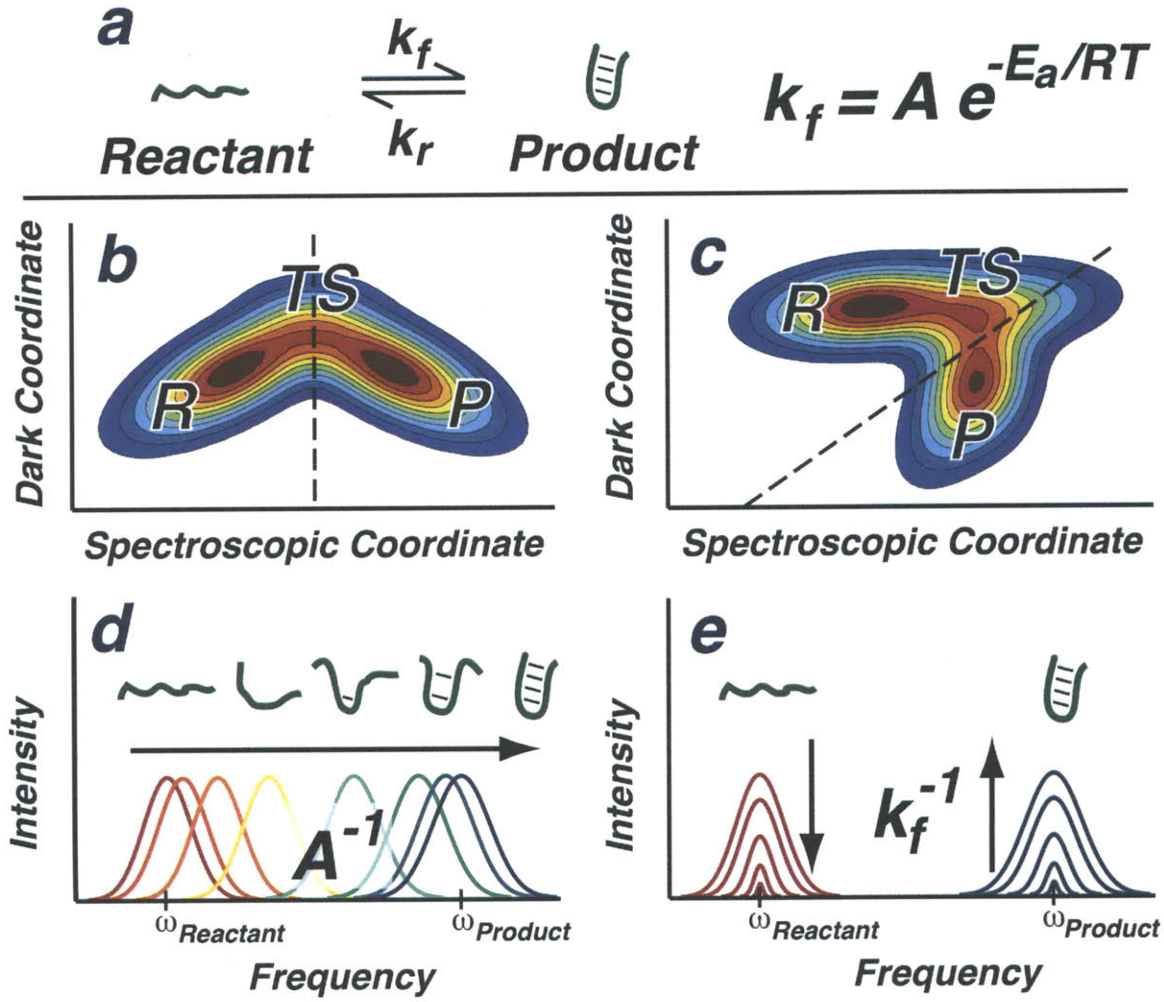

a
Reactant
k_f
k_r
Product
$k_f = A e^{-E_a/RT}$
b
TS
R
P
Dark Coordinate
Spectroscopic Coordinate
c
R
TS
P
Dark Coordinate
Spectroscopic Coordinate
d
Intensity
A^{-1}
$\omega_{Reactant}$
$\omega_{Product}$
Frequency
e
Intensity
k_f^{-1}
$\omega_{Reactant}$
$\omega_{Product}$
Frequency

Figure 1-1: Illustration of considerations for a dynamics experiment. (a) A folding reaction with two structurally distinct equilibrium states that is characterized by the rates k_f and k_r. (b) A spectroscopic coordinate that distinguishes between the reactant (R), product (P), and transition state (TS) with a dividing line on the free energy surface. (c) A free energy surface where the product well and transition state are not separable from the reactant without inclusion of an unobservable, dark coordinate. (d) A schematic of a *dynamics experiment* that directly reveals time-resolved structural changes and the barrier-free rate. (c) Time-resolved populations observed in a *kinetics experiment* that measure the overall rate, k_f, and require additional assumptions to reveal dynamical information.

The requirements of structural sensitivity and time-resolution can come into conflict; since spectroscopy typically infers structure using energy gaps, temporal and spatial resolution are effectively conjugate variables, and a compromise between both is required. Without structural resolution, product and reactant are indistinct. As temporal resolution is reduced, only the average structure is characterized.

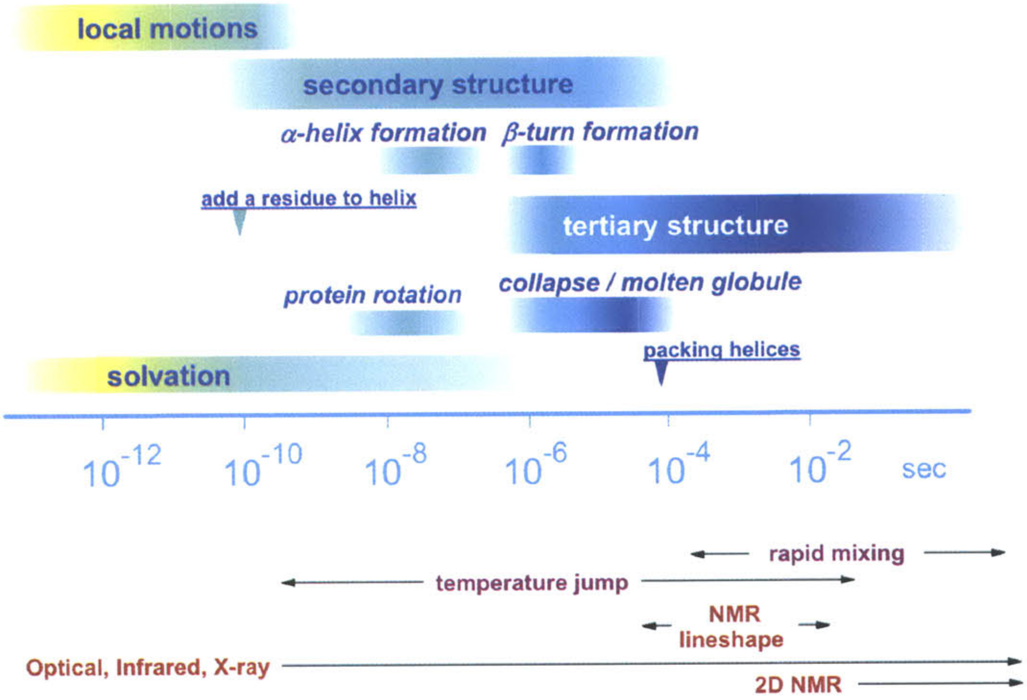

local motions
secondary structure
α-helix formation
β-turn formation
add a residue to helix
tertiary structure
protein rotation
collapse / molten globule
packing helices
solvation
10^{-12}
10^{-10}
10^{-8}
10^{-6}
10^{-4}
10^{-2}
sec
rapid mixing
temperature jump
NMR lineshape
Optical, Infrared, X-ray
2D NMR

Figure 1-2: Overview of time scales for protein dynamics and kinetics and directly accessible time scales by experiments.

In summary, there are three requirements for a successful dynamics experiment of the type described above:

1. **Temporal Resolution** The time resolution of the probe must be faster than the dynamics of interest. Figure 1-2 shows the time scales for a variety of biomolecular motions. There is an approximate inverse relation between the length-scale for the structural rearrangement and its rate; the time scales range from fast motions correlated only among a handful of atoms to slow motions for global structural changes; biophysically relevant conformational changes include short-range fluctuations of protein side-chains, torsions, and hydrogen bonds ($\sim 10^{-13} - 10^{-11}$ s); protein reorientation, chain diffusion, nucleation and folding

of secondary structure ($\sim 10^{-9} - 10^{-6}$ s); domain folding and tertiary contact formation ($\sim 10^{-6}$ s); and folding, binding or aggregation kinetics through activated barrier crossing ($> 10^{-3}$ s). From investigations for a protein folding "speed limit," there is evidence that a time resolution of 1 μs or shorter is required, which is set by the diffusion-limited passage time across protein folding transition states.[21–25]

2. **Structural Resolution** Protein folding and binding dynamics can occur over several disparate coordinates including solvation, reorientation, secondary structure formation (α helices, β turns and sheets), hydrophobic association of side chains, and spatial diffusion. Each coordinate probed in a dynamics experiment reveals a projection of the reaction coordinate, and while a full characterization requires a complement of tools, consideration must be given to the tradeoff between spatial and temporal resolution. Multidimensional NMR provides atomic level structures, but its utility for following folding is restricted to long times.[26,27] As the motionally narrowed limit is approached, the ability to extract fast kinetics becomes dependent upon modeling assumptions. Time-resolved x-ray diffraction probes atomic-level structural changes in crystals,[28,29] and has been applied to proteins in solution with only coarse structural information.[30,31] IR absorption,[32,33] Raman,[34,35] and UV circular dichroism[36,37] have been used with success to reveal the rates of folding and unfolding processes on picosecond to millisecond time scales, but these methods are challenged by ambiguous relations between observables and folding coordinates. Single-molecule methods offer unparalleled sensitivity to heterogeneity, but have historically been limited to $\approx$ms time resolution and coarse structural information from fluorescence.[38,39] While the time resolution of single-molecule experiments is improving,[39] poor structural resolution and difficulty sampling rare events continue to limit their dynamical information content.

3. **Rapid Reaction Initiation** Because the flux of proteins undergoing folding and binding barrier crossing events is typically unobservable in equilib-

rium bulk experiments, rapid initiation techniques are required to observe the fastest biomolecular motions. Microfluidic rapid mixing devices offer a versatility of possible denaturants, but are only capable of 10 μs time resolution.[40,41] The fastest, generally applicable technique has been unfolding with a laser-excited temperature jump. In these experiments, the solvent is rapidly heated via absorption of an optical dye[42] or by direct excitation of water in the infrared. For the latter, nanosecond temperature jumps ($\approx$10-20 K) are routinely reported[25,43], and changes as fast as 3-100 ps have also been observed.[44,45] Shorter time scales are yet possible, although non-thermal distributions[46] are created once the pulse length is shorter than the $\approx$10 ps thermalization timescale of water. Sufficient time-resolution to observe protein folding speed limits has also been reported using a 0.7 μs pressure jump.[47]

1.2 Thesis Outline

Two-dimensional infrared spectroscopy (2D IR) is the central tool developed in this thesis to meet the challenges addressed above for protein dynamics experiments due to its combination of time and structural resolution. The year this research began, a feature article was published by our group highlighting the ability of 2D IR to characterize the potential energy surface and transition dipoles for a system of two coupled oscillators, $Rh(CO)_2C_5H_7O_2$.[5] This thesis chronicles a transition in the use of 2D IR from studying few-atom systems to dynamics experiments on proteins, macromolecules with so many degrees of freedom that vibrational spectroscopy is often considered ambiguous.

The experimental and calculation tools required to probe protein conformational dynamics were formulated to treat the folding of trpzip2 and ubiquitin, and are described in Chapters 2-6. The technical challenges associated with performing third-order infrared spectroscopy on proteins are discussed in Chapter 2, and system-specific details appear in Chapters 5, 6, and 8. Chapter 5 examines structural heterogeneity in the β-hairpin trpzip2. These experiments and modeling provided evidence for two turn geometries, which argues against the consensus that the native state of trpzip2 consists of one well-ordered structure. In Chapter 6, transient temperature-jump 2D IR was used as a probe of the unfolding dynamics of ubiquitin. Time-resolved structural changes manifested as a frequency blue-shift were compared against transient 2D IR calculations to interpret the unfolding mechanism as a sequential loss of strands in the β-sheet.

One general problem was repeatedly addressed in this thesis: how can 2D IR spectra best be modeled to reveal atomistic structural information? The key feature that now sets 2D IR apart from other fast protein probes is that the data can readily be calculated from an atomistic structure or molecular dynamics simulation using the methods developed in this thesis work. The theoretical background for these calculations is described in Chapter 3, and applications are presented for the amide I-II spectroscopy of NMA (Chapter 3); the amide I'-II' spectroscopy of poly-L-lysine (Chapter 3); the D-Arg peptide (Chapter 4), trpzip2 (Chapter 4 and 5); ubiquitin (Chapter 4 and 6); and insulin monomers and dimers (Chapter 8). These applications demonstrate the entire range of currently available theoretical models for 2D IR spectra ranging from phenomenological (Chapter 3) descriptions; models suited for idealized, periodic systems (Chapter 3); calculations that reach the limit of currently available numerical accuracy (Chapter 5); the calculation of transient 2D IR spectra (Chapter 6); and approximations for calculating the 2D IR spectra of larger protein systems (Chapter 8). In addition to calculating 2D IR spectra, these simulations can be used to examine the vibrational mode composition (Chapter 8), directly learn about solvation (Chapter 4 and 5), and characterize unfolding mechanisms (Chapter 6). The emerging paradigm is to interpret 2D IR spectra with the aid of an atomistic, molecular dynamics simulation.

Chapters 7-9 describe the application of these tools to study protein binding. The monomer-dimer transition of insulin was first considered as a model system for the biophysics of coupled folding and binding in this work. It was chosen for our first protein dissociation dynamics experiments because it displays the coupling between binding and folding that has been described as the major challenge for models of protein-protein interactions. Moreover, it is a small (51 amino acid) protein that diffuses across average intermolecular separations during time scales that are relevant to structural changes, and is amenable to molecular dynamics simulations. While 2D IR spectroscopy has undergone rapid maturing, there are no formulaic tools for interpreting the data, and each experiment inspires new ways to visualize and model the data. Chapter 7 presents an introduction of the biophysics of protein binding and folding, a critical review of the currently used tools, and a framework for designing dynamics experiments aimed at protein association. Chapter 8 describes the monomer-dimer equilibrium of insulin, and how it can be tuned with by the protein concentration, temperature, and solvent studied using 2D IR, related nonlinear spectroscopies, and molecular dynamics simulations. As a complement to the structural information provided by 2D IR, Chapter 9 describes dye-labeling and intrinsic tyrosine fluorescence experiments on insulin that provide a more coarse-grained sensitivity to the monomer-dimer equilibrium.

1.3 Introduction to Two-Dimensional Infrared Spectroscopy

Two-dimensional infrared spectroscopy is one of a rapidly expanding class of new ultrafast coherent vibrational spectroscopies[48–55] that are finding broad use in studies of molecular structure and dynamics that probe peptides,[56] proteins,[1–3,57] DNA,[58] chemical exchange kinetics,[59,60] hydrogen bonding,[61,62] and rapidly initiated chemical reactions.[63–65] Inspired by pulsed NMR techniques, 2D IR spreads a vibrational spectrum over two frequency axes to reveal vibrational couplings through cross peaks. A 2D IR spectrum correlates the frequency of vibrational excitation ω_1 (referred to by others as ω_τ or ω_{pump}) with the frequency of detection ω_3 (or ω_t or ω_{probe}). Diagonal peaks can be assigned to chemically distinct normal vibrational modes or eigenstates. The presence and splitting of cross peaks characterizes the anharmonic couplings between vibrations. Connectivity, distance, or orientation between chemical bonds can be extracted by modeling vibrational couplings.[5] Positive and negative amplitude features in 2D IR spectra, corresponding to induced absorption or stimulated emission processes in the detection step, characterize the vibrational anharmonicity. Since the measurement is made with a picosecond or faster shutter speed, it captures this structural information on a faster time scale than the evolution of most protein dynamics. 2D IR diagonal and anti-diagonal line widths report on inhomogeneous and homogeneous broadening, respectively, and can be analyzed in more detail to describe variance in structural parameters.[66]

1.3.1 Challenges in Interpreting Protein Vibrational Spectra

The appeal of using vibrational spectroscopy to understand chemical reaction dynamics arises from its generality, its strong connection to the internuclear geometries and forces that drive reactions, and its potential for femtosecond time resolution. In a semi-classical picture for any spectroscopy, the resonance coupled to the light field must oscillate enough times to give the desired frequency resolution; because vibrational oscillation periods are $\approx$ 1 - 50 fs (4000 - 100 cm^{-1}), they can be used to resolve subpicosecond frequency shifts. However, the exquisite sensitivities that IR spectroscopy offers are inextricably linked to line broadening process that obscure spectral information; in the condensed phase, the generality of vibrational chromophores implies that they will often be present at high concentrations, and plenty of relaxation pathways will be available. Generally speaking, the more sensitive an oscillator is to its environment, the faster it will dephase, and the broader its line shape will be. Protein vibrations arising from the amide backbone are known to be particularly sensitive to secondary structure, often display broad, featureless bands, and arise from a large density of states- a protein of N amino acids will have $(N-1)$ amide I eigenstates, all within 1600-1700 cm^{-1}.

1.3.2 Amide I Vibrations

Amide group vibrations of the backbone receive the most attention in protein IR spectroscopy because they are native to all proteins and report on secondary conformation and solvation. These include amide I (primarily CO stretch), amide II (CN stretch and NH in-plane bend), amide III (CN stretch and NH bend and CO in-plane bend), and amide A (NH stretch). The amide I band (1600-1700 cm^{-1}) is by far the most studied because its line shape is sensitive to the type and amount of secondary structures and is not strongly influenced by sidechains.[67] Well-established empirical structure-frequency correlations, summarized in Fig. 1-3, find that β sheets have a strong absorption band near 1630-1640 cm^{-1} and a weaker band at high frequencies (>1680 cm^{-1}). The peaks for α helices and random coils are located at 1640-1660 and 1640-1650 cm^{-1}, respectively.

The secondary structure sensitivity of amide I results from coupling between amide I oscillators that leads to vibrational states delocalized over large regions of the protein.[12,68,69] To illustrate, the IR-active amide I vibrational modes of ideal antiparallel (AP) β sheets and α helices are visualized in Figure 1-3. The color-coded diagrams indicate the vibrational amplitude and phase of the individual peptide oscillators within the normal modes of these secondary structures. Antiparallel β sheets are predicted to have two dominant IR active modes. For the intense lower frequency $\nu_\perp$ mode, oscillators are in-phase perpendicular to the β strands and out-of-phase with their bonded neighbors. These relative phases are flipped for the $\nu_{||}$ mode. α helices have two bright states corresponding to the modes shown in Fig. 1-3. Since the carbonyls are aligned with the helix axis, most of the oscillator strength is carried by the ν_A mode with all oscillators in-phase and less by the degenerate ν_{E1} modes whose phase varies with a period of 3.6 residues.

1.3.3 Amide I 2D IR of Secondary Structure

Two-dimensional infrared spectroscopy provides an extra level of discrimination by spreading congested FTIR spectra over an additional dimension to correlate different

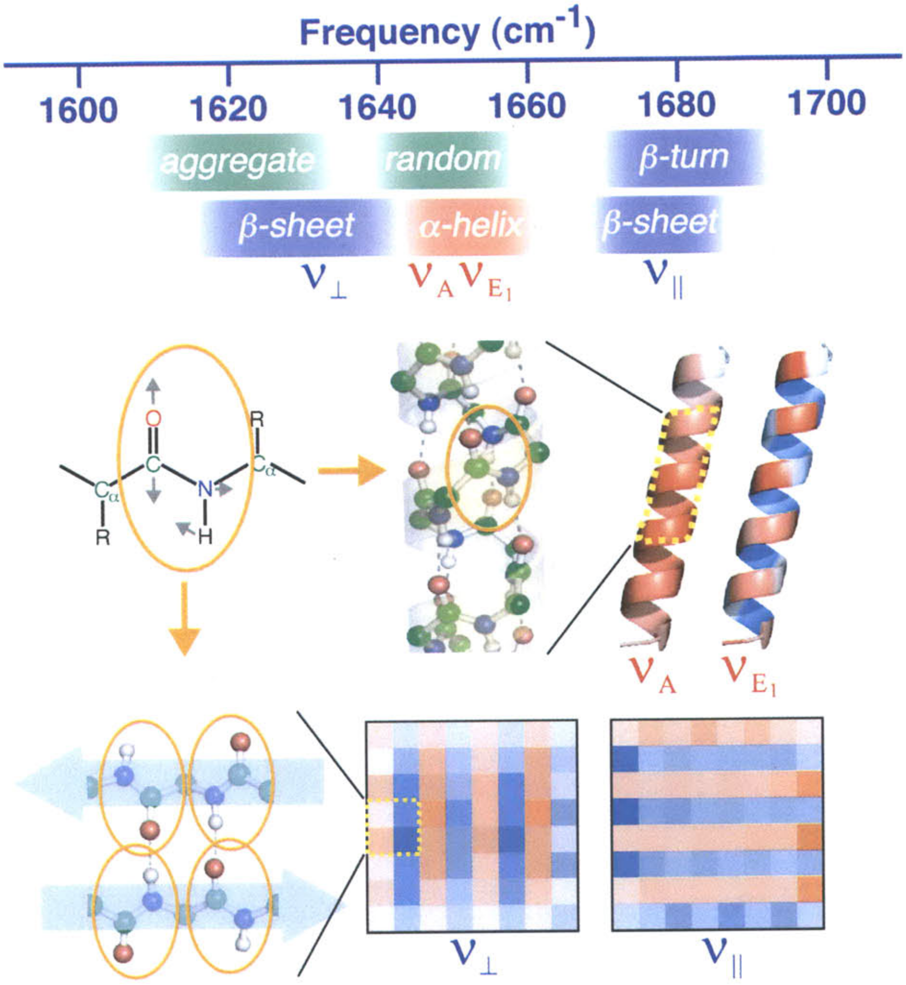

Frequency (cm^{-1})
1600
1620
1640
1660
1680
1700
aggregate
random
β-turn
β-sheet
α-helix
β-sheet
$\nu_\perp$
ν_A ν_{E_1}
$\nu_\parallel$
O
C
N
H
R
Cα
R
Cα
ν_A
ν_{E_1}
$\nu_\perp$
$\nu_\parallel$

Figure 1-3: The Amide I Mode in Proteins. (Top) Empirical protein structure-frequency relationships in the amide I region. (Bottom) A color-coded visualization of the IR active β-sheet and α-helix states where the shading intensity is proportional to the participation ratio of a unit oscillator and the color denotes its phase; blue and red are 180° out-of-phase. Figure excerpted from Ganim et al.[1]

spectral features. To demonstrate, FTIR and 2D IR spectra for the three most common secondary structure motifs are shown in Fig. 1-4. The AP β-sheet, α-helix and random coil states of poly-L-lysine have served as basis spectra for circular dichroism and exhibit FTIR peaks with the canonical secondary structure assignments.

For AP β sheets, the amide I FTIR spectrum shows the strong $\nu_\perp$ mode at ~1620 cm^{-1} and a weaker intensity $\nu_\parallel$ mode at ~1680 cm^{-1}. The corresponding 2D IR spectrum has $\nu_\perp$ and $\nu_\parallel$ peaks along the diagonal and two off-diagonal cross peaks.

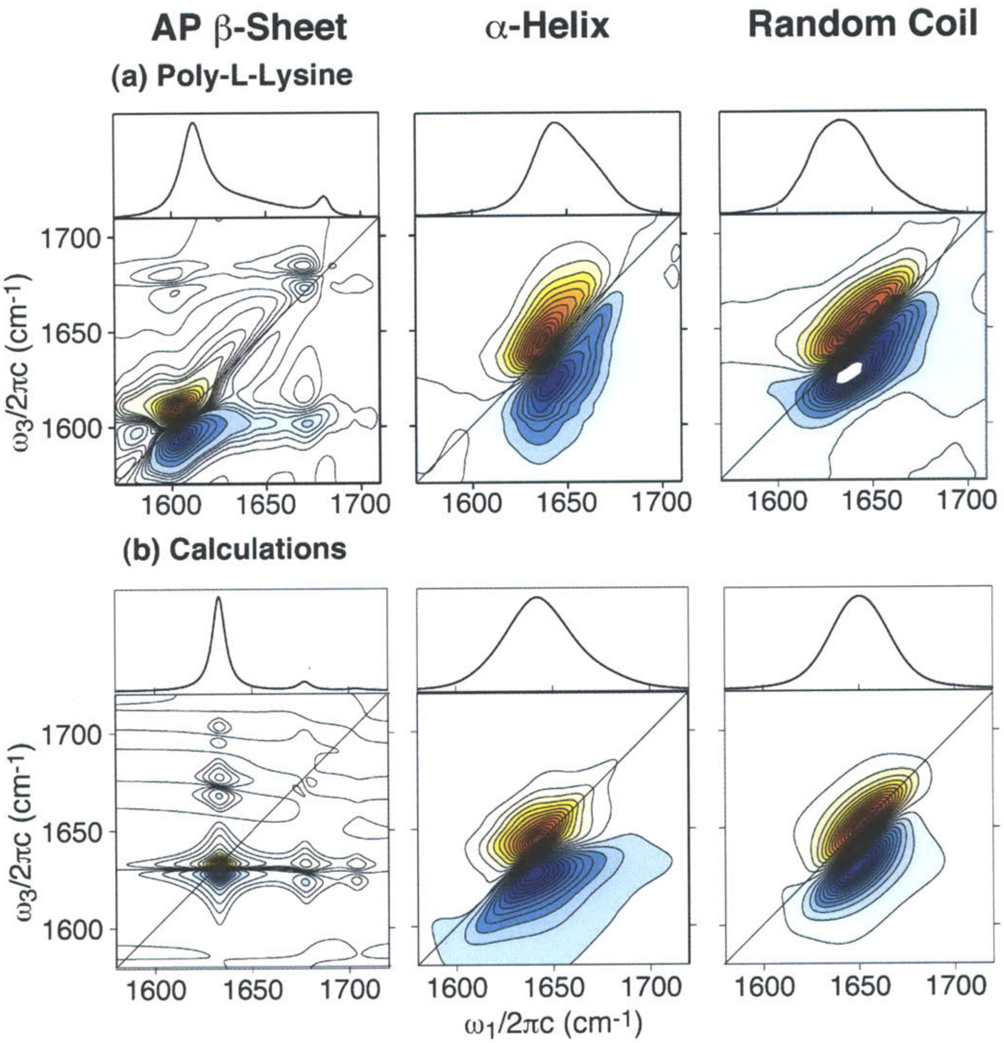


Figure 1-4: Amide I FTIR and 2D IR spectra demonstrating secondary structure sensitivity. (Top) FTIR and 2D IR spectra of the three forms of poly-L-lysine compared to calculated FTIR and 2D IR spectra of idealized structures (Bottom). Gaussian random site energies are sampled around a mean of 1650 cm^{-1} for the α helix, and random coil (σ = 10 cm^{-1} and 12 cm^{-1}) with a homogeneous line width of γ=12 cm^{-1} for all. Figure adapted from Ganim et al.[1]

The frequency and intensity of $\nu_\perp$ is sensitive to the size of the β sheet,[70] which causes the $\nu_\perp/\nu_\parallel$ splitting to increase as the sheet grows. Due to vibrational anharmonicity, induced absorption (positive peaks) and stimulated emission (negative peaks) appear displaced along ω_1, which causes 2D IR spectra to be asymmetric about the diagonal.

In FTIR spectra, both α helices and random coil regions appear as a single peak. For α helices, this peak is at ~1650 cm^{-1}. The corresponding 2D IR spectra show a flattened figure-8 line shape due to their composition of a ν_A mode at ~1639 cm^{-1} and doubly degenerate ν_{E1} modes at ~1652 cm^{-1} that are typically irresolvable with natural line widths of ~15 cm^{-1}. Extended 3_{10} helices are predicted to have similar ν_A and ν_{E1} modes; however this assignment is not straightforward in realistic 3_{10} helices that typically appear as short segments capping the ends of α helices.[71]

Unstructured proteins and random coils display a symmetric peak in the FTIR spectrum at ~1640 cm^{-1} that overlaps with the α-helical region. The 2D IR signature of this disorder is diagonal elongation caused by the stochastic variation in structure and hydrogen-bonding environments. The diagonal line width reports on inhomogeneous broadening, whereas the anti-diagonal gives the homogeneous line width. Disordered regions in the experimental AP β sheet yield a random coil feature in between the $\nu_\perp$ and $\nu_\parallel$ bands.

1.3.4 Protein 2D IR Spectra

Features from the amide I 2D IR spectra of idealized β sheets, α helices, and random coils presented in the previous section can be seen in the variety of protein spectra in Fig. 1-5. The two extended β sheets in concanavalin A (con A) give rise to a large ~40 cm^{-1} splitting between $\nu_\perp$ and $\nu_\parallel$ and a well-defined cross peak in the top left quadrant that is similar to those of the β-sheet form of poly-L-lysine (Fig. 1-4a). An overall Z-shaped contour profile appears for proteins with AP β sheets due to diagonal elongation of $\nu_\perp$ with disorder, diagonal peaks at ~1650 cm^{-1} and constructive interference with cross peaks in the ω_1 dimension.

With decreasing β-sheet content, the splitting between $\nu_\parallel$ and $\nu_\perp$ decreases and the $\nu_\perp/\nu_\parallel$ cross-peak becomes a ridge extending from the $\nu_\parallel$ band. This trend in splitting continues with β-lactoglobulin and RNase A, which retain AP β-sheet features despite twisted sheets. β-lactoglobulin, which contains an eight-stranded β barrel, shows distinct bifurcation of the $\nu_\perp$ band. The spectrum for RNase A shows an α-helix peak along the diagonal at 1650 cm^{-1}. The β hairpin trpzip2 has a clear $\nu_\parallel$ peak and a strong ridge indicative of its stability.[4] Ubiquitin shows the same Z-shape, although with smaller splitting, even though its five-stranded sheet contains a mixture of AP

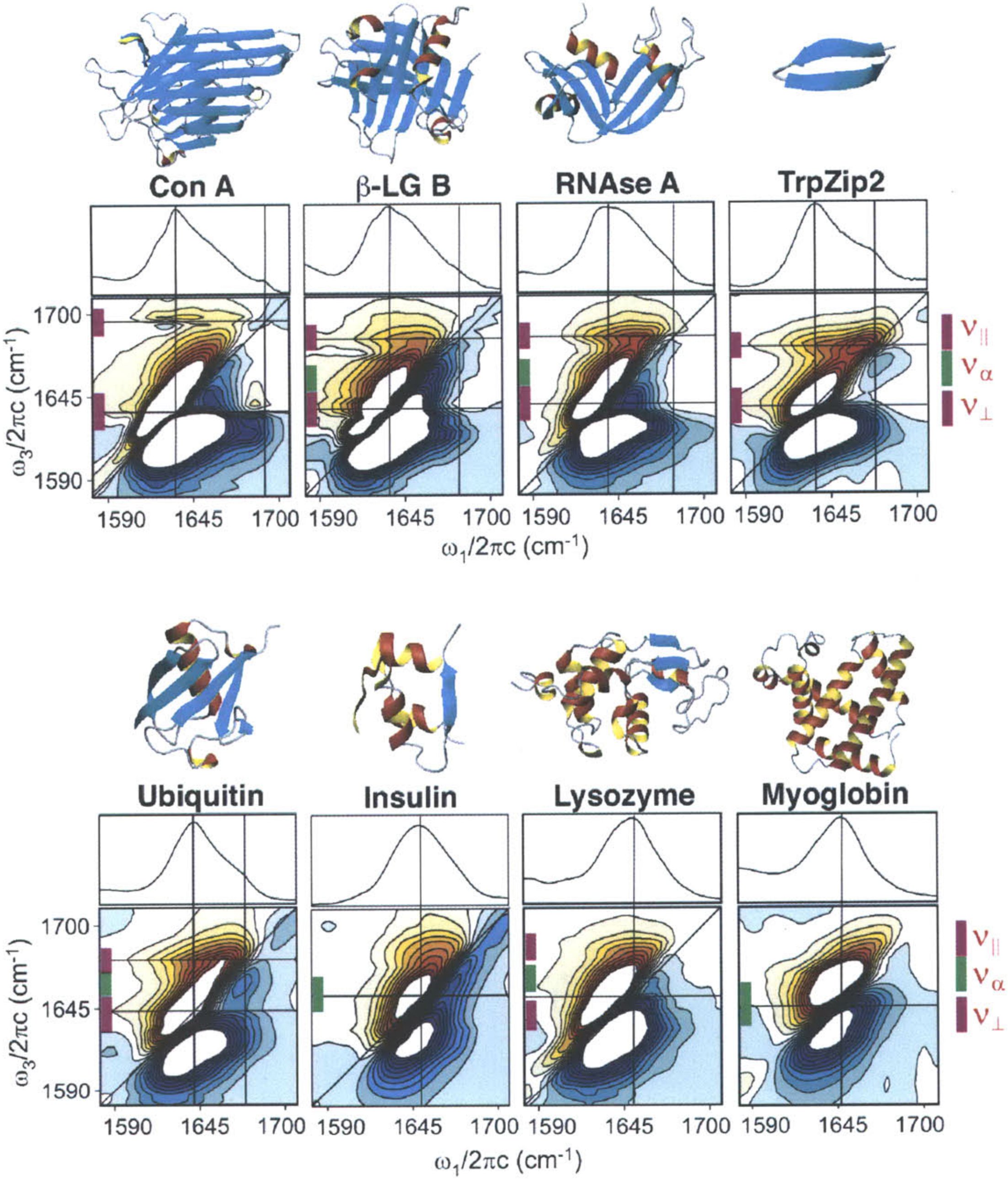


Figure 1-5: FTIR and 2D IR spectra for proteins with a varying amounts of β sheet and α helix secondary structure. The data was compiled from Demirdöven et al.,[2] Chung et al.,[3] Smith et al.,[4] and Ganim et al.[1]

and parallel contacts. Lysozyme, insulin, and myoglobin display nearly identical FTIR spectra, but a comparison of their 2D IR spectra shows subtle plateaus for lysozyme both above and below the diagonal that originate in its small β sheet.

These qualitative interpretations of amide I 2D IR correlation spectra can only provide limited information. Different representations of the 2D data and 1D representations, such as the heterodyne-detected dispersed vibrational echo (HDVE) are presented in Chapter 2 along with a discussion of general features of 2D IR spectra. Chapter 3 demonstrates how integrating cross-peak regions can be used to compare a set of spectra. Chapter 6 shows a set of metrics for analyzing transient 2D IR spectra including changes in intensity, frequency, and antidiagonal line width. Finally, Chapter 8 presents a series of analysis techniques and demonstrates how 2D IR and HDVE can be used to quantify binding equilibrium constants and thermodynamic parameters.

1.4 Afterword

Sections of text in this chapter were excerpted from "Amide I Two-Dimensional Infrared Spectroscopy of Proteins," by Ganim et al.[1]

Chapter 2

Third-Order Infrared Experimental Methods

Because two-dimensional infrared spectra provide a complete characterization of the third-order response, 2D IR spectroscopy is becoming an increasingly popular technique for problems previously studied using IR absorption, pump-probe, photon echo, and transient grating spectroscopies. By correlating the detection frequency with the excitation frequency, 2D IR spreads vibrational spectra over two frequency axes. Viewing the third-order response as a function of these two frequency dimensions is the best way to extract vibrational couplings through cross peaks. The methods used to acquire, calibrate, and phase 2D IR spectra in this thesis have been described elsewhere,[5,63,72] but they will be reviewed here. Different components of the third-order response function will be shown to selectively accentuate diagonal and cross peaks. One dimensional projections and slices of the 2D data will be presented that facilitate comparison of a set of 2D IR spectra or to monitor selected spectral features as a function of another external parameter.

2.1 Implementations of 2D IR Spectroscopy

Two-dimensional infrared spectroscopy (2D IR) has historically been performed two different ways, which reflect two different conceptual pictures on how the signal is created. In both cases, the goal of the experiment is to acquire a correlation spectrum that reflects the joint probability distribution of an oscillator being excited at a frequency ω_1 and probed at a frequency ω_3 after a waiting time, τ_2. In its first manifestation[73] ("the double resonance method"), 2D IR was performed as a modified pump-probe experiment; frequency resolution in ω_1 was achieved by filtering the pump to a tunable $\approx$10 cm^{-1} slice, which was followed by a broadband probe ($\approx$1575 - 1725 cm^{-1}) that was measured using an IR spectrometer for ω_3 resolution. This can be contrasted to the typical, broadband pump-probe experiment, which places no such filter in the ω_1 dimension and effectively integrates over all ω_1 frequencies. The first 2D IR spectrum was composed of a set narrow-band pump, broadband probe spectra.[73]

The 2D IR spectra in this thesis were acquired using the time-domain analog, Fourier transform 2D IR. The approach has also been called the "heterodyned photon echo" method, which reveals its historical development. In a three-pulse IR echo experiment, a signal can be emitted in a unique direction relative to the input pulses that reflects a rephasing process. The principles of an echo experiment can be understood by considering a simplified version of the third-order response for a two-level system (g and e),

$$R_{-}^{(3)}(\tau_1, \tau_2, \tau_3) \propto \left\langle exp\left[-i\int_0^{\tau_1} dt\omega_{eg}(t) + i\int_{\tau_1+\tau_2}^{\tau_1+\tau_2+\tau_3} dt'\omega_{eg}(t')\right]\right\rangle. \tag{2.1}$$

This expression represents an experiment where phase matching was used to select only the $R_{-}^{(3)}$ signal where the phase evolution during τ_1 is conjugate to τ_3. In this case, a recurrence in the signal is produced when the conjugate phases cancel and the argument of the exponential is nearly zero; this recurrence is called an echo because it reproduces the dipole and synchronized oscillation induced at the beginning of the experiment, when $\tau_1 = \tau_2 = \tau_3 = 0$, and the argument of the exponential is also zero.

In such a photon echo experiment, the time at which this recurrence occurs, and how that changes when additional delay (τ_2) is introduced between the τ_1 and τ_3 coherence periods can be analyzed to reveal timescales for loss of phase memory.[74–78]

However, these are time-domain experiments for which standard methods of interpretation can only function in the case of dilute, isolated chromophores.[79] As the upcoming sections will detail, 2D IR spectroscopy is a more general form of this experiment, which can be performed by interferometry with an additional, local oscillator (LO) field to provide phase resolution in the τ_3 dimension. Both the echo (now called the *rephasing* signal) and its phase matching complement, the *non-rephasing* signal are summed, which is critical for obtaining absorptive, 2D IR lineshapes. Cervetto et al. have compared the 2D IR spectra resulting from double resonance and heterodyned photon echo experiments,[80] and concluded that while the same peaks arise in both experiments, the former causes distortions in the ω_1 dimension of the line shapes. Recently, a new method for acquiring 2D IR spectra using a collinear pulse pair for the first two interactions was reported.[81] This method combines some of the simplicity of the double resonance method because a boxcar geometry is not used, and retains many of the advantages of the pulsed Fourier transform method.

The challenges for acquiring 2D IR spectra in the time domain include generating short pulses in the mid-IR; introducing the pulses to the sample with correct focusing, timing, and polarization; spatially isolating the signal; overlapping the signal with local oscillator; and aligning the signal and local oscillator through the polychromator onto the two-stripe array detector.

2.2 Acquiring 2D IR Spectra

2.2.1 Generating 6 μm Light

The front end of our 2D IR spectrometer consists of a Spectra-Physics Tsunami Ti:sapphire oscillator (791.5 nm, fwhm=16 nm, 84 MHz, 600 mW) pumped by a Spectra-Physics Nd:YAG Millenia (532 nm, CW, 6 W). The Tsunami provides seed pulses for amplification in a Spectra-Physics Spitfire amplifier, which was pumped by a Spectra-Physics Nd:YLF Evolution or Empower (532 nm, 10W). The Spitfire regularly produced a 1 kHz train of 100 fs pulses with 1 mJ/pulse.

Infrared pulses centered at 1660 cm^{-1} ($\lambda = 6.02 \mu m$) were produced by pumping a Spectra-Physics OPA-800 optical parametric amplifier with 720 mW of the amplifier output. In the first stage, a 1% reflection was used to create a white light seed in sapphire for the parametric conversion of 20% of the 800 nm light into 1.84 μm (signal) and 1.4 μm (idler) by a type I process in 1 mm β barium borate (BBO). A dichroic mirror was used to isolate the idler for an efficient seed to the second pass conversion (remaining 79%) in the same BBO. The signal and idler were focussed into 1 mm $AgGaS_2$ and the 6 μm light was isolated from the signal and idler using a 3 mm Ge plate oriented at Brewster's angle.

After generation, the infrared pulses were collimated to 1 cm (diameter of 80% intensity). They were spatially overlapped and mode matched with a He:Ne laser for ease of further alignment. Collimation was confirmed using a home-made beam profiler consisting of a razor blade mounted to a micrometer; the intensity profile was compared to an integral of a two-dimensional Gaussian to determine the radius, which was measured at two points separated by 5 m.

2.2.2 2D IR Spectrometer

A five-beam interferometer (Figure 2-1) was used to split the pulses into five replicas, and control timing and polarization.Three pulse (a, b, and c in Figure 2-1) were aligned into a 1" boxcar geometry, and used to generate the third-order signal by focusing into the sample using 3" off-axis gold parabolic reflectors. A fourth pulse (d in Figure 2-1) propagates along the signal direction, which was used for coarse signal alignment and for two-beam pump-probe experiments. The fifth pulse (e in Figure 2-1) forms the local oscillator (LO), which may be chosen to pass through the sample or propagate externally before being combined with the signal to allow heterodyne detection.

In the interferometer, the polarization of the beams was controlled using 2 mm thick MgF_2 have-wave plates and 2mm thick ZnSe wire-grid polarizers, and calibrated by monitoring the transmission through another polarizer placed at the sample point to account for any polarization shift due to focusing by the off-axis parabolic. The

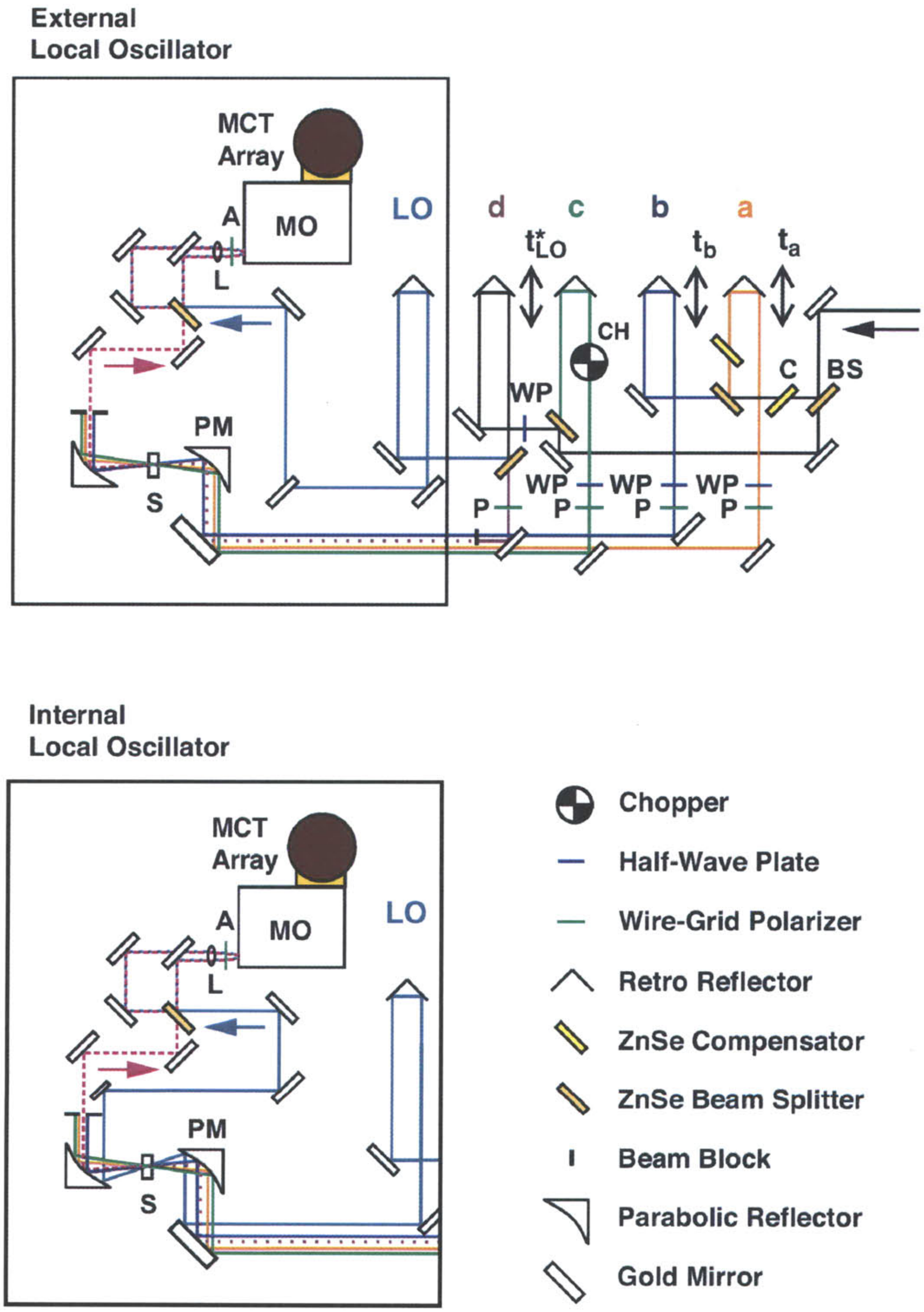

External
Local Oscillator
MCT
Array
MO
A
L
LO
d
c
b
a
t^*_{LO}
t_b
t_a
CH
C
BS
WP
WP
WP
WP
P
P
P
P
PM
S
Internal
Local Oscillator
MCT
Array
MO
A
L
LO
PM
S
Chopper
Half-Wave Plate
Wire-Grid Polarizer
Retro Reflector
ZnSe Compensator
ZnSe Beam Splitter
Beam Block
Parabolic Reflector
Gold Mirror

Figure 2-1: 2D IR Inteferometer, Sample, and Detection Schematic. Figure adapted from Khalil et al.[5] t^*_{LO}- The local oscillator (LO) and tracer (d) are delayed using the same translation stage, and are thus alternately used.

2D IR spectra were collected for parallel (ZZZZ) or perpendicular (ZZYY) conditions depending on the relative polarization between the first two excitation beams and the third/local oscillator.

Spatial alignment of the pulses was constrained by propagation parallel to the optical table in a 1" boxcar geometry at the exit of the interferometer and optimal transmission through a 50 μm pinhole placed at the focus. Time zero was defined by beam c. Timing overlap relative to beam c was set with ±5 fs resolution by non-collinear sum frequency autocorrelation in $AgGaS_2$ (1 mm) in which a glass slide is used to preferentially pass the doubled light at 3 μm. The autocorrelation was detected using a single-channel MCT detector. The autocorrelation was modulated at 500 Hz by chopping beam c and acquired with a lock-in amplifier.

The detection optics were aligned using a third-order signal generated by overlapping the three excitation pulses. Typical alignment samples were 20 mg/mL diglycine in D_2O, 0.1% *N*-methylacetamide in D_2O, or the non-resonant signal from a 300 μm thick piece of ZnSe. The samples were held in a home-built brass cell equipped with 1 mm thick CaF_2 windows and a 50 μm thick Teflon spacer.[82] The temperature was controlled by coupling the brass sample cell to a recirculating water bath. This cell was also used to collect infrared absorption spectra in a Nicolet 380 FT IR spectrometer.

2.2.3 2D IR Experiments

The heterodyne-detected signal was acquired as a function of τ_1 coherence times for two different experiments: a rephasing experiment and non-rephasing experiment. These titles refer to the relative phase evolution in the first and third coherence times: $e^{-i\Omega t_1}e^{+i\Omega t_3}$ and $e^{+i\Omega t_1}e^{+i\Omega t_3}$, respectively (as described in Section 2.1). Schematics of these experiments can be seen in Figure 2-2. Labels a, b, and c refer to the spatial position of the beams in the boxcar geometry, and match the schematic in Figure 2-1. Two signals are simultaneously generated in the $k_{sig} = -k_a + k_b + k_c$ and $k_{sig} = +k_a - k_b + k_c$ directions, which respectively correspond to rephasing and non-rephasing experiments when the pulses interact in the a-b-c order. Since only the former of these two signals is collected, and the boxcar geometry was fixed, the

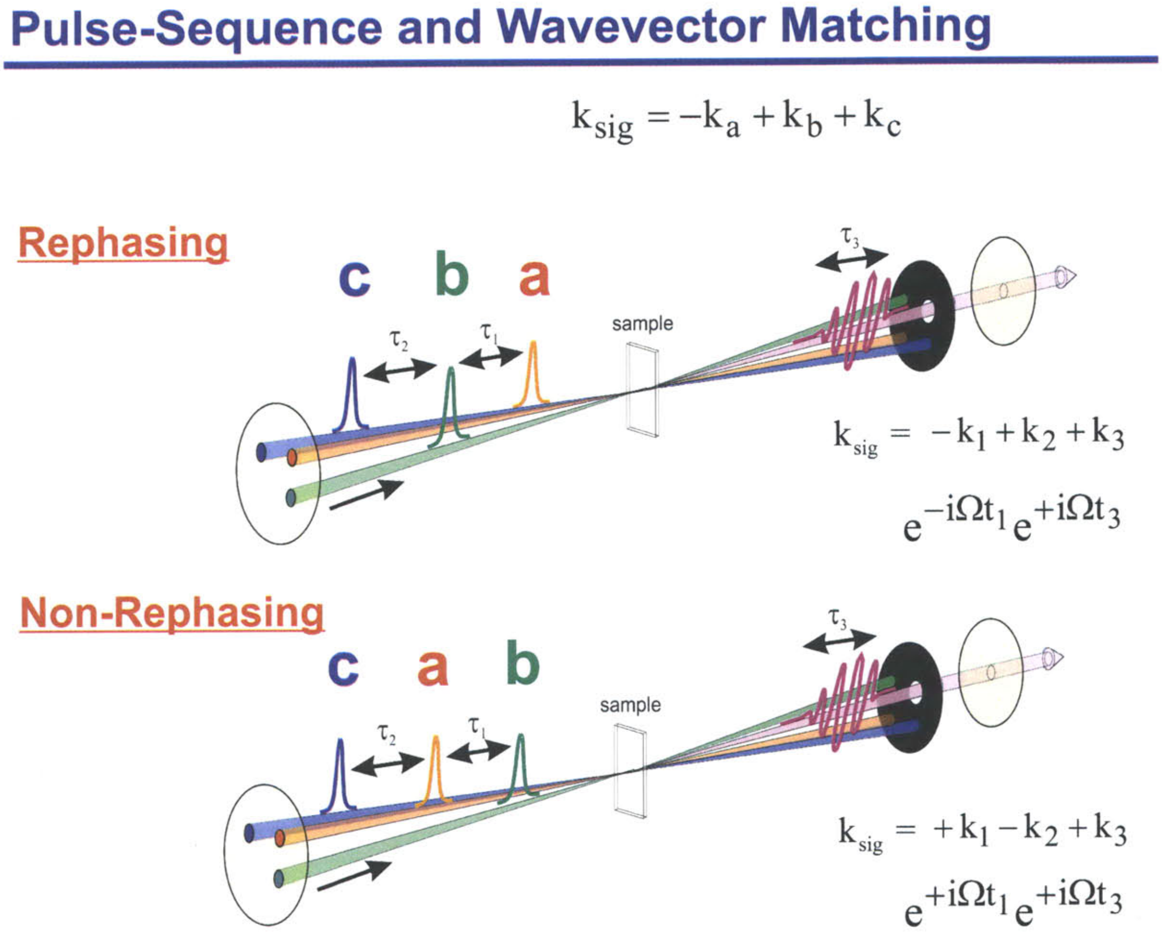

Pulse-Sequence and Wavevector Matching
$k_{sig} = -k_a + k_b + k_c$
Rephasing
c b a
τ_2
τ_1
sample
τ_3
$k_{sig} = -k_1 + k_2 + k_3$
$e^{-i\Omega t_1} e^{+i\Omega t_3}$
Non-Rephasing
c a b
τ_2
τ_1
sample
τ_3
$k_{sig} = +k_1 - k_2 + k_3$
$e^{+i\Omega t_1} e^{+i\Omega t_3}$

Figure 2-2: Pulse Sequence and Wavevector Matching for Rephasing and Non-Rephasing Experiments. Figure adapted from Khalil et al.[5]

time-ordering of the pulses was switched to b-a-c to collect non-rephasing data. The pulses are also designed 1, 2, 3 based on the order they interact with the sample. The coherence times (τ_1) for rephasing and non-rephasing experiments were scanned in 4 fs steps until the signal to noise was less than 1, typically to 4 ps and 2.5 ps, respectively. The waiting time, τ_2, was zero in all spectra unless otherwise indicated.

The signal was isolated with a 1.5 cm aperture spatial mask placed in the phase matched direction, which was the corner of the 1" boxcar with no excitation beam. The signal was combined with the local oscillator on a beamsplitter, which yielded both $Signal - LO$ and $-Signal + LO$ combinations owing to the phase shift upon reflection. In the typical detection scheme, these combinations of the signal and local oscillator are focused onto two stripes (S1 and S2) of a 2x64 element MCT detector (Infrared Systems, Winter Park, FL). The chopper operating at 500 Hz was used to block beam c in every other shot, which removes the signal to detect only the LO and any scattering from beams a and b. The differential signal,

$$\begin{aligned} S &= \langle (S1_O - S2_O) - (S1_C - S2_C) \rangle && (2.2) \\ &= \left\langle |E_{sig} - E_{LO}|^2 - |- \ E_{sig} + E_{LO}|^2 - 2|E_{LO}|^2 \right\rangle && (2.3) \\ &= 2 \left\langle E_{sig} E_{LO} \right\rangle , && (2.4) \end{aligned}$$

was averaged for 500-3000 laser shots per time point (0.5 - 3 s).

Alternatively, the gain on the local oscillator was also detected,

$$\begin{aligned} S &= \left\langle \frac{S1_O}{\langle S1_C \rangle} - \frac{S2_O}{\langle S2_C \rangle} \right\rangle && (2.5) \\ &= \left\langle \frac{|E_{sig} - E_{LO}|^2}{\langle |E_{LO}|^2 \rangle} - \frac{|- E_{sig} + E_{LO}|^2}{\langle |E_{LO}|^2 \rangle} \right\rangle && (2.6) \\ &= 2 \left\langle \frac{E_{sig} E_{LO}}{|E_{LO}|^2} \right\rangle && (2.7) \end{aligned}$$

In both cases, the difference signal from both stripes subtracts any residual homodyne signal ($|E_{Sig}|^2$) or scattering from any of the unchopped beams. The detection scheme in Eq. 2.5 partially corrects for distortions in the ω_3 dimension of the spectrum due to finite excitation bandwidth and linear sample absorption of beams c and LO. The frequencies of the detector channels, and therefore the ω_3 axis, were calibrated using spectral interferometry between the local oscillator and a third-order signal, such as a non-resonant response from ZnSe. In this process, the interferogram resulting from the local oscillator and signal was acquired as a function of the local oscillator time delay. The time-domain interferogram is frequency-resolved by passing through a polychromator and detected using the array for spectral interferometry. Figure 2-3 shows the Fourier transformed set of interferograms. At each channel of the array, the signal is peaked at the detection frequency, representative slices of which are also shown in Figure 2-3. The width of the signal in these slices determines the

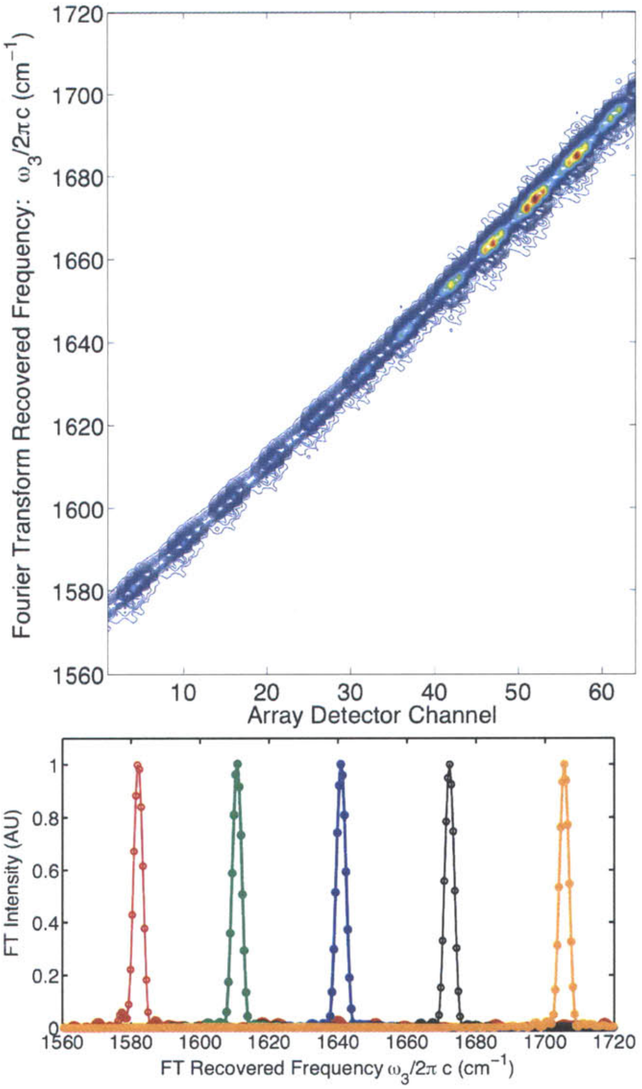


Figure 2-3: Calibrating the ω_3 axis using spectral interferometry.

uncertainty in each channel, which is governed by the width of the entrance slit to the polychromator (typically 0.1 mm), the alignment and focusing onto the array, and the length of time that the signal is scanned; for 1 cm^{-1} resolution, the interferogram must be scanned for 5.3 ps. The full width half max was 3.4 cm^{-1} in this measurement.

2.3 2D IR Data Workup

Figure 2-4 shows an example of two raw $\tau_1 - \omega_3$ surfaces for a non-rephasing and rephasing experiment on insulin. The surfaces show tilted vertical striations because each channel records oscillations at a different frequency. Slices of these signals are shown in the right panel. The rephasing signal envelope peaks at around 170 fs, which is a measure of the inhomogeneity in this system as it would be measured in a photon echo peak shift experiment. In contrast, the non-rephasing envelope peaks at $\tau_1 = 0$ since phase memory is only lost as this time is scanned.

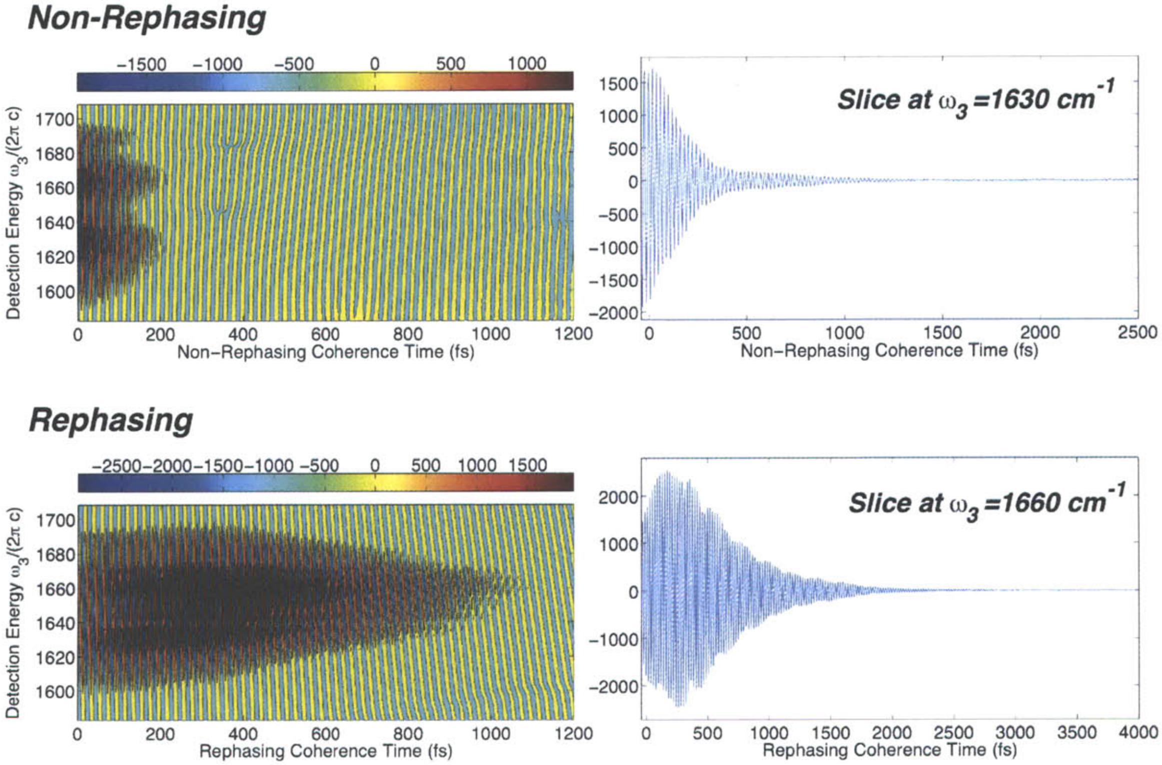


Figure 2-4: Rephasing and non-rephasing t_1-ω_3 surfaces for the insulin dimer 2D IR surface shown in Figure 8-2A.

To measure the real and imaginary parts of the rephasing and non-rephasing surfaces, the data must be corrected for errors in timing stage positioning. An error in timing appears as a phase shift in the frequency domain that mixes the real and imaginary parts, and prevents absorptive line shapes from being obtained. Phasing must be done for the τ_1 and τ_3 dimensions, which occurs in three steps. First, τ_1 for rephasing and non-rephasing experiments are set to 0±5 fs using the autocorrelations. Next, τ_3 is set to 0±5 fs by by comparison to a pump-probe using the procedure described below with Equation 2.15. Final, sub-fs adjustment of the τ_1 timing can be done using an application of the projection-slice theorem.[83] $\tilde{S}(\omega_1, \omega_3)$ and $S(\tau_1, \tau_3)$ (referring to either the rephasing or non-rephasing signal components) are defined by the Fourier transform,

$$\tilde{S}(\omega_1, \omega_3) = \int_0^\infty d\tau_1 \int_0^\infty d\tau_3 \; S(\tau_1, \tau_3) \, e^{i(\omega_1\tau_1 + \omega_3\tau_3)}. \tag{2.8}$$

In a dispersed pump-probe experiment, the first two interactions arise from the same field ($\tau_1 = 0$) and the resulting signal is optically Fourier transformed by frequency-dispersion and collection on a time-integrating detector,

$$\tilde{S}_{DPP}(\omega_3) = \int_{-\infty}^\infty d\tau_3 \; e^{i\omega_3\tau_3} S(\tau_1 = 0, \tau_3). \tag{2.9}$$

A τ_1 point can be isolated from the $\tau_1 - \tau_3$ surface using the Kronecker delta function, $\delta(x)$,

$$\tilde{S}_{DPP}(\omega_3) = \int_{-\infty}^\infty d\tau_3 \; e^{i\omega_3\tau_3} \left[\int_{-\infty}^\infty d\tau_1 \; \delta(\tau_1) \, S(\tau_1, \tau_3) \right], \tag{2.10}$$

which can be replaced by its Fourier transform,

$$\begin{aligned}
\tilde{S}_{DPP}(\omega_3) &= \int_{-\infty}^\infty d\tau_3 \; e^{i\omega_3\tau_3} \left[\int_{-\infty}^\infty d\tau_1 \; \left(\int_{-\infty}^\infty d\omega_1 \; e^{i\omega_1\tau_1} \right) S(\tau_1, \tau_3) \right] && (2.11)\\
&= \int_{-\infty}^\infty d\tau_3 \; e^{i\omega_3\tau_3} \left[\int_{-\infty}^\infty d\omega_1 \; \left(\int_{-\infty}^\infty d\tau_1 \; e^{i\omega_1\tau_1} S(\tau_1, \tau_3) \right) \right] && (2.12)\\
&= \int_{-\infty}^\infty d\tau_3 \; e^{i\omega_3\tau_3} \left[\int_{-\infty}^\infty d\omega_1 \; S(\omega_1, \tau_3) \right] && (2.13)\\
&= \int_{-\infty}^\infty d\omega_1 \tilde{S}(\omega_1, \omega_3) && (2.14)
\end{aligned}$$

This proves that when $\tau_1 = 0$, the ω_1 projection of the $\tilde{S}(\omega_1, \omega_3)$ surface is equal to the dispersed pump-probe slice. Phasing the data to obtain $\tau_1 = 0$ can proceed by comparing this projection to an independently collected, two-beam pump-probe for which $\tau_1 \equiv 0$. However, the final phasing step for data in this thesis was done by insuring that

$$S\left(\tau_1 = 0, \omega_3\right) = \tilde{S}_{DPP}\left(\omega_3\right) \tag{2.15}$$

without use of the projection slice theorem. Figure 2-5 shows the agreement obtained when the zero times for the non-rephasing and rephasing experiments were adjusted by 1.08 fs and -3.37 fs, respectively. (Since this correction may be positive or negative, the coherence times are always scanned beginning from -40 fs.) It is assumed that small (<5 fs) corrections to τ_3 can be compensated using τ_1 phasing. In the top panel of Figure 2-5, it can be seen that the rephasing and non-rephasing time zero slices match each other as well as the pump-probe. Further evidence that the data is phased correctly comes from considering points around time zero, as shown in the bottom panel of Figure 2-5; the non-rephasing signal at 4 fs matches the rephasing signal at -4 fs. Although this agreement is not rigorously valid in the impulsive limit, it allows for an additional check due to the finite pulse lengths, which cause a timing ambiguity near $\tau_1 = 0$ and produce symmetry in the data for short time delays,

After phasing, negative time points are removed, the data is zero padded to 2^{14} τ_1 points and fast Fourier transformed. Because the rephasing and non-rephasing interferograms are asymmetric around zero, both real and imaginary components of the Fourier transform are obtained. Typically only ω_1 points within the observed ω_3 range are retained.

2.4 Representations of 2D IR Spectra

Because 2D IR spectra can characterize the entire third-order response in the observed frequency region, many representations of the data are possible. The complex-valued 2D IR spectrum is obtained from a sum of the non-rephasing ($\tilde{S}_{NR}$) and mirrored rephasing spectra ($\tilde{S}_R$)

$$\tilde{S}_{2D}\left(\omega_1, \omega_3\right) = \tilde{S}_{NR}\left(\omega_1, \omega_3\right) + \tilde{S}_R\left(-\omega_1, \omega_3\right) \tag{2.16}$$

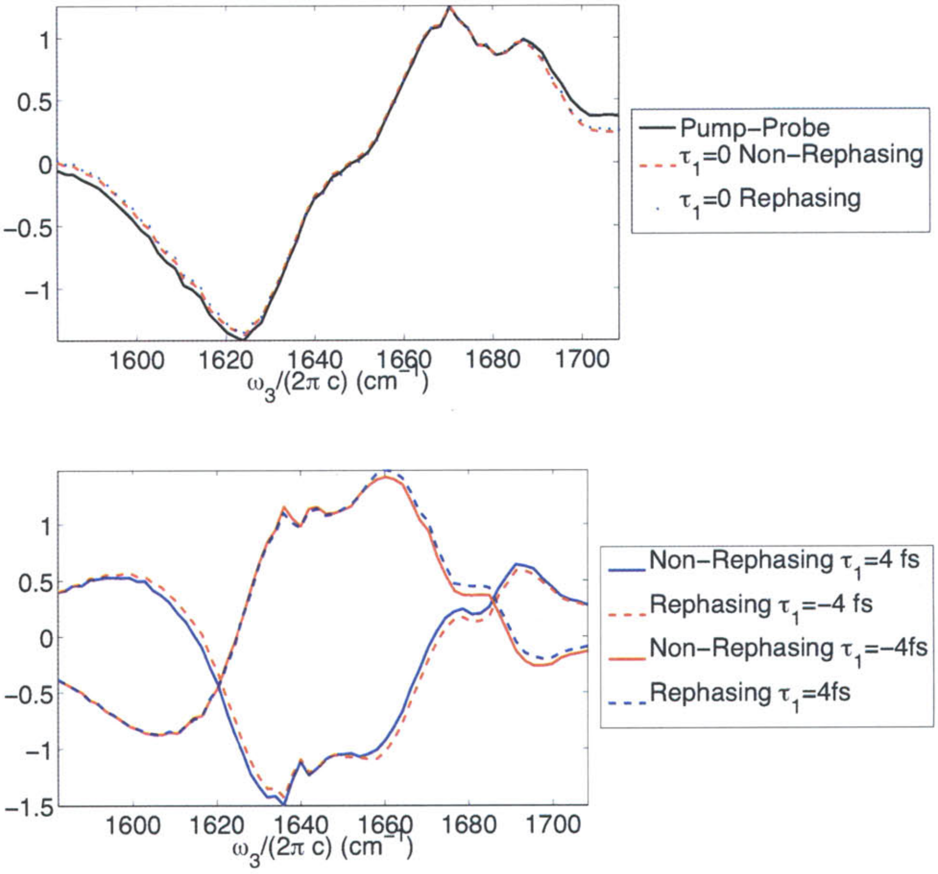


Figure 2-5: Slices of correctly phased data from Figure 2-4, and comparison to the pump-probe.

As explained by Khalil et al.[84], summing the spectra in this way allows for the correlation spectrum with purely absorptive features to be extracted,

$$\tilde{S}_C(\omega_1, \omega_3) = Re\left[\tilde{S}_{NR}(\omega_1, \omega_3) + \tilde{S}_R(-\omega_1, \omega_3)\right]. \tag{2.17}$$

Real and imaginary parts of the complex-valued spectra can be analyzed separately or as power spectra,

$$\tilde{S}_{2DPS}\left(\omega_1,\omega_3\right) = \left|\tilde{S}_{2D}\left(\omega_1,\omega_3\right)\right| \tag{2.18}$$

$$\tilde{S}_{RPS}\left(\omega_1,\omega_3\right) = \left|\tilde{S}_{R}\left(-\omega_1,\omega_3\right)\right| \tag{2.19}$$

$$\tilde{S}_{NRPS}\left(\omega_1,\omega_3\right) = \left|\tilde{S}_{NR}\left(\omega_1,\omega_3\right)\right|, \tag{2.20}$$

where 2DPS stands for the full, two-dimensional power spectrum, RPS stands for the rephasing power spectrum, and NRPS stands for the non-rephasing power spectrum.

2.5 Basic Features of 2D IR Spectra

2.5.1 Diagonal Peaks

Considering the schematic 2D IR spectrum in Figure 2-7, there are four diagonal and four off-diagonal peaks. Each oscillator gives rise to a positive-going, diagonal peak (ω_1=ω_3) located roughly at the v=0-1 energy gap (ω_A or ω_B for this model) resulting from induced absorption of the fundamental transition (plotted red). There is also an oppositely phased, negative-going peak (plotted blue) associated with each diagonal peak, arising from stimulated emission from the overtone. These peaks are almost always redshifted in the ω_3 direction due to vibrational anharmonicity sensed in v=2-1 transition, roughly Δ_A or Δ_B. In the absence of any anharmonic effects that shift the overtone peaks or cause the line shapes to differ from the fundamental peaks, the two cancel perfectly and no nonlinear response is observed. Because high frequency oscillators ($\hbar\omega \gg k_BT$) predominantly populate the ground state, all of the peaks appear with ω_1 values that roughly correspond to the transitions observed in the FT IR spectrum; peaks may be shifted due to interference effects.

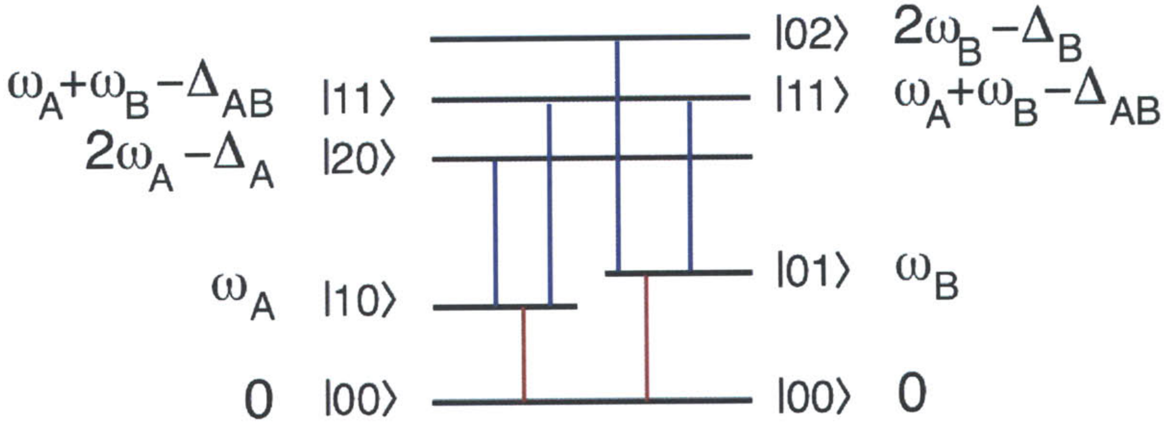


Figure 2-6: One- and two-quantum dipole allowed transitions (in red and blue, respectively) and energy levels for a two-dimensional anharmonic oscillator.

2.5.2 Cross Peaks

The correlation spectrum in Figure 2-7 also shows positive-going cross-peaks at roughly $(\omega_1,\omega_3) = (\omega_A,\omega_B)$ and (ω_B,ω_A) because energy pathways exist to excite oscillator A (and B) and emit from oscillator B (and A). Doublets arise again due to oppositely phased stimulated emission processes, which are shifted due to anharmonicity in the two-quantum combination band $(\omega_{AB} = \omega_A + \omega_B - \Delta_{AB})$

Cross peaks indicate coupling. Coupling is a general term for an energy transfer pathway- not in the eigenbasis. The transition energies in Figure 2-6 describe eigenstates, which *do not couple.* In the eigenstate basis for molecular vibrations, coupling is manifested as a common ground state and nonzero combination band anharmonicity, i.e., the field-matter interactions drive the system through a set of coherent states such that the overall phase can oscillate at ω_A during t_1 and ω_B during t_3. For example, these eigenstates could have arisen due to a coupling of β in a local basis Hamiltonian,

$$\hat{\boldsymbol{H}} = \begin{bmatrix} \omega_X & \beta \\ \beta & \omega_Y \end{bmatrix} = \hat{\boldsymbol{\Lambda}} \begin{bmatrix} \omega_A & 0 \\ 0 & \omega_B \end{bmatrix} \hat{\boldsymbol{\Lambda}}^{-1}. \tag{2.21}$$

Examples of physical mechanisms for the couping β are kinetic or potential coupling

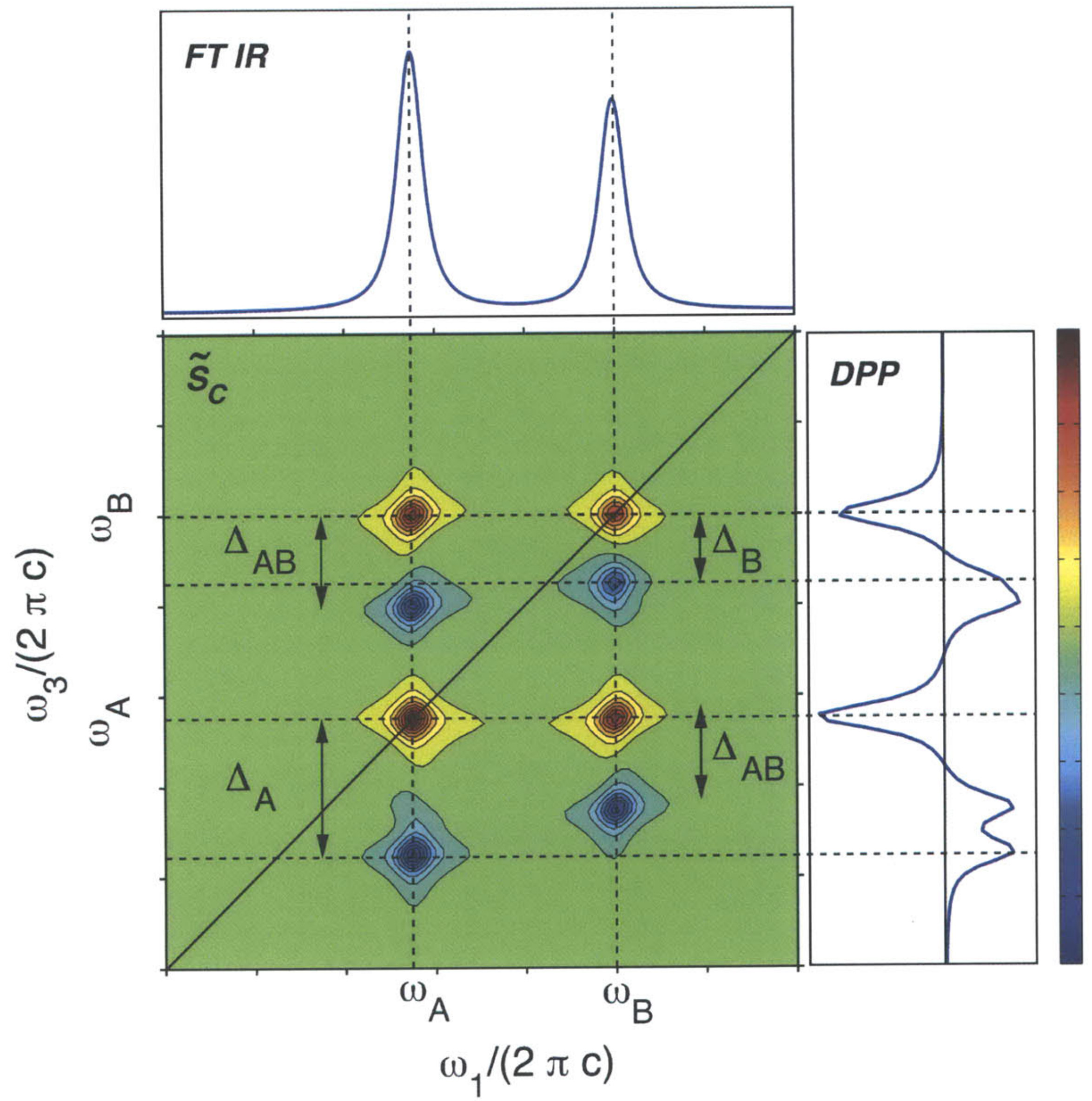


Figure 2-7: Sample 2D IR spectrum for the six-level system.

between oscillators, Förster energy transfer, or chemical exchange. In the weak coupling limit, the word coupling is often used loosely because the eigenstates resemble the respective local modes; for example, coupling between amide I (~1650 cm^{-1}) and amide II (~1550 cm^{-1}) local modes generates amide I+amide II and amide I-amide II eigenstates, but since the coupling is not strong enough to distort the character of these modes, the eigenstates retain the labels amide I and amide II.

2.6 Line Shapes and Intensity

Other representations of 2D IR spectra are explained using a six-level system of two coupled oscillators, whose transition energies and line shape parameters appear in Table 2.1. (The code to generate these spectra is included in Appendix A.) This system contains an inhomogenously broadened mode at 1635 cm^{-1} coupled to a homogenously broadened mode at 1680 cm^{-1}. Both the rephasing and non-rephasing are considered separately along with the correlation spectrum. Three different representations are plotted- real, imaginary, and power spectra. Two different polarization conditions are also plotted, ZZZZ and ZZYY, as explained in Section 2.2.2.

Parameter	**Symbol**	**Value**
Vibrational frequency	ω_a	1635 cm^{-1}
Vibrational frequency	ω_b	1680 cm^{-1}
Transition Dipole Ratioμ_B/μ_A	0.7	
Spectral diffusion time	τ_{aa}	1000 fs
Spectral diffusion time	τ_{bb}	500 fs
Spectral diffusion time	τ_{ab}	200 fs

Table 2.1: Six level system parameters. Parameters for the six-level system are explained in Chapter 3, Section 3.2. The code to calculate the 2D IR spectra appears in Appendix A

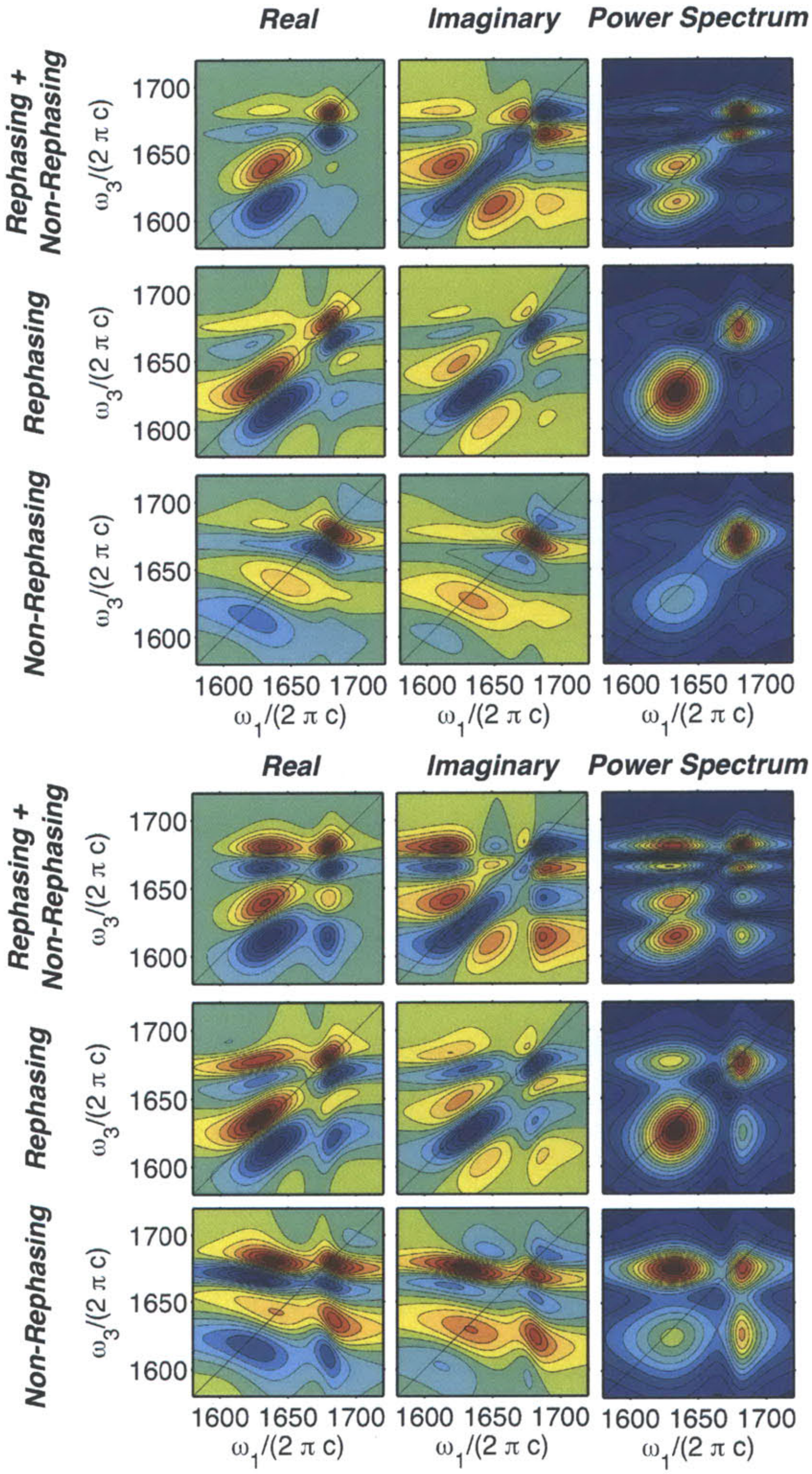


Figure 2-8: ZZZZ (Top) and ZZYY (Bottom) Real, Imaginary, and Power Spectra of Rephasing, Non-Rephasing, and Summed 2D IR spectra

2.6.1 Polarization-Selectivity

A comparison of the top (ZZZZ) and bottom (ZZYY) panels of Fig. 2-8 shows that the spectra appear similar, except the cross peaks are enhanced by the ZZYY polarization. In the model, the angle between the transition dipoles was set to 70°, which implies that perpendicular polarization between the pump beams and probe results in more efficient energy transfer between the two modes than if the polarizations were all parallel. The ZZYY polarization also reduces the likelihood of all transitions occurring on the same oscillator, which reduces the intensity of diagonal features. For a six-level system such as this, as long as the angle between the vibrations, θ_{ab}, is between 54.7° and 125.3° ($acos(\frac{1}{\sqrt{3}})$ and $acos(-\frac{1}{\sqrt{3}})$), then it will be enhanced in the ZZYY spectrum.

2.6.2 Phase Twist in 2D IR Line Shapes

A general observation from the set of 2D IR spectra is that any representation of the pure rephasing or non-rephasing spectrum contains phase-twisted line shapes, which arise from a combination of absorptive and dispersive features. For example, comparing the diagonal peak at 1635 cm^{-1} in the ZZZZ real rephasing spectrum to its representation in the absorptive spectrum shows that it has a long tail in the -ω_1 direction whose contour extends off the edge of the plot. (The same observation can be made of linear absorptive and dispersive spectra.) As originally explained by Khalil et al., purely absorptive 2D IR spectra can be constructed by summing real components of the rephasing and non-rephasing spectra with equal weights (in the absence of effects that distort the relative weights).[84] In doing so, the resulting purely absorptive 2D IR spectra have lines that are as sharp as possible, which minimizes spectral congestion. To quantify the linewidth, one may choose to fit its width along the ω_1 and ω_3 dimensions. Focusing on the $\omega_1 = 1635\ cm^{-1}$, $\omega_3 = 1680\ cm^{-1}$ cross-peak, its ω_1 and ω_3 linewidths vary by the representation. Figure 2-9 illustrates this effect by showing slices at $\omega_1 = 1635\ cm^{-1}$.

The types of interference present in rephasing and non-rephasing spectra are

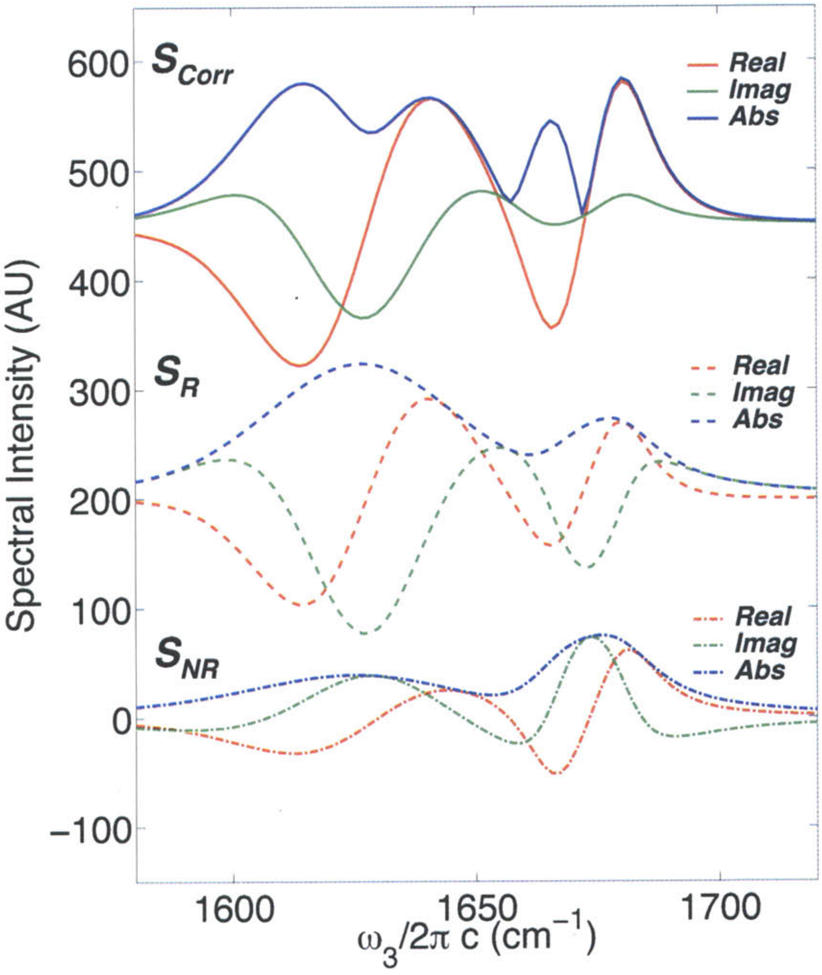


Figure 2-9: Slices of the ZZYY 2D IR surfaces in Figure 2-8 at ω_1=1635 cm^{-1}.

shaped by the phase-twist angles. In non-rephasing spectra, positive and negative lobes of the lineshape are displaced along the diagonal. An important consequence of this tilt is that inhomogeneous features are partially cancelled; for example, consider the ZZZZ real rephasing spectrum in Fig. 2-8. The positive and negative features along the diagonal arising from one resonance tend to destructively interfere with any neighboring diagonal peaks. A static distribution of oscillator frequencies would give rise to such diagonal elongation, which is why the 1635 cm^{-1} peak is weaker in the non-rephasing spectra than the rephasing spectra of Fig. 2-8, but the homogenous peak at 1680 cm^{-1} is roughly equal in both. This is also evident in the RPS and

NRPS.

Another consequence of the phase twist in non-rephasing spectra is that it gives rise to constructive interference in the cross-peak region. This is clearest in a comparison of the cross peaks in the ZZYY rephasing and non-rephaisng real and power spectra (bottom panel). The positive lobes of diagonal peaks extend in the $\pm\omega_1$ direction, which constructively interfere with the positive lobes of the cross peaks. Thus, out of all the spectra plotted in Fig. 2-8, the ZZYY RPS has the best contrast between diagonal and cross peaks.

2.6.3 2D IR Peak Positions

One of the principal pieces of information from a 2D IR spectrum are the (ω_1,ω_3) locations of the peaks. The 1635 cm^{-1} fundamental mode actually appears at the (ω_1,ω_3) positions located in Table 2.6.3. The key observation here is that features of the potential energy surface can rarely be directly read off the 2D IR spectrum due to interference effects between the many positive and negative features (the exception is in the large anharmonicity limit when the peaks are very well-resolved[85]). In all amide I 2D IR spectra, the fundamental peaks appear blue-shifted in the ω_3 direction due to interference with the overtone transition. To properly use 2D IR spectra to quantify a potential energy surface, these interference effects must be considered; a model for the line shapes is required, even if the goal is just to extract energy gaps! In the amide I-II NMA experiments presented in Chapter 3, this was accomplished by iteratively optimizing energy gaps, line shape parameters, and transition dipole magnitudes to obtain agreement between the model and experimental 2D IR spectra.

Peak Max	$\tilde{S}_C$	$Re[\tilde{S}_R]$	$Im[\tilde{S}_{NR}]$	$\tilde{S}_{2DPS}$	$\tilde{S}_{RPS}$	$\tilde{S}_{NRPS}$
ω_1	1634	1629	1635	1634	1634	1635
ω_3	1640	1636	1627	1640	1626	1628

Table 2.2: Peak positions for the 1635 cm^{-1} fundamental mode in the ZZZZ spectra of Fig. 2-8. Spectra are defined in Equations 2.18.

2.7 Slices and Projections: One-Dimensional Representations of 2D IR Spectra

One-dimensional representations of 2D IR spectra are useful for both illustrative and quantitative purposes. Due to the difficulties in overlaying 2D IR spectra (see Figure 5-3 for an example), 1D representations allow for features from several 2D spectra to be compared. In Figure 2-9 of the last section, 1D slices were used to illustrate the line widths in different 2D IR representations. Frequency domain 1D representations have been used to extract melting curves[3, 72, 86](see also Figure 8-5), as a transient measure of unfolding[25, 87–89], as a single-shot measurement[72], to compare dynamics during the waiting time[90, 91], to isolate diagonal and cross peaks (see Figure ??), and to extract a binding constant (Figure 8-3).

Figure 2-10 shows the FT IR spectrum for the six-level system shown in Figure 2-8, and two projections of the ω_1-ω_3 2D IR surfaces, the amplitude of the heterodyne-detected dispersed vibrational echo (HDVE) and the pump-probe. (As described by Jones et al.,[72] the HDVE spectrum characterizes both the real and imaginary parts of the response, and the real part is equivalent to the pump-probe. Unless indicated otherwise, in this section the amplitude HDVE is used and will be referred to as the HDVE.) As projections, these spectra can be obtained by performing the integrations,

$$\begin{aligned} S_{HDVE}\left(\omega_3\right) &= \int_0^\infty \tilde{S}_{2DPS}\left(\omega_1, \omega_3\right) d\omega_1 &(2.22)\\ S_{Pu-Pr}\left(\omega_3\right) &= \int_0^\infty \tilde{S}_C\left(\omega_1, \omega_3\right) d\omega_1. &(2.23)\end{aligned}$$

The projection-slice theorem[83] indicates that these frequency domain projections have analogs as slices of the conjugate, τ_1-ω_3 surfaces,

$$\begin{aligned} S_{HDVE}\left(\omega_3\right) &= \left|S\left(\tau_1 = 0, \omega_3\right)\right| &(2.24)\\ S_{Pu-Pr}\left(\omega_3\right) &= Re\left[S\left(\tau_1 = 0, \omega_3\right)\right]. &(2.25)\end{aligned}$$

The advantage of the first method is that the data is effectively Fourier-filtered for noise outside of the bandwidth of the laser in the workup process. In the second method, the spectra can be acquired without scanning any time delays. Pump-probe spectra have traditionally been acquired in a two-beam setup where the pump-induced changes to the probe spectrum are isolated using a chopper in the pump beam with differential detection. The box-car geometry that was used to acquire 2D IR spectra can simply be converted to a two beam pump-probe experiment simply by blocking two of the beams; however, this makes poor use of the IR power generated. Another method is to acquire the complex HDVE spectra using spectral interferometry with an extrinsic local oscillator, which yields the pump-probe. As explained by Gallagher et al.[92] and elaborated for nonlinear IR signals by Jones et al.,[72] the phase and amplitude can be recovered using a combination of local oscillator phase modulation and Kramer-Kronig-like Fourier transform relations. This technique utilizes the fact that both of these signals arise from the frequency-dispersed third-order response when the three excitation pulses are coincident, and can be related to one another if a phase-sensitive measurement is carried out.

A comparison of the ZZZZ- and ZZYY-extracted projections and slices shown in Figure 2-10 shows that these 1D measures are sensitive to cross-peaks, whose modulation relative to diagonal peaks is the primary effect induced by the polarization change (Figure 2-8). As such, they allow smaller changes to be detected than using FT IR alone[3,89] (see also Figure 8-3). Sensitivity to the diagonal peaks can be enhanced by taking diagonal slices of the 2D IR spectra, as shown by the lower slices in Figure 2-10. These slices appear similar to the FT IR spectrum, but they differ due to the $\sim \mu^4$ scaling of 2D IR peaks and due to interference. There is no simple time-frequency domain experiment that corresponds to a diagonal slice of a 2D IR spectrum, but a frequency-frequency single-shot experiment analogous to that of DeCamp et al.[93] can be used to isolate any slice of a 2D IR spectrum. (For example, a slice parallel to the diagonal can be isolated by spatially dispersing pump and probe beams, and focussing them into the sample using a cylindrical lens such that each position in the sample encodes a different frequency.)

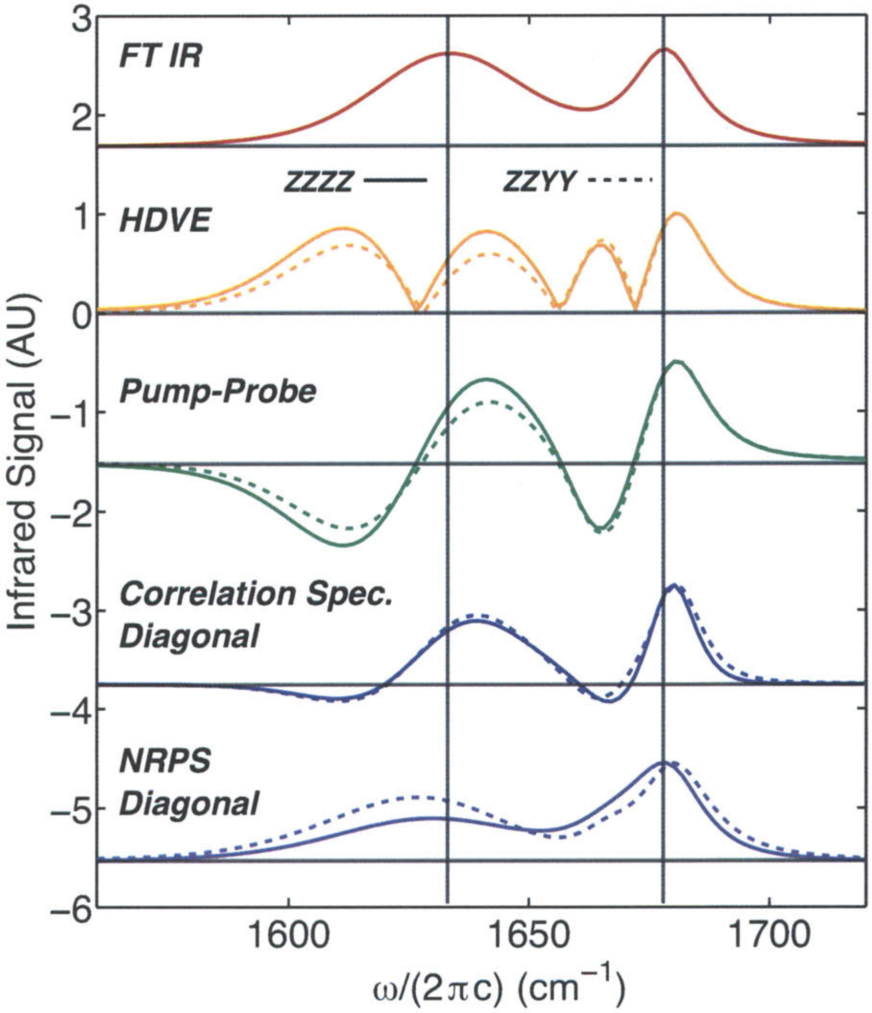


Figure 2-10: FT IR Spectrum, HDVE Spectra, Pump-Probe Spectra, and 2D IR Diagonal Slices for the Six-Level System in Figure 2-8. FT IR spectrum for the six-level system shown in Figure 2-8, projections (HDVE and pump-probe), and slices (Correlation Spectrum and NRPS Diagonal).

Chapter 3

Numerical and Analytical Methods for Evaluating Third-Order Response Functions

In the previous chapter, methods for experimentally measuring and representing $R^{(3)}$ were presented. This chapter focuses on numerical methods for evaluating $R^{(3)}$ for amide oscillator-containing systems using the amide I mode as a case example. An outline of the flow for this chapter is presented in Fig. 3-1. Along with the description of each method, I will outline examples of the type of analysis and insight provided into the system by the modeling. The first two methods presented describe the amide I - amide II spectroscopy of *N*-methylacetamide and of idealized secondary structure motifs. These methods allow for calculating spectra with a small number of parameters, which may be constrained by experimental data. The two methods that follow use atomistic molecular dynamics simulations (MD) to describe the frequency shifts and timescales of the system, and allow for spectra to be calculated in a structure-based fashion without any additional empirical parameters. The first application of the atomistic MD method uses the folding model peptide trpzip2, and applies the atomistic spectral modeling to the current level of theoretical sophistication. The second application considers the thermal unfolding of the protein ubiquitin, which demonstrates various approximations.

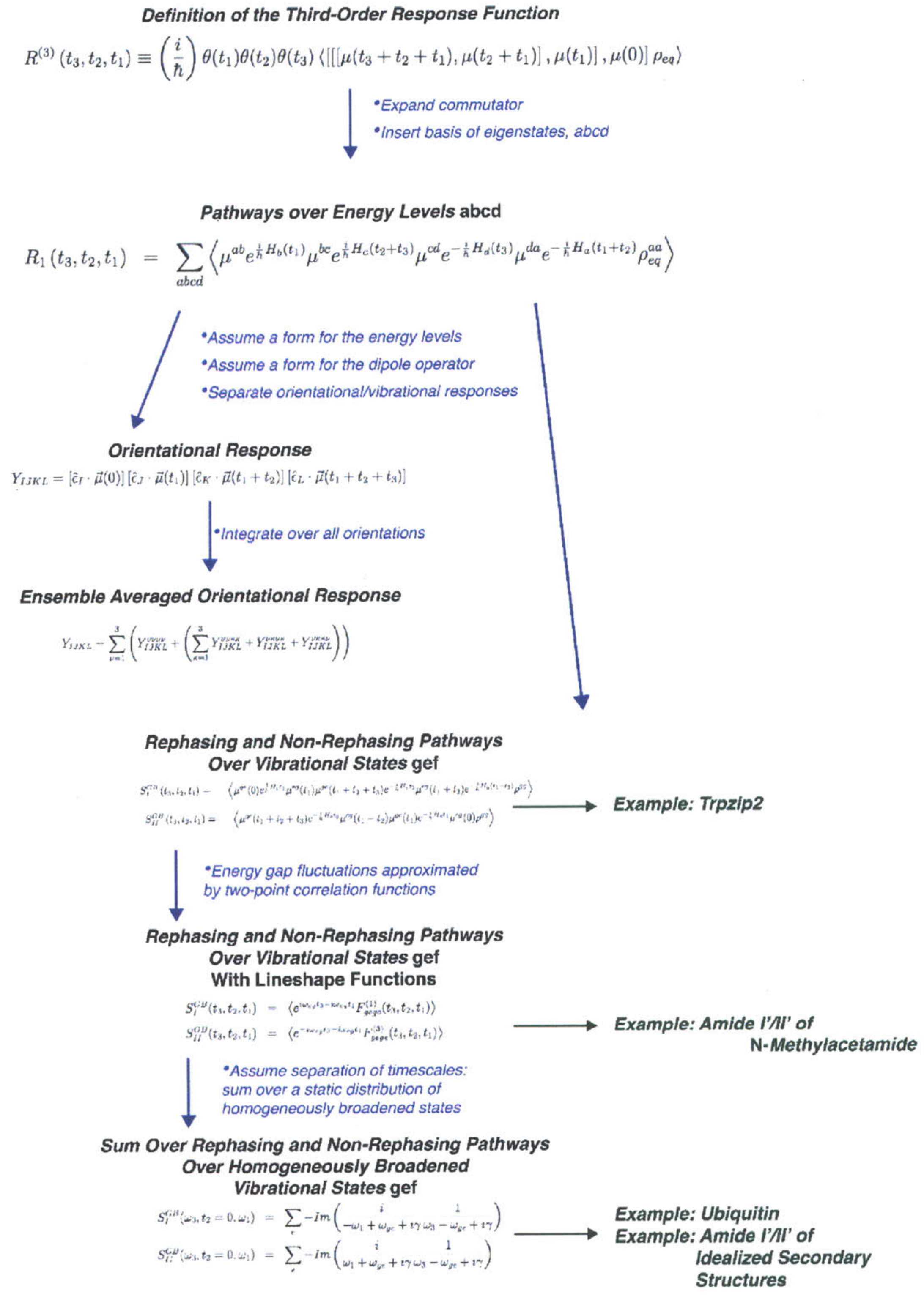

Definition of the Third-Order Response Function
$R^{(3)}(t_3,t_2,t_1) \equiv \left(\frac{i}{\hbar}\right)\theta(t_1)\theta(t_2)\theta(t_3)\langle[[[\mu(t_3+t_2+t_1),\mu(t_2+t_1)],\mu(t_1)],\mu(0)]\,\rho_{eq}\rangle$
•Expand commutator
•Insert basis of eigenstates, abcd
Pathways over Energy Levels abcd
$R_1(t_3,t_2,t_1) = \sum_{abcd}\left\langle \mu^{ab}e^{\frac{i}{\hbar}H_b(t_1)}\mu^{bc}e^{\frac{i}{\hbar}H_c(t_2+t_3)}\mu^{cd}e^{-\frac{i}{\hbar}H_d(t_3)}\mu^{da}e^{-\frac{i}{\hbar}H_a(t_1+t_2)}\rho_{eq}^{aa}\right\rangle$
•Assume a form for the energy levels
•Assume a form for the dipole operator
•Separate orientational/vibrational responses
Orientational Response
$Y_{IJKL} = [\hat{\epsilon}_I\cdot\vec{\mu}(0)]\,[\hat{\epsilon}_J\cdot\vec{\mu}(t_1)]\,[\hat{\epsilon}_K\cdot\vec{\mu}(t_1+t_2)]\,[\hat{\epsilon}_L\cdot\vec{\mu}(t_1+t_2+t_3)]$
•Integrate over all orientations
Ensemble Averaged Orientational Response
$Y_{IJKL} = \sum_{\nu=1}^{3}\left(Y_{IJKL}^{\nu\nu\nu\nu} + \left(\sum_{\kappa=1}^{3} Y_{IJKL}^{\nu\nu\kappa\kappa} + Y_{IJKL}^{\nu\kappa\nu\kappa} + Y_{IJKL}^{\nu\kappa\kappa\nu}\right)\right)$
Rephasing and Non-Rephasing Pathways
Over Vibrational States gef
Example: Trpzip2
•Energy gap fluctuations approximated by two-point correlation functions
Rephasing and Non-Rephasing Pathways
Over Vibrational States gef
With Lineshape Functions
Example: Amide I'/II' of N-Methylacetamide
•Assume separation of timescales: sum over a static distribution of homogeneously broadened states
Sum Over Rephasing and Non-Rephasing Pathways
Over Homogeneously Broadened
Vibrational States gef
$S_I^{GB}(\omega_3,t_2=0,\omega_1) = \sum_e -Im\left(\frac{i}{-\omega_1+\omega_{ge}+i\gamma}\frac{1}{\omega_3-\omega_{ge}+i\gamma}\right)$
$S_{II}^{GB}(\omega_3,t_2=0,\omega_1) = \sum_e -Im\left(\frac{i}{\omega_1+\omega_{ge}+i\gamma}\frac{1}{\omega_3-\omega_{ge}+i\gamma}\right)$
Example: Ubiquitin
Example: Amide I'/II' of Idealized Secondary Structures

Figure 3-1: Layout of the approximations made to derive the third-order response functions presented in this chapter.

3.1 Response Function Formalism

Each of the methods for modeling protein vibrational spectra can be traced to the response function formalism for describing third-order nonlinear spectroscopy. The Hamiltonian is the sum of a time-dependent system Hamiltonian, $H_S(t)$, and a term describing the field-matter interaction through the dipole of the system, $\boldsymbol{\mu}$ and the electric field vector of the light, $\boldsymbol{E}$,

$$H(t) = H_S(t) - \boldsymbol{\mu} \cdot \boldsymbol{E}(t). \tag{3.1}$$

This treatment differs from the common partitioning into Hamiltonians for the system (time-*independent*), bath, and system-bath (time-*dependent*), which facilitates analytical solutions for the line shapes.[94] Here it is understood that the system Hamiltonian only contains the states coupled by the field-matter interaction, and it acquires time-dependence through interactions with the bath, although no explicit form is specified. This approximation is valid insomuch as the timescales for energy transfer within the considered subspace are much faster than for redistribution out of the band.

A central quantity is the system time evolution operator,

$$U = exp\left(-\frac{iH_S(t)t}{\hbar}\right). \tag{3.2}$$

If the Hamiltonian were time-independent, its solution would be

$$U(t,0) = exp\left(-\frac{i}{\hbar}H_S\Delta t\right). \tag{3.3}$$

To use this form, the time evolution operator can be expressed as the product of arbitrarily short time propagators,

$$U(t,0) = \prod_{i=1}^{N} U\left(i\Delta t, \left(i-1\right)\Delta t\right); \Delta t = \frac{t}{N}, \tag{3.4}$$

where the number of steps, N, is chosen to be large enough that the Hamiltonian is well-approximated as time-independent for Δt. The matrix exponential is straightforward to evaluate if the transformation, T, that produces a diagonal Hamiltonian, Λ, is known,

$$\begin{aligned} H &= T\Lambda T^{-1} && (3.5) \\ e^{H} &= Te^{\Lambda}T^{-1} && (3.6) \end{aligned}$$

because the exponential of a diagonal matrix is obtained by exponentiating every element of the diagonal. For the calculations in this thesis, the time-evolution operator contained one- and two-quantum states, which were treated separately due to the absence of off-diagonal terms coupling the one- and two-quantum manifolds, i.e., the Hamiltonian is block diagonal.

Within the dipole approximation, the definition of the third-order response function is[79],

$$\overleftrightarrow{\boldsymbol{R}}^{(3)}(t_3,t_2,t_1) \equiv \left(\frac{i}{\hbar}\right)\theta(t_1)\theta(t_2)\theta(t_3)\left\langle\left[\left[\left[\boldsymbol{\mu}(t_3+t_2+t_1),\boldsymbol{\mu}(t_2+t_1)\right],\boldsymbol{\mu}(t_1)\right],\boldsymbol{\mu}(0)\right]\rho_{eq}\right\rangle, \tag{3.7}$$

which describes the time-evolution of the density matrix originally in thermal equilibrium, ρ_{eq}, in response to four field-matter interactions. Causality is represented by the Heaviside step functions, $\theta(t)$.

This response gives rise to a third-order polarization described by,

$$\begin{aligned} P_I^{(3)}(t_3,t_2,t_1) = \int_0^\infty\int_0^\infty\int_0^\infty \quad & \overleftrightarrow{\boldsymbol{R}}^{(3)}(\tau_3',\tau_2',\tau_1')\boldsymbol{E}_{3J}(t_3-\tau_3')\boldsymbol{E}_{2K}(t_3+t_2-\tau_3'-\tau_2') \\ & \boldsymbol{E}_{1L}(t_3+t_2+t_1-\tau_3'-\tau_2'-\tau_1')d\tau_1'd\tau_2'd\tau_3', \end{aligned} \tag{3.8}$$

and the indices $IJKL$ denote Cartesian directions in the lab frame (X,Y,Z) that the electric field vectors lie upon, such as $\boldsymbol{E}(t) = E(t)\hat{X}$. In all of the calculations in this thesis, the electric fields are taken to be in the impulsive limit ($E(t) = \delta(t)$), and the convolution integral of Eq. 3.8 is trivial.

Response functions of this type can be used to describe third-order spectroscopy at any molecular energy regime. Only the subset of molecular energy levels coupled by the light field were chosen as the system in Eq. 3.1; for most of the experiments described in this thesis, the amide I band of this response function is probed, which defines the bounds of $\omega_1, \omega_3 \in$ 1600-1700 cm^{-1}, the Fourier conjugate variables of t_1 and t_3.

Expanding the commutator yields,

$$\overleftrightarrow{\boldsymbol{R}}^{(3)}(t_3, t_2, t_1) = \sum_{i=1}^{4} \left(\overleftrightarrow{\boldsymbol{R}}_i (t_3, t_2, t_1) - \overleftrightarrow{\boldsymbol{R}}_i^* (t_3, t_2, t_1) \right) \tag{3.9}$$

$$\overleftrightarrow{\boldsymbol{R}}_1 (t_3, t_2, t_1) = \langle \boldsymbol{\mu}(0) \boldsymbol{\mu}(t_1) \boldsymbol{\mu}(t_1 + t_2 + t_3) \boldsymbol{\mu}(t_1 + t_2) \rho_{eq} \rangle \tag{3.10a}$$

$$\overleftrightarrow{\boldsymbol{R}}_2 (t_3, t_2, t_1) = \langle \boldsymbol{\mu}(0) \boldsymbol{\mu}(t_1 + t_2) \boldsymbol{\mu}(t_1 + t_2 + t_3) \boldsymbol{\mu}(t_1) \rho_{eq} \rangle \tag{3.10b}$$

$$\overleftrightarrow{\boldsymbol{R}}_3 (t_3, t_2, t_1) = \langle \boldsymbol{\mu}(t_1 + t_2 + t_3) \boldsymbol{\mu}(t_1 + t_2) \boldsymbol{\mu}(t_1) \boldsymbol{\mu}(0) \rho_{eq} \rangle \tag{3.10c}$$

$$\overleftrightarrow{\boldsymbol{R}}_4 (t_3, t_2, t_1) = \langle \boldsymbol{\mu}(t_1) \boldsymbol{\mu}(t_1 + t_2) \boldsymbol{\mu}(t_1 + t_2 + t_3) \boldsymbol{\mu}(0) \rho_{eq} \rangle \tag{3.10d}$$

This is cast into a different form by assuming that in thermal equilibrium, the density matrix is in a mixed state. The definition of time-evolution in the interaction picture is inserted along with a full set of eigenstates, *abcd*: $\boldsymbol{\mu}(t) = \sum_{ab} exp\left(\frac{i}{\hbar} H_a t\right) \boldsymbol{\mu}^{ab} exp\left(-\frac{i}{\hbar} H_b t\right)$ (or equivalently $\boldsymbol{\mu}(t) = \sum_{ab} U_a^\dagger(t, 0) \boldsymbol{\mu}^{ab} U_b(t, 0)$). Note that no restriction on the form of $\boldsymbol{\mu}$ has been made yet.

$$\overleftrightarrow{\boldsymbol{R}}_1^{abcd}(t_3,t_2,t_1) = \left\langle \boldsymbol{\mu}^{ab} e^{\frac{i}{\hbar}H_b(t_1)} \boldsymbol{\mu}^{bc} e^{\frac{i}{\hbar}H_c(t_2+t_3)} \boldsymbol{\mu}^{cd} e^{-\frac{i}{\hbar}H_d(t_3)} \boldsymbol{\mu}^{da} e^{-\frac{i}{\hbar}H_a(t_1+t_2)} \rho_{eq}^{aa} \right\rangle \tag{3.11a}$$

$$\overleftrightarrow{\boldsymbol{R}}_2^{abcd}(t_3,t_2,t_1) = \left\langle \boldsymbol{\mu}^{ab} e^{\frac{i}{\hbar}H_b(t_1+t_2)} \boldsymbol{\mu}^{bc} e^{\frac{i}{\hbar}H_c(t_3)} \boldsymbol{\mu}^{cd} e^{-\frac{i}{\hbar}H_d(t_2+t_3)} \boldsymbol{\mu}^{da} e^{-\frac{i}{\hbar}H_a(t_1)} \rho_{eq}^{aa} \right\rangle \tag{3.11b}$$

$$\overleftrightarrow{\boldsymbol{R}}_3^{abcd}(t_3,t_2,t_1) = \left\langle e^{\frac{i}{\hbar}H_a(t_1+t_2+t_3)} \boldsymbol{\mu}^{ab} e^{-\frac{i}{\hbar}H_b(t_3)} \boldsymbol{\mu}^{bc} e^{-\frac{i}{\hbar}H_c(t_2)} \boldsymbol{\mu}^{cd} e^{-\frac{i}{\hbar}H_d(t_1)} \boldsymbol{\mu}^{da} \rho_{eq}^{aa} \right\rangle \tag{3.11c}$$

$$\overleftrightarrow{\boldsymbol{R}}_4^{abcd}(t_3,t_2,t_1) = \left\langle e^{\frac{i}{\hbar}H_a(t_1)} \boldsymbol{\mu}^{ab} e^{\frac{i}{\hbar}H_b(t_2)} \boldsymbol{\mu}^{bc} e^{\frac{i}{\hbar}H_c(t_3)} \boldsymbol{\mu}^{cd} e^{-\frac{i}{\hbar}H_d(t_1+t_2+t_3)} \boldsymbol{\mu}^{da} \rho_{eq}^{aa} \right\rangle \tag{3.11d}$$

Equations 3.11 correspond to the Feynman diagrams illustrated in Fig. 3-2 containing labels *abcd*. (See other sources for a description of Feynman diagrams.[79,95]) Note that the complex conjugate diagrams are explicitly shown because they give rise to distinct peaks in multilevel systems. The full response functions are a sum over all possible pathways,

$$\overleftrightarrow{\boldsymbol{R}}_i(t_3,t_2,t_1) = \sum_{abcd} \overleftrightarrow{\boldsymbol{R}}_i^{abcd}(t_3,t_2,t_1) \tag{3.12}$$

Equations 3.11 are very general, but unfortunately not useful because they do not contain information about which transitions between energy levels are allowed. To arrive at the next set of equations, a few approximations are made that correspond to assumptions about $\boldsymbol{\mu}$,

1. **Rotating wave approximation**. Anti-resonant terms (absorption with deexcitation or emission with excitation) are ignored.

2. **Exciton-like energy level structure**. It will be assumed that field-matter interactions couple the ground state to a manifold of one-quantum states or the manifold of one-quantum states to the two-quantum states.

3. **Only the ground state is populated in thermal equilibrium**. Terms in

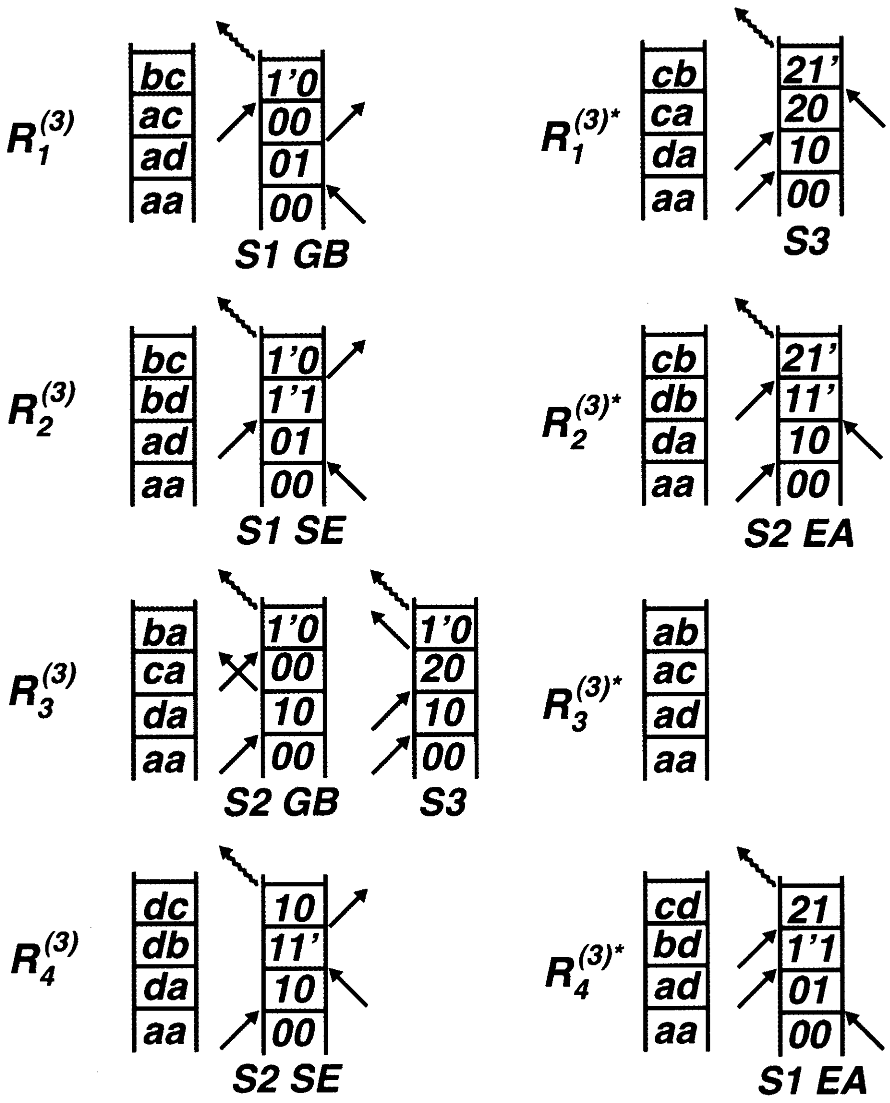

$R_1^{(3)}$
bc
ac
ad
aa
1'0
00
01
00
S1 GB
$R_1^{(3)*}$
cb
ca
da
aa
21'
20
10
00
S3
$R_2^{(3)}$
bc
bd
ad
aa
1'0
1'1
01
00
S1 SE
$R_2^{(3)*}$
cb
db
da
aa
21'
11'
10
00
S2 EA
$R_3^{(3)}$
ba
ca
da
aa
1'0
00
10
00
S2 GB
1'0
20
10
00
S3
$R_3^{(3)*}$
ab
ac
ad
aa
$R_4^{(3)}$
dc
db
da
aa
10
11'
10
00
S2 SE
$R_4^{(3)*}$
cd
bd
ad
aa
21
1'1
01
00
S1 EA

Figure 3-2: Feynman Diagrams corresponding to third-order signals for a multilevel system.

the transition dipole operator not corresponding to excitation from the ground state are set to zero in the first interaction.

4. **The dipole operator is split into terms propagated by classical and quantum-mechanical Hamiltonians.**

Due to the first three approximations, it is worth illustrating the matrix representation of the dipole operator for a six-level system of two coupled oscillators,

$$\overleftrightarrow{\boldsymbol{\mu}(t)} = \left[\begin{array}{c|c|c|c|c|c} 0 & \boldsymbol{\mu}_{g,e1} & \boldsymbol{\mu}_{g,e2} & 0 & 0 & 0 \\ \hline \boldsymbol{\mu}_{g,e1} & 0 & 0 & \boldsymbol{\mu}_{e1,f1} & \boldsymbol{\mu}_{e1,f2} & \boldsymbol{\mu}_{e1,f3} \\ \hline \boldsymbol{\mu}_{g,e2} & 0 & 0 & \boldsymbol{\mu}_{e2,f1} & \boldsymbol{\mu}_{e2,f2} & \boldsymbol{\mu}_{e2,f3} \\ \hline 0 & \boldsymbol{\mu}_{e1,f1} & \boldsymbol{\mu}_{e2,f1} & 0 & 0 & 0 \\ \hline 0 & \boldsymbol{\mu}_{e1,f2} & \boldsymbol{\mu}_{e2,f2} & 0 & 0 & 0 \\ \hline 0 & \boldsymbol{\mu}_{e1,f3} & \boldsymbol{\mu}_{e2,f3} & 0 & 0 & 0 \end{array}\right] . \qquad (3.13)$$

The electric field of light interacts with the system through the dipole operator to couple the ground state, g to two one-quantum excited states, *e1* and *e2*. The one-quantum excited states are coupled to two-quantum excited states, *f1*, *f2*, and *f3*.

Atomistic, structure-based calculations of 2D IR spectra would not be possible without partitioning the system Hamiltonian into quantum mechanical and classical components,

$$H_S(t) = H_S^{vib}(t) + H_S^{rot}(t). \qquad (3.14)$$

The classical Hamiltonian (H^{rot}) governs the orientation ($\hat{\mu}(t)$) and magnitude ($|\boldsymbol{\mu}(t)|$) of the transition dipoles, which are usually derived from the structure, and may be influenced by molecular reorientation and conformational changes. For amide I modes, the Condon approximation holds, and the length of the transition dipole vector is well approximated by a frequency-independent constant. (See Refs. [96,97] for a discussion on the breakdown of the Condon approximation for HOD in D_2O.)

The phase and amplitude of density matrix components are propagated with a quantum mechanical Hamiltonian (H^{vib}), which may be parameterized by classical quantities such as the electric field, electrostatic potential, and bond vectors. The exponential terms describe the vibrational phase accumulated, which is a function of the frequencies of the oscillators that may be time dependent. For simplicity of notation, μ will be taken to mean $|\boldsymbol{\mu}|$.

Since the orientational and vibrational degrees of freedom are separable, these response functions can be further factored as

$$\overleftrightarrow{\boldsymbol{R}}_i^{abcd}(t_3, t_2, t_1) = \sum_{IJKL} Y_{IJKL}^{abcd}(t_3, t_2, t_1)\, R_i^{abcd}(t_3, t_2, t_1)\,, \tag{3.15}$$

where the vector nature of the dipole operator can give rise to polarization-sensitive changes and the labels $IJKL$ were defined in Eq. 3.8. The orientational response,

$$Y_{IJKL}^{abcd}(t_3, t_2, t_1) = \left[\hat{\epsilon}_I \cdot \boldsymbol{\mu}^{ab}(0)\right]\left[\hat{\epsilon}_J \cdot \boldsymbol{\mu}^{bc}(t_1)\right]\left[\hat{\epsilon}_K \cdot \boldsymbol{\mu}^{cd}(t_1 + t_2)\right]\left[\hat{\epsilon}_L \cdot \boldsymbol{\mu}^{da}(t_1 + t_2 + t_3)\right] \tag{3.16}$$

has an analytical form for two coupled oscillators if it is assumed that there are no preferred orientations for the molecular reference frame,

$$\begin{gathered} Y_{IJKL}^{abcd}(t_3, t_2, t_1) = \\ \textstyle\sum_{\nu=XYZ}\Big\{ Y_{IJKL}^{\nu\nu\nu\nu}\left[\boldsymbol{\mu}^{ab}(0)\cdot\hat{\nu}\right]\left[\boldsymbol{\mu}^{bc}(t_1)\cdot\hat{\nu}\right]\left[\boldsymbol{\mu}^{cd}(t_1+t_2)\cdot\hat{\nu}\right]\left[\boldsymbol{\mu}^{da}(t_1+t_2+t_3)\cdot\hat{\nu}\right] \\ +\textstyle\sum_{\kappa=XYZ}\left\{ Y_{IJKL}^{\nu\nu\kappa\kappa}\left[\boldsymbol{\mu}^{ab}(0)\cdot\hat{\nu}\right]\left[\boldsymbol{\mu}^{bc}(t_1)\cdot\hat{\nu}\right]\left[\boldsymbol{\mu}^{cd}(t_1+t_2)\cdot\hat{\kappa}\right]\left[\boldsymbol{\mu}^{da}(t_1+t_2+t_3)\cdot\hat{\kappa}\right]\right. \\ +Y_{IJKL}^{\nu\kappa\nu\kappa}\left[\boldsymbol{\mu}^{ab}(0)\cdot\hat{\nu}\right]\left[\boldsymbol{\mu}^{bc}(t_1)\cdot\hat{x}_\kappa\right]\left[\boldsymbol{\mu}^{cd}(t_1+t_2)\cdot\hat{x}_\nu\right]\left[\boldsymbol{\mu}^{da}(t_1+t_2+t_3)\cdot\hat{\kappa}\right] \\ \left.+Y_{IJKL}^{\nu\kappa\kappa\nu}\left[\boldsymbol{\mu}^{ab}(0)\cdot\hat{\nu}\right]\left[\boldsymbol{\mu}^{bc}(t_1)\cdot\hat{x}_\kappa\right]\left[\boldsymbol{\mu}^{cd}(t_1+t_2)\cdot\hat{\kappa}\right]\left[\boldsymbol{\mu}^{da}(t_1+t_2+t_3)\cdot\hat{\nu}\right]\right\}\Big\}. \end{gathered} \tag{3.17}$$

The indices ν and κ refer to Cartesian directions in the lab frame. If reorientation is treated numerically (rather than through analytical expressions involving orientational correlation functions), then tensor components from Eq 3.17 can be found in Table 3.1.[98]

Polarization Component	$Y^{\nu\nu\nu\nu}$	$Y^{\nu\nu\kappa\kappa}$	$Y^{\nu\kappa\nu\kappa}$	$Y^{\nu\kappa\kappa\nu}$
ZZZZ	$\frac{1}{9}(1+\frac{4}{5})$	$\frac{1}{9}(1-\frac{2}{5})$	$\frac{1}{15}$	$\frac{1}{15}$
ZZYY	$\frac{1}{9}(1-\frac{2}{5})$	$\frac{1}{9}(1+\frac{1}{5})$	$-\frac{1}{30}$	$-\frac{1}{30}$
ZYYZ	$\frac{1}{15}$	$-\frac{1}{30}$	$\frac{1}{4}(\frac{1}{5}-\frac{1}{3})$	$\frac{1}{4}(\frac{1}{5}+\frac{1}{3})$
ZYZY	$\frac{1}{15}$	$-\frac{1}{30}$	$\frac{1}{4}(\frac{1}{5}+\frac{1}{3})$	$\frac{1}{4}(\frac{1}{5}-\frac{1}{3})$

Table 3.1: Tensor components of the orientational response.

With these assumptions, Eqs. 3.11 become

$$S_I^{GB}(t_3,t_2,t_1) = \left\langle \mu^{ge}(0)e^{\frac{i}{\hbar}H_e t_1}\mu^{eg}(t_1)\mu^{ge}(t_1+t_2+t_3)e^{-\frac{i}{\hbar}H_e t_3}\mu^{eg}(t_1+t_2)\rho^{gg}\right\rangle \tag{3.18a}$$

$$\begin{aligned} S_{III}(t_3,t_2,t_1) = & -\Big\langle \mu^{ge}(0)e^{-\frac{i}{\hbar}H_e t_1}\mu^{ef}(t_1)e^{-\frac{i}{\hbar}H_f(t_2+t_3)} \\ & \mu^{fe}(t_1+t_2+t_3)e^{\frac{i}{\hbar}H_e t_3}\mu^{eg}(t_1+t_2)\rho^{gg}\Big\rangle \end{aligned} \tag{3.18b}$$

$$\begin{aligned} S_I^{SE}(t_3,t_2,t_1) = & \Big\langle \mu^{ge}(0)e^{\frac{i}{\hbar}H_e(t_1+t_2)}\mu^{eg}(t_1+t_2)\mu^{ge}(t_1+t_2+t_3) \\ & e^{-\frac{i}{\hbar}H_e(t_2+t_3)}\mu^{eg}(t_1)\rho^{gg}\Big\rangle \end{aligned} \tag{3.18c}$$

$$\begin{aligned} S_{II}^{EA}(t_3,t_2,t_1) = & -\Big\langle \mu^{ge}(t_1)e^{-\frac{i}{\hbar}H_e(t_1+t_2)}\mu^{ef}(t_1+t_2+t_3)e^{-\frac{i}{\hbar}H_f t_3} \\ & \mu^{fe}(t_1+t_2)e^{\frac{i}{\hbar}H_e(t_2+t_3)}\mu^{eg}(0)\rho^{gg}\Big\rangle \end{aligned} \tag{3.18d}$$

$$S_{II}^{GB}(t_3,t_2,t_1) = \left\langle \mu^{ge}(t_1+t_2+t_3)e^{-\frac{i}{\hbar}H_e t_3}\mu^{eg}(t_1+t_2)\mu^{ge}(t_1)e^{-\frac{i}{\hbar}H_e t_1}\mu^{eg}(0)\rho^{gg}\right\rangle \tag{3.18e}$$

$$\begin{aligned} S_{III}(t_3,t_2,t_1) = & \Big\langle \mu^{ge}(t_1+t_2+t_3)e^{-\frac{i}{\hbar}H_e t_3}\mu^{ef}(t_1+t_2)e^{-\frac{i}{\hbar}H_f t_2} \\ & \mu^{fe}(t_1)e^{-\frac{i}{\hbar}H_e t_1}\mu^{eg}(0)\rho^{gg}\Big\rangle \end{aligned} \tag{3.18f}$$

$$\begin{aligned} S_{II}^{SE}(t_3,t_2,t_1) = & \Big\langle \mu^{ge}(t_1)e^{\frac{i}{\hbar}H_e t_2}\mu^{eg}(t_1+t_2)\mu^{ge}(t_1+t_2+t_3) \\ & e^{-\frac{i}{\hbar}H_e(t_1+t_2+t_3)}\mu^{eg}(0)\rho^{gg}\Big\rangle \end{aligned} \tag{3.18g}$$

$$\begin{aligned} S_I^{EA}(t_3,t_2,t_1) = & -\Big\langle \mu^{ge}(0)e^{-\frac{i}{\hbar}H_e t_2}\mu^{ef}(t_1+t_2+t_3)e^{-\frac{i}{\hbar}H_f t_3} \\ & \mu^{fe}(t_1+t_2)e^{\frac{i}{\hbar}H_e(t_1+t_2+t_3)}\mu^{eg}(t_1)\rho^{gg}\Big\rangle, \end{aligned}$$

Note that the ground state is taken as the zero energy reference and its associated propagator (the identity matrix) was dropped. One can glean quite a bit from the Eqs. 3.18 because they now describe physical processes, and they are grouped accordingly. Ground state bleach terms (GB) monitor the vibrational relaxation that fills the ground state hole. Stimulated emission (SE) terms monitor vibrational relaxation from the first excited state. Excited state absorption (EA) terms reveal information about the one- to two-quantum energy gap. There are also diagrams that lead to rephasing (S_I) and nonrephasing (S_{II}) contributions, which were discussed in Section 2.6.2. As shown in Figure 3-2, the S_{III} diagrams describe non-rephasing pathways that only occur when all of the pulses are overlapped because the first two interactions must be positive-signed. Figure 3-2 shows how assuming a form for the energy levels and dipole operator lead to translation of the *abcd* state diagrams into transitions between 0, 1, and 2 quantum states.

Equations 3.18 and 3.17 are all that are required to simulate 2D IR spectra if trajectories for the Hamiltonian and transition dipole operator are available. Sung and Silbey showed that these equations can be simplified if the energy gap fluctuations can be assumed to satisfy two-point correlation functions.[94] The 2D IR spectra calculated for N-methylacetamide in 3.2 used an implementation of this theory.

In the case that the correlation function has a simple exponential form, the 2D IR lineshapes simplify to two-dimensional Lorentzian bands,[57]

$$\begin{aligned} S_I(\omega_1, t_2, \omega_3) = -Im \sum_{d,b} \Bigg\{ & \boldsymbol{\mu}_d^* \boldsymbol{\mu}_d \boldsymbol{\mu}_b \boldsymbol{\mu}_b^* \frac{i}{-\omega_1 + \omega_d + i\gamma} \frac{1}{\omega_3 - \omega_b + i\gamma} + \\ & \boldsymbol{\mu}_d^* \boldsymbol{\mu}_b \boldsymbol{\mu}_d \boldsymbol{\mu}_b^* \frac{i}{-\omega_1 + \omega_d + i\gamma} \frac{1}{\omega_3 - \omega_b + i\gamma} e^{i\omega_{db} t_2} + \\ & \sum_c \boldsymbol{\mu}_d^* \boldsymbol{\mu}_b \boldsymbol{\mu}_{cb} \boldsymbol{\mu}_{cd}^* \frac{-i}{-\omega_1 + \omega_d + i\gamma} \frac{1}{\omega_3 - \omega_{cd} + i\gamma} e^{i\omega_{db} t_2} \Bigg\} \end{aligned} \qquad (3.19)$$

$$\begin{aligned} S_{II}\left(\omega_1, t_2, \omega_3\right) = -Im \sum_{d,b} \Bigg\{ & \boldsymbol{\mu}_d \boldsymbol{\mu}_d^* \boldsymbol{\mu}_b \boldsymbol{\mu}_b^* \frac{i}{\omega_1 - \omega_d + i\gamma} \frac{1}{\omega_3 - \omega_b + i\gamma} + \\ & \boldsymbol{\mu}_d \boldsymbol{\mu}_b^* \boldsymbol{\mu}_b \boldsymbol{\mu}_d^* \frac{i}{\omega_1 - \omega_d + i\gamma} \frac{1}{\omega_3 - \omega_b + i\gamma} e^{-i\omega_{db} t_2} + \\ & \sum_c \boldsymbol{\mu}_d \boldsymbol{\mu}_b^* \boldsymbol{\mu}_{cd} \boldsymbol{\mu}_{cb}^* \frac{-i}{\omega_1 - \omega_d + i\gamma} \frac{1}{\omega_3 - \omega_{cb} + i\gamma} e^{-i\omega_{db} t_2} \Bigg\}, \end{aligned} \qquad (3.20)$$

where d and b are one quantum states, c is the set of two-quantum states, and double indices denote difference frequencies, i.e., $\omega_{cb} = \omega_c - \omega_b$. These expressions are used in Chapter 4 and 6, and adapted for use in Chapter 8 using block diagonalization and the time-averaging approximation described in Section 3.1.

Block Diagonalization

The purpose of the block diagonalization procedure is to break up a large Hamiltonian into smaller Hamiltonians that can be diagonalized separately while ignoring the minimum amount of information. The assumption is that sufficiently small couplings lead to peak splittings that are inherently unresolvable due to the short lifetime of amide I modes, ~1 ps. This procedure was used in Chapter 8 in conjunction with the time-averaging approximation of Jansen and Ruszel[99] and Auer and Skinner.[100] The algorithm is run on each time step separately. Initially, each site, i, comprises its own block. A list is made of all the off-diagonal elements β_{ij} that couple sites i and j, which are larger in magnitude than the cutoff. This list of off-diagonal elements is used to join blocks that contain sites i and j. The list is iteratively used to join blocks until the block definitions converge. Different cutoffs were tested; the current value of 4 cm^{-1} was used because the resulting spectra appeared identical to those from the full calculation. Physically, a 4 cm^{-1} coupling corresponds to energy transfer rate between sites on a timescale, (1.3 ps)$^{-1}$, that is slower than the dephasing time, and is similar to the experimental resolution for distinguishing two peaks. The time-averaging approximation requires two time-averaged Hamiltonian trajectories: one that is forward-averaged and one that is backward-averaged. The trajectory is split into the two time-averaged trajectories, then separated into blocks according to the breakdown of the forward-averaged Hamiltonian trajectory. Since the couplings vary little over the 170 fs window (the dominant effect of time-averaging is on the site energies), the obtained blocks would be nearly identical for both trajectories.

3.2 Amide I' - Amide II' 2D IR Spectroscopy of *N*-methylacetamide

The first example shows the type of analysis that was applied to study the solvent- and isotopologue-dependence of the amide I' and amide II' vibrations of a simple model for protein vibrational spectroscopy, *N*-methylacetamide (NMA). It has been demonstrated that polarization-selective 2D IR spectroscopy can be used to characterize the anharmonic vibrational potential of coupled oscillator systems.[5] The most detailed example of this work described the anharmonic vibrational potential of the symmetric and antisymmetric carbonyl stretches in rhodium dicarbonyl. The analysis on NMA build upon this work and allowed for the extraction of energy levels, transition dipole magnitudes and relative orientation, and line broadening parameters of this two-site system. A discussion of the experiment and the accompanying *ab initio* calculations was described by DeFlores et al.[6] Figure 3-3 shows the 2D IR spectra and fits for NMA-d$_7$ in D_2O.

To obtain the fit parameters from the data, the 2D IR spectra were calculated using the response function formalism of Sung and Silbey[94] to describe resonant transitions between six vibrational levels in terms of the energy levels, transition dipoles, and frequency autocorrelation functions, which are diagrammed in Figure 3-4.

The lineshape function was derived from the correlation function given by the generalized Kubo model,[101] which allows for spectral diffusion to take place on two arbitrary timescales,

$$C(t) = \Delta_1^2 e^{-t/\tau_1} + \Delta_2^2 e^{-t/\tau_2} \tag{3.21}$$

The time scales from the correlation function determine line shapes in the 2D IR spectrum. If frequency fluctuations are fast enough that the frequency ω_3 sampled

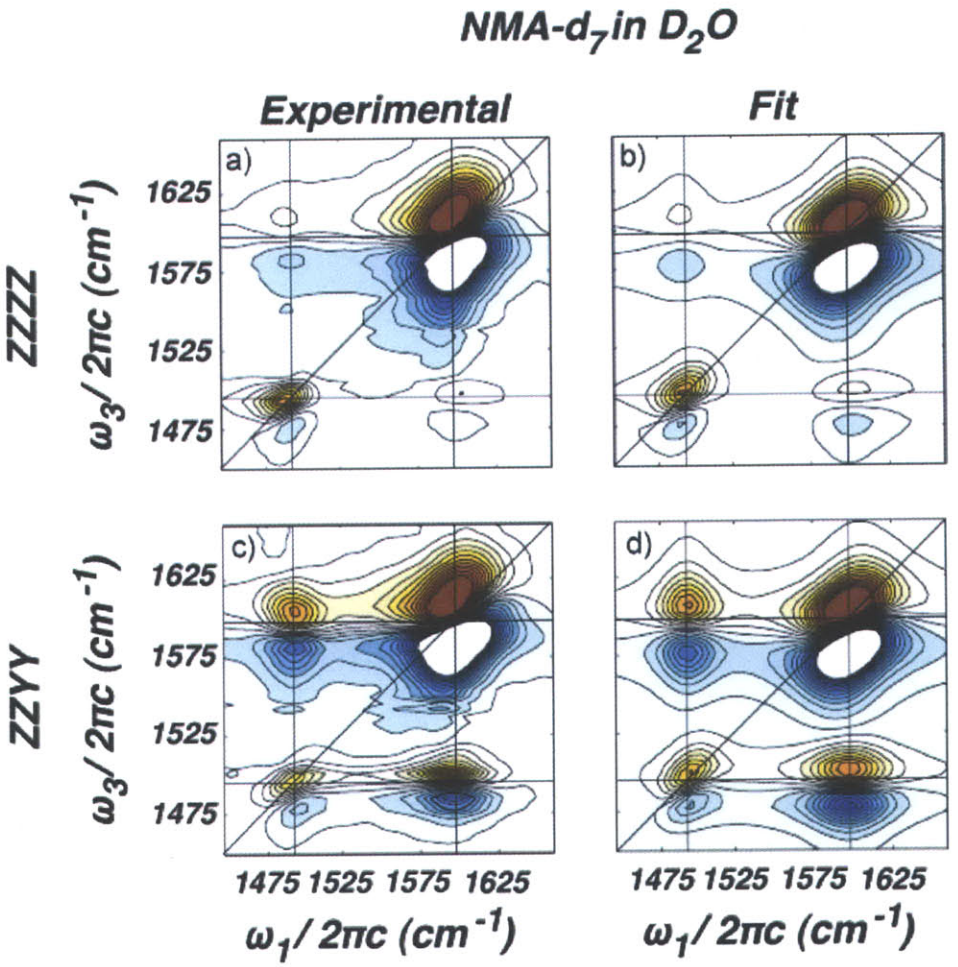


Figure 3-3: Simulated and experimental data for NMA-d_7 in D_2O. Adapted from DeFlores et al.[6].

during t_3 is uncorrelated to the frequency ω_1 sampled during t_1, then the line shape will appear symmetric and homogenous (see discussion in Section 2.6.2). Here "fast" refers to a comparison with the vibrational lifetime or the t_2 waiting time. The parameters are varied to minimize the error using the Fletcher-Reeves conjugate gradient algorithm.[?] The error was defined as

$$err = \left[\sum_{\omega_1, \omega_3 \; grid \; points} \left(Re[S^{Expt}(\omega_1, \omega_3)] - Re[S^{Calc}(\omega_1, \omega_3)] \right)^2 \right]^{1/2}, \quad (3.22)$$

and the results appear in Tab. 3.2.

The cross-peaks in the 2D IR spectra are a direct indication of coupling between

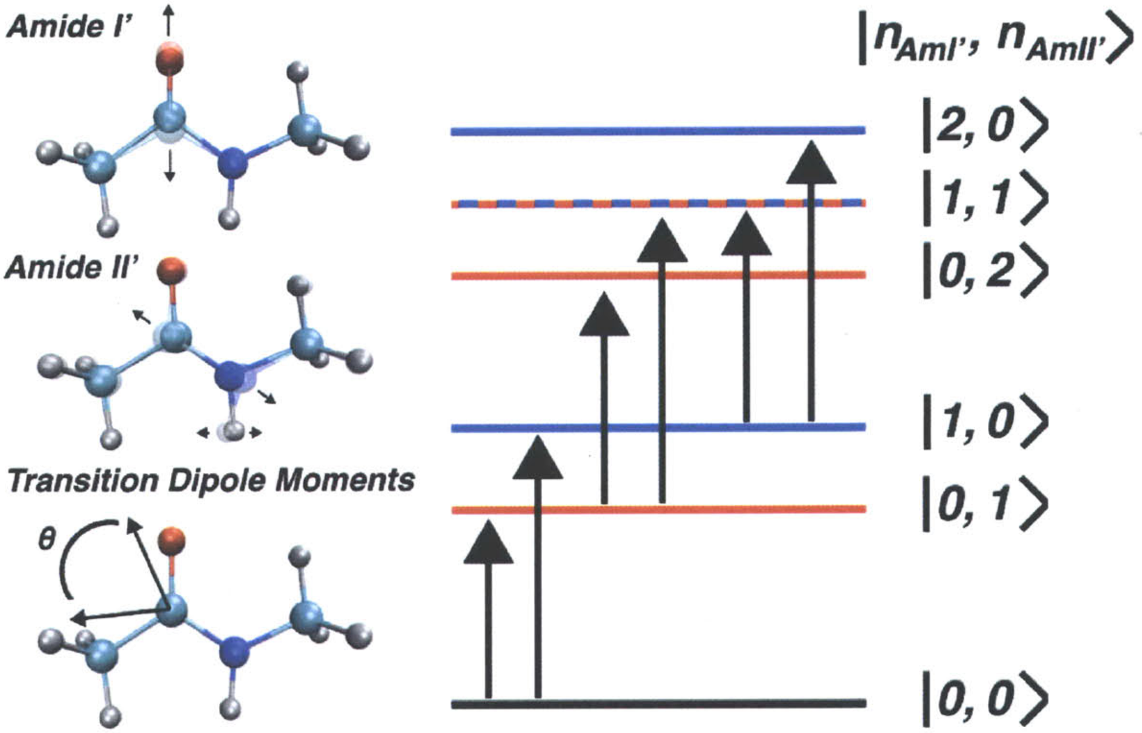


Figure 3-4: Six level system for NMA-d7 in D_2O

the amide I' and amide II' modes. Since these oscillators share atomic composition, it was not surprising to see this coupling. However, a common approximation in modeling amide I modes has been to ignore amide II, which implies that the coupling is weak, or

$$\Delta\omega \gg 4V_{12}, \tag{3.23}$$

where $\Delta\omega$ is the splitting between the oscillator energy levels and V_{12} is the bilinear coupling, in the zero-order, localized basis. Using the eigenstate frequencies obtained from the fits for NMA-d_7 in D_2O, a transformation to the zero-order basis yielded $V_{12} = 39$ cm^{-1} and $\Delta\omega = 64$ cm^{-1}. Thus, the weak coupling approximation is not appropriate for the amide I'/II' system. (This was also manifested as coherent beats in the amide I'-II' pump-probe and oscillation in the cross-peak intensities during the waiting time, τ_2.) Another unexpected observation from the fitting was the deviation from harmonic scaling in the transition dipole magnitudes, which can be seen by

Parameter	**Symbol**	**Value**
Vibrational frequency	$\omega_{AmI'}$	1598 cm^{-1}
Vibrational frequency	$\omega_{AmII'}$	1495 cm^{-1}
Diagonal anharmonicity	$\Delta_{AmI'}$	15 cm^{-1}
Diagonal anharmonicity	$\Delta_{AmII'}$	11 cm^{-1}
Off-diagonal anharmonicity	$\Delta_{AmI'-AmII'}$	11 cm^{-1}
Fundamental transition dipole	$\mu_{AmI'}$	1.00 AU
Fundamental transition dipole	$\mu_{AmII'}$	0.58 AU
Overtone transition dipole	$\mu_{2AmI',AmI'}$	1.51(1.41) AU
Overtone transition dipole	$\mu_{2AmII',AmII'}$	0.68 (0.82) AU
Angle between transition dipoles	$\theta_{AmI',AmII'}$	75°
Spectral diffusion time	$\tau_{AmII'}$	750 fs
Spectral diffusion time	$\tau_{AmI'}$	500 fs

Table 3.2: Numerical parameters for six-level system model of NMA-d7 in D_2O. Values in parentheses indicate those expected from harmonic scaling.

comparing the magnitudes of absorptive and emissive peaks in a doublet. In NMA, deviations up to 17% were observed, indicating significant electrical anharmonicity. For comparison, in rhodium dicarbonyl, the deviation from harmonic scaling was no more than 4%.

By comparing the extracted parameters for NMA-d_7 in D_2O and NMA-h_7 in DMSO, it was observed that the angle between the amide I' and II' transition dipoles shifts from 75° to 40°. *Ab initio* calculations were able to reproduce these values (72° and 41°, respectively) and were analyzed to show isotopic substitution is largely responsible for this effect by increasing the NH in-plane bend component when NMA is protonated.

This model was suitable for extracting the energy level diagram and transition dipole orientation for NMA. However, the model was not sophisticated enough to include coherent effects, such as were observed in the pump-probe and 2D IR cross-peaks by preparing a linear combination of amide I and II with a short pulse. The preparation of coherences and also population relaxation, such as from amide I into amide II, can be included by modeling the density matrix for the system[90] or by introducing these effects into the analytical model for the lineshapes.[102]

The strengths of this model is that it can reproduce the data very well in terms of a small number of parameters by direct fitting of the experimental data. However, such analysis can only take place in the limit of weak spectral congestion. The remaining examples demonstrate models that are useful for congested vibrational spectra of proteins and peptides.

3.3 Modeling the Amide I'-II' 2D IR Spectra of Idealized Secondary Structures

The observation of moderate-strong coupling between the amide I'/II' oscillators in NMA motivated a study on the amide I'-II' spectroscopy of polypeptide systems that more closely approximate the amide spectroscopy of proteins. Since it was known that the 2D IR lineshapes of amide I' modes encode secondary structure sensitivity, it was worth exploring whether the 2D IR lineshapes of amide II' modes or amide I'-II' crosspeaks also display sensitivity to structure due to the strong coupling. Amide I'-II' spectra of three forms of poly-L-lysine (β sheet, α helix, and unstructured) were collected and are shown in Figure 3.3.1.

3.3.1 Experimental Results

After acquiring the data shown in Figure 3.3.1, several unexplained observations were made. It was noted that the amide II' diagonal lineshapes do vary by secondary structure. In the ZZZZ spectra (Figure 3.3.1A), the amide II' region of β sheets shows a broad fundamental and overtone lineshape, and an additional, weaker peak at higher ω_3, 1483 cm^{-1}. In the amide II' lineshapes of α helices, two peaks were also seen, but the lower frequency, 1440 cm^{-1} peak was weaker. The amide II' lineshape of unstructured poly-L-lysine appeared diagonally elongated and similar to the amide I' lineshape. Was this structure in the amide II' band due to coupling with amide I' or due to inherent secondary structure sensitivity in amide II'?

Moreover, all of the amide II' lineshapes showed a polarization dependence. Comparing the ZZZZ and ZZYY spectra (Figure 3.3.1A and B), the more intense higher

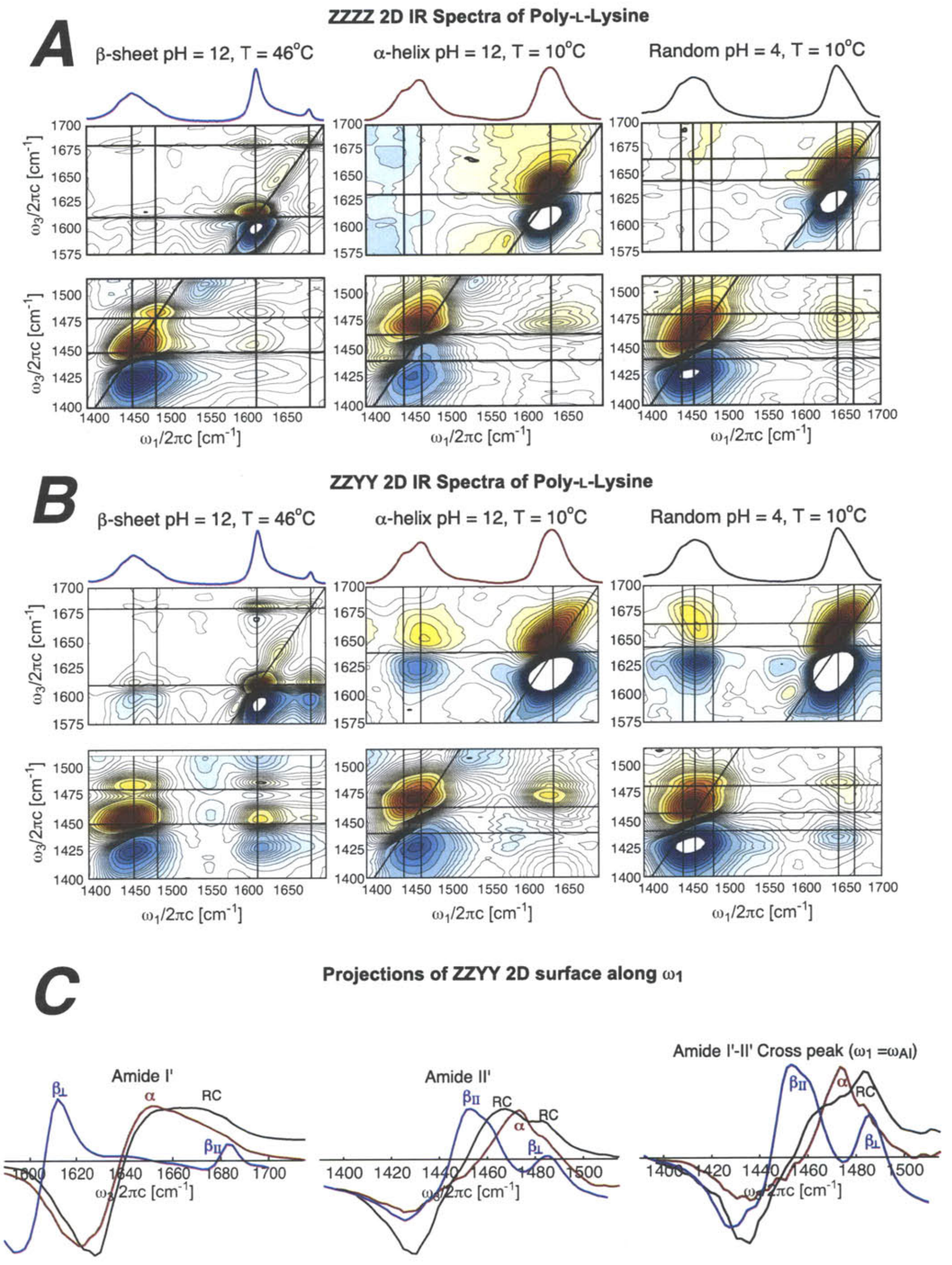


Figure 3-5: Experimental amide I'-II' spectra of poly-L-lysine in its three forms. 2D IR spectra in (A) ZZZZ and (B) ZZYY polarization are plotted along with (C) ω_1 projections. Contours are plotted from ±80% of the amide I' normalized peak at a 5% spacing. For the β-sheet conformation, contours are plotted at 1.25% between ±20% to enhance subtle features of the cross peak region.

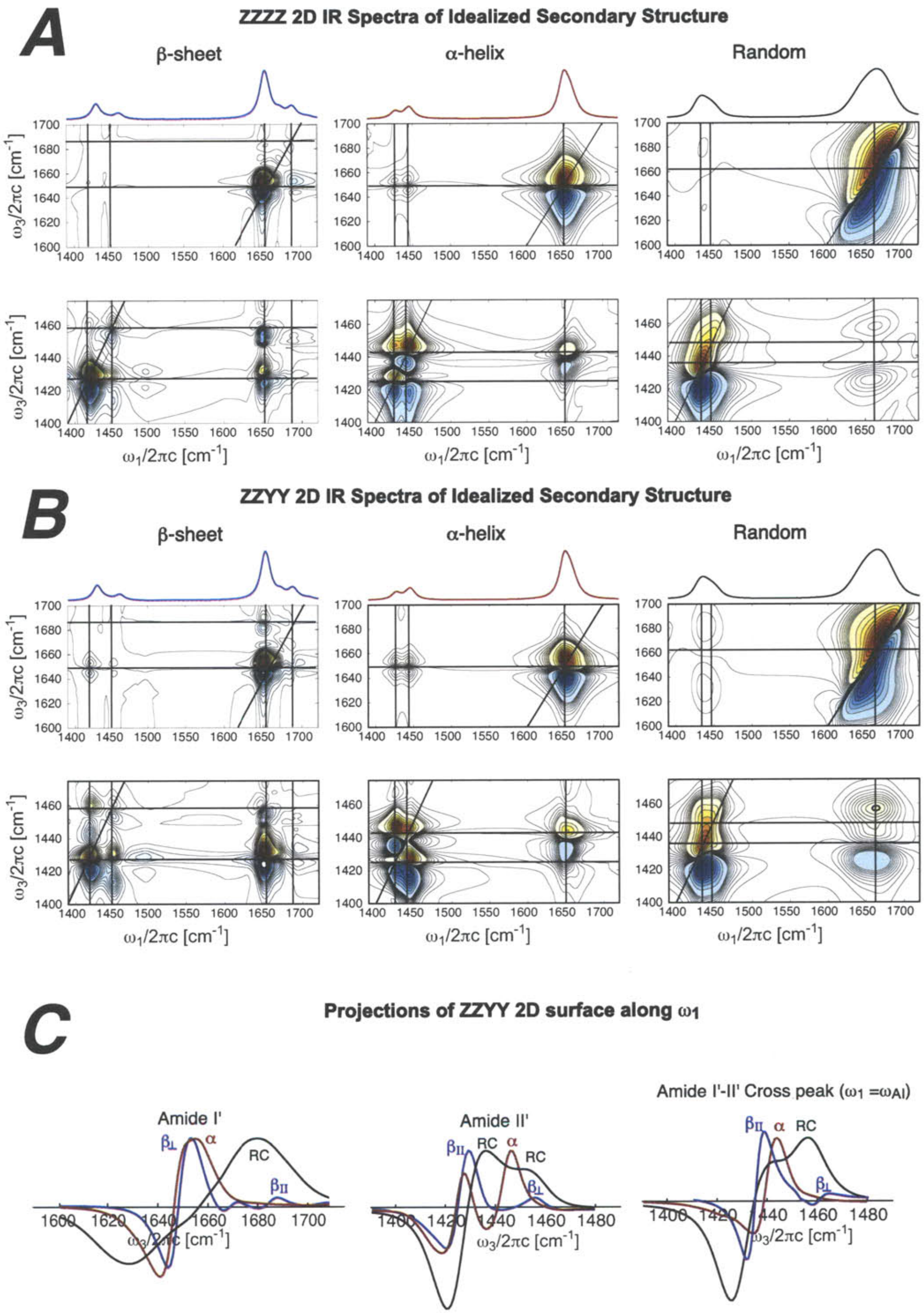


Figure 3-6: Calculated amide I'-II' spectra of three idealized secondary structures. 2D IR spectra in (A) ZZZZ and (B) ZZYY polarization are plotted along with (C) ω_1 projections. Contours are plotted from ±100% of the amide I' normalized peak at 7% spacing and 1.25% spacing for ±15%.

frequency amide II' β-sheet feature shifted from ω_1=1475 cm^{-1} to 1450 cm^{-1}. The ω_3-elongated overtone of the α-helix lineshape became more symmetric and the lower frequency diagonal peak became weaker in the ZZYY spectra. Most surprisingly, the amide II' lineshape of unstructured poly-L-lysine because ω_1-elongated at ω_3 =1680 cm^{-1} in the ZZYY spectra, which is typically indicative of underlying vibrational structure. Upon retrospective analysis, the amide I' lineshape also displayed ω_1-elongation. How could this be possible for an ensemble of unstructured polypeptides?

All of the 2D IR spectra show cross-peaks in the amide I'-II' region. The 2D IR spectra of unstructured poly-L-lysine was also puzzling due to the large splitting between the positive and negative cross-peaks- approximately 10 cm^{-1} larger than in the corresponding α-helix and β-sheet spectra.

3.3.2 Amide I'-II' Model

The amide I'-II' model built upon earlier work to describe the amide I' 2D IR spectra of idealized secondary structures, such as parallel β sheets, anti-parallel β sheets, and α helices. In these models, the site energies were degenerate, or sampled from a normal distribution with a set mean and standard deviation. (See Section 4.2 for a detailed description of the amide I Hamiltonian.) Since the structures are periodic, identification of a unit cell allows for a small number of coupling values to model the full structure. The coupling values within a unit cell can be obtained by *ab initio* calculations, electrostatic models (such as transition dipole coupling), or constrained by experiment.

One deficiency that is inherent to the simplifying approximations in this model is the description of the lineshapes. It is assumed that each transition is only homogeneously broadened (see Section 2.6.2 and Figure 4-1 for a description of line broadening). Inhomogenous broadening can be added by summing the spectra resulting from stoichastic variations in one of the parameters, such as in the site energy for each site. Such a separation is equivalent to the static limit, in which structures never interconvert. To describe the effects of vibrational dynamics that are not infinitely fast or infinitely slow, or to predict the results of a waiting-time experiment, a more sophisticated lineshape model is required.

To extend these models to describe amide II', a block-structured Hamiltonian was constructed, which is diagrammed in Fig. 3-7. The amide I' and II' modes each comprise a block of the Hamiltonian, and each block contains site energies and couplings for the intra-band coupling; amide I'-II' couplings would lie on the off-diagonal elements bridging the two blocks. The amide I' block was treated as before. From the results on NMA described in Section 3.2, a value for the on-site amide I'-II' coupling was obtained and an orientation for the amide II' transition dipole. Comparison to the acquired 2D IR spectra was used to constrain values for amide I'-II' and amide II'-II' couplings and amide I'-II' couplings in the simplest way that captured the essential experimental observations.

Fig. 3-8 shows the amide II' diagonal lineshape resulting from a calculation that was used to discern the physical effect responsible for the amide II' splitting in anti-parallel β-sheet spectra. In panel A, the amide II' block of the Hamiltonian was tridiagonal to describe a manifold of degenerate amide II' modes that are coupled to their nearest neighbors (NN) and the amide I'-II' coupling was set to zero. In panel B, the amide II' Hamiltonian is diagonal with zero coupling values, however, the on-site amide I'-II' coupling was set to the value from NMA. This would potentially allow for the spitting within the β-sheet amide I' band to cause structure in the amide II' band. From these data, it was concluded that the large amide I'-II' energy gap ($\approx$ 100 cm^{-1}) prevents the mode structure of amide I' from inducing amide II' splitting. However, introducing simple, nearest neighbor coupling causes spectral changes that are qualitatively similar to the experimental amide II' data.

Surprisingly, simply adding nearest-neighbor amide II' coupling produced the amide II' lineshapes shown in Figure 3.3.1, which capture the experimentally observed features seen in Figure 3.3.1. The magnitude of the coupling was set by modeling the peak splittings. The sign of the coupling was set by comparing the peak intensities in ZZZZ and ZZYY spectra, which are sensitive to relative transition dipole orientation. The FTIR spectra capture the two mode structure of amide II' with the different

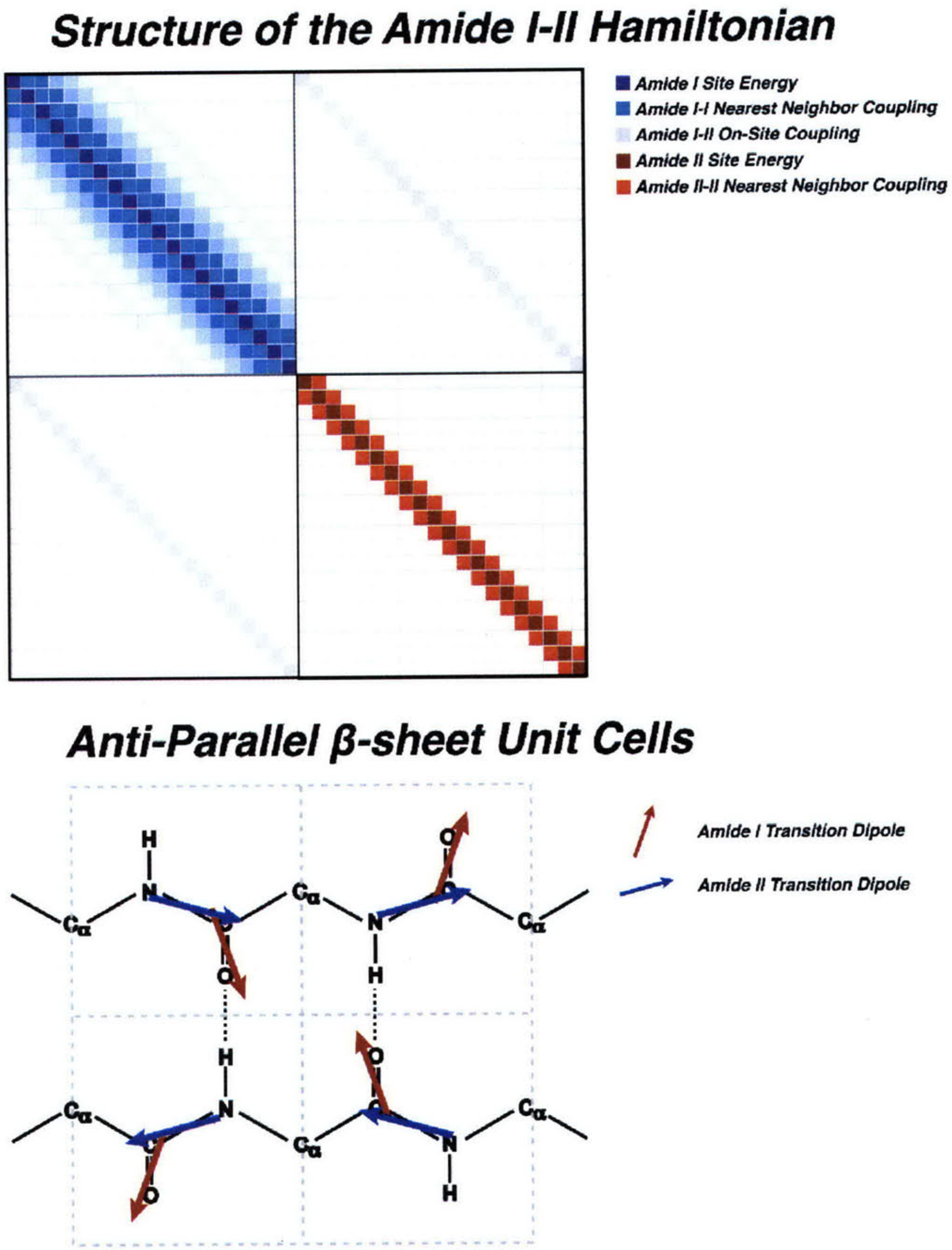

Structure of the Amide I-II Hamiltonian
Amide I Site Energy
Amide I-I Nearest Neighbor Coupling
Amide I-II On-Site Coupling
Amide II Site Energy
Amide II-II Nearest Neighbor Coupling
Anti-Parallel β-sheet Unit Cells
Amide I Transition Dipole
Amide II Transition Dipole
H
N
O
Cα

Figure 3-7: The Amide I-II Model Hamiltonian and Transition Dipoles. (Top) A color-coded visualization of the amide I-II Hamiltonian. Each square on the diagonal represents a site and each off-diagonal square represents a coupling value. The top left quadrant is the amide I manifold and the bottom right quadrant is the amide II manifold; the coupling between the amide I and amide II blocks is diagonal, which indicates that amide I-II couplings are only non-zero if the oscillators are at the same backbone site. (Bottom) A visual representation of the transition dipoles.

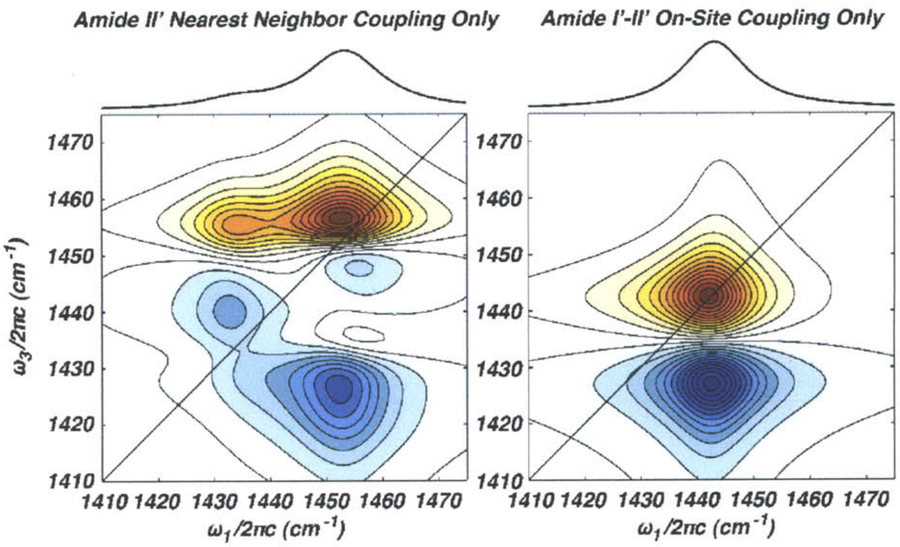


Figure 3-8: Comparison of Amide II' Coupling Mechanisms- Nearest Neighbor Coupling and Mutual Coupling to Amide I'.

intensity variations for sheets and helices. General trends in the position and amplitude of diagonal and cross peaks, and the interference patterns they participate in are largely reproduced. The comparison in Figure ?? also shows projections along ω_1. For the unstructured region spectra, line shapes and intensities of positive and negative contributions are well-reproduced by the calculation, indicating that the two peak structure is not an indicator of a random contour to the chain, but of through-bond coupling that has weak dependence on conformation. The intensity of the low frequency amide II α-helical peak is exaggerated in the calculation and appears as only a shoulder in the experiment. The splitting and intensities of the β sheet projections match the experiment reasonably well. All together these results explain the intensity variation between positive and negative lobes of the peaks without invoking electrical anharmonicity, but rather the result of interference. Agreement between the experiment and calculations demonstrates two points. First, only nearest neighbor coupling is required to reproduce the splitting between amide II vibrations. Second, the sign and magnitude of this nearest neighbor coupling can be unambiguously obtained from the intensity ratio and splitting observed in the amide I-II 2D IR spectra. These parameters are summarized in Table 3.3.2.

Parameter	Amide II' Value	Amide I' Value
Mean Site Energy	1450 cm^{-1}	1650 cm^{-1}
Site Energy St. Dev. σ	10 cm^{-1}	10 cm^{-1}
Anharmonicity	10 cm^{-1}	16 cm^{-1}
Transition Dipole Magnitude	0.58	1.00
Transition Dipole Angle with CO Bond	110	20
α helix NN Coupling	7.5 cm^{-1}	-8.7 cm^{-1}
β sheet NN Coupling	0.67 cm^{-1}	-8.5 cm^{-1}
Unstructured Regions NN Coupling	TDC [103]	-5 cm^{-1} and σ= 3 cm^{-1}
Non-Nearest Neighbor Coupling	0 cm^{-1}	TDC[103]
Amide I' II' On-Site Coupling	39 cm^{-1}	

Table 3.3: Amide I'-II' Modeling Parameters.

This type of model is useful for a first-order description of the spectroscopy. It describes the symmetry of the most important eigenstates for a chosen secondary structure and what patterns of diagonal and cross-peaks are expected. These features can be examined as a function of a small number of parameters. It can be used to predict the size-dependence of the spectra. In some cases, the spectra of real proteins resemble those of idealized structures. For example, concanavalin A and myoglobin display features similar to those of ideal anti-parallel β sheets and α helices. The limitation of this type of model is in its accuracy. It is questionable whether even poly-L-lysine forms idealized secondary structures. Proteins typically contain more than one type of secondary structural motif and plenty of disorder. The next chapter shows a method for improving this description by numerically calculating 2D IR spectra from atomistic molecular dynamics simulations.

3.4 Afterword and Acknowledgements

Sections of this chapter were excerpted from "The Anharmonic Vibrational Potential and Relaxation Pathways of the Amide I and II Modes of *N*-methylacetamide"[6] and "Amide I'-II' 2D IR Spectroscopy Provides Enhanced Protein Secondary Structural Sensitivity" by Deflores et al.[104] I would like to thank Lauren P. DeFlores, Rebecca

Nicodemus, and Hoi Sung Chung for performing the experiments.

Chapter 4

Simulating 2D IR Spectra Using MD Simulations: Spectral Signatures of Heterogeneous Protein Ensembles

4.1 Introduction

Amide I is the backbone CO stretching region of a protein infrared spectrum (1600 – 1700 cm^{-1}). Its large transition dipole moment ($\sim$4 $D\text{Å}^{-1}amu^{-1/2}$ per oscillator)[105] both makes it the strongest feature in a protein infrared spectrum and causes the eigenstates to be delocalized by long range electrostatic coupling. As a result of this delocalization and the inherent fast time resolution of infrared spectroscopy, the amide I lineshape of proteins and peptides has been used as both a rough structural analysis tool and a fast dynamic probe of protein secondary structure. Yet, the amide I lineshape has been difficult to model on the basis of protein structure in a way that incorporates structural plasticity and fluctuating hydrogen bonds.

The amide I region contains roughly as many eigenstates as there are peptide linkages, but with a natural linewidth of $\sim$10 cm^{-1}[106, 107], spectra become congested very quickly with increasing peptide size. This dense set of states is sensitive to many intramolecular structural variables as well as intermolecular solvation forces all evolving on timescales from tens of femtoseconds to nanoseconds, and so the observed FTIR lineshape is broad with few features- typically one pronounced peak and perhaps a shoulder. This mismatch between input parameters and observables has complicated the ability to understand the relative importance of coupling and solvent effects. Thus, except in the case of very small peptides,[108] amide I FTIR of proteins has largely been limited to interpreting group frequencies whose assignment is often made by unphysically motivated Fourier band deconvolution.[109–111]

The first observation of the structual sensitivity of amide I[112], namely that α-helical systems peak about 20 cm^{-1} higher in energy than β-sheet systems, led to theories attributing the difference in frequency to local hydrogen bonding environment[113] and strong coupling[114]. Subsequent work answered the original question with the conceptually simple transition dipole coupling model[115], but studies of NMA in different hydrogen bonding motifs [116–120] showed that both are important. An empirical correction to the amide I frequency of NMA due to hydrogen bonding was proposed and widely used[106]. The observation that the frequency shift is additive with increasing hydrogen bonding partners[119] suggested a purely electrostatic cause with negligible effects of induction, covalency, or dispersion, and led to the development of several more accurate electrostatic models correlating the amide I frequency with the potential or electric field at the amide atomic sites[121–125]. This conveniently suggests a concerted treatment of molecular dynamics (MD) and amide I IR spectroscopy calculation, as the forces and structural variations from MD are direct parameters in an imposed spectroscopic Hamiltonian. This ability to calculate observables from MD simulations provides an important link and test of the wealth of dynamical information from MD that has gone largely untested- most of the tests of MD simulations are thermodynamic (binding and folding free energies), kinetic (rate constants), or structural (comparison to x-ray or NMR structures), rather than mechanistic. Where coherent reaction initiation is possible, transient spectra can be measured and calculated from non-equilibrium MD trajectories, providing a rigorous test of the predicted mechanism.

The recent developments in two-dimensional infrared spectroscopy (2DIR) allow a new understanding of the amide I lineshape based on its ability to distinguish between natural line widths and heterogeneous ensembles and the appearance of structurally-sensitive cross peaks. It will be shown that the approximations sufficient to yield good agreement in calculated FTIR spectra predict 2DIR lineshapes that disagree with experiment. This is because the 2DIR spectrum contains additional information; inhomogeneous broadening appears as a diagonal ($\omega_1 = \omega_3$) elongation of the lineshapes and correlated energy shifts between any two eigenstates A (ω_A) and B (ω_B) are revealed in the A-B cross peak shape at (ω_A, ω_B) and (ω_B, ω_A)[126]. Thus, it is possible in FTIR, but not 2DIR, to use unphysically broad homogeneous lineshapes to mask a lack of knowledge of the correct inhomogeneous broadening in the system. This sensitivity of 2DIR to ensemble heterogeneity provides a rigorous test of equilibrium MD trajectories to correctly sample disorder in the system. There is a lot of interest on this front and other researchers are also pursuing structure-based models of FTIR and 2DIR spectra from MD simulations[127–129].

It is our goal in this work to test a model with no fitting of IR spectra for calculating the amide I lineshape that is consistent with both observed FTIR and 2DIR data. The only external parameters that enter are obtained from other IR experiments or MD simulations. We work within the established framework of a subspace of amide I local modes[106] and evaluate site energy models with MD simulations to incorporate fluctuations in the solute and solvent. We demonstrate the modeling with three test systems that survey different secondary structure motifs- a 12 residue β-hairpin (trpzip2), a 19 residue α-helix (D-Arg), and a 76 residue $\alpha + \beta$ protein (ubiquitin). We find that including electrostatic contributions from the solvent, side chains, and backbone residues are important in establishing the correct redshift to the overall amide I band position. We find that disorder in the site energies, but not couplings, are critical for experimental agreement with calculated 2DIR spectra, which implies that disorder in site energies is critical for calculating physically meaningful FTIR spectra. These methods, in spite of the approximations made, yield good results with lineshapes slightly broader than observed. Systematic improvements are also suggested.

4.2 Methods

4.2.1 The Amide I Subspace

The main simplifying assumption that allows calculation of the amide I lineshape is the description of amide I eigenstates as linear combinations of local amide I vibrations, which ensures that for a polypeptide comprised of $(N+1)$ amino acids and N peptide groups, there will only be N relevant local modes to consider. This is valid insomuch as the local amide I vibration is localized to a single backbone unit and was first introduced by Miyazawa[114] with important contributions from Krimm[130,131] and Torii and Tasumi[105,119,132], who used the GF matrix formalism. This local amide I Hamiltonian picture showed that the FTIR lineshape could be well approximated by considering local oscillator force constants (diagonal elements) and coupling between local oscillators (off-diagonal elements) without considering side chains, solvent, or disorder. Because the G matrix is diagonal, an isomorphic representation can be derived by directly considering parameters of the Hamiltonian in units of cm^{-1}, an approach that was suggested with the excitonic modeling of 2DIR spectra by Hamm and Hochstrasser and subsequently expanded[106,133]. We follow this approach, which means that our site energies (diagonal elements) reflect both the local oscillator force constant and reduced mass.

Site Energy Models

We considered ten different site energy models. Eight of these models are based on electrostatics and their salient features are summarized in Tab. 4.1. We also consider two other models: the degenerate case (all sites = 1650 cm^{-1}), which was the starting point for Torii and Tasumi's work on globular proteins[105], and an empirical heuristic site energy model based on probable hydrogen bonds[12] from the crystal structure. For simplicity at this point, we ignore non-electrostatic contributions to the site energy[134]. We find that many field and potential models compare favorably with one another, and therefore representative results from four of the ten site energy models are plotted in Figure 4-3, and the complete set is included in Figures 4-6 and 4-7. In each of the electrostatic models, the site energy for each peptide group, I, was parameterized by correlation of the calculated amide I frequency to the electrostatic environment evaluated at n various sites i (typically C, O, N, D) in N-methylacetamide (NMA). We calculate the electrostatic environment by considering the protein-solvent system as a cluster of point partial charges perturbing one amide group. In the most general form, the electrostatic models we study define the amide I frequency for site I, ω_I, as

$$\omega_I = \omega_0 + \sum_{i=1}^{n} \left[\lambda_i \phi_i + \sum_{\alpha} \chi_{i,\alpha} E_{i,\alpha} + \sum_{\alpha > \beta} \gamma_{i,\alpha\beta} E_{i,\alpha\beta} \right]. \tag{4.1}$$

The site energy is linearly proportional to the scalar electrostatic potential,

$$\phi_i = \sum_{m \neq \{i\}}^{M} \frac{q_m}{4\pi\epsilon_0} \frac{1}{r_{mi}} \tag{4.2}$$

the vector electric field,

$$E_{i,\alpha} = \sum_{m \neq \{i\}}^{M} \frac{q_m}{4\pi\epsilon_0} \frac{\vec{r}_{mi} \cdot \hat{r}_\alpha}{r_{mi}^3} \tag{4.3}$$

and the tensorial gradient of the electric field,

$$E_{i,\alpha\beta} = \sum_{m \neq \{i\}}^{M} \frac{q_m}{4\pi\epsilon_0} \left[\frac{\delta_{\alpha\beta}}{r_{mi}^3} - \frac{3\left(\vec{r}_{mi} \cdot \hat{r}_\alpha\right)\left(\vec{r}_{mi} \cdot \hat{r}_\beta\right)}{r_{mi}^5} \right], \tag{4.4}$$

from all atoms $m \in M$ outside the peptide group. A site energy model is defined by the terms in Eq.4.1 which are included and the correlation coefficients λ_u, χ_u, and γ_u that dictate frequency shifts and the initial ω_0, which is typically the gas phase value but may be parameterized to include a static shift[121, 122, 124]. In Eqs.4.1, 4.3, and 4.4, Greek indices run over the amide group space ($\alpha, \beta \in \{x, y, z\}$). The CO bond of each amide group defines a y-axis. The x-axis is perpendicular to the y-axis and points towards the amide N, in the CON plane, and $\hat{z} = \hat{x} \times \hat{y}$. Although all of these electrostatic site energy models are *acausal* by construction, that is the electrostatics of an amide group are fit to a frequency shift without reference to the underlying mechanism, the coefficients λ_u, χ_u, and γ_u can be understood as vibrational Stark tuning rates[135,136]. In analogy to work on HOD/D$_2$O where $\Delta\omega \approx -qE\left(Q_{11} - Q_{00}\right)/\hbar$[137], the coefficient for each atom can be roughly identified as a measure of the partial charge times the mean displacement in amide I.

In the parameterization of these site energy models, the electrostatics of the solvent are correlated to the amide I frequency of NMA, not to a peptide group in a protein. We extend the original definition of the solvent to include not only the water, but also the side chains and neighboring backbone. The chromophore is defined to be a peptide group in a chain. The sums over m that define ϕ and E for a peptide group do not include the nitrogen or alpha carbons flanking the carbonyl nor the associated backbone protons. This definition maintains electrostatic neutrality of the system in CHARMM27 and ENCAD forcefields, as well as several others.

We also consider a heuristic site energy model based on structure[12]. We start with an average solvated frequency of 1688 cm^{-1} and apply a set of redshifting criteria,

$$\omega_I = 1688\ cm^{-1} - \delta\omega_I^{CO} - \delta\omega_I^{NH}. \tag{4.5}$$

For backbone hydrogen bonding on the carbonyl side,

$$\delta\omega_I^{CO} = 30\ cm^{-1}\left(2.6\text{Å} - r_{H-O}\right) \tag{4.6}$$

where r_{H-O} is the hydrogen-oxygen backbone hydrogen bond distance[106]. For water hydrogen bonding on the carbonyl side, $\delta\omega_I^{CO} = 20\ cm^{-1}$. For any hydrogen bond on the amino side, we redshift by $\delta\omega_I^{NH} = 10\ cm^{-1}$. For no hydrogen bonding, $\delta\omega_I^{NH} = 0$ and/or $\delta\omega_I^{CO} = 0$.

Coupling Model

Historically, transition dipole coupling was used to explain the splitting of the amide I band observed in β-sheets, but subsequent calculations[138] highlighted the failure of the dipole approximation at short distances and the importance of through-bond effects. In light of this, we use a DFT look-up table[139] for coupling between bonded neighbors, which is a function of the backbone torsion angles between amide units, commonly referred to as (ϕ,ψ). This coupling map can be roughly understood as the second derivative of the total calculated electronic energy with respect to local amide I coordinates Q_I and Q_J in a model dipeptide.

$$\beta_{IJ}^{DFT} = \frac{\partial^2}{\partial Q_I \partial Q_J} V_{IJ}^{DFT} (Q_I, Q_J) \tag{4.7}$$

This implicitly treats the quantum mechanical, through-bond effects most relevant for covalently bonded amide groups as well as electrostatics to all order. All other interactions are treated with transition charge coupling[121, 132, 138]. The TCC coupling equation represents the interaction energy between a cluster of partial charges on peptide groups I and J as a function of local coordinate displacements, which circumvents the dipole approximation of TDC and includes polarizability effects by allowing the charges to change magnitude. The TCC coupling is defined by

$$\beta_{IJ}^{TCC} = \frac{1}{4\pi\epsilon_0} \frac{\partial^2}{\partial Q_I \partial Q_J} \sum_{i,j}^{n} \frac{\left(q_i^0 + q_i^1 Q_I\right)\left(q_j^0 + q_j^1 Q_J\right)}{\left|\vec{r}_i\left(Q_I\right) - \vec{r}_j\left(Q_J\right)\right|} \tag{4.8}$$

where Q_I is the dimensionless local coordinate,

$$Q_I = \sqrt{\frac{m_I \omega_I}{\hbar}} Q'_I \tag{4.9}$$

and q^0 and q^1 are the parameters from a linear fit to the calculated Mulliken charge at each atomic center as a function of local amide I coordinate displacement on N-methylacetamide, with the methyl hydrogen partial charges summed to the methyl carbons. Calculations to parameterize TCC were carried out using the Gaussian 98 implementation of B3LYP/6-31+G*[140]. The transition dipole is obtained by numerical differentiation and found to lie in the NCO plane, 25° off the CO bond axis towards the nitrogen. The magnitude and orientation with respect to the amide group are fixed[125]. The index $i(j)$ runs over all atoms of the peptide group $I(J)$.

4.2.2 MD Simulations

In the present case MD simulations are used to generate an ensemble of structures representative of the experiment. With static averaging (or "instantaneous normal modes"), FTIR and 2DIR spectra are summed for each structure in a set representative of the equilibrium ensemble. In a dynamical picture, this corresponds to a slow interconversion of spectroscopically distinct structures on a timescale longer than the experiment ($\sim$2 *ps*). This approximation leads to artificially broader lineshapes if there are any spectroscopically relevant fast fluctuations. This is undoubtedly the case to some degree as water undergoes many fast motions that can modulate the amide I frequency. However, in the interest of computational and conceptual simplicity, this work operates purely in the static average picture.

We choose three systems to test the generality of the models that span different secondary structure motifs and sizes. Trpzip2 is an extensively studied[4, 141–146] 12 residue peptide that forms a stable, anti-parallel β-hairpin in water. D-Arg is a 19 residue alanine-rich α-helical peptide, based loosely on the Marquesee-Baldwin water-soluble α-helical peptide[147, 148]. In addition to an α-helix and a β-hairpin, we consider a 76 residue $\alpha + \beta$ protein ubiquitin to test on a realistic protein.

The initial trpzip2 structure is taken from the published solution NMR structures (PDB 1LE1 [143]). MD simulations are carried out in the CHARMM 30b1 package[149]. All acidic protons are deuterated to match the experimental conditions. This structure is briefly energy minimized for 1000 steps in the CHARMM27 force field. Thereafter in the simulation, the bond lengths are constrained with the SHAKE algorithm. The peptide is then placed in the center of a preequilibriated (298K, 1g/cc) box of 2048 SPC/E water molecules with periodic, cubic boundary conditions. Waters with oxygens within 2Å of the peptide are removed. Because at pH 7.0 the peptide has a formal charge of +1, a Cl^- counterion was added at a random place in the box. Electrostatics are handled by particle mesh Ewald sums and VDW interactions are shifted to reach zero at the truncation length of 14Å. Ten trajectories are spawned from this initial structure with different initial seeds and each equilibrated for 1 ns in the NPT ensemble (Berendsen method, 298 K, 1 atm), which is long enough to allow the water to reorient and let the box length converge. Dynamics are continued in the NPT ensemble and the structures are saved at 20 fs intervals for analysis for 1 ns total of trajectory.

The initial D-Arg structure was constructed assuming the secondary structure generated an ideal helix (repeating ϕ =-90^o and ψ =-45^o) for the residues NH_3-$YGG(KAAAA)_3$-$(D\text{-}R)$-$CONH_2$, where the Arg is of right-handed chirality. The remaining procedures are identical to trpzip2, except five randomly placed Cl^- counterions were added for charge neutrality.

Ubiquitin structures are generated from equilibrium MD performed by Alonso and Daggett[150] and sampled at 300 ps for IR spectral calculations. Spectra were calculated from the protonated structures assuming that all acidic protons can be deuterated without significantly affecting the dynamics.

4.2.3 IR Spectroscopy

From the MD simulation, each $(N+1)$ residue, solvated protein or peptide structure defines a vector of N local transition dipole vectors and an $N \times N$ Hamiltonian composed from the site energies and couplings as described above in the local amide I basis. Diagonalization of this Hamiltonian yields a set of energies, ω_k. The Hamiltonian diagonalization transformation is used to transform the local amide I transition dipoles to the transition dipoles, $\vec{\mu}_k$.

An FTIR stick spectrum for a structure is generated from the eigenstate energies and transition dipoles and then lifetime broadened and summed over the ensemble:

$$I(\omega) = \left\langle \sum_{k=1}^{N} \mathrm{Im} \frac{-|\vec{\mu}_k|^2}{\omega - \omega_k + i\gamma} \right\rangle \tag{4.10}$$

where ⟨⟩ indicates the equilibrium Boltzmann average. While in some studies the HWHM broadening parameter γ is fit, we set it to 5 cm^{-1} for all systems corresponding to a lifetime of ~1 ps[106]. Note additionally that inhomogeneous broadening arises naturally by the sum over structures.

In practice, the spectra are summed as stick spectra on a grid of 1 cm^{-1}, then convolved with the Lorentzian lineshape function afterwards for an increase in computational efficiency by avoiding continuous recalculation of identical lineshapes. The two methods are mathematically identical and indistinguishable with the chosen grid spacing.

$$I\left(\omega_{grid}\right) = \left\langle \sum_{k=1}^{N} \left|\vec{\mu}_k\right|^2 \delta\left(\text{round}\left(\omega_k\right), \omega_{grid}\right) \right\rangle \otimes Im\left[\frac{-1}{\omega_{grid} + i\gamma}\right] \tag{4.11}$$

Two dimensional infrared spectra are calculated by taking the undiagonalized one-quantum Hamiltonian described above for FTIR, H, and defining a scaled two-quantum Hamiltonian, $H^{(2)}$, to include two-quantum local states and couplings, which is zero unless

$$H^{(2)}_{ij,ij} = H_{i,i} + H_{j,j} - \delta_{i,j}A \tag{4.12a}$$

$$H^{(2)}_{ii,jk} = \sqrt{2}\left(H_{i,k}\delta_{i,j} + H_{i,j}\delta_{i,k}\right); ii \neq jk, j \neq k \tag{4.12b}$$

$$H^{(2)}_{ij,jk} = H_{i,k}; i \neq k, \tag{4.12c}$$

where $\delta_{i,j}$ is the Kronecker Delta symbol, A is the overtone anharmonicity, set to 16 cm^{-1}[106], and the indices commute: $H^{(2)}_{ij,kl}$=$H^{(2)}_{ji,kl}$=$H^{(2)}_{ij,lk}$. Part of the increased sensitivity of 2DIR is seen in Eq.4.12a; the two-quantum Hamiltonian is sensitive to correlations in frequency shifts of the fundamental frequencies in a manner that is difficult to guess *a priori.*

The local transition dipoles are also harmonically scaled to produce two-quantum transition dipoles,

$$\vec{\mu}_{ij} = \sqrt{2}\vec{\mu}_i\delta_{i,j} + \vec{\mu}_j\left(1 - \delta_{i,j}\right). \tag{4.13}$$

The two-quantum Hamiltonian is diagonalized to yield a set of $\left(N^2 + N\right)/2$ energies, $\omega_k^{(2)}$, which are used to transform the local transition dipoles and calculate the ZZYY (first two light fields are perpendicularly polarized to the third and detection) rephasing, S_1, and non-rephasing, S_2, signals[70], which when summed yield the purely absorptive 2DIR spectra (Equations 4.14 and 4.15),

$$\begin{aligned} S_1\left(\omega_1,\omega_3\right) = -\operatorname{Im}\Bigg\langle \sum_{d,b}^{N} \Bigg\{ & \frac{\vec{\mu}_d^*\vec{\mu}_d\vec{\mu}_b\vec{\mu}_b^*}{-\omega_1+\omega_d+i\gamma}\frac{1}{\omega_3-\omega_b+i\gamma} \\ & + \frac{i\vec{\mu}_d^*\vec{\mu}_b\vec{\mu}_d\vec{\mu}_b^*}{-\omega_1+\omega_d+i\gamma}\frac{1}{\omega_3-\omega_b+i\gamma} \\ - \sum_{c}^{\left(N^2+N\right)/2} & \frac{i\vec{\mu}_d^*\vec{\mu}_b\vec{\mu}_{cb}^{(2)}\vec{\mu}_{cd}^{(2)*}}{-\omega_1+\omega_d+i\gamma}\frac{1}{\omega_3-\omega_{cd}^{(2)}+i\gamma} \Bigg\}\Bigg\rangle \end{aligned} \tag{4.14}$$

$$\begin{aligned} S_2\left(\omega_1,\omega_3\right) = -\operatorname{Im}\Bigg\langle \sum_{d,b}^{N} \Bigg\{ & \frac{\vec{\mu}_d\vec{\mu}_d^*\vec{\mu}_b\vec{\mu}_b^*}{\omega_1-\omega_d+i\gamma}\frac{1}{\omega_3-\omega_b+i\gamma} \\ & + \frac{i\vec{\mu}_d\vec{\mu}_b^*\vec{\mu}_b\vec{\mu}_d^*}{\omega_1-\omega_d+i\gamma}\frac{1}{\omega_3-\omega_b+i\gamma} \\ - \sum_{c}^{\left(N^2+N\right)/2} & \frac{i\vec{\mu}_d\vec{\mu}_b^*\vec{\mu}_{cd}^{(2)}\vec{\mu}_{cb}^{(2)*}}{\omega_1-\omega_d+i\gamma}\frac{1}{\omega_3-\omega_{cd}^{(2)}+i\gamma} \Bigg\}\Bigg\rangle . \end{aligned} \tag{4.15}$$

In a manner completely analogous to the FTIR, the rephasing and non-rephasing signals are summed independently as stick spectra, then convolved with the appropriate lineshape functions and summed to yield the 2DIR correlation spectrum, as shown in Equations 4.16 and 4.17,

$$\begin{aligned} S_1\left(\omega_{1,grid},\omega_{3,grid}\right) = \Bigg\langle \sum_{d,b}^{N} \Bigg\{ & \left(\vec{\mu}_d^*\vec{\mu}_d\vec{\mu}_b\vec{\mu}_b^* + i\vec{\mu}_d^*\vec{\mu}_b\vec{\mu}_d\vec{\mu}_b^*\right)\delta\left(\operatorname{round}\left(\omega_b\right),\omega_{3,grid}\right) \\ - \sum_{c}^{\left(N^2+N\right)/2} & i\vec{\mu}_d^*\vec{\mu}_b\vec{\mu}_{cb}^{(2)}\vec{\mu}_{cd}^{(2)*}\delta\left(\operatorname{round}\left(\omega_{cd}^{(2)}\right),\omega_{3,grid}\right)\Bigg\}\delta\left(\operatorname{round}\left(\omega_d\right),\omega_{1,grid}\right)\Bigg\rangle \\ & \otimes\left[\frac{-1}{-\omega_{1,grid}+i\gamma}\frac{1}{\omega_{3,grid}+i\gamma}\right] \end{aligned} \tag{4.16}$$

$$S_2\left(\omega_{1,grid}, \omega_{3,grid}\right) = \left\langle \sum_{d,b}^{N} \left\{ \left(\vec{\mu}_d \vec{\mu}_d^* \vec{\mu}_b \vec{\mu}_b^* + i\vec{\mu}_d \vec{\mu}_b^* \vec{\mu}_b \vec{\mu}_d^*\right) \delta\left(\text{round}\left(\omega_b\right), \omega_{3,grid}\right) \right.\right.$$

$$\left.\left. - \sum_{c}^{\left(N^2+N\right)/2} i\vec{\mu}_d \vec{\mu}_b^* \vec{\mu}_{cd}^{(2)} \vec{\mu}_{cb}^{(2)*} \delta\left(\text{round}\left(\omega_{cd}^{(2)}\right), \omega_{3,grid}\right) \right\} \delta\left(\text{round}\left(\omega_d\right), \omega_{1,grid}\right) \right\rangle$$

$$\otimes \left[\frac{-1}{\omega_{1,grid} + i\gamma} \frac{1}{\omega_{3,grid} + i\gamma}\right] \quad (4.17)$$

The indices d and b run over the one-quantum states, and the index c runs over two quantum states. The frequency $\omega_{ci}^{(2)} = \omega_c^{(2)} - \omega_i$. The eigenstate transition dipoles are used to calculate ZZYY (first two fields polarized perpendicular to third and detection) spectra as described in Equation 3.17.

4.2.4 Experimental

The experimental methods employed here are identical to those described previously[3,4]. Trpzip2 (*SWTWENGKWTWK*) and D-Arg[147] (*YGG*(*KAAAA*)$_3$-(*D-R*), where the Arg is of right-handed chirality) were synthesized at the MIT Biopolymers Lab (Cambridge, MA) using Fmoc solid-phase synthesis as C-terminal amide peptides, HPLC purified, then lyophilized against 50 mM DCl to remove residual trifluoroacetic acid. The α-helical character of D-Arg was verified with UVCD (data not shown). BPTI (structure from PDB 1BPI [7]) was purchased from Sigma (St. Louis, MO) and used without further purification.

4.3 Results and Discussion

4.3.1 Information Content of FTIR vs. 2DIR

The increased information content of 2DIR can be used to confirm or reject approximations made in calculating the amide I FTIR lineshape. In Fig.4-1, we show the calculated FTIR and 2DIR spectra for the protein BPTI using the heuristic site energy model based on one crystal structure (1BPI [7]) with several values of the homogeneous broadening parameter, γ, and no inhomogeneous broadening. By comparing

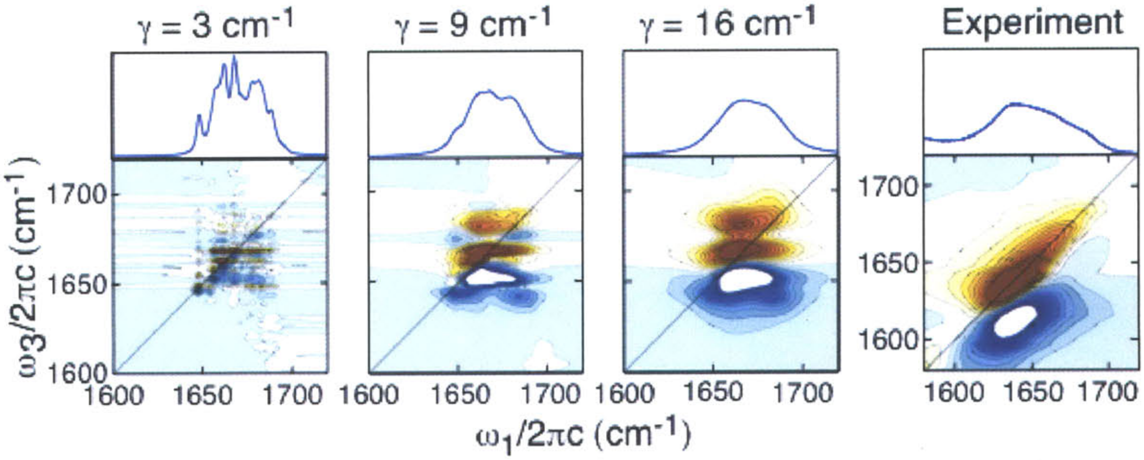


Figure 4-1: Comparison of Homogeneous and Inhomogeneous Broadening Mechanisms. As the Lorentzian HWHM parameter γ is increased, the predicted FTIR linewidth begins to match experiment, but the predicted tilt in the 2DIR peaks is qualitatively incorrect. The spectra are for bovine pancreatic trypsin inhibitor (BPTI) calculated using the heuristic site energy model from the PDB structure 1BPI[7]. 21 evenly spaced contours are plotted from ±50% of the maximum for each spectrum.

the bandshapes, it can be seen that a value of $\gamma \approx 16\ cm^{-1}$ reproduces the shoulder and FWHM of the observed FTIR shape. However, the 2DIR spectrum implied by this homogeneous broadening is qualitatively very different than the observed 2DIR. The measured 2DIR lineshapes are diagonally elongated, a signature of inhomogeneity in the system.

It can be seen in the calculations that when an unphysically narrow linewidth eliminates crowding ($\gamma = 3\ cm^{-1}$), many distinct off-diagonal cross peaks compose the observed lineshapes. For example, a series of cross-peaks extending in the $+\omega_3$ and $+\omega_1$ direction from the $\sim$1650 cm^{-1} fundamental peak arises from anharmonic coupling between these vibrations.

In the homogeneous limit, these cross peaks are round. In the presence of inhomogeneity, correlated frequency shifts between the $\sim$1650 cm^{-1} peak and any of these other peaks will be reflected in the cross-peak shape. For example, a change in the angle between hydrogen-bonded peptide groups will modulate the coupling and can cause a correlated red- and blue-shift in a doublet of peaks. Water penetration into a pair of β-strands can simultaneously redshift many vibrations. The naturally short lifetime of molecular vibrations broadens these peaks beyond resolution and obscures this information in uniquely shaped ridges and bands. None of this is accounted for in a model that neglects disorder.

Although the agreement between the calculated and experimental FTIR spectra can be improved, the qualitative features are reproduced with only homogeneous broadening. Clearly, this Lorentzian approximation is incompatible with the observed 2DIR lineshapes and thus the agreement with the calculated FTIR is reached by overestimating the homogeneous broadening to compensate for neglecting inhomogeneity. These observations motivate the direction of this work; we choose to work towards a model that captures the experimental heterogeneity demonstrated in the 2DIR lineshapes rather than including more parameters to accurately fit the FTIR lineshape.

4.3.2 Static Averaging for 2DIR

In the static averaging technique, a representative ensemble of structures is required to calculate a 2DIR spectrum that sums the individual contributions from each structure. Spectroscopically relevant fluctuations can be expected in two places: site energies (diagonal Hamiltonian elements) and couplings (off-diagonal Hamiltonian elements). The coupling fluctuations arise from flexible secondary structure that changes the orientation and distance between amide I oscillators. The site energy fluctuations arise from the evolving hydrogen bonding environments at each site. This nicely suggests a concerted treatment of dynamics and spectroscopy, since MD simulations simultaneously provide a set of structures as well as a description of the electrostatics.

Figure 4-2 shows how the observed 2DIR spectrum arises from the average of 2DIR spectra of individual structures. In this picture, each component structure generates a distinct 2DIR spectrum with many homogeneously broadened diagonal and cross peaks. The pattern of peaks in each component spectrum is a sensitive indicator of the couplings and site energies for a particular structure. In 2DIR spectra, each homogeneous peak appears as an oppositely signed doublet. When averaged over the

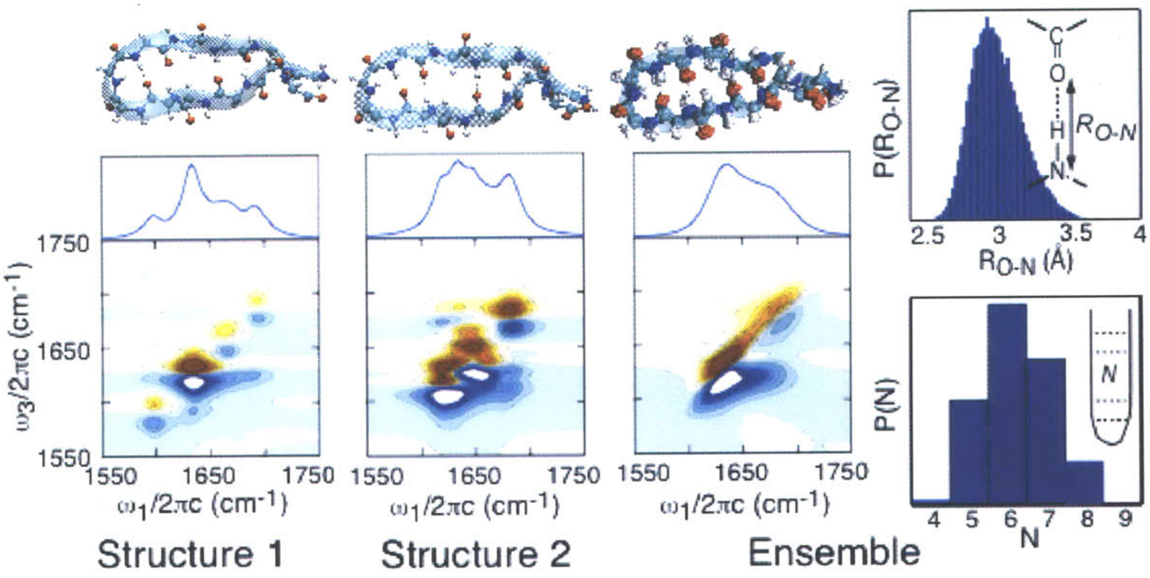


Figure 4-2: Averaging over structural heterogeneity in 2D IR lineshapes. Two sample spectra and structures are shown for slightly different hydrogen bonding environments calculated using the Skinner 6P site energy model. Note that the fine structure disappears and the diagonal elongation of the 2DIR peaks arises from the static average over the equilibrium distribution of hydrogen bonding environments characterized by N, the number of internal hydrogen bonds, and R_{O-N}, the oxygen-nitrogen hydrogen bond distance. 21 evenly spaced contours are plotted from ±50% of the maximum for each spectrum.

equilibrium ensemble, much of the off-diagonal structure disappears as a result of shifting negative and positive amplitudes. The remaining structure reflects the constructive addition of peaks along the diagonal axis, characteristic of inhomogeneous broadening, and ridges of constructively interfering cross peaks that stretch along ω_1[151]. Figure 4-2 shows that although the structures of the ensemble at first glance seem very similar, they vary considerably in the quantities that influence the 2DIR spectrum most: the number and geometry of hydrogen bonds.

Figure 4-3 shows the comparison of calculated 2DIR and FTIR spectra for trpzip2, D-Arg, and ubiquitin against experiment. The agreement in the experimental and calculated spectra from electrostatic site energy models are an improvement over the degenerate model and the heuristic model presented in Figure 4-1 and 4-7, which

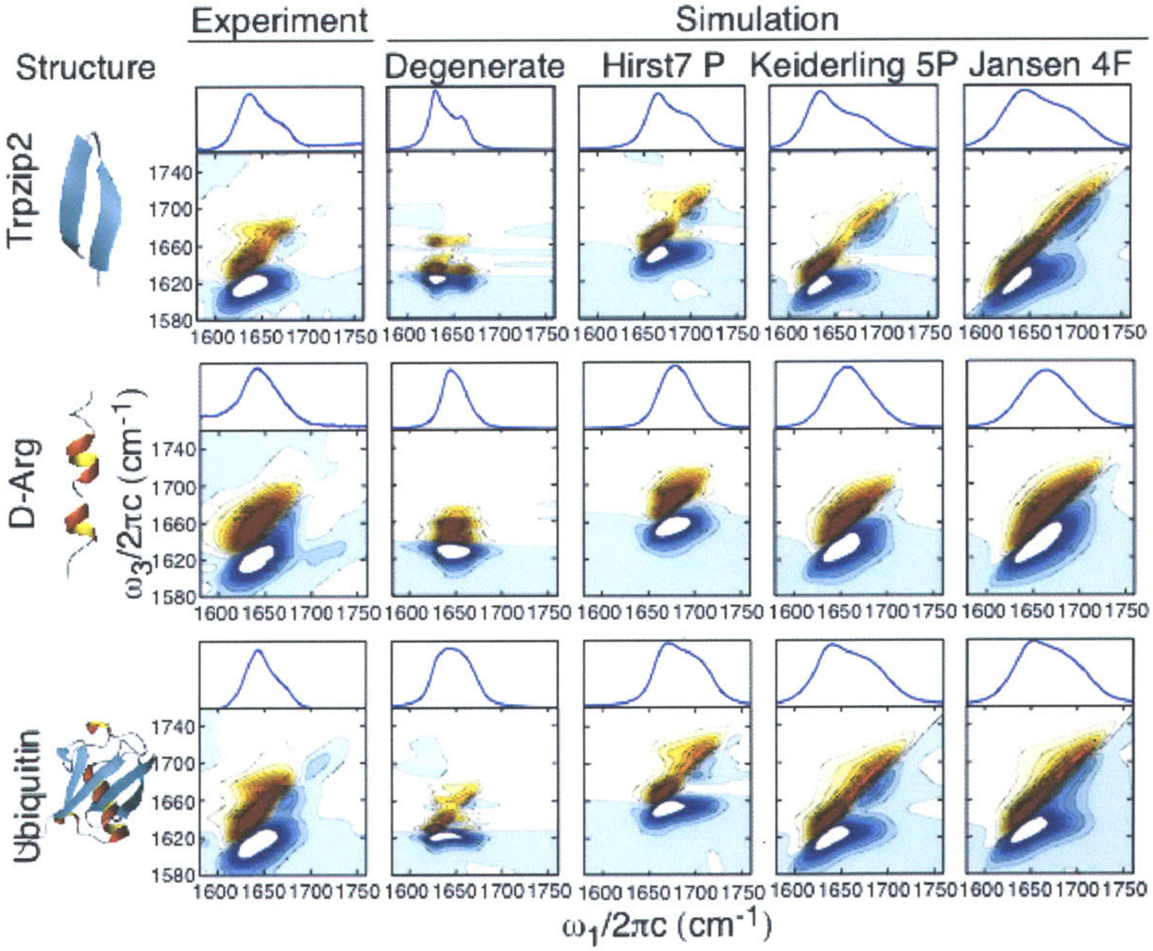


Figure 4-3: Comparison of Calculated and Experimental FTIR and 2DIR for Trpzip2, D-Arg, and Ubiquitin. 21 evenly spaced contours are plotted from ±50% of the maximum for each spectrum.

also assumes fixed (yet different) site energies. The diagonal elongation is much better represented by averaging over many structures in all of the data and using a site energy model sensitive to fluctuating hydrogen bonds. A common deficiency in the spectra calculated from electrostatic models is the overall blueshift relative to experiment, which can be traced to the site energy parameterizations, many of which underestimate the redshift that occurs upon aqueous solvation of NMA from the gas phase. All of the predicted spectra (FTIR and 2DIR) are also too broad, which may result from neglecting motional narrowing effects. It may also be an artifact of the force field parameterization, which implicitly treats electronic polarizability by overestimating partial atomic charges to fit quantum mechanical calculations[152,153]. The field-based models seem to predict even broader lineshapes for reasons discussed in the next section, which may indicate that motional narrowing effects are more important there and that a static averaging technique should use the potential-based models.

In the trpzip2 FTIR spectra, the high frequency shoulder and correct relative peak intensity is observed in all of the calculations, which has been noted as a signature of anti-parallel β-sheet structure. Many subtle features match up in the calculated and predicted 2DIR spectra. The cross peak ridge extending horizontally to the red from the high frequency peak is observed in each of the electrostatic models, but it is weaker than the experiment in every case. The ridge extending horizontally to the blue from the most intense overtone peak is reproduced, as is the node between the fundamental and overtone peaks. Interestingly, the sharp red-side extension from the low frequency fundamental peak at $\omega_3 \approx 1640\ cm^{-1}$ is observed in each potential model and the experiment, but not in the field models. The narrowest spectrum with degenerate site energies appears qualitatively similar to the FTIR, but the same model fails to predict the observed 2DIR lineshapes.

All of the D-Arg FTIR spectra show a singly-peaked and roughly symmetric lineshape in agreement with the experiment. The calculated 2DIR lineshapes show the broad, diagonally elongated peak, but also show a bit of extra structure unobserved in the experiment extending from the fundamental peak in the $-\omega_1$ direction. The substantial extension of the overtone in the $+\omega_3$ direction at high frequencies is underestimated in each calculation.

The ubiquitin spectra differ from trpzip2 in the intensity between the high and low frequency anti-parallel β-sheet peaks arising from the α-helix and random coils. Although the calculation correctly predicts structure in this region, it is overestimated leading to a mismatch in the relative intensity of the two main peaks. The cross peak ridges are reproduced, as well as the sharp extension from the low frequency fundamental. The slope of the node is also predicted in good agreement. In ubiquitin like trpzip2 and D-Arg, the calculated lineshapes from electrostatic site energy models are broader than experiment.

4.3.3 The Origin of Frequency Shifts in Site Energy Models

Each considered site energy model was parameterized to describe the fluctuating amide I frequency of N-methylacetamide. All but Jansen 4F were parameterized with water. In extending these models to describe the amide I site energy fluctuations for various sites in a polypeptide chain, several questions arise. Should only the electrostatics of the water be included? Will the models work at all if the original definition of "solvent" is now extended to include surrounding protein? What role do the side chains and other parts of the backbone play? These questions are investigated by looking at the average site energies and which structures contribute to the redshift from the gas phase (Tab.4.1). In this table, the brackets $\langle\rangle$ imply an average over site and ensemble. The average site energy is given by $\langle\omega\rangle$, which tells about the typical hydrogen bonding environment which causes a redshift from the gas-phase value of 1717 cm^{-1}. This cumulative shift is broken down into the average shift from water, backbone, and side-chains in $\langle\delta\omega_{Water}\rangle$, $\langle\delta\omega_{Backbone}\rangle$, and $\langle\delta\omega_{Sidechain}\rangle$ respectively. Also note that three of the models parameterize a static redshifting contribution[121, 122, 124]. The average frequency standard deviation is given by $\langle\sigma\rangle$.

First, we observe that by summing the electrostatics from the water, side chains, and remaining backbone, the redshift comes close to the expected region of ~1655 cm^{-1}. The majority of the redshift is obtained from the water, showing that a correct parameterization of the solvent electrostatics is the most important determinant of the site energies. As noticed by Skinner, et. al.[124], differences in the potential models reflect how many water molecules were used in the NMA/water cluster calculations to parameterize the model- the Keiderling 5P model used the largest clusters and the Hirst 4P and 7P models used the smallest. Not enough models are present to compare this in the field models. The field models also underestimate the contribution of the backbone by ~10 cm^{-1} relative to the potential models, causing them to differ in the relative importance of backbone and sidechain contributions; the field-based models include more of a contribution from the side-chains than backbone which is opposite in the potential-based models.

The other striking difference between the field- and potential-based models are the significantly larger field site energy standard deviations, the effects of which are also manifested as broader lineshapes in the field models. This can be rationalized by the distance sensitivity of the field $\left(\propto \frac{1}{r^2}\right)$ and the potential $\left(\propto \frac{1}{r}\right)$ formulations and the $\frac{\sqrt{n}}{n}$ scaling properties of signal to noise. More atoms are significant to the potential than the field because the cutoff is slower and so the random solvent modulations will be smaller. The difference in distance scaling may also explain the inconsistency in relative backbone contributions and difference in dephasing properties. In illustration of the latter, site energy correlation functions $C_{\epsilon\epsilon}(t) = \langle \Delta\epsilon(t)\Delta\epsilon(0)\rangle / \langle \Delta\epsilon(0)^2\rangle$ (where $\langle\rangle$ indicate an ensemble and site average) are plotted in Fig.4-4 for the total shift and water-only contribution to a representative field and potential model. A bioexponential fit to reveal the general timescales shows that the short-time component is faster in the Jansen 4F model than the Keiderling 5P model (180 fs and 500 fs respectively), which is indicative of the quicker randomization of frequencies induced by the larger modulations. This was also seen in the work of Jansen and Knoester[125], which was also manifested as a faster decay with fast oscillations in the frequency correlation function for NMA for the field parameterization. Similar work on NMA by Skinner, et. al.[124] did not reveal a difference in the standard deviations between field and potential, but did note a shorter T_2^* dephasing time for the field-based model than for the potential based models.

4.3.4 Relative Importance of Coupling and Site Energies

With the results described above in hand it is important to ask what the delocalized eigenstates are more sensitive to: equilibrium backbone structural disorder causing changes in the coupling or hydrogen bonding and solvent disorder causing changes in the site energies. In Fig.4-5, we computationally address this. We calculate the FTIR and 2DIR spectra of trpzip2 in six different ways using the Keiderling 5P potential model:

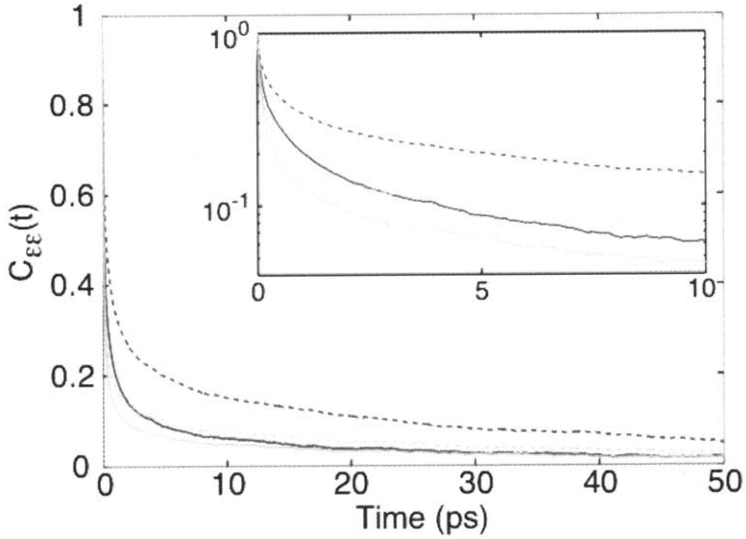


Figure 4-4: Trpzip2 site energy correlation functions. The normalized site energy fluctuation correlation function (solid line), calculated with the Keiderling 5P (dark) and Jansen 4F (light) models (a representative potential and field site energy model, respectively) are shown, in addition to the contribution from water (dashed line) for both models.

(A) Assuming that the Hamiltonian is equal to its ensemble averaged value.

(B) Assuming that only the site energies are fixed equal to the ensemble averaged values, while the couplings are sampled.

(C) Assuming that only the coupling elements are equal to the ensemble averaged values, while the site energies are sampled.

(D) Allowing each element to be sampled at each step (the same as the full treatment in the previous section).

(E) Assuming that the cross-correlation between site-energies is negligible and modeling the spectra based on Gaussian fluctuating site energies with the mean and standard deviation obtained from the MD simulation.

Comparing (A) and (D) shows that approximating trpzip2 as one average structure- including a static distribution of different hydrogen bonding environments at each site is invalid, similar to the example in Fig.4-1. Comparing (B) and (C) to

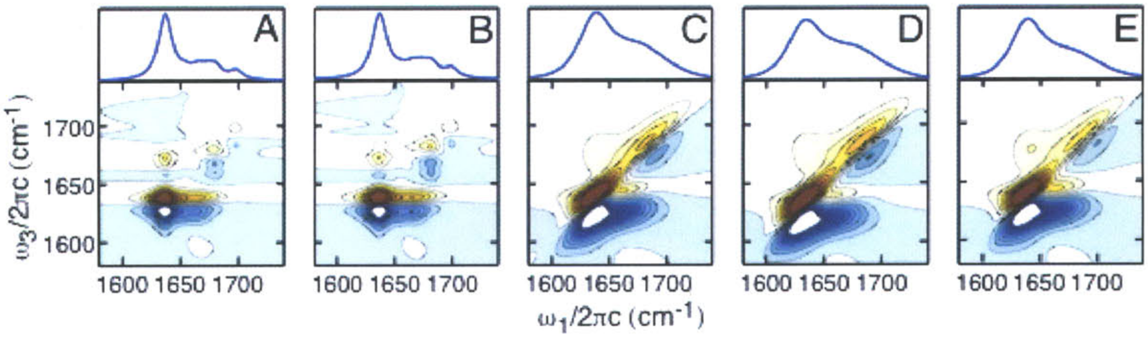


Figure 4-5: Site energy fluctuations give rise to diagonal broadening. Six exploratory calculations are presented in an attempt to identify the physical origins of spectral features in trpzip2, (A) All elements of the Hamiltonian set to their average value, (B) Site energies set to their average value, couplings sampled from MD, (C) Couplings set to their average value, site energies sampled, (D) All elements sampled, (E) Couplings set to their average value, site energies sampled from a Gaussian distribution with center and width corresponding to MD site energy statistics. 21 evenly spaced contours are plotted from ±50% of the maximum for each spectrum.

(D) reveals that it is the site energy fluctuations that causes (A) to be such a poor approximation. Trpzip2 has a notably stable structure, which biases this picture to a site-energy based one, but these results are also verified in D-Arg and ubiquitin. The spectral signatures of equilibrium variations in coupling are nearly nonexistent in these systems.

This emphasizes the critical importance of fluctuating site energies in modeling 2DIR, and therefore FTIR, spectra. The results are striking in that they indicate that the site energy fluctuations are far dominant over coupling fluctuations. By averaging out the site energy fluctuations, the spectra appear nearly indistinguishable from the entirely static spectrum. By only surveying fluctuations in the site energies, a spectrum nearly indistinguishable from the totally averaged spectrum is obtained. The minor differences attributable to fluctuations in coupling include a slight extension of the ridge and some minor (~1 cm^{-1}) line broadening, and are more noticeable in the unconvolved stick spectra (not shown). This is both computationally fortunate and physically revealing; of the N^2 elements of the Hamiltonian, the fluctuations in merely N of them largely determine the disorder shown in a 2DIR spectrum.

The dominance of disorder in the site energies and the relative stability of the off-diagonal elements implies that one may reduce the complexity in calculating 2DIR spectra by sampling site energies from independent random distributions, which further assumes that frequency shifts between sites are uncorrelated. In Fig.4-5E, we compare a calculation of the 2DIR spectrum obtained by modeling the spectrum based on a Gaussian random fluctuations of site energies in which the mean and standard deviation of the distribution is obtained for each site from the MD simulation. The excellent agreement between Figs.4-5D and 4-5E indicate that the independent site approximation is valid. The implication is that knowledge of a representative equilibrium structure and its fluctuations can be used to model the infrared spectroscopy. However, the sensitivity of the mean site energies to hydrogen bonding indicates that methods must be validated to assign mean site energies in the absence of explicit solvent in a trial structure.

4.4 Conclusions

We have shown that the information that is typically sufficient to model FTIR spectra does not correctly predict 2DIR spectra. This is demonstrated in the difference between experimental and predicted 2D lineshapes from one structure and arises chiefly from the sensitivity of 2DIR to different broadening mechanisms. This suggests that although the agreement between an experimental FTIR and one calculated from a single structure appears good, the interpretation may be flawed by the disparity in parameters and observables. This modeling can be refined by examining the richer, 2D lineshapes which are more sensitive to disorder.

We have pursued this by calculating FTIR and 2DIR spectra from MD simulations, while testing a set of proposed site energies models. We obtain the best prediction of gross and subtle features in protein 2DIR spectra to date in spite of the empirical force fields used to calculate electrostatics and neglecting motional narrowing effects by using static averages. Strong features such as cross-peak ridges and even many subtle lineshape features are reproduced faithfully. The most important step forward is the agreement in the diagonal elongation of peaks arising from disorder, which is typically ignored in modeling FTIR spectra. We track the origin of this disorder and find it to be based in fluctuating site energies, not fluctuating couplings. In turn, these site energies are ≈50% driven by the water solvent. The potential-based site energy models predict that the remainder is mostly caused by the backbone, whereas the field-based site energy models favor the side chains for most of the remaining fluctuations. These two pictures are inconsistent, but may be resolved by comparing isotopically shifted (^{13}C and/or ^{18}O) single sites bordering charged and hydrophobic residues.

Because 2DIR data can be acquired with the resolution to capture the fastest timescales of biomolecular rearrangement, our modeling provides the link that allows mechanistic non-equilibrium MD predictions of protein folding and aggregation pathways to be directly tested[150, 154, 155]. The delocalized vibrational eigenstates provide mesoscopic structural variables that are more meaningful than rate constants, which may need to be rescaled based on anomalous solvent diffusion, or equilibrium constants, which are sensitive to slight errors in energy calculation in a way not necessarily indicative of an incorrect mechanism. This utility is immediately applicable to debates in the β-hairpin literature over the relative importance of forming the turn-region, hydrophobic core, and backbone hydrogen bonds, with the possibility of off-register hydrogen bonds[141, 142, 144, 146, 155, 156]. Furthermore, the sharp sensitivity of 2DIR to solvent electrostatics suggests a way beyond thermodynamic comparisons to test the next generation of implicit solvent models that reproduce structural transformations such as concurrent core collapse and desolvation[157–161] and anharmonic force fields[162] that preclude the need for a separate spectroscopic Hamiltonian.

The static averaging we employ for computational simplicity produces lineshapes that are broader than measured. By taking into consideration time-dependence in the adiabatic approximation, we can effectively increase the frequency resolution of the simulated spectra and sense more subtle features. A step beyond this would be to test at which point the non-adiabatic effects observed for tri-alanine[163] disappear in larger systems- to what extent will they broaden the spectra of hairpins and proteins? Further refinements in the calculation of spectroscopy are also possible, including accounting for rotation of the local transition dipoles, fluctuating values of the anharmonicity, and extension to other vibrational modes[125, 164]. Finally, although we have shown the effects of disorder in IR spectra, the effects of disorder on FTIR (one-quantum) and 2DIR (two-quantum) eigenstates remain to be seen and can be tested with the recently revisited bright-state analysis[12].

4.5 Acknowledgements

We wish to thank Hoi Sung Chung and Adam Smith for data and help with data acquisition, Darwin Alonso and Valerie Daggett for the ubiquitin simulations, Peter Hamm for assistance with transition charge coupling, and Minhaeng Cho for the nearest neighbor coupling map. We also wish to thank Thomas Jansen, Kevin Jones, and Rebecca Nicodemus for helpful comments. This work was supported by the National Science Foundation (Grant CHE-0316736).

4.6 Afterword

This chapter was adapted from "Spectral Signatures of Heterogeneous Protein Ensembles Revealed by MD Simulations of 2DIR Spectra," by Ganim and Tokmakoff.[8]

Property Correlated	Model Name	Parameterized Sites (m)	$\langle\omega\rangle$	$\langle\sigma\rangle$	$\langle\delta\omega_{Water}\rangle$	$\langle\delta\omega_{Backbone}\rangle$	$\langle\delta\omega_{Sidechain}\rangle$
Potential	Keiderling 5P[123]	$C_\alpha\ C\ O\ N\ D$	1660	16.6	-27.5	-15.8	-13.2
Potential	Cho 6P[122]	$C_\alpha\ C\ O\ N\ H\ C_\alpha$	1649	15.6	-28.6	-17.7	-11.0
Potential	Cho 4P[121]	$C\ O\ N\ H$	1657	15.2	-26.0	-13.4	-10.3
Potential	Hirst 7P[165]	$C\ CO_{mid}\ O\ CN_{mid}\ N\ NH_{mid}\ H$	1686	14.7	-12.3	-9.3	-8.9
Potential	Hirst 4P[165]	$C\ O\ N\ H$	1696	7.1	-9.9	-6.5	-4.4
Potential	Skinner 6P[124]	$C_\alpha\ C\ O\ N\ D\ C_\alpha$	1660	17.4	-24.6	-19.2	-12.8
Field	Skinner 6F[124]	$C_\alpha\ C\ O\ N\ D\ C_\alpha$	1675	24.2	-18.3	-2.0	-14.5
Field to 2nd Order	Jansen 4F[125]	$C\ O\ N\ D$	1670	24.8	-26.3	-1.9	-18.5

Table 4.1: Site Energy Models and Fluctuation Statistics for Trpzip2. Listing of site energy models tested, comparison of electrostatic property correlated, site correlated, nomenclature, and fluctuation statistics for Trpzip2. $\langle\omega\rangle$ is the average site frequency, which is broken down to contributions from water, backbone, and sidechain respectively in $\langle\delta\omega_{Water}\rangle$, $\langle\delta\omega_{Backbone}\rangle$, and $\langle\delta\omega_{Sidechain}\rangle$. $\langle\sigma\rangle$ is the average site standard deviation. All quantities are in cm^{-1}.

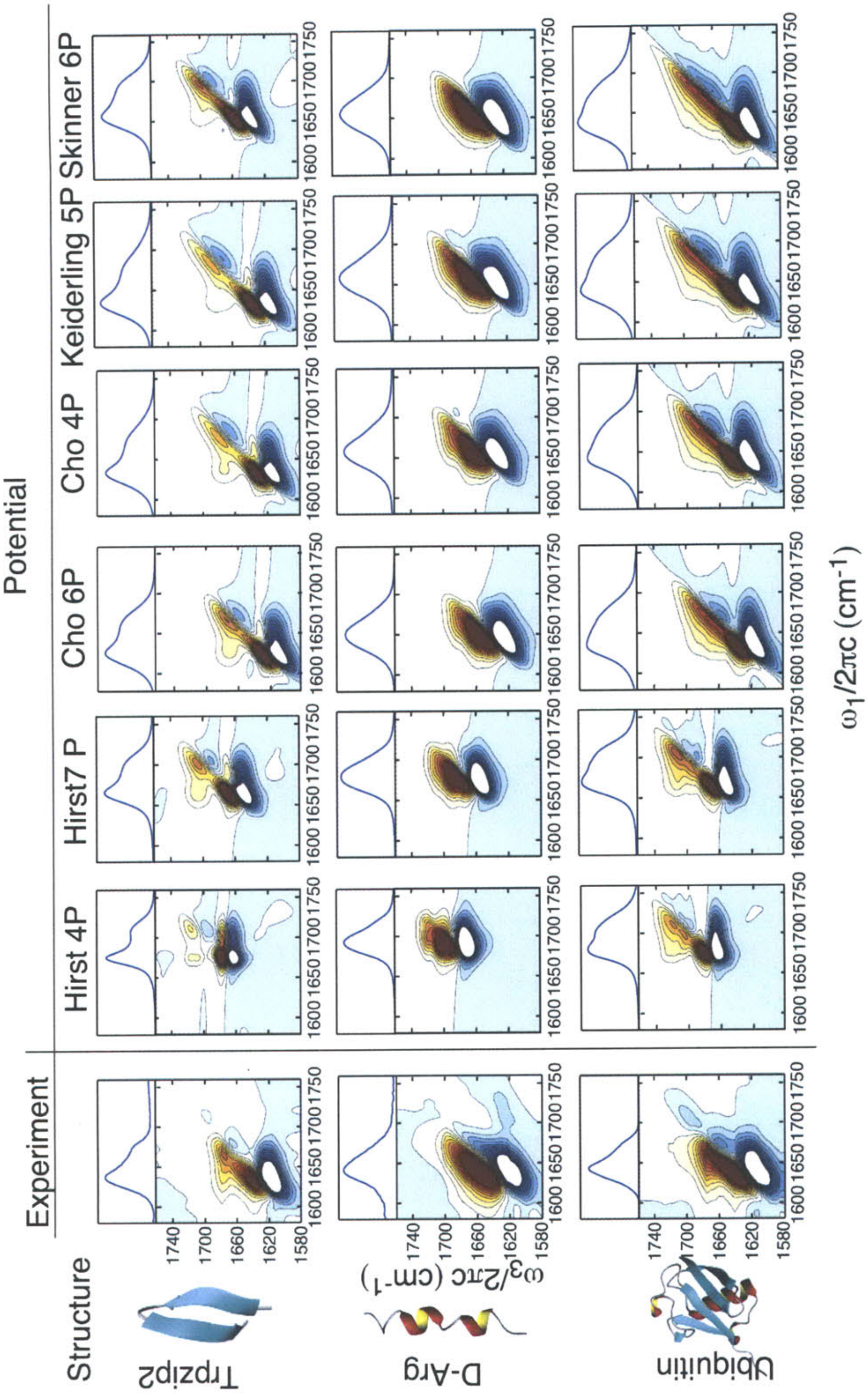


Figure 4-6: Comparison of Calculated and Experimental FTIR and 2DIR for Trpzip2, D-Arg, and Ubiquitin for all electrostatic potential-based site energy models. Calculated spectra from the manuscript of Ganim et al.[8] 21 evenly spaced contours are plotted from ±50% of the maximum for each spectrum.

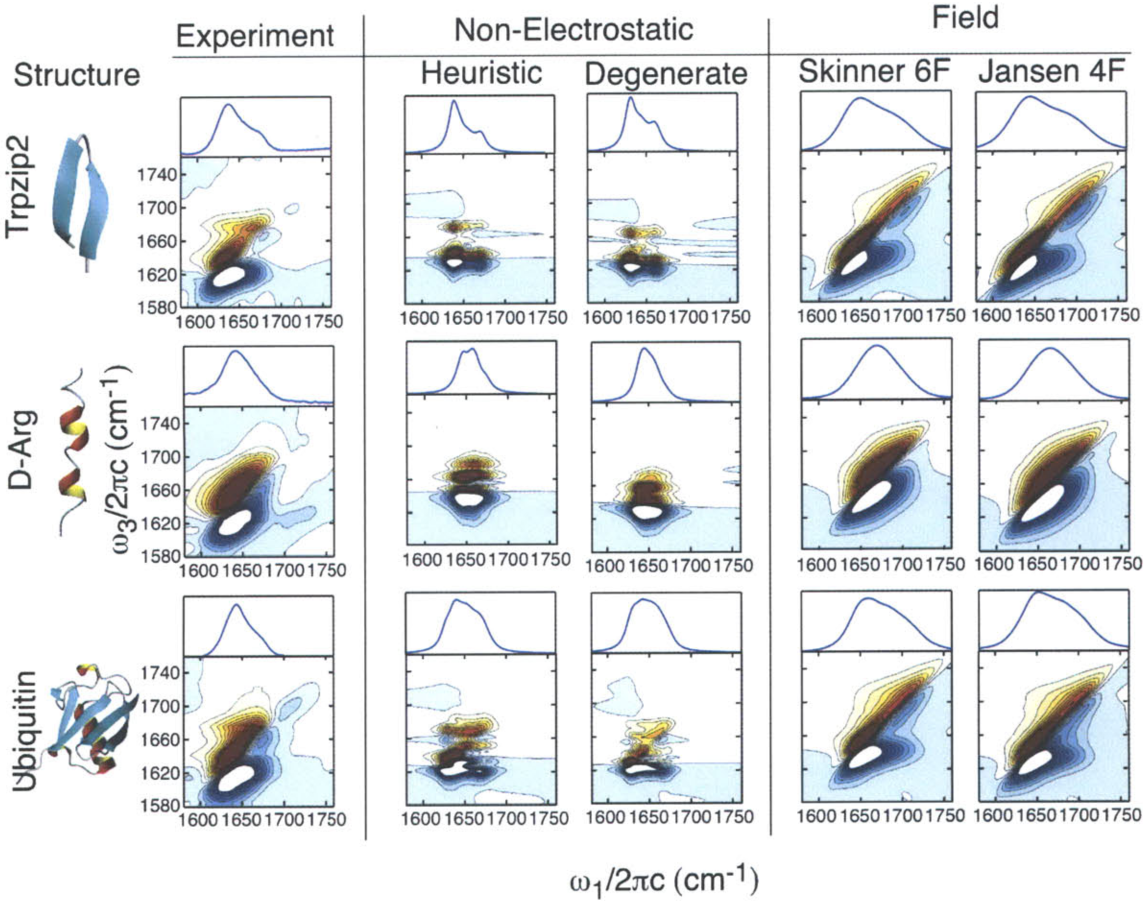


Figure 4-7: Comparison of Calculated and Experimental FTIR and 2DIR for Trpzip2, D-Arg, and Ubiquitin for all non-electrostatic and electrostatic field-based site energy models. Calculated spectra from the manuscript of Ganim et al.[8] The heuristic model uses PDB structures 1LE1 for trpzip2, 1UBQ for ubiquitin, and a random equilibrated MD structure for D-Arg. 21 evenly spaced contours are plotted from ±50% of the maximum for each spectrum.

Chapter 5

Melting of a β-hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations

This chapter describes an investigation of the structure of the β-hairpin peptide trpzip2 using isotope-edited 2D IR to site-specifically investigate the conformation of its backbone. In this mode, the isotope-shifted amide I band can be used to infer local structure and dynamics through variations in its vibrational frequency or couplings to other vibrations. 2D IR spectra of a series of peptide isotopologues are compared with calculated spectra for atomistic structural assignments.

Trpzip2 (TZ2) is a 12 residue peptide that forms a stable anti-parallel β hairpin in water (sequence SWTWENGKWTWK, diagrammed in Figure 5-1). Since it is one of the smallest models for an anti-parallel β sheet, it has attracted significant attention as a model system for protein folding and computation. Owing to its small size, it has also become a test system for calculations of amide I spectra. Trpzip2 is a non-native peptide that was designed by Cochran, et al.[166] The tripzip peptides were engineered with four tryptophan residues arranged with two side chains on each of the opposing β strands that lead to a remarkably stable system through the favorable perpendicular stacking of the indole rings. Often, folding of this peptide is described within the framework of a two-state folding model, although its validity and the nature of the two states is unclear. To obtain residue-level structural insight in the thermal denaturation of TZ2, amide I 2D IR was used in combination with ^{13}C and ^{18}O isotope labels, which red shift the amide I vibrational frequency away from the main band. Four isotopologues were used to probe cross-strand contacts and isolate peptide units in the turn (K8), termini (S1), and central region of the peptide (T10 and T3T10).

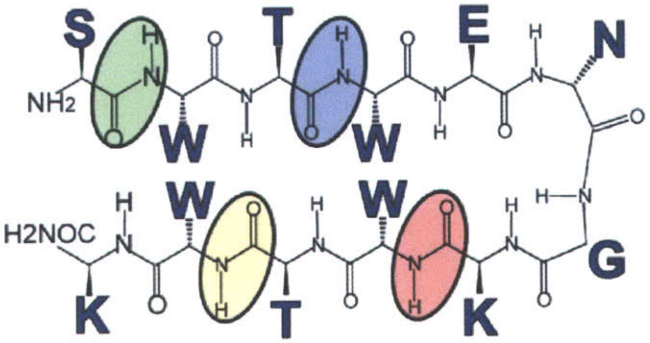


Figure 5-1: Sequence, backbone structure, and isotope labeling sites for trpzip2 experiments.

5.1 Isotope Labeling Experiments

Experimental details for the synthesis and spectroscopy of the isotope labeled TZ2 peptides appear in the full manuscript.[167] The NMR structure of the native state of TZ2 shows a tightly folded, single secondary structural conformation. From these results, it would be expected that any isotope label in the peptide would yield a single, shifted peak whose central value would reflect the intrinsic frequency of the labeled oscillator and an additional ≈ 40 cm^{-1} redshift. Figure 5-2 shows 2D IR spectra TZ2 isotopologues with labels at positions K8, T10, and T3/T10. Several features were observed in these spectra that defied a simple analysis. The most surprising spectrum was that of K8 whose single isotope label produced two isotope shifted peaks (1 and 2). The dual labelled peptide, T3/T10, showed a single peak. In contrast, the T10 peptide showed no distinct peak. How should these features be assigned to learn about the conformational heterogeneity of TZ2?

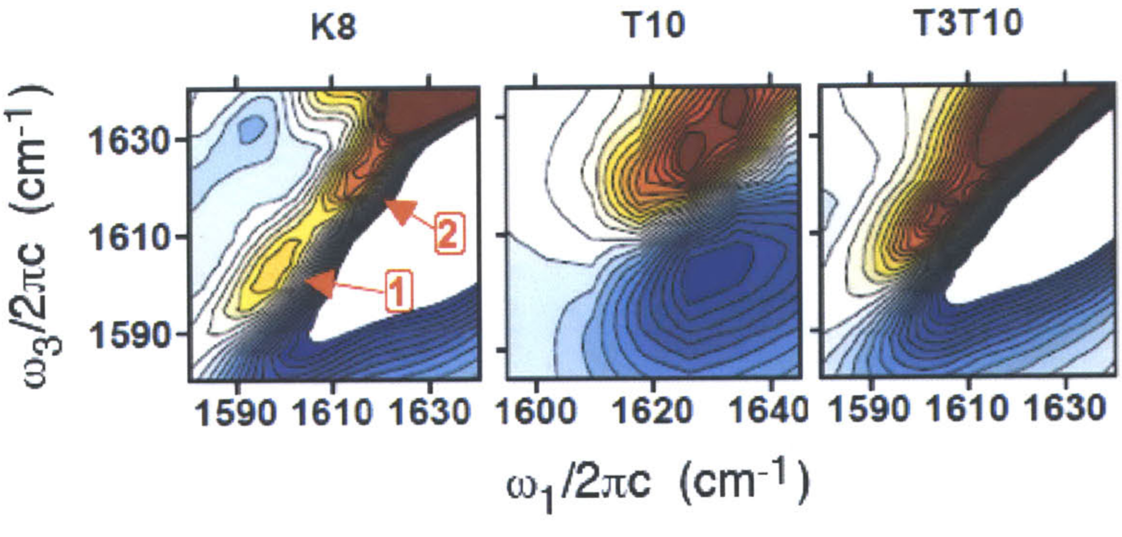


Figure 5-2: 2D IR spectra of trpzip2 isotopologues focusing on the isotope-shifted peaks.

Empirical observations[168] and DFT calculations[123] suggest that the intrinsic frequency of an amide I oscillator is mostly sensitive to the local hydrogen bonding environment. Interpreting the TZ2 isotope-labeling experiments in light of this evidence would suggest that the two peaks in the K8 spectrum arise from different hydrogen bonding environments. Similarly, if the frequency of the T10 oscillator was higher or lower than expected, the isotope labeling may not produce a clear peak. However, these interpretations are not trivial because the amide I site energy is also sensitive to the dihedral angles of the peptide and a coupling-induced shift cannot be ruled out for any of these experiments. The isotope-labeling experiments on TZ2 were assigned by drawing upon the expansive, 3.23 μs simulations of Pitera et al.[169] and the Markov state decomposition of Chodera et al.[170]

5.2 Modeling Trpzip2 Spectroscopy using Hidden Markov Model Macrostates

The 2D IR spectra were simulated using five kinetically metastable macrostates from the 40 Markov states identified.[170] Each macrostate was comprised of 30 structurally related microstates with atomistic coordinates. Each structural model was subjected to ≈ 5000 steps of a steepest descent energy minimization to conform to the OPLS/AA force field using the particle mesh Ewald method to treat long-range electrostatics. The peptides were solvated in a box of rigid SPC/E water that extended 1.5 nm from the solute, and the water was allowed to equilibrate for 10 ps in the NPT ensemble (300 K, 1 atm) while the peptide was fixed. The simulation was continued with no constraints, except on the bond lengths, for 1 ns of equilibration and 100 ps of data acquisition. Structures were saved every 20 fs to yield 5000 frame trajectories.

The 40 TZ2 Markov states are grouped by patterns in backbone hydrogen bonds and side chain packing. Because amide I spectra are most sensitive to differences in secondary structure contacts, the five chosen Markov states survey different backbone structural conformations of TZ2:

- **Folded** (FO, 250851) is a well-ordered ensemble that has an average of four backbone hydrogen bonds, which correspond to those in the NMR structure.[166] The S1, T3, K8, and T10 carbonyls are all oriented for cross strand hydrogen bonds, and all four tryptophan side chains are packed to one face of the peptide.

- **Bulged Turn** (BT, 214369) is a compact coiled structure with a well-solvated, bulged turn region and an average of one cross-strand hydrogen bond that appears near the strand termini. This state contains S1, T3, K8 and T10 peptide carbonyls that are solvent exposed, and tryptophan side chains make rare contacts.

- **Disordered** (SR, 271154) is a disordered ensemble whose common feature is a bulged turn and an average of one backbone hydrogen bond with a misaligned registry near the C terminus. Tryptophans are poorly organized.

- **Frayed** (FR, 11131) has a compact, properly formed type I' turn region with an average of two backbone hydrogen bonds, but disordered strand termini. The T3 and T10 are oriented toward the opposite strand, and tryptophan side chains are disordered.
- **Extended Disorder** (ED, 64336) is an extended and well-solvated state with substantial disorder and rare backbone hydrogen bonds.

Here the number identifying each state refers to the designation in supplementary materials to the work of Chodera et al.[170]

For each frame, a Hamiltonian was calculated as described in 4.2, using previously described models for the amide I frequencies and coupling values.[125,171] The isotope-labeled sites were given a 41 cm^{-1} redshift to correct for systematic errors in the site energy model. The 2D IR spectra were calculated using the numerical integration of the Schrödinger equation (NISE) scheme.[172,173] In this procedure the time-evolution matrices needed in the nonlinear response functions that govern the 2D IR spectra are calculated by solving the time-dependent Schrödinger equation using Equations 3.18 and 3.17. The resulting spectra were summed over each structure in the macrostate to produce spectra that reflected the disordered ensembles.

5.2.1 Time-Correlation Methods vs. Static Averaging Methods for Calculating 2D IR Spectra

The trpzip2 simulations in this chapter were calculated by numerical integration of the Schrödinger equation. It is instructive to compare these calculations to simplified versions, assuming that the amide I energy gap correlation functions take a simple exponential form. The frequency domain manifestation of this assumption is that the 2D IR spectrum will be composed of two-dimensional Lorentzian bands, as described by Equations 3.19 and 3.20. Figure 4-1 in Chapter **??** shows an example demonstrating that the inhomogeneous lineshapes observed in amide I 2D IR spectra of proteins cannot be reproduced with any one static structure. The FTIR spectra alone are misleading because inhomogeneity causes correlated frequency shifts that cause clear diagonal elongation in 2D IR spectra; the FTIR spectra are not sensitive to this effect.

In the static averaging method, inhomogeneous broadening arises primarily from averaging the 2D IR spectra from a distribution of structures. Figure 5-3 shows a comparison between this method of introducing inhomogeneity and the more accurate method of numerically integrating the Schrödinger equation. In all of the static average spectra, the diagonal line width is much broader than the NISE spectra. If a line width of 10 cm^{-1} is chosen, the antidiagonal width is too narrow. A line width of 15 cm^{-1} better approximates the aspect ratio between the diagonal and antidiagonal widths, but the spectrum is much too broad and the higher frequency peak cannot be resolved. The overlay allows for a frequency-dependent comparison of the line widths.

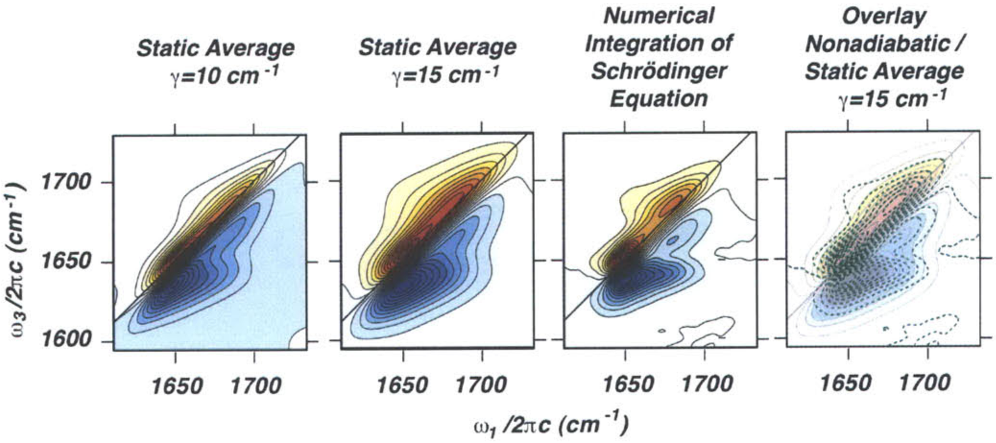


Figure 5-3: Comparison between numerical integration of the Schrödinger equation and static average methods for calculating 2D IR spectra. 2D IR spectra are plotted for the trpzip2 bulged Markov state (BT) using the static averaging method with different values of the Lorentzian linewidth parameter, γ, and the NISE method.

5.3 Results for Trpzip2

Simple comparisons to the experiment can be drawn by inspecting the site energy distributions. Figure 5-4 shows histograms of the site energies for each amide I oscillator separated by the Markov state. It can clearly be seen that site 8, corresponding to isotope label K8, displays at least two different distributions: (1) a narrower (30 cm^{-1} full width half max) one peaked at 1631 cm^{-1} for the the native-like and frayed, and (2) a disordered and extended disordered states (42 cm^{-1} full width half max) that is peaked at 1625 cm^{-1}. In contrast, oscillators such as site 6 show almost no sensitivity to the differences between the Markov states. This observation supports the hypothesis that two different environments exist for the K8 label since the site energy distributions naturally do not include *any* coupling, although the experimental splitting between the peaks is 18 cm^{-1}. Moreover, site 1 peaks at a higher energy (1645 cm^{-1}) than any of the other experimentally probed sites (T3- 1636 cm^{-1}, T10- 1635 cm^{-1}), which may explain why the isotope label did not provide a sufficient enough shift to produce a distinct peak. Calculations of this type can guide the choice of isotope label (saving weeks of labor) by allowing sites to be chosen that can provide clear evidence in favor or against structural models. Figure 5-5 also shows

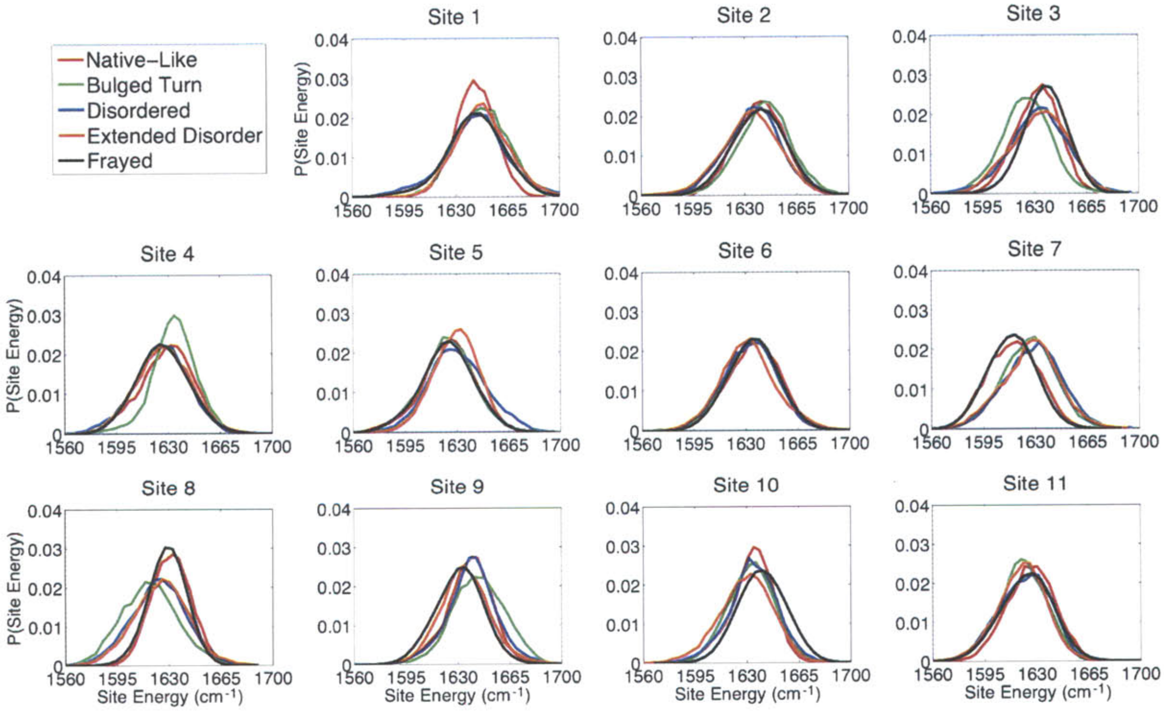


Figure 5-4: Histograms of the site energies for each trpzip2 site separated by the five different Markov states.

that amide I site energies can be correlated with the number of hydrogen bonds (HB). The local electric field and dihedral angles to neighboring residues are inputs to the amide I site energy model; since HBs are largely electrostatic, amide I site energies are parametrically dependent on hydrogen bonding. The hydrogen bonding criteria was $r_{X...O} \leq 4.5$Åand $\theta_{X...HO} < 35°$. Each site was found to display this correlation, although an HB to an amide carbonyl proton or to water was found to produce the same shift. When fully solvated by water, the amide carbonyl has the ability to form multiple hydrogen bonds, which gives a larger average red shift. From analysis of both types of HBs, it was observed that on average, one HB leads to an 11 cm^{-1} red shift, however the dispersion in this relationship (σ = 25 cm^{-1}) leads only to a modest correlation (ρ = -0.50). These results can be compared with the results of *ab initio* calculations on isolated clusters, which reveal a 20 cm^{-1} red-shift for each HB to oxygen and an additional 10 cm^{-1} red shift for hydrogen bonds donated by the peptide N-H.[174]

The calculated TZ2 2D IR spectra appear in Fig 5-6 for the five chosen Markov states. The corresponding FT IR spectra appear in Figure 5-8. In general, there are qualitative similarities between several of the experimental and simulated spectra, which are discussed in terms of peak positions, amplitudes, and linewidths and lineshapes. The simulated 2D IR spectra of the unlabeled peptide for FO, BT, and FR conformers each display two diagonal peaks and Z-shaped contours that resemble the experiment at low temperatures. For the disordered SR state, two bands are also resolved although in this case, the high-frequency band carries most of the intensity. For the disordered ED state, the amide I band has little structure and reflects the diagonally stretched resonance of an inhomogeneous lineshape.

The simulated 2D IR spectra were used to assign the two isotope-shifted peaks of TZ2-K8 as peptides with two different turn conformations. Each of the five K8 simulations give rise to one resonance for the labeled site, although the peak position and linewidth is correlated with its solvent exposure. In the FO and FR states, the carbonyl oxygen of the K8 amide group forms a cross-strand hydrogen bond with the W4 amide proton, as expected for a type I' β turn. In their calculated spectra, the ν_{K8} peak is split from the amide I maximum by 25-30 cm^{-1}, and the diagonal linewidth is a compact 5-6 cm^{-1}. In addition, a distinct cross peak is observed between the

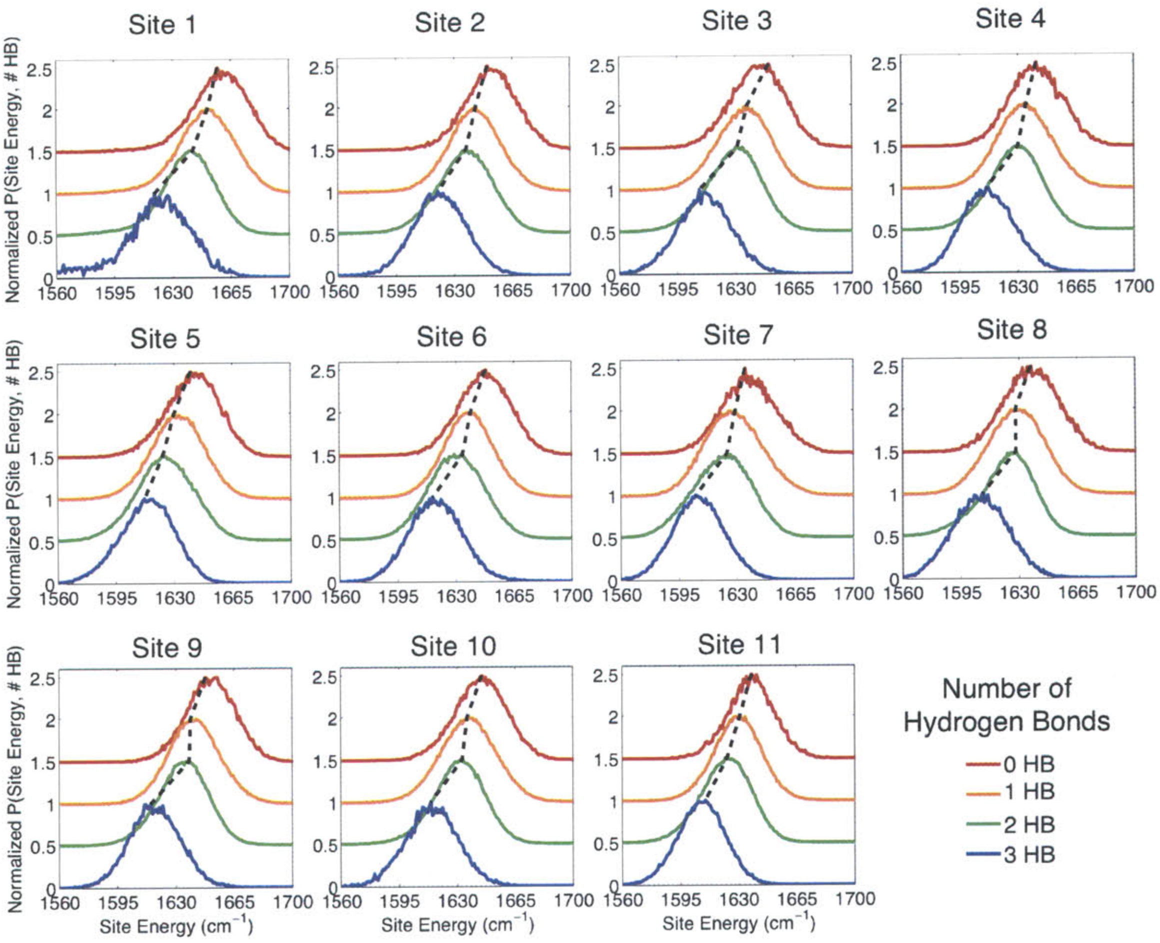


Figure 5-5: Histograms of the site energies for each trpzip2 site separated by the number of hydrogen bonds.

ν_{K8} and $\nu_{\perp}$ bands in the calculated spectra for these states. In each of the other structures (SR, BT, and ED) the K8 oxygen is more exposed to water with a higher variance in the number of hydrogen bonds formed. In these spectra, the splitting between ν_{K8} and $\nu_{\perp}$ is 30-36 cm^{-1} and the labeled resonance has a distinctly broader diagonal linewidth (10-14 cm^{-1}) compared to the FO and FR states. The calculated BT spectrum shows the largest shift and linewidth.

The observed differences can be attributed to variation in hydrogen bonding to the K8 carbonyl. Fully solvated carbonyls will on average participate in two hydrogen bonds to water, as opposed to one cross-strand hydrogen bond in the case of properly folded β turns. For the K8 carbonyl, each hydrogen bond contributes to an average

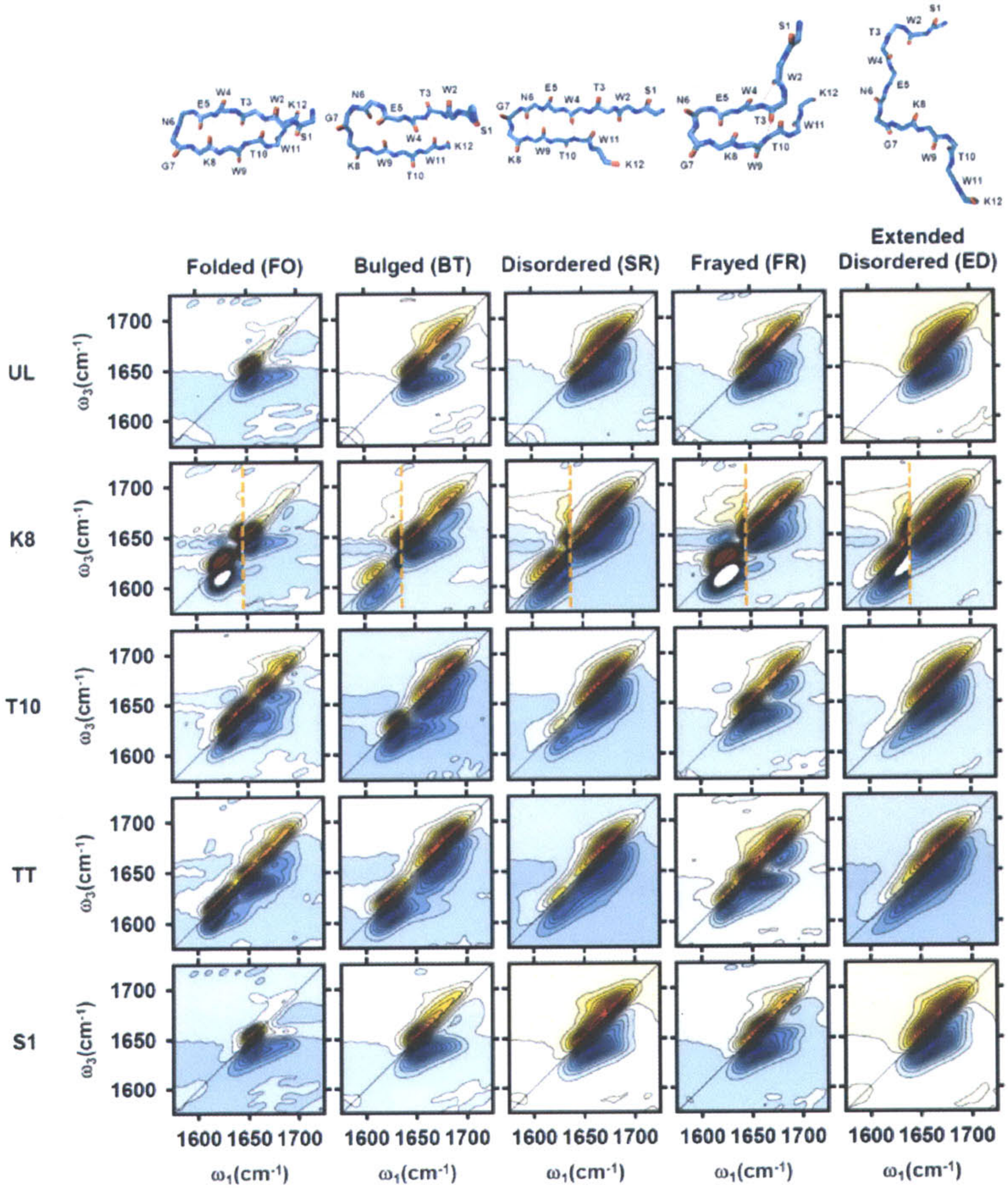


Figure 5-6: Calculated 2D IR spectra of trpzip2 and isotopologues for five different Markov states.

16 cm^{-1} red shift in amide I frequency; however, the correlation between hydrogen bond number and red-shift is mild (ρ = -0.53). Therefore, the solvent-exposed K8 conformers should exhibit an additional red shift relative to FO, but the conforma-

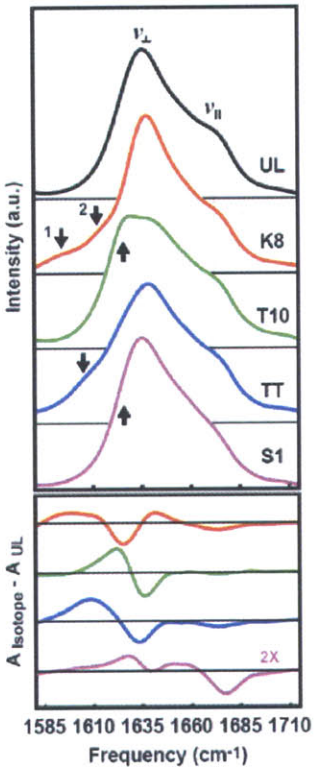


Figure 5-7: FT IR and FT IR difference spectra of trpzip2 isotopologues. Equilibrium FTIR spectra are shown for each TZ2 isotopologue at 25°C and pH = 2.5. Each spectrum is baseline corrected with a linear subtraction, and area normalized for comparison. Difference data (bottom) are obtained by subtracting the unlabeled spectrum (TZ2-UL) from each of the spectra. The S1 difference spectra is presented at 2x magnification for clarity. Arrows have been drawn to emphasize the isotope labelled peaks.

tional disorder and fluctuations in peptide backbone and water hydrogen bonds will also lead to a larger inhomogeneous width. Based on these observations, it was concluded that the ν_{K8-1} peak represents peptides with non-native turns, and the higher frequency, ν_{K8-2} peak reports on structures with native type I' β turns.

The T10 peptide carbonyl adopts a variety of conformations in the Markov states simulated. In the case of the FO and FR states it is oriented for a cross-strand hydrogen bond, for BT it is entirely solvent exposed, and for SR and ED it samples a variety of configurations. The conformers with a solvent-exposed T10 carbonyl give rise to a structured excitonic spectrum and a red-shifted and spectrally distinct

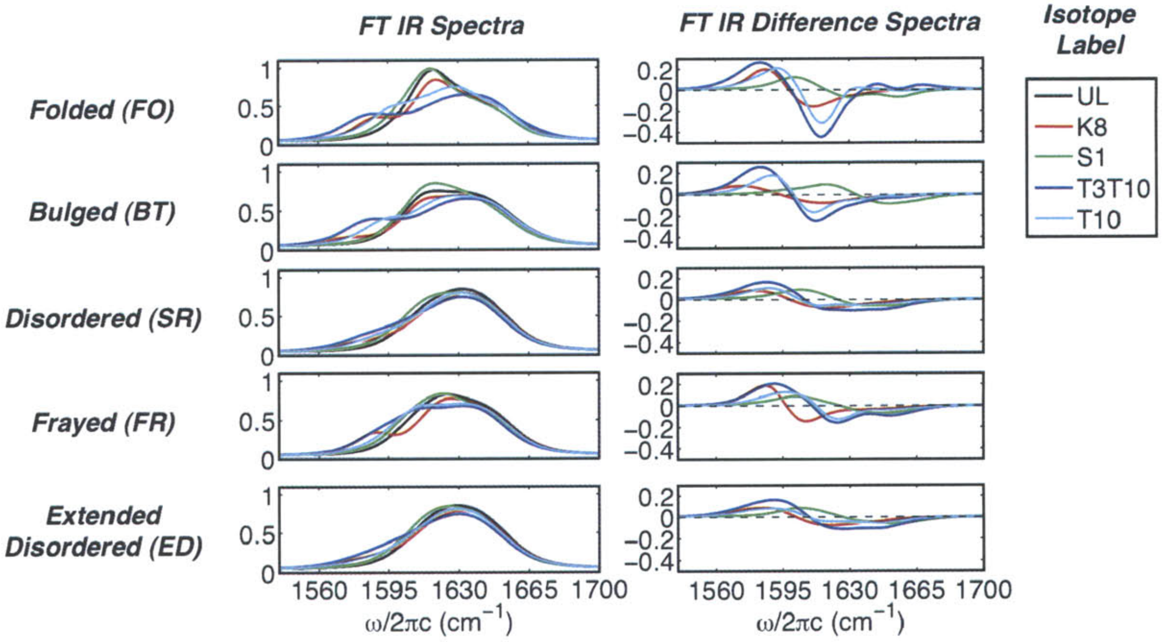


Figure 5-8: Calculated trpzip2 FT IR and FT IR Difference Spectra. The difference spectra are relative to the unlabeled peptide for each Markov state.

T10 peak; internally hydrogen bound T10 sites produce a peak merged with the excitonic band, which was observed in the experiment. For the SR and ED states, the intensity of the labeled transition is greatly diminished over the intense band seen in the completely water-exposed state. Of all simulations, the FR state corresponds best to the low temperature experimental spectra. These observations demonstrate that the T10 experimental spectra provide evidence of a T10 carbonyl that participates in a single cross-strand hydrogen bond over the temperature range sampled.

For 2D simulations of TT (T3T10), all conformers except ED have a resolvable red-shifted peak arising from the TT label. In the case of the disordered states (SR and ED), there is little change in the spectrum between the T10 and TT simulations, which reflects the limited correlation between the positions of their T3 and T10 peptide units. For the FO and FR conformers, which have close T3-T10 contacts, a distinct ν_{TT} peak is observed although the T10 spectrum showed no distinct T10 isotope-shifted feature. This is a result of the additional red-shift of the ν_{TT} band as a result of T3-T10 amide I coupling, which varies between 6-7 cm^{-1} for the FO conformers and leads to a $\nu_{TT} - \nu_{\perp}$ splitting of >30 cm^{-1}. Comparing the T10 and TT 2D spectra for the BT state shows few differences besides additional inhomogeneous broadening of the ν_{TT} transition. Each of the simulated spectra containing proximal T3 and T10 sites bear similarity to the experimental spectra, but those that compare favourably to calculated T10 and TT spectra are FR and FO. Finally, it was noted that the intensity of the simulated ν_{TT} feature scales with the strength of coupling between the two threonine sites. From these observations, it can be concluded that contacts between strands in the central region of the peptide exist throughout the temperature range studied. It is also concluded that out-of-registry structure is not significantly populated in the equilibrium conditions observed here.

Simulations of the S1 spectrum were used to address the lack of distinct peak for the ^{18}O isotope label, and the shifts in intensity and lineshape seen in FTIR and 2D IR spectra. In general, for the Markov states studied, the S1 site is protected from solvent, because the serine side chain and terminal -NH3+ both prefer to be oriented to solvent, and because tryptophan side chains interfere with water penetration to the S1 carbonyl. Thus, the S1 site energies are generally high regardless of conformation. The FO state has many intact S1-W11 hydrogen bonds, the SR has S1 carbonyls that experience a range of solvent and side chain environments, and the FR is the most solvent exposed.

The experimental (S1-UL) FTIR difference spectrum (Figure 5-7) indicates that intensity is lost at 1675 cm^{-1} and 1640 cm^{-1} upon S1 labeling. Simulations of this spectrum for various conformers show that this characteristic double peak loss feature matches those observed in the FO state, whereas other states show a single loss feature (Figure 5-8). Cross-strand hydrogen bonding between the S1 and W11 in the FO state leads to strong coupling across the β strand and a low frequency $\nu_{\perp}$-like mode which is decoupled from the rest of the band and shifts upon ^{18}O isotope labeling. However, the high frequency loss features in FTIR and 2D IR are also present in BT, SR, and FR, which have high site energies and weak coupling to the remaining peptide units. ^{18}O labeling shifts the S1 vibrational frequency for each system from very high energies (1700 cm^{-1}) in the unlabeled spectrum to frequencies just to the blue of the

$\nu_\perp$. The distribution of S1 configurations available in these conformers suggests that the S1 site is generally disordered at all temperatures.

5.4 Conclusions

The picture that emerges from these data provides a detailed view of a heterogeneous thermal fraying process. At 25 °C, roughly 60% of the conformers retain the type of cross-strand contacts seen in the NMR structure, whereas others have mid-strand hydrogen bonding contacts and a bulged loop. The temperature-dependent experiments indicate that central contacts remain within TZ2 and no idealized, unfolded state with a fully solvated backbone is observed. These simulations indicate that the native TZ2 state is heterogeneous, and therefore, that the folding must also be heterogenous.

From a modeling perspective, the trpzip2 calculations demonstrate the power of combining atomic-level simulations with 2D IR spectroscopy. The current level of theoretical sophistication provides enough accuracy to assign frequency trends to variations in conformation and solvent exposure. Once the assignment is made, the simulation can be interpreted in far more detail than experiments currently afford, especially to quantify the structure of water around peptides. For TZ2, more expansive simulations of the entire set of Markov states may be used to try to quantitatively reproduce the IR spectra, and be used to predict the spectral changes resulting from temperature jumps.

5.5 Acknowledgements and Afterword

Sections of this chapter were excerpted from "Melting of a β-hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations," by Smith et al.[167] I would like to thank Adam Smith, Joshua Lessing, Chunte Sam Peng, and Jongjin Kim for synthesis of the trpzip2 peptides and acquiring FT IR and 2D IR spectra. I would also like to thank Santanu Roy and Thomas la Cour Jansen for assistance with the

spectral calculations. I would like to acknowledge Bill Swope and Jed Pitera for providing us with the Markov states used in the modeling. This work benefited from extensive conversations with John Chodera, Vijay Pande, Jed Pitera, and Bill Swope on the simulation and analysis of TZ2 folding mechanisms and heterogeneity. This work was supported by the US National Science Foundation (CHE-0316736, CHE-0616575, CHE-0911107 to AT), the US Department of Energy (DE-FG02-99ER14988 to AT), the David and Lucile Packard Foundation, and the ACS Petroleum Research Fund. I would like to thank the Poitras Predoctoral Fellowship for Fellowship support.

Chapter 6

Transient Two-Dimensional Spectroscopy of Ubiquitin Unfolding: A Comparison of Dynamics from Experiment and MD Simulation

6.1 Introduction

Transient two-dimensional infrared (2D IR) spectroscopy is used as a probe of protein unfolding dynamics in a direct comparison of fast unfolding experiments with molecular dynamics simulations. In the experiments, the unfolding of ubiquitin is initiated by a laser temperature jump, and protein structural evolution from nanoseconds to milliseconds is probed using amide I 2D IR spectroscopy. The temperature jump prepares a subensemble near the unfolding transition state, leading to quasi-barrierless unfolding (the burst phase) before the millisecond activated unfolding kinetics. The burst phase unfolding of ubiquitin is characterized by a loss of the coupling between vibrations of the β-sheet, a process that manifests itself in the 2D IR spectrum as a frequency blue-shift and intensity decrease of the diagonal and cross-peaks of the sheet's two IR active modes. As the sheet unfolds, increased fluctuations and solvent exposure of the β-sheet amide groups are also characterized by increases in homogeneous linewidth. Experimental spectra are compared with 2D IR spectra calculated from the time-evolving structures in a molecular dynamics simulation of ubiquitin unfolding. Unfolding is described as a sequential loss of strands in ubiquitin's β-sheet, using two collective coordinates of the sheet: (i) the native interstrand contacts between adjacent β-strands I and II and (ii) the remaining β-strand contacts within the sheet. The methods used illustrate the general principles by which 2D IR spectroscopy can be used for detailed dynamical comparisons of experiment and simulation.

From one perspective, protein folding is a chemical dynamics problem concerning a description of the interplay of noncovalent interactions that involve the protein and surrounding solvent and exploration of the configurational space that leads to formation of the native structure. Although individually these interactions act on short time and distance scales, collectively they result in nanometer-scale conformational changes observed over time scales from picoseconds to seconds. The vast range of length and time scales, and the collective nature of folding coordinates, ensure that no single technique can time-resolve all relevant structural changes in solution[175–177]. Most experimental methods favor either temporal or structural resolution and are limited to characterizing folding rates (kinetics) rather than mechanism (dynamics). From a computational approach, molecular dynamics (MD) simulations offer detailed dynamical information at atomic resolution for single molecules.[178] Yet, simulation of folding is also challenged by the microsecond or longer time scales required, the need for extensive sampling of the ensemble, and only indirect experimental benchmarks. These challenges have spurred an interest in comparison between experiment and simulation that builds on their combined strengths to address mechanistic questions.[179,180]

As an ultrafast vibrational spectroscopy that probes transient molecular structure in solution, two-dimensional infrared (2D IR) spectroscopy provides an avenue to directly reveal protein folding dynamics. 2D IR spectroscopy achieves its time resolution through the use of femtosecond mid-IR pulse sequences and its structural information by probing vibrational couplings.[5,133,181] It is now being used in several contexts to characterize the structure and dynamics of proteins and peptides,[2,55,73,128,182,183] including unfolding.[3,25,65,86] The target in the majority of these studies is amide I spectroscopy of the polypeptide backbone. Despite lacking atomic resolution, amide I IR spectroscopy is appealing because changes in peak positions and spectral lineshape reflect variations in protein secondary structures.[67] In addition, models to calculate amide I spectra have advanced to the point where it is feasible to directly connect amide I 2D IR spectra with structure.[8,127,128,133,173] Thus, it becomes possible to perform transient 2D IR spectroscopy of protein folding in silico, using MD simulation to interpret the underlying mechanism of unfolding within the framework of experimentally constrained time scales and structural changes. Here we present such a method of revealing dynamics in experiments, characterizing how ubiquitin unfolds after a temperature jump (T-jump) and drawing on MD computer simulations to assist in the detailed dynamical interpretation.

Transient folding studies are a common experimental approach to time-dependent studies of protein folding,[176,184–186] but most experiments characterize kinetics: the rates of crossing between stable minima on a free-energy surface. What has always been desirable, but difficult to achieve, is to describe dynamics: a direct, time-dependent characterization of molecular structure. Observation of dynamics in experiments is typically hindered by barriers along the reaction coordinate. However, there is now evidence that fast-folding experiments initiated with a T-jump can shift equilibrium conditions in such a way that a nonequilibrium state is prepared at or near the folding transition state.[21,22] These downhill unfolding experiments work in a diffusive regime with barriers of ≈kT and show nanosecond-to-microsecond time scale nonexponential relaxation.[185]

Our interpretation of a downhill unfolding experiment is shown by the free-energy profiles in Fig. 6-1. At the initial temperature, equilibrium exists between a folded state and an unfolded state. The abrupt T-jump shifts the barrier and free-energy profile along the observed unfolding coordinate by an amount that depends on the degree of disorder in the native state. A fraction of the originally folded population is trapped near the new unfolding transition state. This subensemble unfolds in a quasi-barrierless manner, whereas the majority of the population equilibrates at the new temperature by longer-time activated barrier crossing. The fast response of this transient subensemble appears as the burst phase signal in conventional protein-folding kinetics. With sufficient time resolution, a structure-sensitive probe such as 2D IR provides an opportunity to describe unfolding dynamics by watching the downhill process, which is temporally isolated from the slower unfolding kinetics.[187]

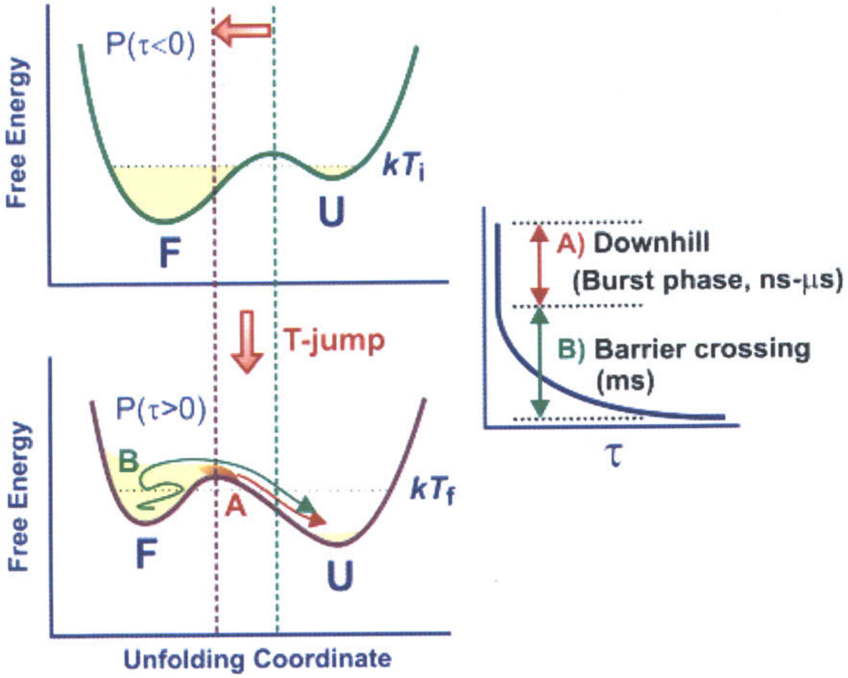


Figure 6-1: Downhill unfolding dynamics. A T-jump induces a barrier shift toward the folded state. A subensemble is trapped at the shifted transition state and unfolds in a downhill manner on the nanosecond-microsecond time scale (A). The downhill unfolding appears as a burst phase in experiments with millisecond time resolution. Subsequently, the excess population in the folded well unfolds across the barrier, which results in a millisecond exponential relaxation (B). The dotted vertical lines indicate barrier positions before and after the T-jump.

In the following, we track the unfolding of ubiquitin from nanosecond to millisecond time scales with 2D IR spectroscopy, to structurally characterize the burst phase and the slow unfolding kinetics. The experiments probe two vibrational modes of ubiquitin's β-sheet and the random coil region. Spectral changes can be interpreted as a sequential unfolding of strands within the sheet, ending with the strand I-II hairpin. (The structure of ubiquitin and a 2D representation of the contacts between strands of the sheet are presented in Fig. 6-2.) The detailed picture emerges from comparison with MD simulation. We demonstrate this by using a spectroscopic model to calculate transient 2D IR spectra from a T-jump unfolding simulation of ubiquitin, which we use to interpret the collective frequency shifts and intensity changes observed in the experimental data. This method provides a general approach to dynamical interpretations of folding experiments and testing of folding simulations.

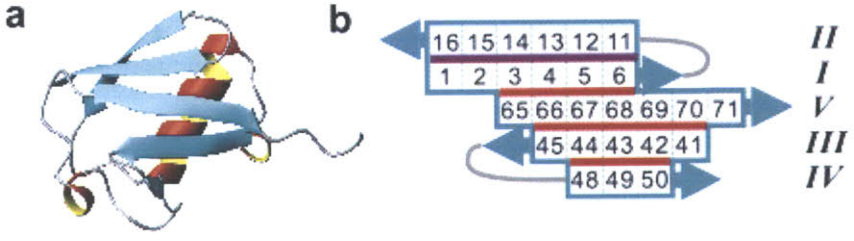


Figure 6-2: Structure of ubiquitin. (a) Crystal structure of ubiquitin[9] rendered with MOLMOL.[10] (b) Projection of the β-sheet of ubiquitin. A square box with a digit n represents a peptide group formed by residues n and $n+1$. Red and purple lines indicate native contacts used in calculating the interstrand distance coordinates R_1 and R_2, respectively, in Fig. ??.

6.2 Materials and Methods

6.2.1 2D IR Spectroscopy

For equilibrium and transient measurements, we used Fourier transform 2D IR spectroscopy employing 90-fs mid-IR pulses resonant with the amide I vibrational band ($\lambda = 6$ μm; FWHM, 160 cm^{-1}). Details of how we perform equilibrium Fourier transform 2D IR spectroscopy are given elsewhere.[5] Briefly, the third-order nonlinear signal from which the 2D IR spectrum is derived is obtained from three sequential pulses (a, b, and c in Fig. 6-3), which are crossed in and focused to a 100-μm-diameter spot at the sample and delayed with respect to each other by time intervals t_1, t_2, and t_3. The nonlinear polarization generated by these pulses emits a signal into the wave-vector matched direction ($\mathbf{k_s}$ = -$\mathbf{k_a}$ + $\mathbf{k_b}$ + $\mathbf{k_c}$). The third-order signal is combined with the LO on either side of a 50:50 beam splitter, yielding two matched pairs of combined fields with a π phase shift between them. Each pair of fields is dispersed by a monochromator and imaged onto one stripe of a dual-stripe mercury-cadmium-telluride (MCT) array detector (2x64). Balanced detection is performed by subtracting signals on the lower stripe from those on the upper stripe to obtain heterodyned components and to minimize the baseline fluctuation by removing the LO intensity and the homodyne echo signal.[188, 189]

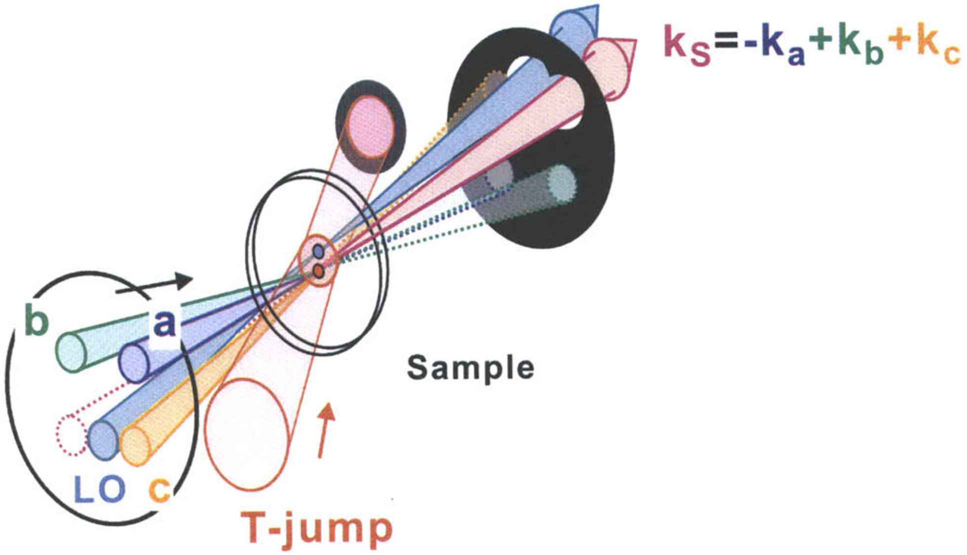


Figure 6-3: Beam alignment and data processing for transient 2D IR spectroscopy. Alignment of beams in the sample region (a). The LO is vertically displaced by 100 μm from the other three beams (a, b, and c) that generate the third-order signal.

The array disperses the heterodyned signal into the ω_3 dimension of the spectrum. The heterodyned signal is collected as a function of t_1, the delay between the first and second 6-μm pulses. Numerical Fourier transformation with respect to t_1 results in the ω_1 dimension. We obtain the complex 2D IR correlation spectrum by summing the rephasing ($\mathbf{k_R}$ = -$\mathbf{k_1}$ + $\mathbf{k_2}$ + $\mathbf{k_3}$) and nonrephasing ($\mathbf{k_{NR}}$ = $\mathbf{k_1}$ - $\mathbf{k_2}$ + $\mathbf{k_3}$) spectra (4). Here we present the absorptive or real part of the correlation spectrum , as described previously.[5] Both the equilibrium and transient 2D IR spectra were obtained by undersampling in the t_1 dimension with a step of 14 fs from 0 to 2.1 ps and 0 to 1.2 ps for rephasing and nonrephasing spectra, respectively. The actual frequency ω_1 is obtained by reflecting the transformed frequency (ω_{1u}) to the Nyquist frequency (ω_N) as $\omega_1 = 2\omega_N - \omega_{1u}$. The combination of balanced detection and undersampling provides the needed improvement in data acquisition time for T-jump measurements. All beams have polarizations controlled with wire-grid polarizers, allowing parallel (ZZZZ) and perpendicular (ZZYY) spectra to be acquired.

6.2.2 Transient 2D IR spectroscopy

Here we provide a summary of the methods necessary to surmount three technical challenges to obtain the transient data: (i) synchronization of the high-energy 20-Hz nanosecond T-jump laser with the 1-kHz 2D IR femtosecond laser system, (ii) acquisition of interferometric measurements through a T-jumped sample region, and (iii) signal-to-noise improvements to overcome the reduced repetition rate for data acquisition (described above). Full experimental details and technical challenges of the T-jump 2D IR spectrometer required for these measurements were be published by Chung et al.[190]

6.2.3 Temperature Jump Synchronization

For transient 2D IR spectroscopy, changes in the 2D IR spectrum are monitored after the preparation of a nonequilibrium state by an abrupt T-jump. A 7-ns T-jump laser pulse (λ= 2 μm, 4.2 mJ) is obtained from a beta barium borate (BBO)-based optical parametric oscillator (OPO) pumped by the second harmonic of a 20-Hz, Q-switched Nd:YAG laser and focused to 500 μm in the sample. Weak absorption of the T-jump pulse by O-D overtone stretching vibrations of water raises the temperature by 9° from T_i to T_f on the same timescale of the T-jump pulse duration. The transient temperature profile is characterized by the small change of solvent transmission at 6 μm. The temperature remains constant at T_f from 10 ns to $\approx$ 100 ms and then re-equilibrates to T_i by 10 ms (see Fig. 6-6f).

Synchronization between the mid-IR probe pulses and the T-jump pulse relies on subsequent division of the 82-MHz photodiode signal from the Ti:sapphire oscillator pulse train to 1 kHz and 20 Hz, respectively. For each T-jump pulse, 50 mid-IR pulses probe the sample at 1-ms intervals. The delay, t, between the T-jump pulse and the first probe pulse is controlled electronically between 10 ns and 1 ms. The 49 following probe pulses detect millisecond structural changes, leading to the ability to probe time delays between 100 ns and 50 ms. In this paper, we display the transient difference correlation spectrum, $\Delta S = \tilde{S}'(\tau, T) - \tilde{S}'_0(T_i)$, between the transient response and a reference spectrum at the initial temperature T_i, which, in practice, is obtained from the mid-IR pulse preceding the T-jump. To maximize signal strength, transient spectra were acquired in the parallel (ZZZZ) polarization geometry.

For T-jump measurements, to keep the relative phase between the LO and the third-order signal constant at the T-jump point in the sample, the LO is sent through the same region as the other beams. However, to avoid pump-probe signals from the LO, it is vertically displaced by 100 μm from the focal spot of the other three beams (see Fig. 6-3). Because the focal spot size of the T-jump laser (500 μm in diameter) is much bigger than that of the 6-μm beams (100 μm in diameter), the temperature at the two spots is similar and the relative phase between the LO and the third order signal does not change.

6.2.4 Sample Preparation

Ubiquitin was purchased from Sigma Aldrich (U6253; St. Louis, MO). The concentration of the ubiquitin sample was 30 mg/ml in 0.35% (wt/wt) DCl/D2O solution (pH* ≈ 1). The solution was placed between two CaF_2 windows separated by a 50-μm-thick Teflon spacer. The windows are mounted in a brass sample cell, whose temperature was controlled at T_i to $\pm 0.1^\circ$ C with a circulating water bath. IR spectroscopy of the sample reveals a sigmoidal unfolding curve with a melting temperature of $64.0\ ^\circ$ C.

6.2.5 Simulating 2D IR Spectra

Our calculations of 2D IR spectra require two elements: (i) an atomistic molecular dynamics simulation with explicit water and (ii) a structure-based spectroscopic model that draws on the simulation trajectory to model 2D IR spectra. We use a MD simulation of ubiquitin unfolding following an instantaneous T-jump from the crystal structure to 498 K, which was originally presented as trajectory D2 by Alonso et al.[150] The 1.6-ns trajectory was provided by Valerie Daggett (Department of Medicinal Chemistry, University of Washington, Seattle, WA), with snapshots of the protein structure sampled at 1 ps. Although the trajectory was supplied without water molecules, structures were resolvated for spectral calculations because the solvent is important for reproducing the spectra.

MD Simulations

The ubiquitin unfolding trajectory was obtained with atomic coordinates for the protein. Protein snapshots were resolvated by inserting them into cubic boxes of equilibrated SPC/E D_2O encasing the protein at least 10 Å in each direction and removing solvent molecules closer than 2.2 Å. This resulted in 5,000-7,000 water molecules. Although the denaturation simulation was performed at low pH, resolvation was done at neutral protein charge to eliminate the slow counterion diffusion degrees of freedom. By using the CHARMM 30b1 package and the CHARMM22 force field, the system was energy-minimized with steepest descents and cubic periodic boundaries for 500 steps, during which the protein was constrained in space with a 24-kcal/mol $Å^2$ harmonic restoring force. Particle mesh Ewald sums with $k = 0.32$ were used for the electrostatic energies. Van der Waals interactions were shifted to truncation at 14 Å. With all bond lengths fixed using the SHAKE algorithm, the water was then allowed to equilibrate around the fixed protein structure during 10 ps of dynamics with a 2-fs time step in the NPT ensemble at 300 K and 1 atm. 20-ps trajectories sampling dynamics were run, for which structures were saved each 50 fs and used to calculate spectra.

Spectral Modeling

Because IR spectroscopy characterizes vibrational eigenstates, a model is required in order to translate the atomistic structure into 2D IR spectra. We use a local amide Hamiltonian (LAH) model that is widely used to describe amide I spectroscopy of proteins and peptides.[73, 105, 133] The model assigns a local-mode Hamiltonian on the basis of the protein's peptide units, which contains local-mode frequencies (site energies) and vibrational couplings (through-bond and through-space). The LAH is diagonalized and used to obtain transition dipole eigenvectors and energy eigenvalues for the system, which can then be used to calculate experimental observables from a response function.[8]

Considering only the vibrational subspace of amide I vibrations, each structure from the simulation was used to generate a local amide Hamiltonian. Amide I site energies are sensitive to CO and NH hydrogen bonding and are parameterized through linear correlation coefficients between the amide I energy and the electrostatic potential originating in the surrounding protein and solvent. We used the four-site model of Bour and Keiderling[123] (sampled at C_α, C, O, and N atomic sites), which we have referred to as "Keiderling4"[8]. Coupling elements between these sites depend on their distance and relative orientation.[171] Through-bond coupling is assigned from an ab initio calculated (ϕ,ψ) map that builds in through-bond coupling. Through-space couplings use the electrostatic, transition charge model. A local transition dipole is also set for each site and, after diagonalization, mix to form the eigenstate transition dipole for each corresponding energy. The Hamiltonian is also scaled and anharmonically shifted to provide two-quantum energies and generate two-quantum transition dipoles. With these parameters, a 2D IR spectrum is calculated for each structure snapshot, assuming Lorentzian lineshapes (γ = 10 cm^{-1}) for perpendicular (ZZYY) polarization. Realistic, inhomogeneous lineshapes are generated by assuming the static limit and summing 100 2D IR spectra with different solvent configurations for each protein structure. Protein conformational variation in each region is accounted for by summing spectra from different snapshots, as follows: A (0, 70, 114 ps), B

(150, 160, 168, 169, 170 ps), C (179, 186, 190, 200, 286 ps), D (315, 400, 424, 510, 511, 512 ps), E (523, 524, 525 ps).

A significant distinction from the experimental difference spectra is that the simulated transient spectra only reflect the unfolding dynamics and do not include the signal changes resulting from the transmission change of the D_2O solvent.

6.3 Experimental Results and Discussion

6.3.1 Equilibrium Measurements

The 2D IR spectrum correlates the frequency of initial vibrational excitation (ω_1) with a final detection frequency (ω_3). The frequencies of diagonal peaks can be assigned to chemically distinct vibrational modes. The presence and splitting of cross-peaks characterizes the anharmonic coupling of the vibrations and helps decompose congested spectra. We concentrate on the diagonal and cross-peaks between two vibrational bands of β-sheets ($\nu_\perp$ and $\nu_\parallel$), whose individual amide oscillators vibrate in-phase perpendicular or parallel to the β-strands, respectively.[3, 12, 70, 86] The splitting between these modes and the frequency of $\nu_\perp$ in particular are indicators of the size of the folded β-sheet[70] and provide an important signature in transient experiments.

Fig. 6-4 shows equilibrium 2D IR spectra taken at 63 and 72° C, the initial and final temperatures of the T-jump experiment. Absorptive spectra were acquired with parallel (ZZZZ) and perpendicular (ZZYY) probing polarizations. For ubiquitin, the $\nu_\perp$ and $\nu_\parallel$ β-sheet modes are observed on the red (1642 cm^{-1}) and blue (1676 cm^{-1}) sides of the amide I spectrum. At 63° C, these two transitions are not clearly resolved because of inhomogeneous broadening but appear as a broad diagonal peak both for the fundamental transition ($\text{v} = 0 \rightarrow 1$, positive) and for the overtone transition ($\text{v} = 1 \rightarrow 2$, negative). The overtone transition lies below the fundamental along the ω_3 axis because of the anharmonicity of the vibrational potential. Because the $\nu_\perp$ and $\nu_\parallel$ modes have nearly orthogonal transition moments, the cross-peaks are small in the parallel polarization geometry but are enhanced and form a cross-peak ridge in the upper left corner (purple arrow) in the perpendicular polarization geometry. Loss of negative intensity (a positive change) in the lower right corner (green arrow) also indicates the presence of a positive cross-peak in this region. The overall Z-shape of the perpendicular spectrum, which arises from interference effects between $\nu_\perp$ and $\nu_\parallel$ diagonal and cross-peaks, is a characteristic signature of the β-sheet structure.[2]

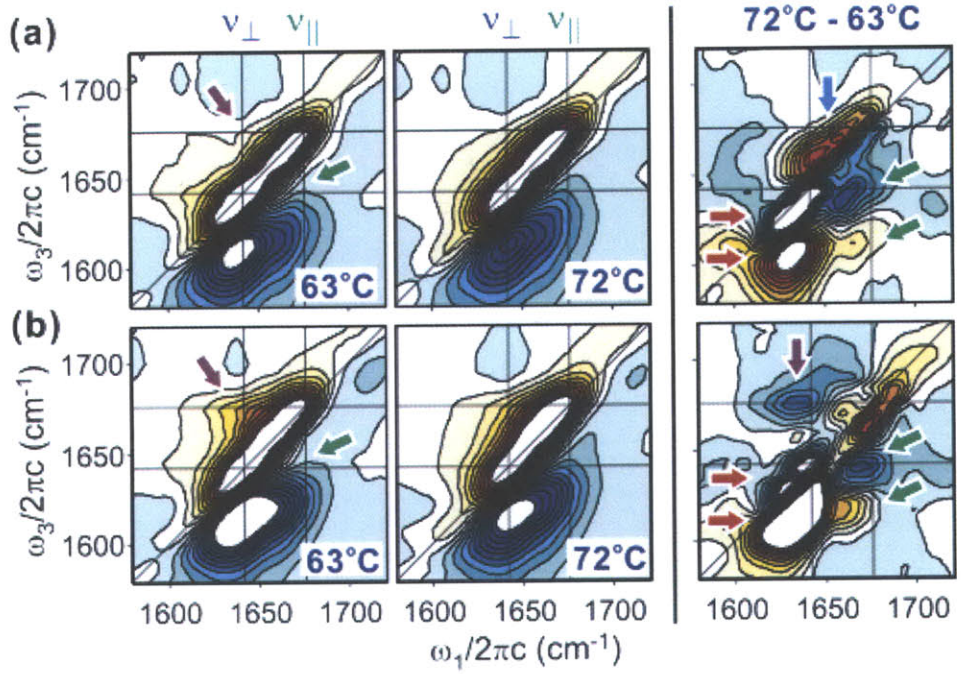


Figure 6-4: Equilibrium thermal unfolding of ubiquitin monitored by 2D IR spectroscopy. Parallel (ZZZZ) (a) and perpendicular (ZZYY) (b) polarization geometries. Spectra are normalized to the maximum of the 63 ° C spectrum. Twenty-one contours are plotted for ±60% of the spectra at 63 and 72 ° C and for 15% of the difference spectra (Right). Positive and negative peaks are indicated by red and blue. Green and purple arrows represent cross-peaks. In the difference spectra, red and blue arrows indicate the diagonal peaks on the red and blue sides of the ω_3 axis, respectively. $\nu_\perp$ and $\nu_\parallel$ transitions are marked with lines.

As the temperature is raised, the $\nu_\perp$ transition blue-shifts along the diagonal, which is observed as a negative/positive doublet in the difference spectrum (marked with red arrows in Fig. 6-4a Right and b Right). Concomitantly, in the off-diagonal region (lower right corner), the negative/positive doublet (marked with green arrows) appears, indicating the depletion of the $\nu_\perp/\nu_\parallel$ cross-peak intensity. These changes indicate disruption of the β-sheet. This disruption is accompanied by an increase in diagonal intensity in the random coil region of the amide I spectrum, ν_R, indicating an increase in disordered structures that result from unfolding (blue arrow). Conversely, in the perpendicular difference spectrum (Fig. 6-4b Right), this region is dominated by depletion of the cross-peak intensity (purple arrow) because cross-peaks are enhanced in this polarization geometry. The spectral features of these two difference spectra help in the interpretation of the transient 2D IR difference spectra.

6.3.2 Transient Thermal Unfolding of Ubiquitin

Previously,[25] we demonstrated that the transient unfolding response of ubiquitin can be divided into three stages. The fastest change is observed coincident with the nanosecond T-jump pulse as a result of increased thermal excitation of the solvent. The second stage is the microsecond downhill unfolding of the subensemble trapped at the unfolding transition state, as illustrated in Fig. 6-1. The last stage is millisecond unfolding over a barrier. Transient 2D IR spectra capture these changes as described below.

Transient 2D IR difference spectra (ZZZZ) after a T-jump from T_i = 63° C to T_f = 72° C are shown in Fig. 6-5 for delays between τ = 100 ns and 7 ms. The maximum spectral changes observed are 2.7%. For the fastest response observed, the τ = 100 ns spectrum, the protein conformational change is small and the spectrum is similar to those observed for a T-jump performed at low initial temperature (T_i = 25° C) for which the protein does not unfold (Sec 6.3.4). The difference spectrum shows an increase in the amplitude of the broad positive and negative diagonal peaks as a result of the small transmission increase at 6 μm on temperature elevation. The depletion on the red side of the diagonal region (red ellipse) is attributed to the blue shift of the sheet vibrations caused by a weakening or partial disruption of hydrogen bonds within the sheet. This loss feature is more intense for T_i = 63° C than for T_i = 25° C, indicating that it is not just a result of the sample density decrease. Time-dependent spectral changes observed in the subsequent spectra are different from those observed at T_i = 25° C and originate in unfolding of the protein. This is apparent from the similarity between the transient difference spectrum at 7 ms and the equilibrium difference spectrum shown in Fig. 6-4a.

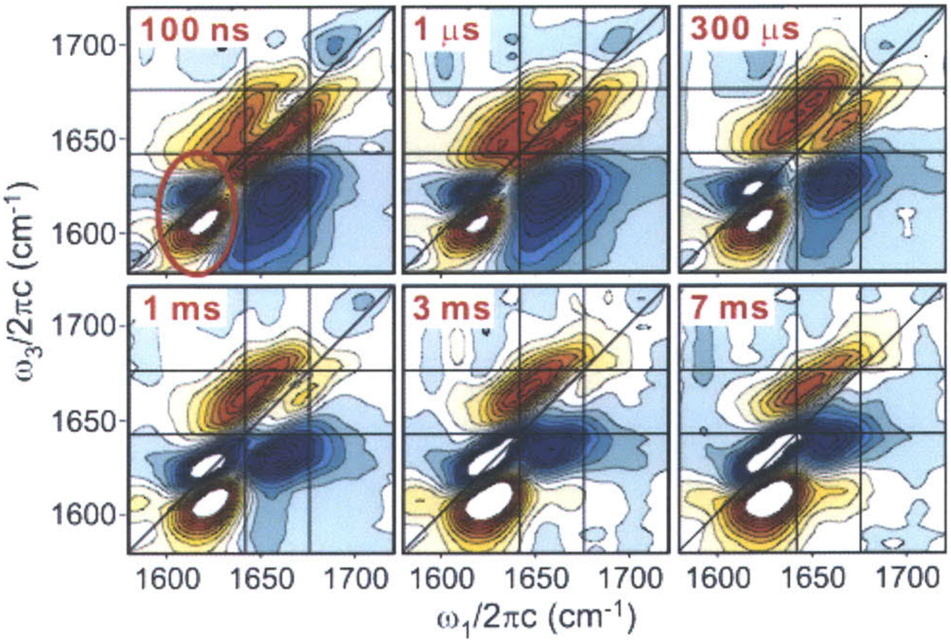


Figure 6-5: Transient 2D IR difference spectra (ZZZZ) after a T-jump from 63 to 72°C. Transient difference spectra are plotted as a function of delay τ. Twenty-one contours are plotted at 1.5% of the maximum of the reference spectrum at $T_i = 63$ °C. The spectra from 1 to 7 ms are obtained from the same data set as the 100-ns spectrum. The red ellipse in Upper Left indicates depletion on the red side of the diagonal region.

In Fig. 6-5, distinct transient spectral changes are observed in three different spectral regions: (i) the low-frequency diagonal region, (ii) the off-diagonal region for $\omega_1 > \omega_3$, and (iii) the high-frequency detection region (upper half of each spectrum). In the low-frequency diagonal region (lower left), the amplitude of the negative/positive doublet of the β-sheet $\nu_\perp$ mode rises over microsecond-to-millisecond time scales and shifts to the blue along the diagonal axis as a result of β-sheet unfolding. The disruption of the β-sheet can also be monitored by the cross-peak intensity in the lower right. Although initially dominated by a negative peak originating from the D_2O thermal transmission increase, with increased delay the negative/positive cross-peak doublet is formed and pushes the negative diagonal peak to the blue. For the upper half of the transient spectrum, the changes are less dramatic. The transmission increase induces a positive change in this region, and the increase of the random coil component by thermal unfolding also appears as a positive change, both of which lead primarily to lineshape variations. The global time-dependent changes in the transient spectra characterized through singular value decomposition (SVD) are shown in Fig. 6-6a. The relaxation curve obtained from the first SVD component shows a gradual nonexponential decay from 100 ns to 100 μs, an abrupt decay on the 2-ms time scale, and finally a return to baseline for $\tau > 5$ ms. The final stage represents the refolding of the protein after the temperature of the sample reequilibrates.

The short time response corresponds to a downhill unfolding path in which interstrand contacts within the β-sheets of a small fraction of the total population are gradually reduced. This can be shown by using slices through the transient spectra for $\omega_1 = 1620$ and 1642 cm^{-1}. The slice at $\omega_1 = 1642$ cm^{-1} (Fig. 6-6b) shows a clear blue shift in the diagonal peak. The difference signal of the $\nu_\perp$ mode, $(\omega_1, \omega_3) = (1642, 1639)$, is positive initially but becomes negative with a time-dependence that contains both the fast downhill and slower millisecond components (Fig. 6-6c). In that slice, the off-diagonal region $(\omega_1, \omega_3) = (1642, 1663)$, which is dominated by the random coil region of the spectrum, is observed to rise gradually over the entire unfolding period. For the slice at $\omega_1 = 1620$ cm^{-1}, the positions of the diagonal peaks do not change but the amplitude increases slightly. However, the positive peak initially formed in the $\nu_\perp$ region at $\omega_3 \approx 1640$ cm^{-1} blue-shifts with delay time on microsecond and millisecond time scales (Fig. 6-6e). Although the shift may be the partial result of temperature-dependent solvent transmission changes, the similarity in the relaxation profiles (Fig. 6-6c and e) indicates that the decrease in intensity of the $\nu_\perp$ diagonal region is correlated with the blue shift of the $\nu_\perp$ assigned peak along the shown $\omega_1 = 1620$ cm^{-1} slice as unfolding of the β-sheet proceeds. The change in intensity indicates a decrease of amide vibrations contributing to the β-sheet mode, whereas its blue shift may indicate a decrease in delocalization of the β-sheet excitation and, therefore, a decrease in its folded size.

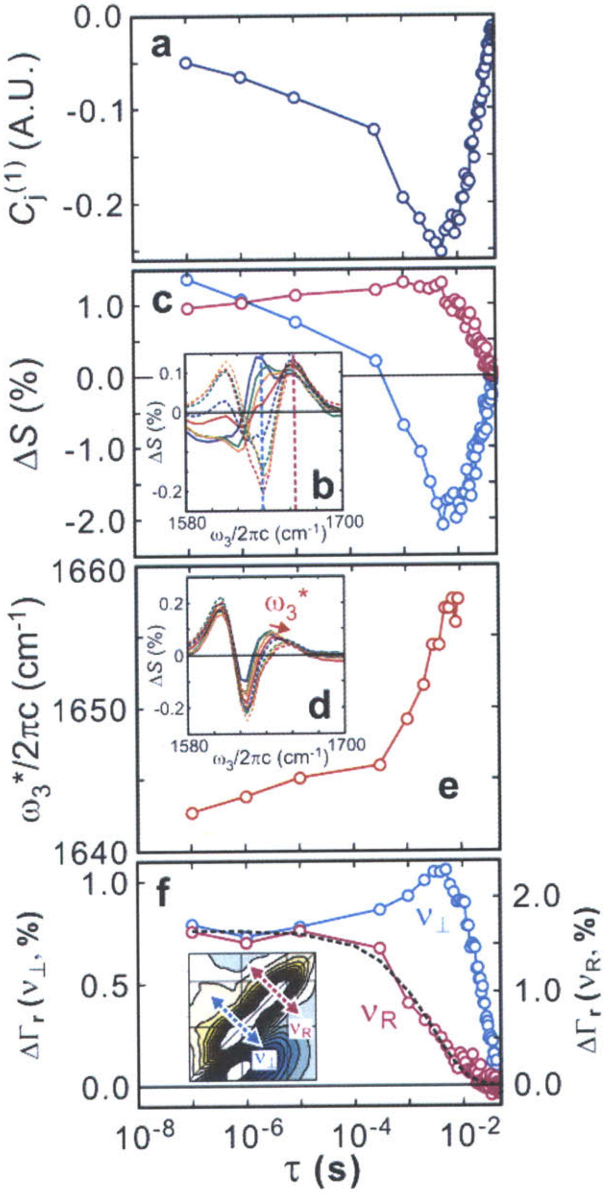


Figure 6-6: Semilog plot of transient changes in 2D IR spectra. (a) Temporal profile of unfolding and refolding of ubiquitin constructed from the first SVD component of the transient difference spectra shown in Fig. 6-5. (b) Transient changes of slices at ω_1 = 1642 cm^{-1} for representative delays between τ = 100 ns (blue solid line) and 7 ms (red dashed line). (c) Relaxation profiles from ω_3 slices in b at ω_3 = 1639 cm^{-1} (light blue) and 1663 cm^{-1} (magenta). (d and e) Transient slices at ω_1 = 1620 cm^{-1} and the corresponding frequency shift of the positive peak ω_3*. (f) Relative changes in the antidiagonal width of the $\nu_{\|}$ (light blue) and random coil (magenta) components marked with arrows in the Inset. The normalized transient temperature relaxation profile (black dashed line) is also shown.

6.3.3 Line Broadening During Unfolding

Variation of the homogeneous vibrational linewidth observed in the T-jump from $T_i = 63^\circ$ C can be used to characterize changes in amide group fluctuations during the unfolding of the protein. In addition to vibrational population relaxation, homogeneous line-broadening of amide I vibrations of proteins depends on solvent perturbations and structural fluctuations of the protein. The correlation time of solvent fluctuations, characterized through the amide I frequency correlation function for N-methylacetamide, has been shown to be $\approx$1 ps (or 5 cm^{-1} of half-width half-max linewidth).[168] In a protein, this contribution changes based on the solvent accessibility of the amide groups. Backbone fluctuations in the protein's local minimum on a $\approx$1-ps time scale also contribute to the homogeneous linewidth by modulating the coupling and hydrogen bonding between amide I oscillators.

For proteins consisting of many amide I oscillators, the lineshape and linewidth of an IR absorption spectrum strongly depend on the distribution of amide I frequencies. Because of this broad, inhomogeneous character, the homogeneous line-broadening cannot be independently observed in the FTIR spectrum. However, the inhomogeneous and homogeneous parts can be separated in the 2D lineshape through diagonal and antidiagonal slices, respectively.[191] Thus, any changes to amide I frequency fluctuations as a result of solvent exposure of those peptide units should be observed as antidiagonal broadening.

The relative change of the homogeneous linewidth ($\Delta\Gamma_r$) induced by a T-jump is defined as

$$\Delta\Gamma_r\left(\tau,\omega_{AD}\right) = \frac{\Gamma\left(\tau,\omega_{AD}\right) - \Gamma\left(0,\omega_{AD}\right)}{\Gamma\left(0,\omega_{AD}\right)}, \tag{6.1}$$

where ($\Gamma\left(\tau,\omega_{AD}\right)$) is the antidiagonal linewidth (FWHM) along the slice $\omega_{AD} = (\omega_1 + \omega_3)/2$ for the transient 2D IR spectrum at delay τ. In Fig. 6-6f, time-dependent changes of the homogeneous width are plotted for two frequency components: the $\nu_\perp$ mode ($\omega_{AD} = \omega_\perp = 1642$ cm^{-1}) and the random coil region ($\omega_{AD} = \omega_R = 1668$ cm^{-1}). The changes in $\Delta\Gamma_r$ in the random coil region ($\omega_{AD} = \omega_R$) track the temperature relaxation. Conversely, for the $\nu_\perp$ mode, the initial change of 0.8% at 100 ns increases to ≈1.1% at 6 ms before reequilibrating with the refolding of the protein.

The variation of homogeneous linewidth in the random coil region follows the transient temperature profile because the coil regions are flexible and solvent-exposed, even in the folded state. This conclusion is supported by the broader linewidth of this region in the equilibrium spectrum. The antidiagonal widths for the $\nu_\perp$ and random coil features in Fig. 6-4a are 10.6 and 12.7 cm^{-1}, respectively, and they increase by 17% and 7%, respectively, between the initial and final temperatures. The increase of the random coil region is smaller, as in the transient case. The changes in the homogeneous linewidth for the T-jump from 25 to 35° C, where structural change is small, track the temperature profile, regardless of the frequency (Fig. 6-9).

6.3.4 Comparison to T-jump 2D IR Spectrum of Ubiquitin at T_i = 25 °C

In our previous observations in the transient DVE experiment,[25] ultrafast responses were seen to arise from solvated regions of the protein on the nanosecond timescale. The ultrafast responses induced by a T-jump in the DVE spectra are characterized by two features: increase of the signal due to increased solvent transmission at 6 μm and depletion of the intensity on the red side of the spectrum (≈1600 cm^{-1}) due to decreased hydrogen bonding with the solvent. 2D IR versions of these responses during the T-jump (25 → 35C) are shown in Fig. 6-7.

To separate the possible temperature-dependent effects of density and transmission changes from protein unfolding signals, we compare the unfolding response to a low-temperature transient response at T_i = 25 ° C, in which the protein does not unfold. In the transient 2D IR difference spectrum at 100 ns, both the positive and negative diagonal peaks become stronger as a result of the increased transmission of the probe beams and the LO. Also, the weakening of solvent hydrogen bonds induces the blue shift of peaks and the loss of intensity of the diagonal region on the red side (marked with a red ellipse in Fig. 6-7). However, very little spectral change is found at 1 ms except for the decrease in the magnitude of the signal. This lack

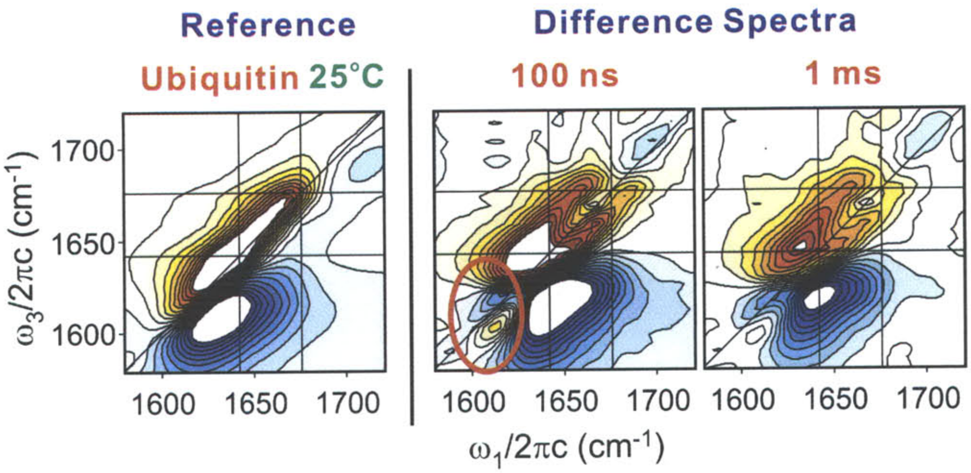


Figure 6-7: Transient 2D IR spectra of ubiquitin (25→35°C). (Left) Reference 2D IR spectrum of ubiquitin at $T_i = 25\ ^\circ$ C in the ZZZZ polarization geometry. Twenty-one contours are plotted in ±60% of the maximum. (Right) Transient 2D IR difference spectra $[\Delta S(t)]$. Twenty-one contours are plotted from ±1.5% of the maximum of the reference spectra. Depletion of the signal on the red side of the diagonal regions at 200 ns is marked with a red ellipse. The guidelines parallel to the two frequency axes (ω_1 and ω_3) mark the two resonances of the β-sheet in ubiquitin.

of spectral change is expected because there are no significant structural changes in this low temperature range. The spectral change is mostly due to the modulation of the transmission of the solvent by a temperature re-equilibration. However, there is a slight difference between the reference spectrum and transient difference spectra. The antidiagonal width of the positive peak in the transient spectra is broader than that in the reference spectrum. Because the antidiagonal width of a 2D IR spectrum represents the homogeneous linewidth,[191] this change indicates the increase in homogeneous broadening with temperature. Quantitative analysis of this effect is found in the main text.

6.3.5 Comparison with DVE Results

A DVE spectrum is mathematically identical to the squared absolute value of the projection of a complex 2D IR spectrum onto the ω_3 axis as

$$\tilde{S}_{DVE}\left(\tau_1 = 0, \tau_2, \omega_3\right) \propto \left| \int_{-\infty}^{\infty} \tilde{S}'\left(\omega_1, \tau_2, \omega_3\right) + i\tilde{S}''\left(\omega_1, \tau_2, \omega_3\right) d\omega_1 \right|^2 \tag{6.2}$$

Therefore, the result of the T-jump DVE experiment can be reproduced from 2D IR spectra by integrating them over the ω_1 axis. Fig. 6-8a and b show the differences in the measured and reconstructed DVE spectra. Although the spectral shape and the amplitude of changes are somewhat different because of the different spectral shape of the femtosecond IR pulses between the two measurements, their time-evolutions describe the same dynamics. At 100 ns, both spectra show depletion of the intensity at $\approx$1600 cm^{-1}. Also, a decrease and blue shift in the $\nu_\perp$ region due to the unfolding of the protein occurs on the microsecond-to-millisecond time scale.

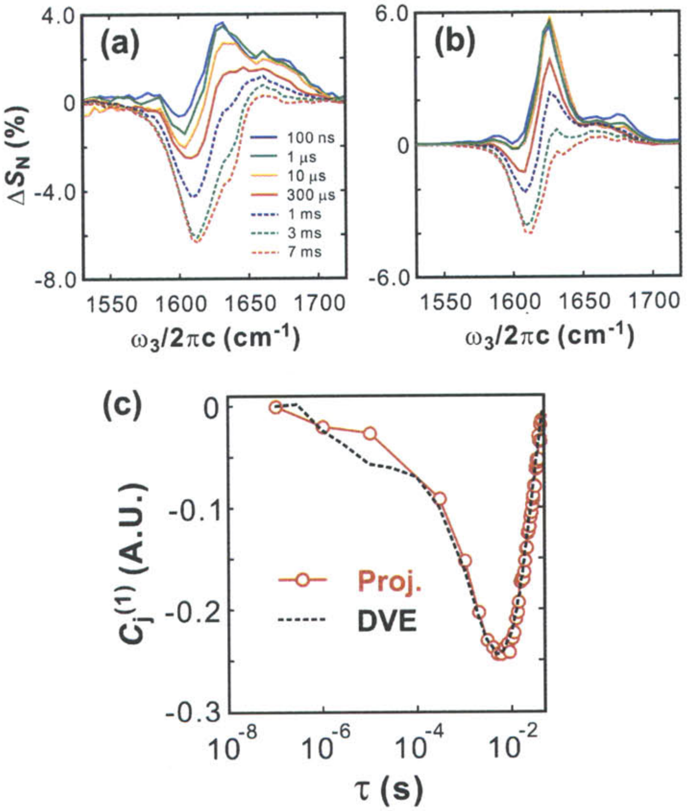


Figure 6-8: Comparison of DVE spectra measured and reconstructed from 2D IR spectra. (a) DVE difference spectra obtained from the T-jump at T_i = 63 ° C. Each spectrum is divided by the maximum of the reference DVE spectrum. (b) DVE difference spectra constructed from the absolute value square of the projection of complex 2D IR spectra onto the ω_3 axis. (c) Comparison of relaxation profiles constructed from the first singular value decomposition component of the $\nu_\perp$ region (1577-1651 cm^{-1}).

For better comparison, we applied singular value decomposition (SVD) to the $\nu_\perp$

region. The relaxation profiles are obtained from the coefficients of the first SVD component and are compared in Fig. 6-8c. As expected, the two relaxation curves match each other well. The two unfolding phases on the microsecond and millisecond timescales and the refolding phase after 7 ms observed in the DVE spectra are successfully reproduced from 2D IR spectra, which confirms that the two different measurements are probing the same dynamics. This consistency provides a very important basis for the comparisons made in further analyses shown in the main text.

6.3.6 Temperature Dependence of Homogeneous Broadening

When the structural change due to a T-jump is small, the line-broadening is dominated by increased solvent fluctuations induced by the rapid temperature change, and the deviation from the thermal profile can be used to separate this effect from conformational changes due to protein unfolding. In Fig. 6-9, time-dependent changes of the homogeneous width for three frequency components (the $\nu_\perp$ mode, the random coil region, and the $\nu_{||}$ mode) during a low-temperature T-jump (25 $\rightarrow$ 35C) are plotted. Changes of all three frequency components track the temperature relaxation profile. Because the magnitude of the change is very small (<2%), width changes can be assumed to be linear with temperature within the relatively small T-jump range of 10 ° C. Therefore, we can conclude that the increase in linewidth at low temperature is fast and solely a result of the increased thermal fluctuations (primarily solvent perturbations), without significant structural change in the protein.

6.3.7 Unfolding Mechanism of Ubiquitin

Ubiquitin has a five-stranded mixed parallel/antiparallel β-sheet and an α-helix (see Fig. 6-2a), and it is one of the most actively studied folding systems. Several experimental and simulation studies have reported sequential folding after the formation of a stable core that includes the strand I-II hairpin and the α-helix.[192–196] The stability of this N-terminal 1-37 fragment has been investigated by fragmentation studies, multidimensional NMR experiments, and MD simulation.[11,197–200] MD simulations

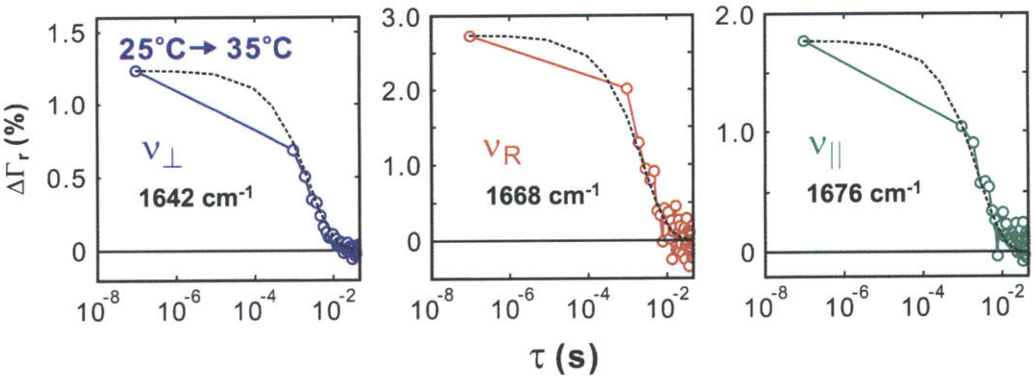


Figure 6-9: Relative changes in the antidiagonal width for a T-jump from $T_i = 25$ °C, plotted with a temperature relaxation profile normalized to the initial change (dashed line).

suggested a transition state with β-strands I, II, and V intact and an unfolding pathway involving strands III-V before unfolding of the hydrophobic core.[11,150]

Our previous T-jump experiments on this system[25] lent further support to the two-state model for ubiquitin unfolding. Those experiments used dispersed vibrational echo (DVE) spectroscopy, a nonlinear spectroscopy that can be related to a power spectrum of the 2D IR signal projected onto the ω_3 axis. We interpreted the DVE transients as sequential unfolding of the less stable strands III-V (3 μs) before unfolding of the protein's hydrophobic core (80 μs). A physical picture to describe the multiphase unfolding process was proposed in terms of a free-energy surface consisting of two coordinates to which the IR experiments are most sensitive: (i) the native contacts between strands I and II and (ii) the remaining native contacts between adjacent strands I, V, III, and IV. These contacts are depicted in Fig. 6-2b. Fluorescence-probed downhill folding induced by a barrier shift at 8° C has been reported for a cold denatured ubiquitin mutant, F45W,[22] with a time scale of 100 μs. The difference in time scale likely originates from the temperature-dependent difference between the folding and unfolding free-energy surfaces.

The results from transient 2D IR spectroscopy capture the previous DVE results and expand on them by disentangling spectral changes along the ω_1 frequency dimension. The separation of the diagonal and cross-peak region dynamics, and the ability to monitor the frequency-dependent homogeneous linewidth, have yielded a unique description of the solvent-exposed random coil regions. The observation that diagonal and off-diagonal $\nu_\perp$ features relax in a correlated fashion also shows that the original interpretation of the DVE experiments was valid. However, because of the delocalized nature of amide I vibrational eigenstates, a complete picture including the order of strands unfolding has relied on qualitative comparisons with other experiments and with MD simulation. By calculating 2D IR spectra from protein structures sampled along an unfolding pathway consistent with our hypothesis, we seek to make a comparison by identifying similarities in intensity changes and frequency shifts.

6.3.8 Comparison with Simulation

Amide I eigenstates are sensitive to subangstrom changes in the distance between β strands and in hydrogen bond lengths,[8, 125, 174] but spectral congestion masks these changes in FTIR spectra. The cross-peaks observed with 2D IR help to recover sensitivity and allow for rigorous comparisons of amplitude changes and peak shifts. To provide an atomistic test of the qualitative picture presented, we analyze the ubiquitin-unfolding MD simulation of Alonso and Daggett[150] by calculating 2D IR spectra, using structures sampled along an experimentally relevant set of reaction coordinates.

We choose structures by projecting the T-jump simulation along two coordinates that are closely related to the β-sheet contacts to which 2D IR is sensitive: summed $C\alpha$ distances between native contacts. These are defined by

$$\begin{aligned} R_1 &= R_{I-V} + R_{V-III} + R_{III-IV}, & (6.3) \\ R_2 &= R_{II-I}, & (6.4) \\ R_{N-M} &= \sum_{(i,j)} \sqrt{(C_{\alpha,i} - C_{\alpha,j})^2}, & (6.5) \end{aligned}$$

where R_{N-M} describes a sum of pairwise C_α distances (i, j) between hydrogen-bonded strands N and M. By using the strand diagram in Fig. 6-2b, R_1 and R_2 are the sum of contact distances illustrated in purple and red, respectively. The simulation trajectory is plotted along these coordinates in Fig. ?? to visualize the steps (A-E) in the progressive unfolding of the protein. Beginning from the crystal structure, the initial reaction to the T-jump is global expansion and increased disorder of the protein (A). The unfolding proceeds through transitions (A$\rightarrow$B$\rightarrow$C$\rightarrow$D), during which contacts between strands III and IV, V and III, and I and V are broken sequentially, yielding a persistent configuration in which the α-helix has rotated to align with strands II and I (D). The final transition breaks these contacts, and only the helix is structured by the end of the simulation (E). This pathway is consistent with the studies guiding our hypothesis, including our T-jump DVE results indicating that β-sheet unfolding occurs on two distinct time scales.

Although the large T-jump ($\Delta T = +200°$ C) and the lack of extensive sampling in the simulation prohibit a thermodynamic and kinetic comparison, we can still search for spectral features that correlate with structural changes by calculating 2D IR spectra from solvated snapshots in each persistent structural region. Sample spectra are plotted in Fig. ??. The equilibrium spectrum, reproduced from our previous work,[8] shows the Z-shape typical of β-sheets, with an intense $\nu_\perp$ transition at 1640 cm^{-1} melding into the random coil region at 1660 cm^{-1}, a high-frequency $\nu_\parallel$ peak at 1677 cm^{-1}, and cross-peak ridges extending along ω_1 between $\nu_\parallel$ and $\nu_\perp$. As the protein unfolds, the spectral features gradually transition to produce an unstructured, blue-shifted, diagonally elongated spectrum. These changes are reflected in the doorway modes,[12] which are shown in Figure 6-12. As strands unfold from ubiquitin's β sheet, the bright state corresponding to the $\nu_\perp$ mode becomes more localized.

For better comparison with the experiment, Fig. ?? also shows the simulated transient difference spectra and quantifies changes in peak intensity and frequency from configurations A to E. The difference spectra show derivative features indicating a correlated drop in $\nu_\perp$ at 1640 cm^{-1} with an increase in the random coil region at 1670 cm^{-1}. This is accompanied by loss of negative $\nu_\perp$, $\nu_\parallel$ cross-peak intensity in the lower-right region of the spectrum and appears as a positive change. The cross-peak ridge loses >50% of its intensity and blue-shifts by 2 cm^{-1} in transition from A to

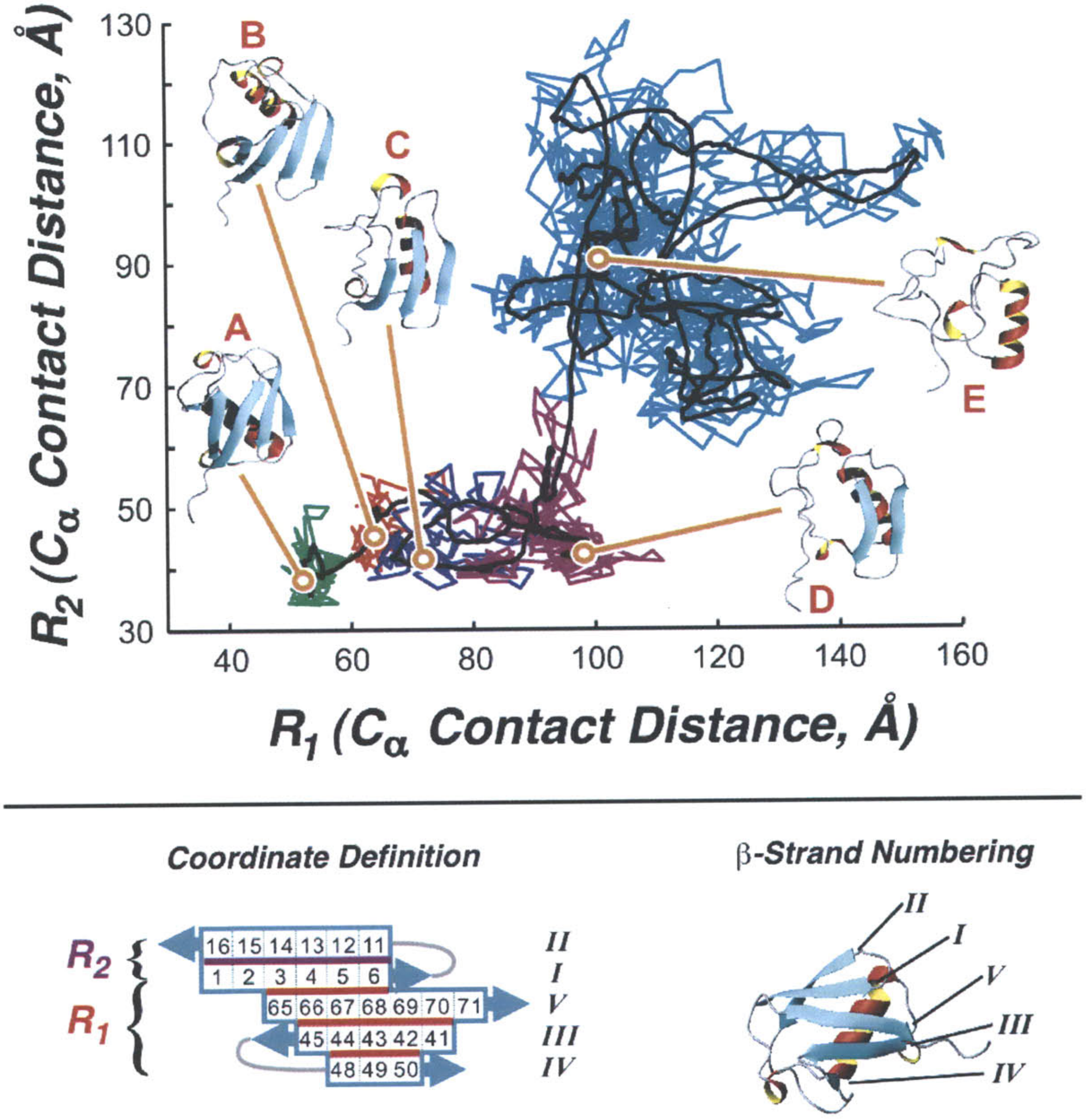


Figure 6-10: Thermal unfolding trajectory of ubiquitin. The unfolding simulation[11] is plotted along the coordinates R_1 and R_1 defined in the text. Protein snapshots correspond to five persistent structural regions (shown in green, red, blue, magenta, and cyan, respectively). The trajectory after a 17-ps window average is plotted in black.

C, where III-V strand contacts are broken. The blue shift is also apparent in the $\nu_\perp$ diagonal region but is less clear because of interference with the broad random coil peak. Blue-shifting of the $\nu_\perp$ peak has been predicted to appear with localization of a vibrational exciton upon loss of vibrational coupling.[70] The dynamics are complete

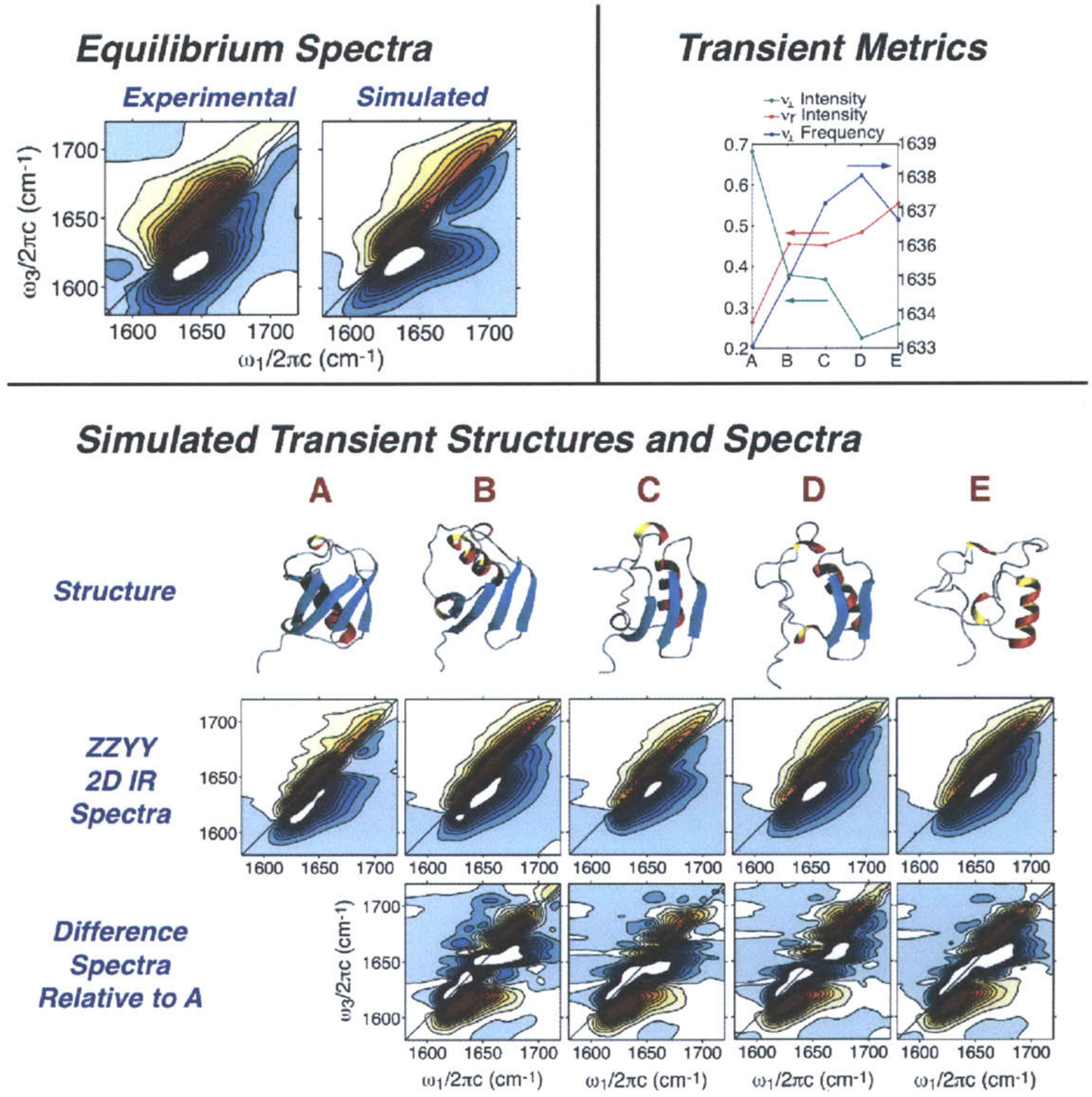


Figure 6-11: Calculated transient 2D IR spectra of ubiquitin unfolding. Ubiquitin 2D IR spectra calculated at equilibrium (25 °C) (adapted from Ganim et al.[8]), for the transient states A, B, C, D, and E, and the difference spectra relative to A (B-A, C-A, D-A, and E-A). The trends in $\nu_\perp$ frequency and intensity and in ν_R intensity were obtained from $\omega_1 = 1640$ cm^{-1} and $\omega_1 = 1662$ cm^{-1} slices, respectively.

by the final spectrum which, like the 7-ms experimental transient, can be compared with the equilibrium difference.

There is a large disparity of time scales between the ≈1-ns simulation and the 100-ns to ≈1-ms experimental responses that cannot be accounted for with the ther-

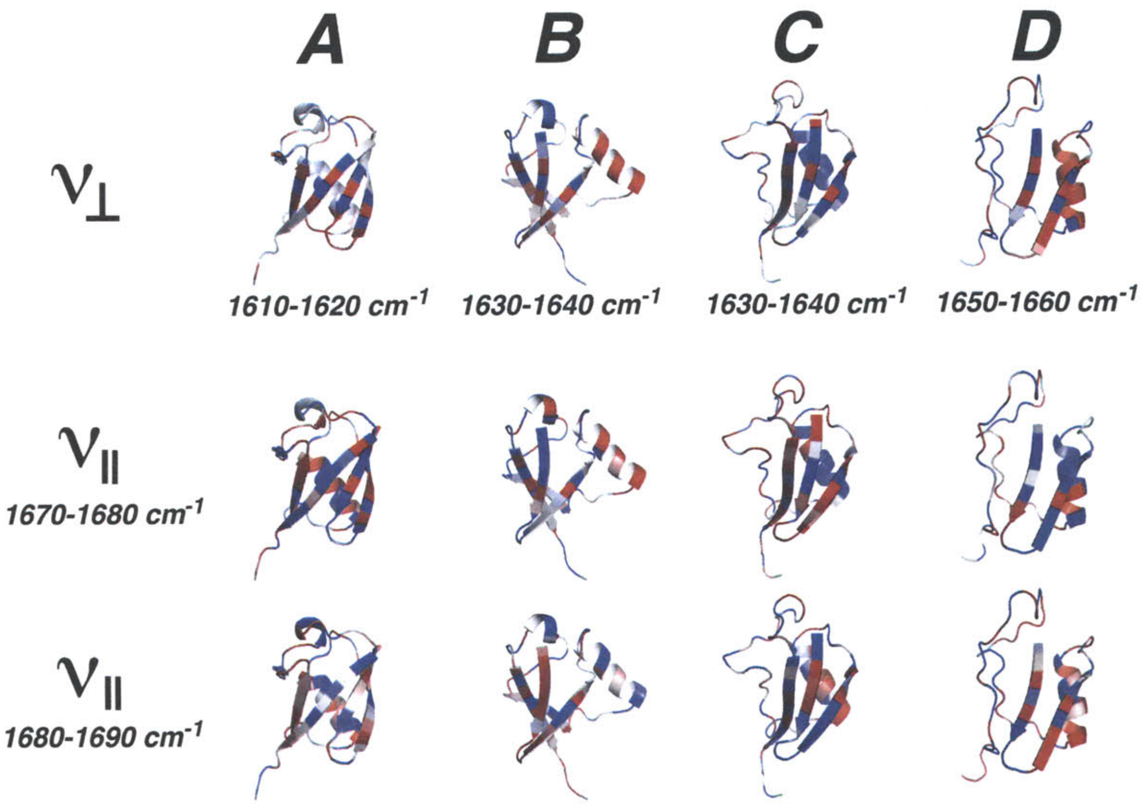

A
B
C
D
$\nu_\perp$
1610-1620 cm⁻¹
1630-1640 cm⁻¹
1630-1640 cm⁻¹
1650-1660 cm⁻¹
$\nu_\parallel$
1670-1680 cm⁻¹
$\nu_\parallel$
1680-1690 cm⁻¹

Figure 6-12: Doorway modes of the transient ubiquitin unfolding states. Doorway modes (calculated as described previously[12]) for selected regions of the $\nu_\perp$ and $\nu_\parallel$ bands.

mal change in diffusion. The difference is due to the different nature of the free-energy surface at 227° C (simulation) vs. 72° C (experiment) and to differences in comparing single-molecule and ensemble representations. The downhill unfolding subensemble prepared in the experiment moves along a corrugated free-energy surface while surmounting and accumulating behind barriers of $\approx$kT. This friction and exploration along orthogonal coordinates delays and disperses the ensemble from a high-temperature limit that would behave similarly to the single protein in this single unfolding trajectory.

6.4 Conclusions

We have presented an approach for characterizing protein folding dynamics that uses 2D IR spectroscopy of burst-phase unfolding interpreted through comparison with MD simulation. The interpretation of transient 2D IR spectra is consistent with calculated spectra from a simulation that shows a progressive unfolding of the β-sheet. The ability to quantitatively model amide I IR spectroscopy on the basis of proposed structures allows for a detailed connection between unfolding dynamics in experiments and simulations. In the future, one can imagine using 2D IR to discriminate between different simulated unfolding pathways, which will enable experimentally verified simulations to report on variables that cannot be probed by experiment.

6.5 Acknowledgements

We thank Valerie Daggett (Department of Medicinal Chemistry, University of Washington, Seattle, WA) for providing the trajectories of ubiquitin unfolding described by Alonso et al.[150] This work was supported by National Science Foundation Grants CHE-0316736 and CHE-0616575 and by The David and Lucile Packard Foundation. Additional support came from Department of Energy Grant DE-FG02-9ER14988.

6.6 Afterword

This chapter was excerpted from "Transient 2D IR Spectroscopy of Ubiquitin Unfolding Dynamics," by Chung and Ganim et al.[63] I would like to thank Hoi Sung Chung and Kevin Jones for collaboratively acquiring and analyzing the data with me.

Chapter 7

The Biophysics of Coupled Protein Folding and Binding

Without exception, proteins perform their function by interacting with one or more other molecules, the ligand(s). In the case that the ligand is a small molecule, approximations resulting from its conformational rigidity will shape a description of the dynamics. The remaining chapters in this thesis are concerned with the biophysics of protein-protein interactions, in which both reactants are flexible and nominally exist in a distribution of states. Predicting the *dissociation* rate, and therefore the equilibrium constant, requires a model for the stability of the bound protein complex, which generally evades a simplistic treatment. However, regardless of the molecular picture of binding, at long distances diffusion is required to bring the reactants in close enough proximity to undergo recognition and create a bound complex, which can be understood without alluding to proteins' internal structure.

7.1 Diffusive Encounter

7.1.1 Single Particle Diffusion

Let us consider typical time and length scales for one diffusing protein in the infinite dilution limit. The protein is treated as a free, Brownian particle with mass m, its position, $x(t)$, obeys the Langevin equation[201]

$$\frac{d^2\vec{x}(t)}{dt^2} = -\frac{\gamma}{m}\frac{d\vec{x}(t)}{dt} + \frac{1}{m}\vec{\xi}(t) \tag{7.1}$$

where γ is the friction coefficient and $\vec{\xi}(t)$ is the random force of the solvent. To avoid being concerned with realizations of $\vec{\xi}(t)$, it is useful to consider the corresponding Fokker-Planck equation for the probability density of the particle, $P(\vec{x}, \dot{\vec{x}}, t)$. Given the diffusion of the center of mass of proteins, it is appropriate to consider the high friction case, $(\frac{d^2x(t)}{dt^2} \approx 0)$, where the velocity distribution is relaxed. In this limit, the Fokker-Planck equation takes a simple form,

$$\frac{\partial P(x,t)}{\partial t} = D\frac{\partial^2 P(x,t)}{\partial x^2}, \tag{7.2}$$

where the translational diffusion constant, $D = \frac{\gamma}{k_B T}$, fully characterizes the system.[201] If the particle position is known to be $\vec{x}^0$ at t=0, its probability distribution at later times will be a spherically expanding Gaussian,

$$P(\vec{x},t) = \sqrt{\frac{1}{4\pi Dt}} exp\left[-\frac{(\vec{x}-\vec{x}^0)^2}{4Dt}\right]. \tag{7.3}$$

This can be combined with the Stokes-Einstein-Debye relation for the translational diffusion constant of a sphere of radius r, which assumes that the solvent is a continuum in thermal equilibrium (at temperature T) that obeys no-slip boundary conditions,[202–204]

$$D = \frac{k_B T}{6\pi\eta r}. \tag{7.4}$$

In this description, the diffusion constant for a relatively small protein (r=1.4 nm, the radius of dimeric insulin) is 17 $\mathring{A}^2/ns$ (or $1.7x10^{-6}$ cm^2/s) at room temperature in water. According to Eq. 7.3, in 30 μs the standard deviation for the protein's probability density function will be 100 nm.

7.1.2 Association Kinetics from Two Particle Diffusion

If free diffusion is the only mechanism driving molecular encounters, the association rate constant can be described by Smoluchowski's expression for two uniformly reactive spheres of radii r_A and r_B with diffusion constants D_A and D_B,[205]

$$k_a = 4\pi \left(D_A + D_B\right)\left(r_A + r_B\right) = 4\pi DR, \tag{7.5}$$

in which $D = D_A + D_B$ is the relative diffusion constant and $R = r_A + r_B$ is the interaction radius. The Stokes-Einstein form of the diffusion constant predicts that $k_a = 7.4x10^9 M^{-1}s^{-1}$ in water at room temperature, regardless of radius. (The radius drops out of this expression because $D \propto \frac{1}{r}$, which indicates that increasing the radius of the particle makes it an easier target, but also slows diffusion and the association rate is unchanged.) Equivalently, $\frac{1}{k_a} = 130\ psM$, and the reaction half-life, $t_{1/2}$, is 130 ns in a 1 mM solution.

Another key timescale describes how long it takes proteins to diffuse a sub-nm distance. Within the approximations of this rigid-body theory, this is the timescale for encounter complex dissociation, or the timescale for forming an encounter complex from two proteins at the edge of feeling mutual molecular forces. For the diffusion constant listed above, it takes 10-250 ps for proteins to diffuse $1 - 5\ \mathring{A}$.

When considering proteins in close proximity, the preferred orientation for reactivity may contribute to the rate. However, it is often the case that rotational diffusion takes place on a much faster timescale than translational diffusion. The relevant Stokes-Einstein-Debye relation for orientation diffusion is,

$$D_r = \frac{k_B T}{8\pi\eta r^3} = \frac{3D}{4r^2}, \tag{7.6}$$

which gives $D_r = 6.3x10^7\ s^{-1}$ for the r=1.4 nm protein under consideration. The according rotational correlation time

$$\tau_r = \frac{1}{6D} \tag{7.7}$$

is 2.6 ns, which is appreciably faster than the timescales described for translational diffusion.

Together, these timescales describe a kinetic theory for rigid, uncharged protein association. The slowest timescale (concentration-dependent, but radius-independent) is for the proteins to diffuse to a separation of a few $\mathring{A}$ (1-2 solvation shells in the atomistic description) at which point intermolecular forces between the proteins may become non-negligible (~130 ns in 1 mM solution). This is referred to as a "macrocollision." During a macrocollision, the proteins may undergo several "microcollisions" (~10-250 ps), defined as making van der Waals contact, before they are likely to drift apart. The timescale for orientational diffusion (~3 ns for small proteins) indicates that proteins are likely to reorient several times during a macrocollision, but that the orientation will be mostly fixed during the many microcollisions.

7.2 Recognition and Binding

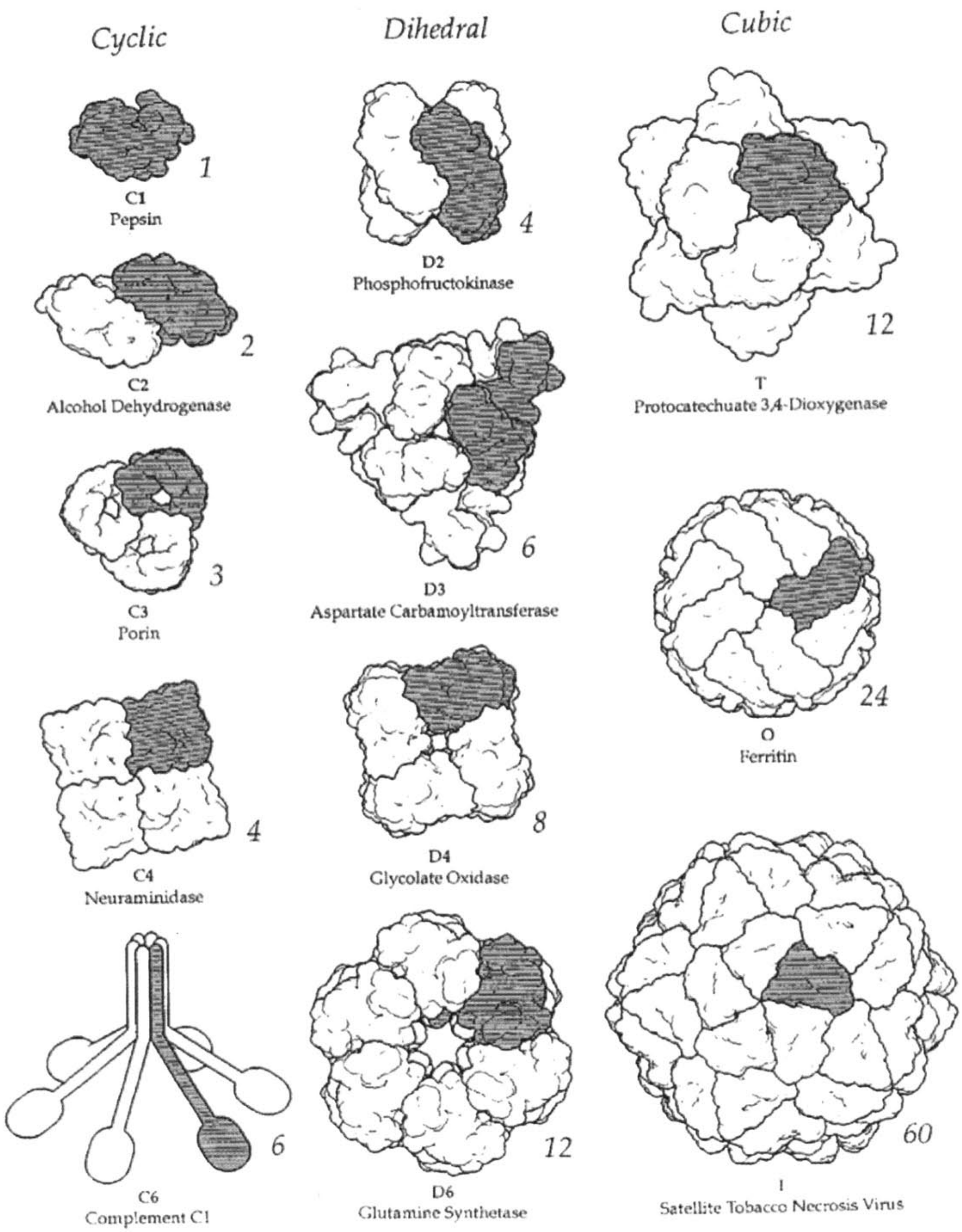
To the extent that proteins are treated as uniformly reactive, uncharged, rigid spheres, the above treatment is the complete story for the kinetics of protein-protein association. One major exception is when strongly charged proteins associate quickly ($k_{on} > 10^8 M^{-1} s^{-1}$),[206] which requires consideration of structure-specific local fields [207] that are affected by pH and ionic strength. The rate constant can also be modified to include orientational constraints on reactivity.

However, the fundamental challenge facing models for the biophysics of protein-protein interactions is that proteins are not infinitely rigid. The accessible states of proteins are determined by a competition between non-covalent interactions such as the formation of hydrogen bonds, hydrophobic packing, solvation, and salt bridges. Individually, these interactions are nonspecific and weak (each may contribute a fraction of k_BT to stability), but may cooperate to stabilize a subensemble of configurations referred to as the native state. It has been reasoned that the interactions stabilizing proteins must be weak to avoid the formation of kinetic traps in non-native states, which would prevent folding altogether.[208] One consequence is that, in ambient conditions, there will be sufficient energy to break a few of these contacts, and some flexibility must be included in the definition of a native state. A related fact that the sum of all these interactions, the protein folding energy, is on the order of $\sim k_BT$ per residue.[209] This is additionally a matter of practical utility, as proteins have a finite lifetime, after which they must be unfolded and recycled. These considerations build a picture where proteins are flexible, and continually fold and unfold; the extent to which this generally constitutes an integral part of the association mechanism or a mere nuisance remains a matter of debate.

7.2.1 Methods for Studying Protein Association Biophysics

Historically, there has been an emphasis on rigid-body methods for studying and modeling protein-protein interactions. This has been driven by the fact that atomistic structural elucidation methods, namely x-ray crystallography and nuclear magnetic resonance, have largely shaped the state of modern molecular biophysics. These techniques have produced a database of protein complexes, which allowed for the first conclusions about the structure of protein interaction sites.[210–212] It has been claimed that recognition is facilitated by "surface patches,"[213] which are responsible for providing both a driving force for binding and specificity. It was observed that protein interfaces are close packed, and the shape of the residues at the interface display steric complementarity (see Figure 7-1). The forces stabilizing protein interactions were found to be (unsurprisingly) the same forces that drive protein folding-hydrophobicity, hydrogen bonding, electrostatics, and van der Waals. To be sufficiently stabilizing, protein binding typically buries at least 800 $\mathring{A}^2$ of the surface[214] (with $\Delta G \approx 25cal/\mathring{A}^2$)[215], which roughly corresponds to the $\sim$ 15 kcal/mol of translational and rotational entropy loss.[210] There is a slightly increased propensity for hydrophobic residues in recognition sites.[216] It has also been repeatedly claimed that only low energy conformational changes take place upon dimerization, such as the rotation of side chains or the ordering of unstructured loop regions.[210,212,217]

While structural tools have informed a majority of the current thinking, protein-protein interactions are studied using a variety of other experimental techniques.



Cyclic
Dihedral
Cubic
1
C1
Pepsin
4
D2
Phosphofructokinase
12
T
Protocatechuate 3,4-Dioxygenase
2
C2
Alcohol Dehydrogenase
6
D3
Aspartate Carbamoyltransferase
3
C3
Porin
24
O
Ferritin
4
C4
Neuraminidase
8
D4
Glycolate Oxidase
6
C6
Complement C1
12
D6
Glutamine Synthetase
60
I
Satellite Tobacco Necrosis Virus

Figure 7-1: Examples of protein oligomer complexes displaying shape complementary and symmetry. Reproduced from Goodsell and Olson.[13]

Separation and thermodynamic methods such as chromatography, centrifugation, and calorimetry have been widely used to discover binding partners, quantify dissociation constants, and study temperature dependence. Researchers have combined these tools with site-directed mutagenesis to probe single-residue contributions to the stability of complexes. Alanine scanning mutagenesis has revealed that often only a few residues are responsible for stabilizing protein-protein complexes. Hot spots are such residues that contribute at least 2.0 kcal/mol of stabilization, which often show a preference for tryptophan, arginine, and tyrosine.[218,219]

Specific protein-protein interactions can be tagged and studied with surface plasmon resonance biosensors[220,221] or FRET (Förster resonance energy transfer),[222] which may also be used *in vivo*. System-specific spectral signatures have been isolated that allow for kinetic measurements of protein-protein interactions, such as UV-vis absorption, fluorescence, circular dicroism, NMR measurements sensitive to conformational exchange such as paramagnetic relaxation enhancement[223], line shape modeling,[224], amide protein exchange rates, and backbone amide S^2 order parameters.[225]

7.2.2 Computational Methods for Predicting Protein Association

There have been significant efforts to codify these observations into models that compute the location of recognition sites, the structure of bound complexes, and the free energy of binding.[14] Typically, these models are *physics-based* (solving equations such as the Boltzmann equation and minimizing an energy functional), but incorporate some *knowledge-based* elements (empirical predictions based on correlations from databases).

The most popular class of physics-based algorithms are the "docking" class of procedures that search for surface complementarity by fitting together two or more monomer structures to predict the structure of complexes - an idea that was first suggested by Francis Crick.[226] The initial structures may result from experimental models (for the unbound protein or, preferentially, the protein bound to a different partner[227]) or by homology modeling. These computational methods bear some similarity to those used to predict folded states of proteins, but have received more attention from developers due to their industrial usage for rational drug design.[227] Incomplete consideration of flexibility is a computational requirement for these models; permissive, flexible docking models allow for side chain flexibility and motion of domains through "shear" and "hinge-bending" transformations.[228] Numerous implementations of these algorithms are available on the internet, which vary in protein representation, conformational space search method, and scoring function, such as RosettaDock,[229] ClusPro,[230] GRAMM-X,[231] ZDOCK,[231] PatchDock, and SymmDock.[232]

The scoring functions for these software packages may include both physics-based, energy minimization terms and knowledge-based terms, including residue-residue pair propensities, van der Waals energies, hydrogen bond geometry, implicit Gaussian desolvation/solvation, side-chain rotamer probabilities, and electrostatic energies. Improvement add-ons to conventional docking algorithms try to reconstruct association pathways,[233] incorporate molecular dynamics and normal mode analysis.[228]

Such calculations typically go through thermodynamic cycles (such as depicted in Fig. 7-2), because it is easier to calculate protein binding entropy change in the gas phase using the ideal gas approximation, and it is typically assumed that the protein-protein terms do not change. By computing the relative solvation energies of the bound and unbound proteins, the cycle is complete. However, the summation of this cycle includes oppositely signed, large energies ($\sim$100 kcal/mol), which must cancel to the 1% or better level for predictive accuracy. These calculations expose errors in the simulation force fields as well as deficiencies due to inadequate long-time sampling.

Knowledge-based algorithms incorporate heuristic conclusions from database analysis of previously obtained experimental data. For both folding and binding problems, there is a consensus that evolutionarily conserved residues are likely to play an important function.[234,235] In fact, it has been recognized that evolution occurs at a slower rate at binding interfaces because substitutions in one monomer are not likely to be passed on without complementary changes in the binding partner; this phenomenon is called co-evolution.[236] Several cases have been observed where a small number of polar residues interspersed into a largely hydrophobic interfaces are key to binding.[235]

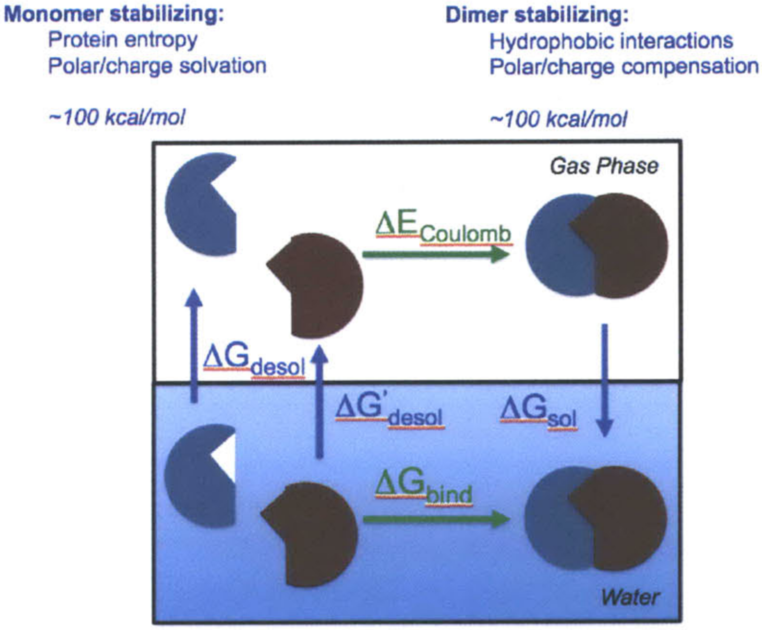

Monomer stabilizing:
Protein entropy
Polar/charge solvation
~100 kcal/mol
Dimer stabilizing:
Hydrophobic interactions
Polar/charge compensation
~100 kcal/mol
Gas Phase
$\Delta E_{Coulomb}$
ΔG_{desol}
$\Delta G'_{desol}$
ΔG_{sol}
ΔG_{bind}
Water

Figure 7-2: Schematic of a thermodynamic cycle used for calculating free energy of association.[14]

This type of information can be used to predict protein-protein interactions based on similarities in atomistic structure or homology to known binding partners. Correlated changes in phylogenetic trees, proximity in the genome, and network analysis of interacting partners may also be used to predict binding.[237]

7.2.3 Conclusions about the Current State of Protein Association Models

CAPRI (Critical Assessment of PRediction of Interactions) is a community wide experiment designed to assess models for predicting structures of bound protein complexes. It asks- given the atomistic structure of two protein monomers, how reliable and accurate will the model for the dimer be? CAPRI is a blind prediction challenge, that relies on unpublished structures obtained using NMR or x-ray crystallography, i.e., the published structure is only revealed after the competition is over. Figure 7-3 shows aggregated results from the CAPRI challenge for the past five years; typically 300 - 400 structures are submitted, 95-100% of which are deemed incorrect based on insufficient accuracy (ratio of true positive and true negative to total predictions), specificity (ratio of true positive to false positive predictions), and sensitivity (ratio of true positive to false negative) using metrics quantifying the number of correct residue contacts at the interface.[238] Several evaluations have acknowledged that conformational changes that occur during binding are a major limitation, and that current attempts to improve the accuracy of energy functionals may not improve the accuracy of models:

- "'Unbound' docking starts from the two independently determined protein structures. Therefore, it must handle the conformational changes that are inevitable on association and make docking a far more difficult task than just searching for an exact fit."[239]
- "Conformational changes, both local and large-scale present enormous complications."[240]
- "the methods are hampered by a lack of a complete understanding of the forces involved and by the conformational changes that often take place upon protein-protein binding"[241]
- "One of the most important difficulties of protein docking is that the interface residues of both the receptor and the ligand may undergo a conformational change on complex formation."[242]
- "Conformational flexibility is one of the most important determinants for docking success, and more effort should be put not only on the development of new algorithms to overcome the rigid-body docking and scoring approach but also on estimating the conformational flexibility upon binding for a given case."[243]
- "modeling induced fit by flexible docking remains a central challenge, and a large portion of current docking research is focused in this area."[244]
- "it appears that extensive interactions between two proteins always induce a major conformational change... Improved scoring and discrimination will not

affect the prediction of such complexes because rigid-body searches of current docking methods will not produce hits. Therefore, despite major progress in the protein-protein docking in the past few years, it will be necessary to develop approaches that do not start with a rigid-body search."[245]

- "In fact, improving the specificity of a scoring function also increases its sensitivity to the differences between bound and unbound protein conformations, and the quality of discrimination does not necessarily improve."[246]

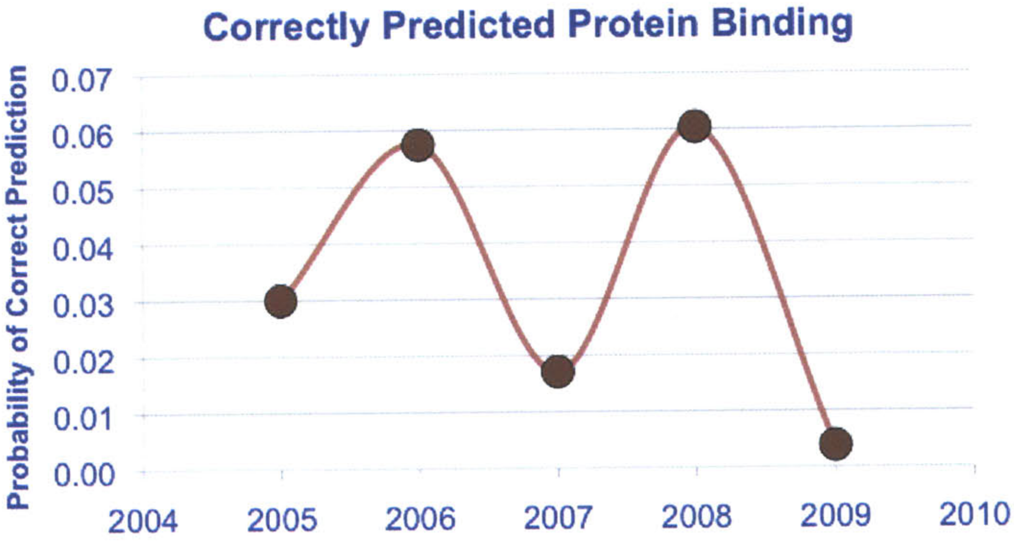


Figure 7-3: Prediction accuracy results of the CAPRI challenge.

False positives plague predictive models for protein-protein interactions because basic chemistry dictates that there will generally be a non-zero attraction between *any* two proteins. Ultimately, it is the inability of the currently available techniques to reliably discriminate against false positives that prevents practical applications of predictive protein docking models.[246] In the end, one finds that the best computational approaches are only potentially useful when there is shape and chemical complementarity with little structural change; this is otherwise known as the "lock and key" limit.

7.3 Free Energy Surfaces for Coupled Protein Folding/Binding Processes

There seems to be no reason to conceptualize binding in separate physical terms than protein folding, and the evidence suggests that models fail to be generally predictive when this simplification is made. Over the past two decades, there has been growing acceptance of the energy landscape view of protein folding.[208,247] However, its implications with regards to binding and other protein functions have not been as extensively explored. An energy landscape for protein binding would synthesize all aspects of binding equilibria and dynamics; it would describe the ensembles of bound and unbound states, the heights of barriers separating them, the relevant reaction coordinates, and an ensemble of reaction pathways for both binding and dissociation.

Models for the biophysics of protein binding have been hampered by the rigid-body lock and key concept that proteins exist in one conformational state. Even a dressed-up lock and key concept, such as the assumption that proteins can exist in *two* conformations- bound and unbound, allows for a very narrow interpretation of possible protein binding mechanisms. A free energy surface is the type of model that synthesizes all the information we have about a particular binding problem. It will show the energetic wells as well as the barriers, summarized along the most relevant coordinates, and it can be used to predict different pathways. Putting together a free energy surface from a dynamics experiment relies on the energetic and structural characterizations, but it uniquely provides information about barriers in contrast to equilibrium measures that mostly sense structures at free energy minima. Moreover, construction of a free energy surface based on dissociation experiments allows for predictions to be made about the association process. Any binding mechanism can be described by a free energy surface with the appropriate coordinates and topology, and the conventional lock and key, conformational selection, and induced fit mechanisms arise as limiting behaviors. Such models may be the only ways of understanding the binding of intrinsically disordered proteins that remain predominantly disordered after binding.[248]

7.3.1 Binding Mechanisms Characterized by Free Energy Surfaces

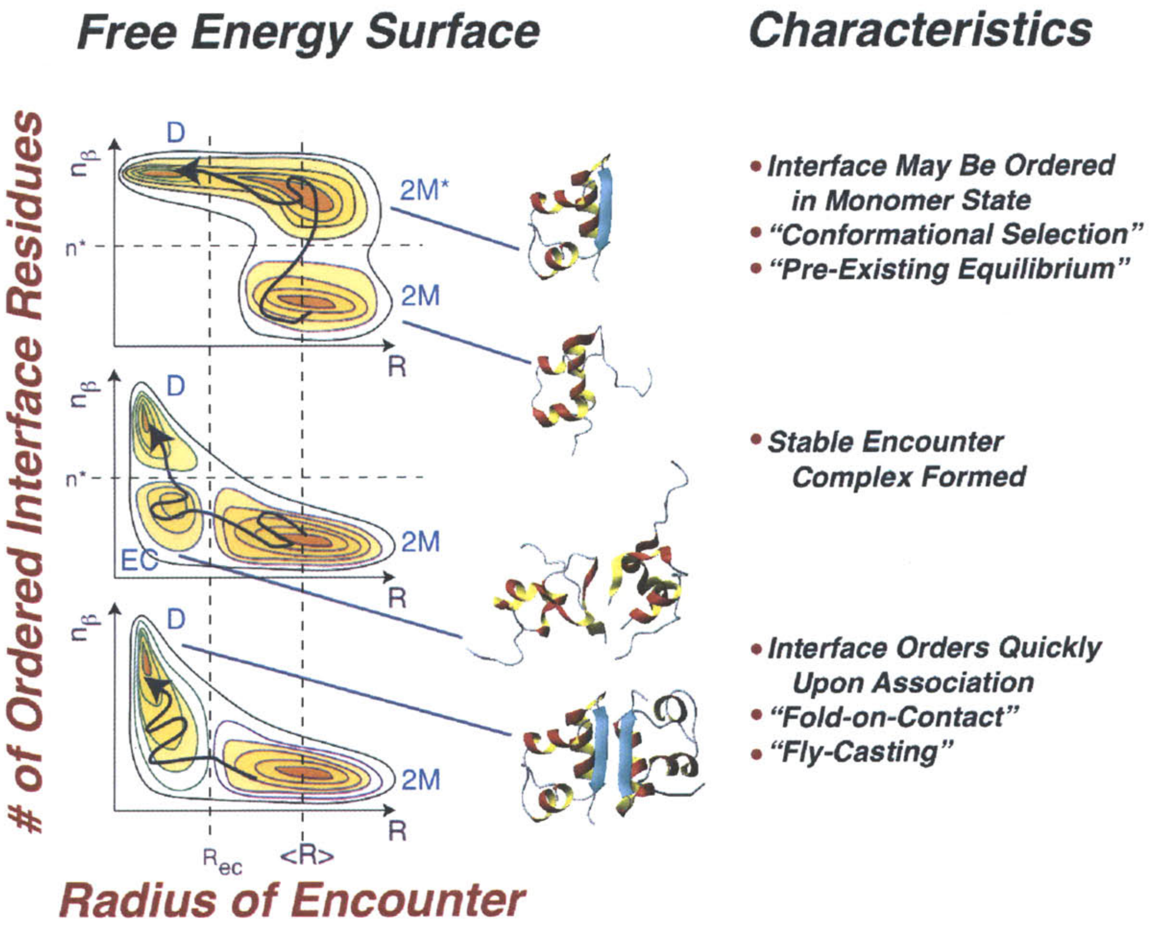

Free Energy Surface
Characteristics
# of Ordered Interface Residues
nβ
D
n*
2M*
2M
R
EC
Rec
<R>
Radius of Encounter
• Interface May Be Ordered in Monomer State
• "Conformational Selection"
• "Pre-Existing Equilibrium"
• Stable Encounter Complex Formed
• Interface Orders Quickly Upon Association
• "Fold-on-Contact"
• "Fly-Casting"

Figure 7-4: Example free energy surfaces for protein dimerization. Insulin monomer and dimer structures are shown as illustrative examples rather than constraints or specific predictions.

To explore the type of information that can be synthesized with a free energy surface, Figure 7-4 shows three examples. Along the independent axis is the radius of encounter, which discriminates associated from dissociated monomers. This coordinate is a key order parameter for describing an association reaction. However, a variety of coordinates may be chosen for the dependent axis, which should describe the folding or conformational change that occurs upon dimerization. Examples include RMSD deviation from the dimer crystal structure, the number of intramolecular contacts, the number of intermolecular contacts, the radius of gyration, or the number of hydrogen bonds to water. Choice of this coordinate reflects preexisting information about the structural changes in the system and the experimental observable being probed. Thus, a decision about the best coordinate will be driven by conclusions from the data, and a thorough investigation would consider several of the aforementioned options. The number of ordered interface residues was chosen as one example that anticipates experiments in Chapter 8.

The three free energy surfaces shown in Figure 7-4 differ by where barriers occur. Is the rate for restructuring the binding interface slower than the rate of protein encounters? If restructuring the interface is the only rate-limiting step, then the first surface may be an appropriate description. Here the only barrier is for folding of the interface, which may occur independently of association. The conformational selection model for association is obtained if only a subensemble of monomers can form ordered dimers. (Since basic chemistry dictates that there will be a non-zero attraction between *any* two proteins, formation of non-native or disordered dimers is always possible, if only transiently.) As a special case of conformational selection, if monomers exist in a discrete number of well-defined states, and only a subset of these can form dimers, then the pre-existing equilibrium model for association is recovered. Furtheremore, the rigid, lock and key model for association can be viewed as the limit of the pre-existing equilibrium model as disordered monomers are destabilized relative to the ordered monomers.

If an ordered dimerization interface is only stable when the monomers are associated, one of the bottom two free energy surfaces in Figure 7-4 may describe binding. The second surface has barriers in both coordinates, which leads to the formation of an encounter complex intermediate state. In this free energy surface, the encounter complex is formed from disordered monomers, and so these binding mechanisms mirror protein folding by hydrophobic collapse. Evidence for this free energy surface would result from characterizing the structure of the encounter complex. Are any specific contacts required to form the encounter complex or is it best characterized as a molten globule-like state? What are the relative heights of the barriers separating the encounter complex from ordered dimers and dissociated monomers?

In the third free energy surface, the only barrier to dimerization occurs along the radius of encounter. These mechanisms can be generically classified as fold on contact scenarios. This limit is appropriate if ordering at the interface occurs much faster than the rate of encounters. Considering molecular trajectories that give rise to this free energy surface results in an intriguing paradox- how do disordered monomers recognize one another so quickly? An ensemble of loose, non-specifically associated monomers is possible, and it may be difficult to rationalize how these may all find the native dimer state much faster than dissociating. (This can be contrasted with the second free energy surface, in which the conformational restrictions on ordering the dimer interface due to association result in an activation barrier.) Shoemaker et al. proposed a solution to this paradox with a "fly-casting" mechanism, originally used to describe DNA-protein binding.[249] In this scenario, the enlarged radius of gyration for disordered proteins enhances the rate of forming long-range, non-specific contacts to DNA. After association, the dimensionality of the problem is reduced, which enhances the rate of folding. This free energy surface can describe other fold on contact mechanisms, such as if the formation of specific intermolecular contacts is required to nucleate ordering of the dimer interface.

The dimerization of insulin was chosen as a model system to develop dynamics experiments for folding and binding. There has been only one time-resolved dimer dissociation experiment on insulin[250], and none of the mechanisms discussed above can be excluded *a priori*. In Chapter 8, the structural changes concomitant with the monomer-dimer transition of insulin are characterized using 2D IR spectroscopy. As a complement to these coordinates, Chapter 9 presents fluorescence experiments aimed at providing long-range sensitivity to the intermonomer separation.

7.4 Insulin Background

Insulin homodimerization is one of the simplest protein-protein binding reactions involving a secondary structural change. Insulin is a 51 amino acid, mostly α-helical protein that dimerizes through the formation of an intermolecular β sheet[16]. It is known that the binding region becomes disordered or partially unfolded when dissociated into its monomeric state.[251–258] Although questions concerning the monomer structure and degree of disorder remain, the fact that dimer is more ordered than monomer raises fundamental questions about coupled folding and binding processes, and stands in contrast to classical rigid-body association models.[259,260] How do two partially disordered proteins encounter one another, locate specific intermolecule contacts, and fold? Despite the volume of research on insulin, there have been no postulated mechanisms for how folding and association couple in the dimerization of insulin.

Inspection of the dimer crystal structure[16] (Fig. 7-5a) shows that insulin has a network of hydrophobic contacts that spans across both monomer units and stabilizes the dimer (Fig. 7-5b). When ordered, the dimer interface is largely flat, and is comprised of aromatic and aliphatic residues. The packing of these nonpolar sidechains is reinforced and given specificity by backbone hydrogen bonds between C-terminal residues on the B chains. The B chain holds a critical set of contacts for dimer formation and the degree to which it is folded largely determines how much ordering is required to form the dimer. Some monomer models also show disordered α helices upon dissociation. Models for the monomer (Fig. 7-5c and 7-5d)[251–258] have provided conflicting evidence on whether the B chain is extended into the solvent or if it resembles its conformation in the dimer. An extended conformation of the B chain would expand the radius of gyration for the protein and allow it to make weak, non-specific contacts. This scenario is consistent with the fly-casting mechanism for reducing the entropic cost for a binding transition state.[249] Alternatively, if the B chain largely resembles its conformation in the dimer, insulin monomers may dimerize by making native contacts with largely ordered interfaces. Thus, placing a collection of experimental constraints on the monomer structure may help suggest important reaction coordinates for describing protein association.

The methods used to study the monomer-dimer equilibrium of insulin have typically focused either on structure or association state, however sensitivity to

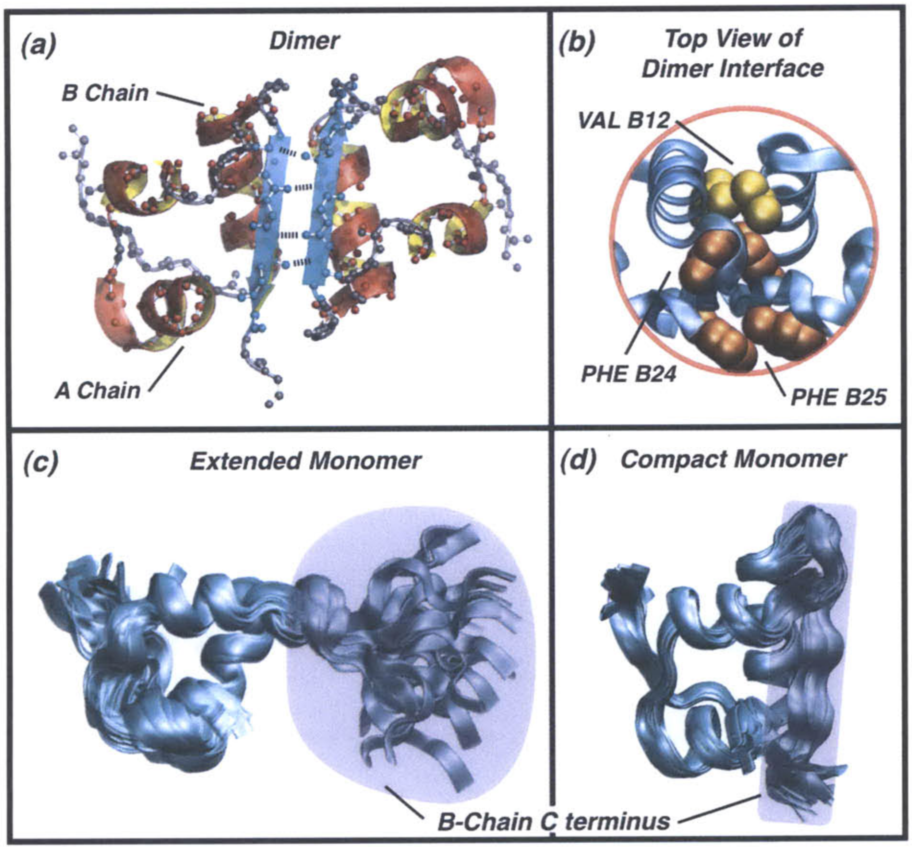

(a)
Dimer
B Chain
A Chain
(b)
Top View of Dimer Interface
VAL B12
PHE B24
PHE B25
(c)
Extended Monomer
(d)
Compact Monomer
B-Chain C terminus

Figure 7-5: Structural visualizations of insulin dimer and monomer models. Strucutral models from PDB IDs 4INS (a,b), 2JV1(c), and 1JCO (d). Structures visualized using VMD[15] and POV-Ray.

both is required for an ideal probe of binding and folding. Assays that have used sedimentation,[261–263] scattering,[264] kinetics,[250] mass spectrometry,[265] ultraviolet absorption,[266] fluorescence depolarization,[267] pulsed field-gradient spin-echo NMR[268] and FRET[269] provide sensitivity to the monomer/dimer ratio and may be used to quantify the dissociation constant, K_D, but they lack secondary structural sensitivity. Circular dichroism[270,271] has been deconvoluted to yield a mix of qualitative structural resolution and sensitivity to the association state. Infrared vibrational spectroscopy has been used to study insulin fibrilization,[272–274] but it has not revealed any changes about the monomer/dimer transition. Since the insulin dimer forms the asymmetric unit, X-ray crystallography has provided atomistic structures for various polymorphs of the insulin dimer and hexamer,[16,275] but has not yielded a native monomer structure. Nuclear magnetic resonance (NMR) spectroscopy has yielded the most information about insulin in solution via a series of hexamer,[276,277] and dimer[278,279] structures. Due to the poor sensitivity of NMR at micromolar concentrations, many investigations of the insulin monomer have relied on mutants or co-solvent addition;[251–253,256,257,278,280,281] while the resulting set of structures have generated growing agreement that insulin is more disordered in the monomer state than in the dimer, there has been no quantitative-level agreement. Moreover, neither crystallography nor 2D NMR have yielded a measurement of K_D.

In Chapters 8 and 9, 2D IR spectroscopy, molecular dynamics simulations, and fluorescence experiments are used to develop insulin dimer dissociation as a model system for investigations of the biophysics of coupled protein folding/binding. Chapter 8 presents the signatures of monomers and dimers using various representations of amide I 2D IR spectroscopy. These signatures were used to quantify the dissociation constant and to parameterize a two-state model for the thermodynamics. These 2D IR spectra were interpreted in atomistic detail using molecular dynamics simulations of insulin drawing on crystal structure and solution NMR models. Chapter 9 complements these IR experiments with fluorescence quenching measurements of the dimer-monomer equilibrium of insulin using dye-labeled proteins and intrinsic tyrosine fluorescence.

Chapter 8

Insulin Dimer Dissociation and Unfolding Revealed by Amide I Two-Dimensional Infrared Spectroscopy

8.1 Introduction

The monomer-dimer transition of insulin has been probed with two-dimensional infrared spectroscopy and related infrared spectroscopies to isolate spectral signatures of the conformational changes concomitant with dissociation. These experiments were atomistically interpreted using 2D IR spectra calculated from an ensemble of monomer and dimer structures including the effects of disorder, which provided a complement and a point of comparison to NMR and x-ray crystallography models. The amide I $\nu_{\perp}$ mode, which is delocalized over both monomer units through an intermolecular antiparallel β sheet, was lost upon dimer dissociation and shifts were observed in the $\nu_{\|}\beta$-sheet and α-helix bands. These spectral changes provided a structurally sensitive probe of dimer dissociation, which was used to measure the binding constant, K_D, and to parameterize a thermodynamic model for the dimer fraction. The solvent conditions surveyed the effects of ethanol and salt addition on the dimer fraction in acidic, deuterated water as a function of temperature. It was found that addition of ethanol had a significant destabilizing effect on the dimer state, and shifted K_D from 70 μM in D_2O to 7.0 mM in 20% EtOD at 22 °C. Simulation of the monomer 2D IR spectra indicates that the B-chain C terminus is partially disordered, although not fully solvated by water.

8.2 Methods

8.2.1 Sample Preparation

Insulin from bovine pancreas was obtained from Sigma-Aldrich (St. Louis, MO) and used in all of the experiments described in this manuscript. The sample was H/D exchanged for infrared studies by twice dissolving it in D_2O at ~1 mg/mL, heating to 60 °C for 1 h, and lyophilizing it. Full H/D exchange was verified by the loss of the protonated amide II band at 1550 cm^{-1}. All of the buffers used were based on 0.27 M DCl in D_2O and some buffers included EtOD. All of the deuterated reagents were obtained from Cambridge Isotope Laboratories, Inc. (Andover, MA), and verified to be zinc free to further inhibit the formation of aggregates.

8.2.2 Window Treatment

At insulin concentrations of 10 mg/mL, the conditions for temperature-dependent spectra without using ethanol co-solvent were aggregation prone and necessitated treating the ordinarily hydrophobic CaF_2 glass sample cell windows with a hydrophilic coating of a polyethyleneglycol (PEG) silane. The window was pretreated by rinsing with 4M NaOH and distilled H_2O. The silane solution was prepared by diluting methoxy(ethyleneoxy)$_{9-12}$ propyltrimethoxysilane, obtained from Gelest, Inc. (Morrisvila, PA), to 2% in a 10 mM acetic acid solution in 95% ethanol and 5% water. The entire window was dipped in the silane solution for 2-3 min, rinsed with ethanol, and cured for 10 min at 110 °C in a hot plate oven. As the coating degraded (typically after 2-3 experiments), the windows were cleaned with NOCHROMIX (Cabin John, MD), and recoated using the aforementioned procedure.

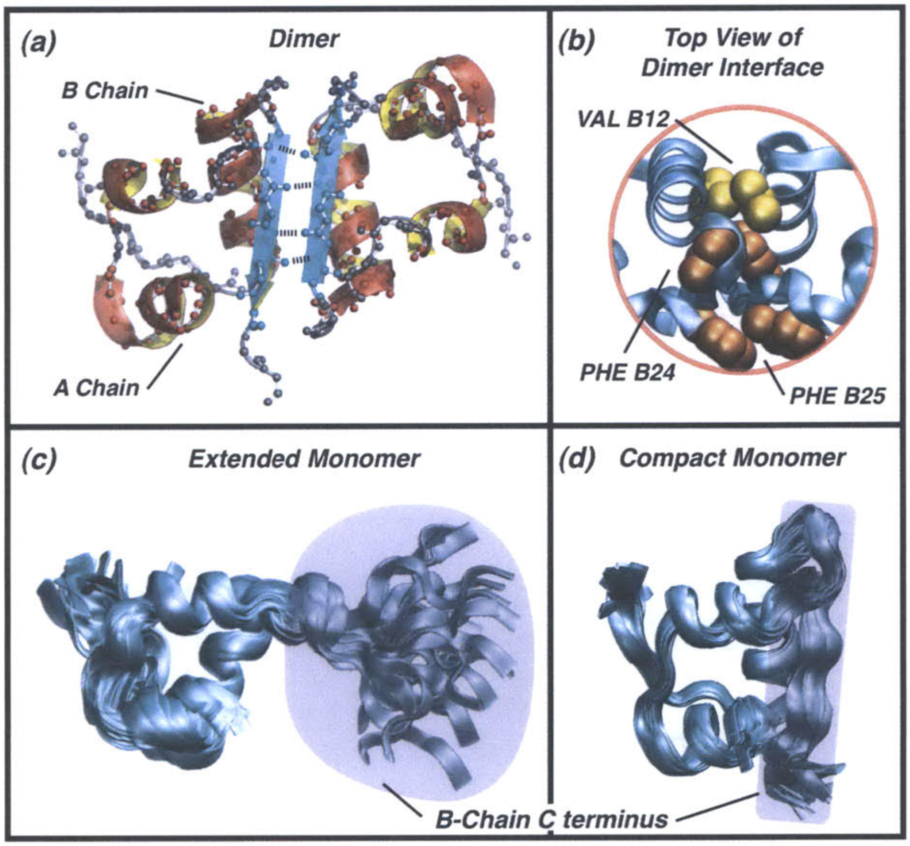

(a)
Dimer
B Chain
A Chain
(b)
Top View of Dimer Interface
VAL B12
PHE B24
PHE B25
(c)
Extended Monomer
(d)
Compact Monomer
B-Chain C terminus

Figure 8-1: Structural visualizations of insulin dimer and monomer models. Strucutral models from PDB IDs 4INS (a,b), 2JV1(c), and 1JCO (d). Structures visualized using VMD[15] and POV-Ray.

8.2.3 Simulations

Molecular Dynamics

To help interpret the experiments, 2D IR spectra were calculated for three structural models derived from the porcine insulin dimer (PDB ID: 4INS)[16], the human insulin monomer in 35% CD3CN (PDB ID: 2JV1)[256], and an engineered, monomeric human insulin mutant (PDB ID: 1JCO)[253] (see Figure 8-1). All of the structures were simulated in GROMACS 3.3.1.[282–288] In the simulation procedure, missing protons were filled in, and the structures were energy-minimized for up to 10000 steps to conform to the OPLS/AA[289–295] force field parameters. The protein structures were solvated[296] with SPC/E water[297] and position-constrained during 100 ps of dynamics in the NPT ensemble at 300 K and 1 atm. After all of these equilibration steps, molecular dynamics were run in the same NPT ensemble, and structures of the entire solvent box were saved each 20 fs for calculating IR spectra. The dimer was simulated for 1 ns and each structure of the monomer ensembles (2JV1 has 50 structures and 1JCO has 25 structures) was simulated for 100 ps.

IR Spectroscopy

The IR absorption spectra and 2D IR spectra were calculated using the previously described code,[8] which has been upgraded to incorporate developments in the models.[99,100,171] The 2D IR spectra were calculated using site energies derived from the electrostatic potential across each amide unit[123] and a combination of electrostatic and DFT-derived coupling between sites.[171] All of the simulated frequencies are corrected for systematic errors with a 20 cm^{-1} redshift. To better approximate the dynamics, the time-averaging approximation[99,100] was used with a Gaussian window of 170 fs. Since the dimer contained 98 sites, a new block diagonalization procedure was employed to make the two-quantum matrix diagonalizations more computationally efficient, which was described in Section 3.1.

Bright State Analysis

The vibrational wavefunctions were visualized using a doorway mode analysis of the simulated spectra, as described previously.[12,68,298] The bright states that characterize the influence of particular amide I vibrations on the infrared spectrum were calculated from the eigenstates of the entire trajectory corresponding to energies within a frequency window. To calculate the mode decomposition, bright states were calculated for a 5 cm^{-1} sliding window across the spectrum. Then, each bright state was decomposed into contributions from the residues in chosen motifs (α helices, β sheets, and unstructured regions) by summing the squared amplitudes for residues in each respective structural motif. In the ideal scenario, the amplitude of the first bright state dominates the total. For all of the calculations in this manuscript, the first bright state was the only one considered, and this was found to carry 0.41-0.44 of the total intensity, with no dependence on frequency or structure.

8.3 Results

8.3.1 Concentration-Dependent 2D IR Spectral Features

To reveal the secondary structural changes associated with insulin dimerization, concentration-dependent 2D IR spectra were acquired in 20% ethanol. Figure 8-2 shows the absorptive and NRPS representations under for monomeric conditions (430 μM) and at higher concentrations where dimer fraction is 50% (6.9 mM). The dimer spectra are consistent with previously observed spectra of mixed α/β proteins, which show a Z-shaped contour lineshape, indicative of a β sheet. (To interpret basic features of 2D IR spectra, the reader is referred to our review[1]).

In each of the dimer 2D IR spectra (Figure 8-2A and 8-2D), three peaks can be seen along the diagonal, which are highlighted in the diagonal slices plotted above each spectrum. Focusing on the absorptive ZZYY spectra, the most intense peak appears at $\omega_1 = 1657$ cm^{-1}, and can be attributed to α-helical vibrations based on empirical assignments. The second most intense peak at 1645 cm^{-1} corresponds to $\nu_\perp$ vibrations of the β sheet, named by the fact that its transition dipole lies perpendicular to the strands of the β-sheet. At 1690 cm^{-1} there is a peak corresponding to its inverted symmetry partner, the $\nu_\parallel$ β-sheet mode, which may also include contributions from β turns. The same vibrations are observed in the NRPS as distinct peaks due to the interference effects inherent to non-rephasing spectra, and are broader due to the inclusion of imaginary contributions to power spectra.

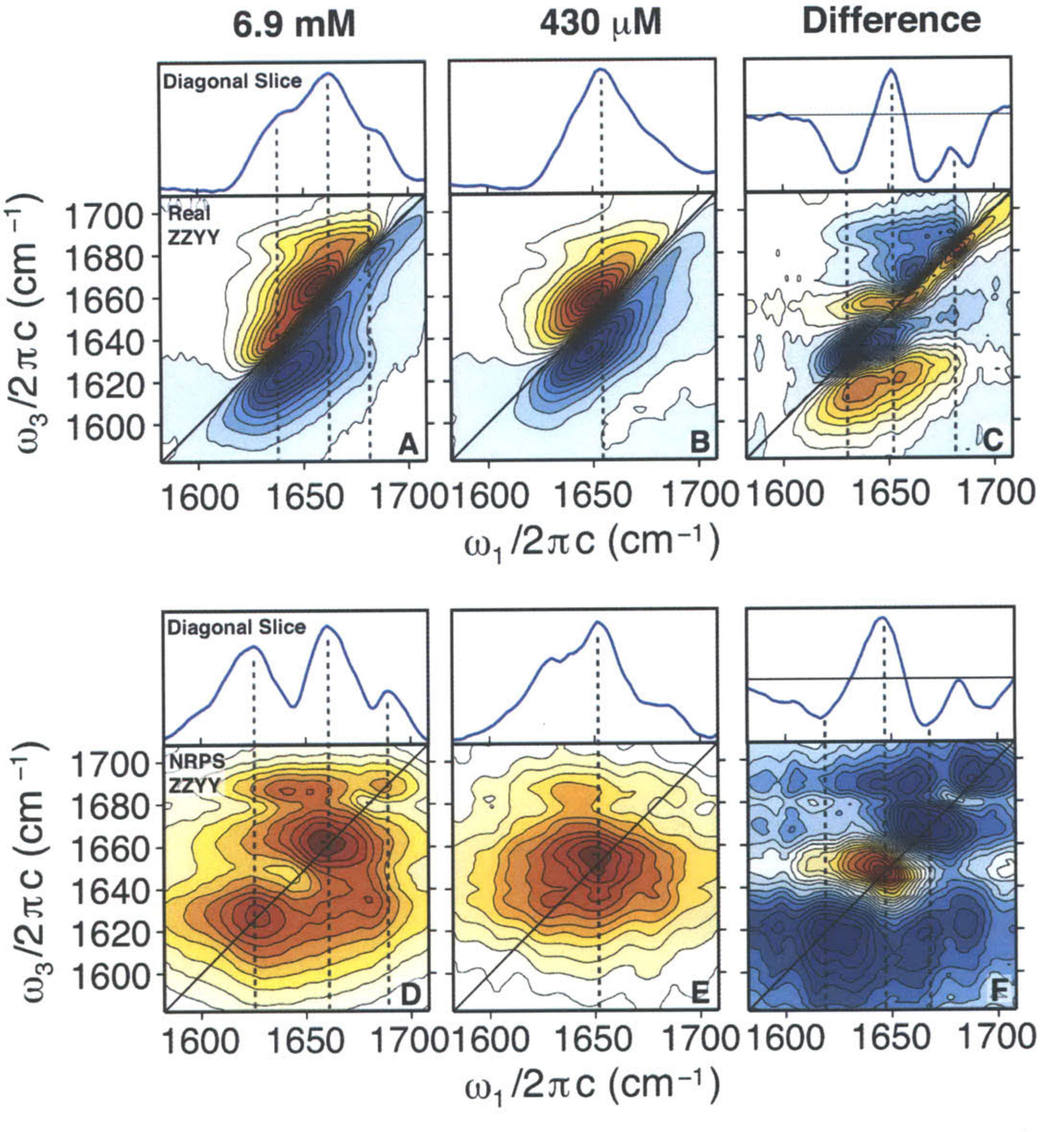


Figure 8-2: Concentration-dependent Insulin ZZYY 2D IR spectra and difference spectra. Concentration-dependent ZZYY 2D IR spectra and difference spectra acquired in 0.27 M DCl 20% (v/v) EtOD-D2O. Top: absorptive spectra; bottom: NRPS. All contours are plotted in 8.3% intervals. Slices along the diagonal ($\omega_1 = \omega_3$) are plotted above each spectrum, except for panel A and B, where the slices chosen to coincide with the peak maximum ($\omega_1 + \Delta = \omega_3$).

The dimer 2D IR spectra also show a pattern of cross peaks among the α-helical and β-sheet vibrations, which are most easily distinguished in the NRPS (Figure 8-2D). A cross peak between the two β-sheet vibrations is observed along the horizontal corresponding to ω_3=1635 cm^{-1} and cross peaks between $\nu_\parallel$ and both $\nu_\perp$ and the α-

helical mode appear at ω_1 and $\omega_3 = 1690$ cm^{-1}. The linewidth of these peaks gives rise to cross-peak ridges, or broadening along the ω_1 dimension. This effect is enhanced in the absorptive 2D IR spectra.

The monomer 2D IR spectra (Figure 8-2B and 8-2E) show one clear peak that is similar to the α-helical feature in the dimer spectra, but it is downshifted to 1650 cm^{-1}. At predominantly monomer concentrations, the $\nu_\perp$ mode and its cross-peaks are greatly suppressed, although weak features arising from the β-sheet modes remain. Inspection of the NRPS diagonal slice shows that the higher frequency, 1690 cm^{-1} peak still appears in the monomer spectra. The difference spectra (Figure 8-2C and 8-2F) summarize the spectral changes that occur as the equilibrium is shifted towards insulin monomers; there is a diagonal narrowing of the spectrum due to the loss of intensity at 1635 cm^{-1} and 1690 cm^{-1} in the diagonal and cross-peak regions, and the peak maximum shifts from 1665 to 1650 cm^{-1}.

8.3.2 K_D Measurements

To provide further evidence that the spectral changes arise from dimer dissociation and to insure consistency with previous investigations, concentration-dependent FTIR and ZZZZ 2D IR spectra were acquired in the range of 88 μM to 6.9 mM in D_2O and 400 μM to 6.9 mM in 20% EtOD:D_2O. Since spectral intensities in both experiments are linear in concentration, changes in concentration-normalized spectra will reflect monomer or dimer fraction. Under these conditions the integrated area in FTIR spectra remained constant, indicating no significant change in transition dipole moments between the monomer and dimer states.

Figure 8-3 shows the analysis of FTIR and 2D IR RPS to obtain the dissociation constant, K_D. Across the concentration range, the changes to the FTIR spectra are barely discernible. The spectra can be described without a priori knowledge of the monomer and dimer spectra by singular value decomposition (SVD) analysis, which reveals a slight red shift with concentration. The dimer population was obtained from 2D IR by integrating the off-diagonal region of the RPS that included β-sheet signatures from the dimer ($\omega_1 =$ 1582-1624 cm^{-1}, $\omega_3 =$ 1664-1708 cm^{-1}). This region,

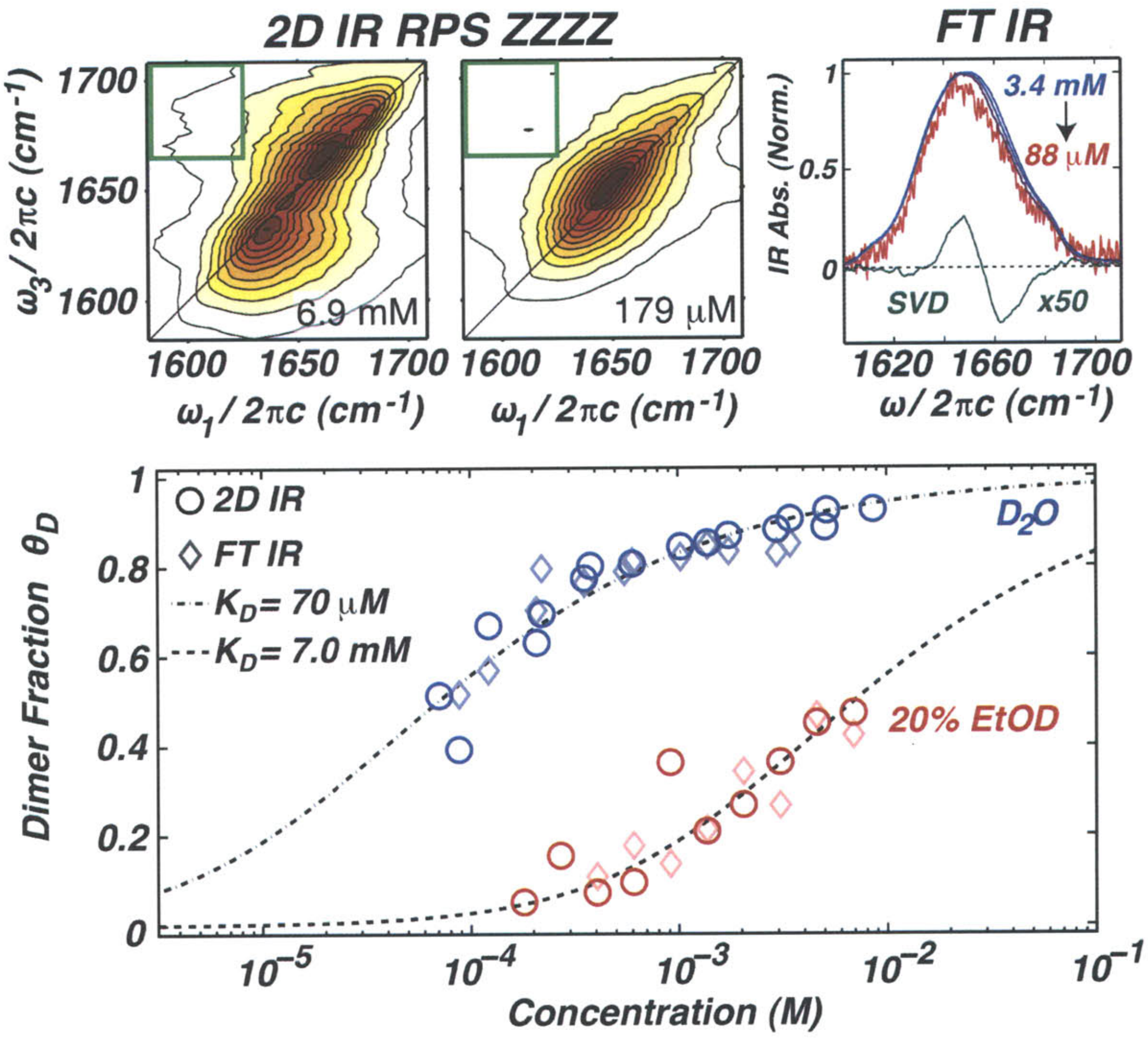


Figure 8-3: Measurement of the insulin dimer dissociation constant, K_D. Extraction of dimer fraction from FTIR and 2D IR rephasing power spectra (RPS). Representative RPS spectra, and the FTIR series and 2nd SVD component are shown on top. Concentration-normalized off-diagonal integration of 2D IR spectra and second component SVD amplitudes of FTIR spectra were fit to the dimer fraction to extract K_D (bottom).

illustrated in Figure 8-3, was chosen for its large monomer/dimer contrast, insensitivity to phasing and non-resonant contributions, and favorable signal-to-noise. The concentration-normalized 2D IR intensities and FTIR second component SVD amplitudes, $I(c)$, were taken to be proportional to the dimer fraction,

$$aI(c) + b = \theta_D(c) \tag{8.1}$$

$$\theta_D(c) = \frac{4c + K_D - \sqrt{K_D^2 + 8K_Dc}}{4c} \tag{8.2}$$

Here c is the total insulin concentration, and a, b, and K_D were the fit parameters. K_D values of 70 μM and 55 μM were extracted from the 2D IR and FTIR measurements in D_2O, respectively. Addition of 20% ethanol shifted K_D to 7.0 mM and 6.0 mM according to 2D IR and FTIR fits. In D_2O, this K_D is consistent with the previously observed values in protonated solvents at similar pHs and ionic strengths, which range from 25 μM to 261 μM. Although the destabilizing effect of ethanol on insulin dimers has been observed,[267,299] these data show the first corresponding K_D measurement.

8.3.3 Temperature-Dependent 2D IR Spectral Features

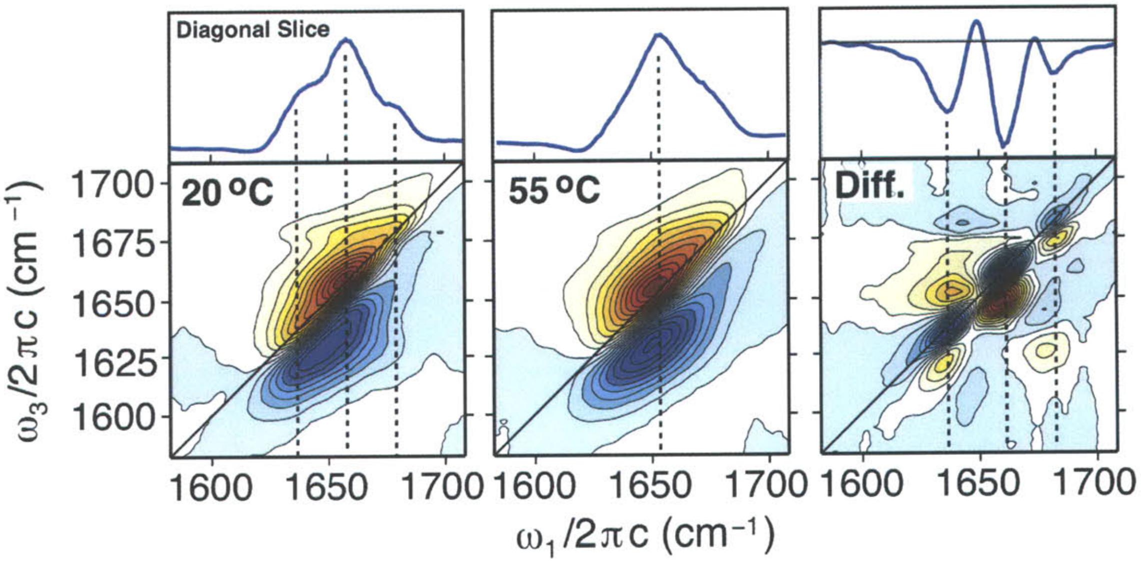


Figure 8-4: Temperature-dependent Insulin ZZYY 2D IR Spectra. ZZYY spectra were acquired for 1.7 mM insulin concentration in 0.27 M DCl, 0.1 M NaCl 20% EtOD-D2O. All spectra show contours plotted in 8.3% intervals.

Temperature can also be used to control the monomer/dimer equilibrium of insulin. Figure 8-4. shows temperature-dependent 2D IR spectra collected at 1.7 mM. At 20°C, the 2D IR spectrum has a single clear peak at 1655 cm^{-1} with high and low frequency shoulders at 1637 cm^{-1} and 1676 cm^{-1}, which are similar to the features associated with increasing dimer content in Figure 8-2A. Moreover, a similar pattern of cross-peak ridges is observed among the β-sheet and α-helix vibrations, most clearly visible at ω_3=1680 and 1620 cm^{-1}. The difference spectra show the loss of distinct peaks along the diagonal and in the cross-peak regions; the loss of both positive and negative lobes of the cross-peak between the $\nu_\perp$ and $\nu_\|$ modes can be seen at ω_1 1680 cm^{-1}. In addition to the temperature-dependent changes induced by dimer dissociation, the 2D IR spectra will reflect thermal changes to the vibrational dynamics and increased solvent transmission. However, the similarity between the dimer and monomer signatures between the temperature- and concentration-dependent 2D IR spectra provides conclusive evidence that increasing the temperature induces dimer dissociation.

8.3.4 Solvent-Dependent Thermal Dissociation and Unfolding

For a more extensive characterization of the insulin binding thermodynamics, the temperature dependence of dimer dissociation was quantified in eleven different solvent conditions using HDVE. Because the HDVE spectrum is formally equivalent to an ω_3 projection of the 2D IR spectrum, its power spectra are a phase-insensitive measurement that retain vibrational coupling information, and can be acquired rapidly without scanning any time-delays. The set of solvents was chosen to test the effects of adding ethanol and sodium chloride salt to the thermal dissociation at 1.7 mM (10 mg/ml). Figure 8-5A shows representative HDVE thermal dissociation data in 0.27 M DCl 30% EtOD:D_2O. The salient changes to the HDVE spectrum are a narrowing with temperature due to a shift of intensity from 1640 and 1690 cm 1 to 1650 cm^{-1}, which is consistent with the 2D IR changes seen in Figure 8-4. The addition of salt had less of an effect on the thermal dissociation than ethanol, and its effects were not

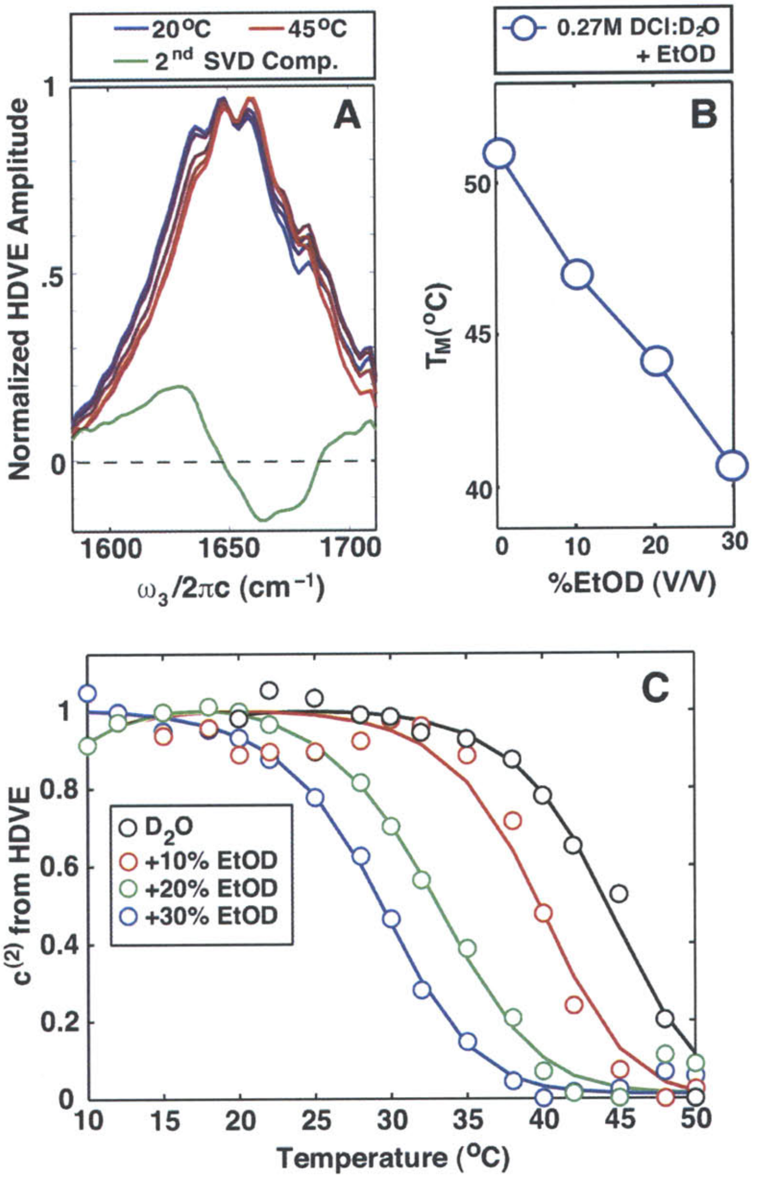


Figure 8-5: Temperature-dependent HDVE spectra were acquired as a function of ethanol co-solvent concentration. Spectra were acquired in 2.5 °C intervals from 10-55 °C. Representative spectra and 2nd SVD component are shown in panel A for 30% EtOD. Extracted melting temperatures and amplitudes of the second SVD are shown in B and C.

additive with ethanol (See Figure 8-6).

The temperature-dependent HDVE spectra in all ethanol-containing solvents were decomposed into a two-state basis derived using SVD analysis on the data set in 0.27 M DCl, 0.1 M NaCl 20% (v/v) EtOD:D_2O. Use of the same basis spectra allowed for a

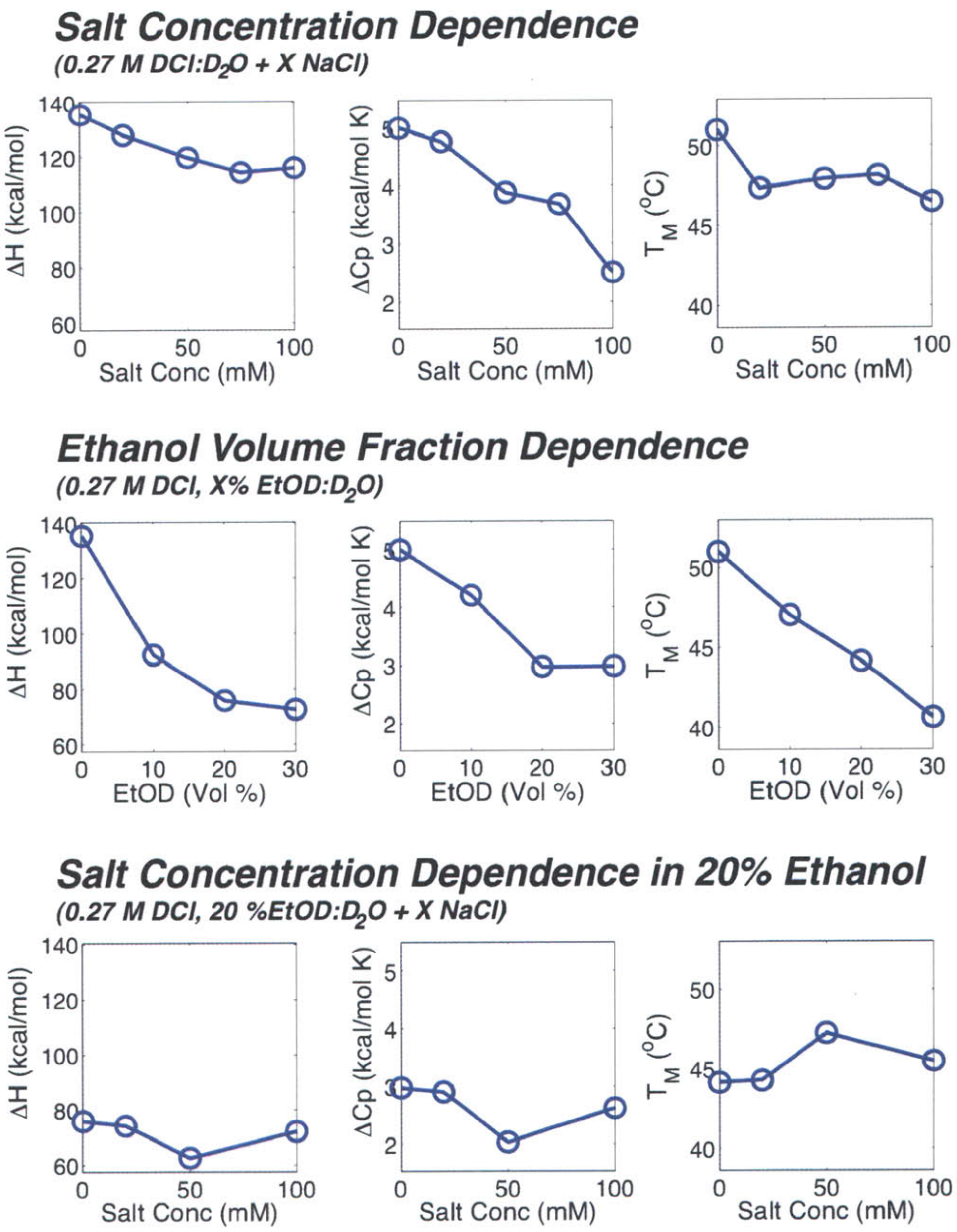


Figure 8-6: Extracted parameters for the two-state insulin dimer dissociation model.

faithful comparison of the temperature-dependence across the solvent conditions. The melting curves in Figure 8-5C show that increasing the ethanol concentration dropped the melting temperature by roughly 3.5°C for each 10% addition of ethanol and reduced aggregation. When the ethanol content was less than 20%, aggregation was a limiting factor for acquiring spectra at temperatures >50°C. Cold-induced aggregation was also seen that may be a sign of cold-induced dissociation, which is predicted by the two-state thermodynamics. One key feature of 2D IR spectroscopy is that the transition dipole scaling makes the distinct 1615 cm^{-1} mode of β-sheet aggregates appear intensely, which allows these spectra to be excluded from the analysis. (See Figure 8-7). The fits were performed over temperatures that were aggregation-free for all samples. An identical analysis was performed on the experiments with salt-containing solvents, although more emphasis will be placed on the ethanol results due to their reduction of aggregation.

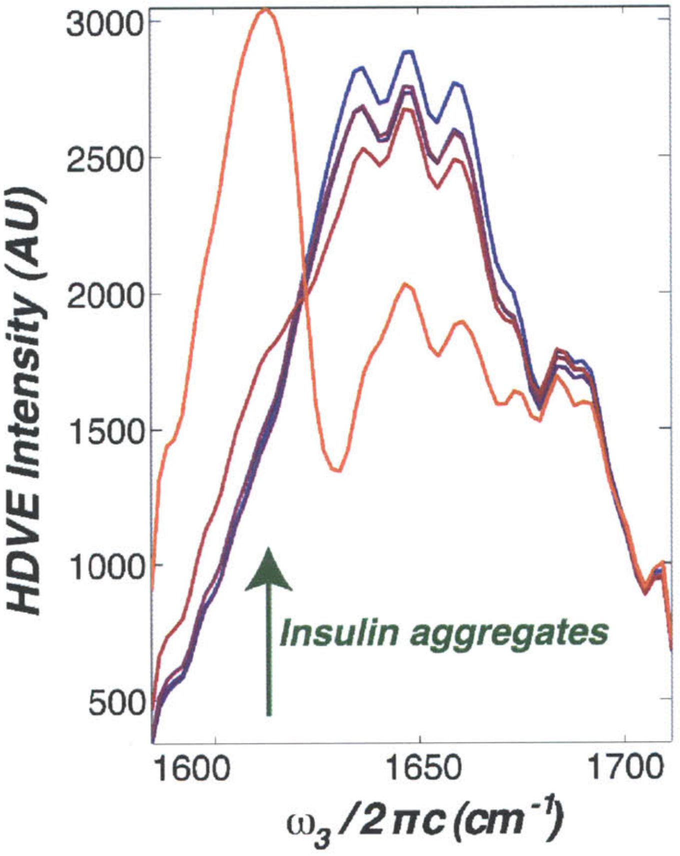


Figure 8-7: Temperature-dependent HDVE series showing the aggregate peak used to exclude spectra from two-state fitting.
Spectra were candidates for exclusion from the fitting procedure if a high or low temperature endpoint in that series clearly displayed the aggregate peak (>20% of the signal). In these sets, the endpoint spectra and adjacent temperatures displaying the aggregate peak (>3% of the signal) were excluded. Aggregation was inhibited by salt concentrations $\geq$ 50 mM and ethanol fractions $\geq$20%.

8.3.5 Thermodynamic Modeling

After decomposing the spectra into a basis derived from SVD analysis, the second component amplitudes were fit using a thermodynamic model that extended two-state models for protein unfolding[300] with concentration dependence,

$$D \leftrightharpoons 2M \tag{8.3}$$

$$\Delta G = \Delta H^0 - T\Delta S^0 + \Delta C_P \left[T - T_M - Tln\left(T/T_M\right)\right] \tag{8.4}$$

$$\Delta G = -RTln\frac{c_M^2}{c_D} \tag{8.5}$$

where ΔG, ΔH^0, ΔS^0, and ΔC_P are the free energy, enthalpy, entropy, and heat capacity changes, and T_M is defined as ΔG=0, which may not be where the normalized melting curves intercept $\frac{1}{2}$. The second component amplitude was assumed to be proportional to the dimer fraction defined as

$$\theta_D(T) = \frac{2c_D}{2c_D + c_M} \tag{8.6}$$

where c_M and c_D are the concentrations of monomers and dimers. To fit the thermal dissociation curves, T_M, ΔH^0, and ΔC_P were varied and the resulting K_D was used to calculate $\theta_D(T)$ using the standard state of c_0=1 M such that ΔG=0 if K_D=1 M. These fits were constrained by our measurement of the equilibrium constants at 22 °C and considered data at temperatures that were aggregation-free. The equilibrium constant measured in 20% ethanol was assumed to be representative of solvents with 10-30% ethanol. For the reference solvent of 1.7 mM insulin in 0.27 M DCl:D_2O, the extracted parameters were T_M=51 °C, ΔH^0=135 kcal/mol and ΔC_P=5 kcal/mol K. At 22 °C, ΔH = -9.6 kcal/mol, ΔS = 51 cal/mol K and ΔG = 5.4 kcal/mol. Upon addition of 30% ethanol, T_M dropped by 11 °C, ΔH^0 was -60 kcal/mol and ΔC_P was -2 kcal/mol K, which indicated greatly reduced stability of the dimer state relative to the monomer. The addition of 100 mM salt caused minor changes relative to ethanol addition; it reduced T_M by 5 °C in the reference solvent and increased T_M by 1.5

°C when added to 20% ethanol conditions. In summary, ethanol was found to be an effective co-solvent for destabilizing the insulin dimer state relative to the monomer without any significant changes to the spectral features (See Figure 8-8).

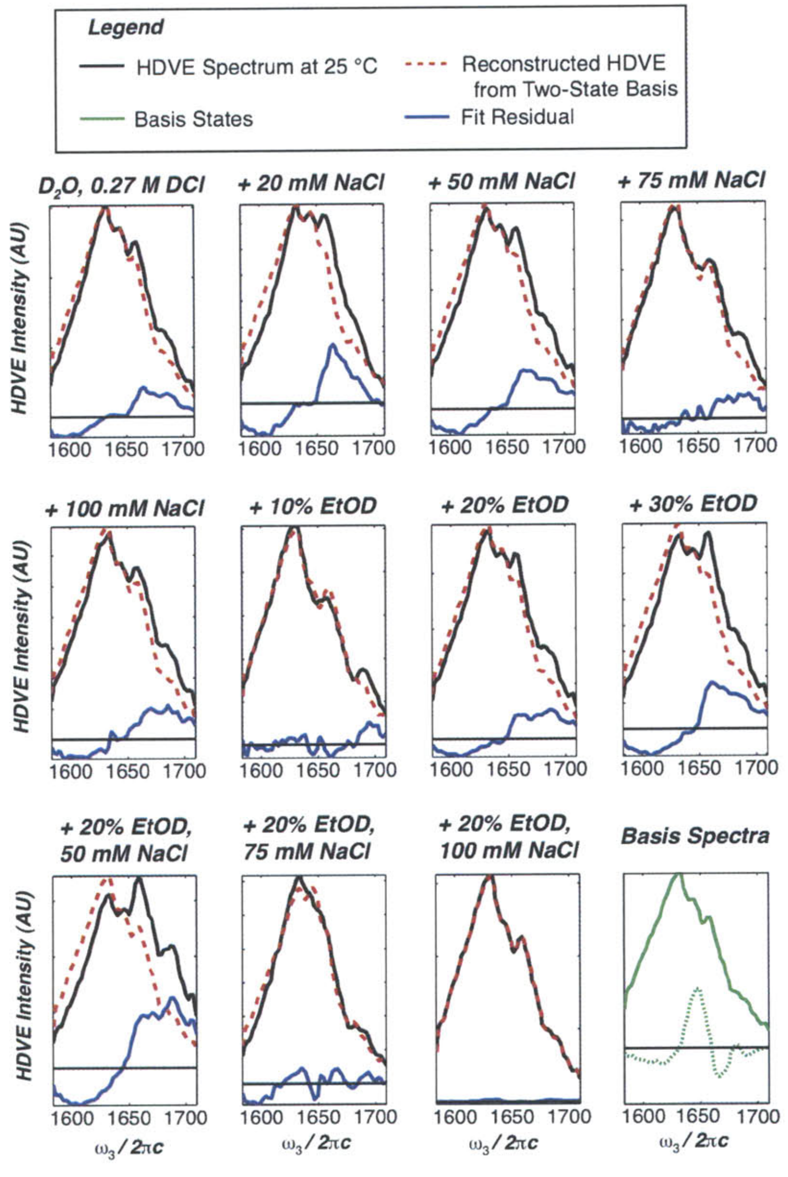


Figure 8-8: Solvent-dependent HDVE spectra. The HDVE spectra are shown for 25°C with reconstructed spectra using a two-state basis derived from 0.27 M DCl, D2O:20% EtOD + 100 mM NaCl. The fit residuals are indicative of experimental irreproducibility, such as shifts in the spectrum of the laser or variations in scattering.

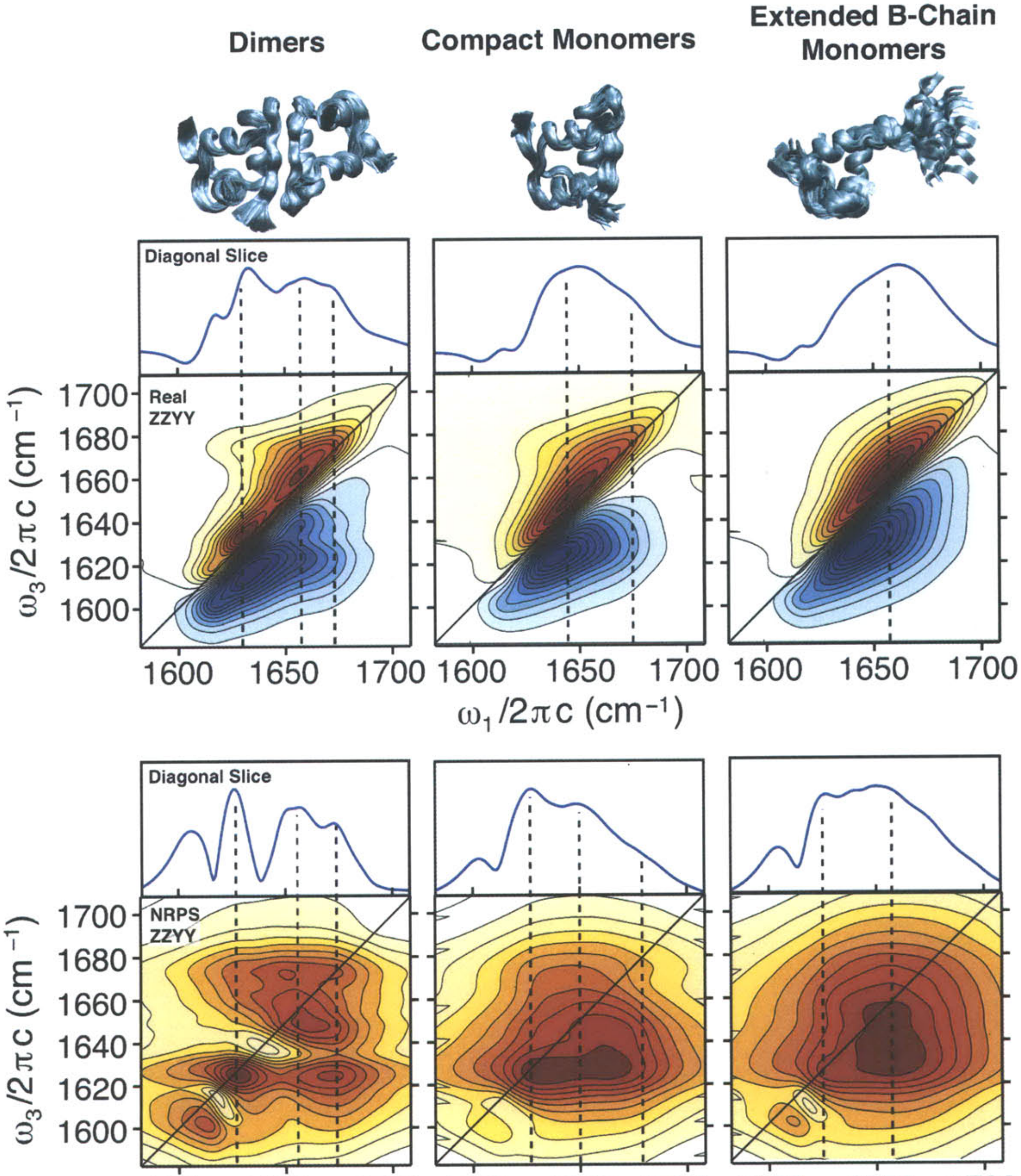


Figure 8-9: Simulated 2D IR spectra for insulin monomers and dimers. Simulated 2D IR spectra for dimers (4INS), compact monomers (1JCO) and extended B-chain monomers (2JV1). The intensity at 1620 cm^{-1} includes an exaggerated contribution from proline. Structures visualized using VMD[15] and POV-Ray.[?]

8.3.6 Structural Modeling of Insulin 2D IR Spectra

A key feature of amide I protein 2D IR spectra is the ability to make atomistic assignments with structure-based modeling. To interpret the 2D IR spectra, MD simulations were performed on structural models obtained from x-ray crystallography or solution NMR spectroscopy and the spectra appear in Figure 8-9. Since the spectra in the amide I region arise almost entirely from backbone vibrations, models for different insulin secondary structures can be compared, even in cases such as these where the primary structure displays minor variation from the experiment. The 2D IR spectra calculated from static structure snapshots in the trajectory displayed fine structure from many peaks. Each structure was solvated and dynamics were run to sample the small-amplitude, sub-ns solvent and protein fluctuations (10,000 - 50,000 frames) to obtain smooth lineshapes consistent with the experiments.

The calculated 2D IR correlation spectrum for the dimer shows three diagonal features: two well-defined peaks at $\omega_1 = 1634$ and 1660 cm^{-1} and a shoulder at 1678 cm^{-1}, which nominally correspond to the $\nu_\perp$ β-sheet mode, α-helical modes, and the $\nu_\parallel$ β-sheet modes assigned for the dimer spectrum in Figure 8-9A (all calculated frequencies are corrected with a systematic 20 cm^{-1} redshift). While it is a subtle feature in the calculated spectra, the high-frequency shoulder is further evidenced by the cross-peak ridge extending horizontally from the diagonal at 1678 cm^{-1}. The positions of the $\nu_\perp$ and α-helix bands match the experiment exactly, but the $\nu_\parallel$ mode is 10 cm^{-1} too red-shifted. Another difference with respect to the experiment is that the intensity of the α-helix peak is 50% too low in the calculation relative to the β-sheet modes. These difference may be due to the structural differences between the solution and crystalline dimer states or errors in the calculation. Examination of the corresponding NRPS shows discrete cross-peaks between $\alpha/\nu_\parallel$ modes at (1655/1676 cm^{-1}) and $\nu_\parallel/\nu_\perp$ modes at (1675/1627 cm^{-1}), with a similar overall structure to the experimental NRPS.

There are varying proposals for the structure of the insulin monomer, which differ in the conformation of the B chain C-terminus that folds and stabilizes dimer contacts.

2D IR spectra have been calculated based on two ensembles resulting from solution NMR experiments (Figure 8-1). These structures place a lower limit on the amount of disorder that may be observed in 2D IR spectra, due to the fact that the ns-μs fluctuations of the B chain terminus would be in the partial-complete motionally narrowed limit for 30-200 ms mixing time NMR experiments, but in the static limit for 2D IR experiments with 4 ps coherence times. (Evidence for poor separation of these timescales has been observed in the anomalous variation of amide line widths in insulin.[301]) One of the monomer ensembles was comprised of 50 compact structures with B chains that strongly resembled the conformation in the dimer (PDB ID: 2JV1). The second ensemble (PDB ID: 1JCO) contained 26 structures with extended B chain C termini, and more disorder in the N-terminal B-chain region and the A-chain helices. The resulting spectra were summed over the respective ensembles and the results appear in Figure 8-9. Each of the calculated monomer spectra shows features that resemble the experimental monomer spectra (Figure 8-2B and 8-2E), but neither one entirely reproduces the experiment. The compact monomer spectrum has its peak in the same position as the experiment, 1650 cm^{-1}, but displays too much off-diagonal structure with its cross-peak ridges at ω_1=1676 and 1630 cm^{-1}. These off-diagonal features are not present in the extended B-chain monomer spectrum, but this spectrum is too diagonally elongated compared to the experiment. Inspection of the NRPS shows a feature at 1630 cm^{-1}, which appears in both monomer spectra, but is more intense for compact monomers. This feature arises from remaining strong vibrational couplings between the B-chain terminus and the helices in the monomer.

8.3.7 Bright State Analysis of the Calculated Spectra

A quantitative structural assignment of the vibrational modes contributing to different regions of the IR spectrum was performed with doorway mode analysis. This procedure obtains the amide I bright states that carry the IR transition dipole intensity within a given spectral window, provides a visualization of the structures that contribute within that window, and allows the contributions to be decomposed by different secondary structures. Doorway-mode analysis was performed on a 5 cm^{-1}

sliding window across the dimer spectrum (Figure 8-10a). The mode decomposition was obtained from the bright states within a spectral window by summing over the squared vibrational amplitudes for oscillators within the motif of interest as,

$$P_{sec\ str} = \sum_{i}^{sec\ str} |a_i|^2 , \tag{8.7}$$

where a_i is the vibrational amplitude of the ith oscillator in the unit-normalized bright state. The β-sheet features were found to peak at 1646 cm^{-1} and 1680 cm^{-1} and it was found that the bright states corresponding to these bands display $\nu_\perp$ and $\nu_\parallel$ symmetry character. These modes are observed in proteins, and here it was noted that they can also arise from intermolecular couplings. Further, the α-helical and β-sheet modes are significantly mixed; each mode appears to be delocalized over the dimer. While the dimer does not rigorously contain a C_2 axis, the α-helix oscillations on each monomer unit are still generally out-of-phase with respect to each other. The direct coupling between α-helices is weak (≤ 2 cm^{-1}), and the dominant coupling mechanism was found to be through mutual interaction with the β-sheet oscillators. By comparing the bright states in the two monomer ensembles, the effects of extending the B-strand on the mode structure can be understood. Two representative structures from the monomer ensembles were chosen for the doorway mode analysis. In the calculated 2D IR spectra of the monomers, it was noted that there was intensity at 1630 cm^{-1} that did not appear in the experimental monomer spectrum. Bright state analysis in this region (not pictured) reveals that this intensity arises from two regions; a hydrogen bond contact between the two main chains (B4 and A11) and, more dominantly, the β turn bridging the B-chain helix and C-terminal strand. Both of these regions are disordered in the ensemble with the extended B-chain strand, and this intensity is reduced. Comparing the high frequency β-sheet regions for the two monomers, two features appear at 1663 cm^{-1} (μ_{M1}) and 1682 cm^{-1} (μ_{M2}), shown in Figure 8-10b and 8-10c, respectively. The μ_{M1} mode is more intense in the compact monomer, and is a combination of α-helix modes strongly mixed with vibrations on the β strand. The μ_{M2} mode is more intense and broader in the extended B-strand monomer, and is localized to the turn with decreasing participation further along the strand. In summary, these results show that adding disorder to the monomer structure and extending the B-chain strand leads to a loss in the 1630 cm^{-1} region and localizes the $\nu_{\|}$ mode to the turn region, which causes it to blueshifts and broaden.

8.4 Discussion

8.4.1 Monomer and Dimer Spectral Signatures Across the Various IR Spectra

There are features common among all of the amide I monomer and dimer spectra that can be interpreted using the structure-based calculations. In all of the IR spectra, a loss is seen at 1630 cm^{-1} upon dimer dissociation. Using the bright state corresponding to this mode in the insulin dimer, it can be assigned to delocalized mode that is dominated by $\nu_{\perp}$-symmetry vibrations on the two-stranded, anti-parallel β sheet. In the 2D IR spectra, loss of this peak on the diagonal is accompanied by the loss of corresponding cross-peaks to the 1690 cm^{-1} β-sheet and the 1665 cm^{-1} α-helix modes.

The simulations can be used to assign intensity in the 1690 cm^{-1} region. There is partial intensity loss in this region upon dimer dissociation, which is evidenced by diagonal narrowing of the monomer 2D IR spectra and the loss of discrete diagonal and cross peaks in the NRPS. However, a cross-peak ridge at ω_1=1690 cm^{-1} remains in the monomer spectra and diagonal slices of the absorptive 2D IR spectra show a peak intensity ratio ($\nu_{\perp}$ β-sheet : α-helix : $\nu_{\|}$ β-sheet) of 6:10:5 in the dimer and 1:10:4 in the monomer. This mode cannot result from dimer contamination, since the dimer signature $\nu_{\perp}$ mode carries a much larger transition dipole than the $\nu_{\|}$ mode. The simulations can be used to assign the 1690 cm^{-1} region in both of the monomer spectra as a high frequency β-strand mode. This mode includes contributions from the turn region and is delocalized along the β strand, with less delocalization when the strand is more disordered.

In both the experimental 2D IR and FTIR spectra, a red shift in the peak of the spectrum is seen as the equilibrium is shifted from dimers towards monomers (7 and 9 cm^{-1} for the 2D IR absorptive and NRPS spectra in Figure 8-2, and 5 cm^{-1} in the FTIR spectra). In principle, this peak shift may be caused by any of three effects: a change in the coupling to the oscillators in the α-helices, an intrinsic change to the frequencies of the α-helix oscillators, and interference, such as the loss (or gain) of a nearby peak arising from uncoupled modes at higher (or lower) frequency. Each of these effects can be tested using the simulated spectra.

By comparing the calculated spectra of the dimer and the compact monomer, it is found that the α-helix-localized modes are coupled to β-sheet-localized modes; loss of the anti-parallel β sheet leads to an 11 cm^{-1} red shift in the α-helix peak, despite the fact that the α helices in the compact monomer strongly resemble those in the dimer. This coupling was quantified by transforming to a basis of α-helix-localized and $\nu_\perp$ β-sheet-localized modes using peak positions from the spectra, and a coupling of 15 cm^{-1} was extracted. The same assumptions applied to the experimental spectra yield a coupling of 9 cm^{-1}. While this comparison provides evidence that the loss of $\nu_\perp$ β-sheet modes can shift the α-helix modes, it does not exclude the other mechanisms.

Evidence for a shift in the α-helix peak due to interference can be seen by comparing the two monomer simulations, in which the extension of B chain is accompanied by a 20 cm^{-1} blueshift to the $\nu_\parallel$ band, which would additionally contribute to a blueshift in the α-helix peak. An analysis of the frequencies of individual amide I oscillators shows no appreciable frequency shift for the α-helices in any of the calculations. Thus, comparison to the simulations has concluded that the α-helix peak shifts in the monomer spectrum relative to the dimer due to loss of coupling to the $\nu_\perp$ mode and perhaps also due to interference with the shifting $\nu_\parallel$ mode.

Our analysis showed that the differences between monomer spectra obtained from concentration-dependence and temperature-dependence were negligible. Any changes must be smaller than thermally induced changes to the properties of water and vibrational dynamics of amide I oscillators. These results are consistent with the results of a calorimetry study, which concluded that the insulin dissociation and monomer unfolding transitions are inseparable,[302] and our analysis did not require a third thermodynamic species.

While many monomer features were reproduced in the spectra calculated from the NMR structures, neither one entirely reproduced the combination of diagonal elongation and cross-peak ridges evident in the experimental monomer spectrum. To more quantitatively model the monomer spectrum, one might consider a sum of the compact and extended monomer spectra or a weighted sum of the spectra comprising each ensemble. However, when comparing the simulations and experiments, one should note that the theory to calculate amide I protein spectra is still under development. Typically, a set of simulated spectra can used to derive trends that may be compared to the experiment. Direct comparisons of relative amplitudes, frequency splittings, and linewidths is still at a qualitative or semi-quantitative level. With these limitations, one needs to carefully isolate the features that are robust against the many approximations in these calculations, such as the MD force field, mapping snapshots from the simulation to site energies and couplings, and assuming that the dynamics are near the static limit.

8.4.2 Solvent and Temperature Effects

The effects of adding ethanol to the thermal dissociation curves can be interpreted by noting its drastic influence on the surface tension of water. While there are many contributions to the free energy for insulin dimer dissociation, there will be a component that is proportional to the hydrophobic surface area and the surface tension. For large solutes (radius $>$ 2 H-bond lengths), this solvation free energy is dominated by enthalpic contributions.[303] It is found that the changes in surface tension upon ethanol addition are consistent with the overall trend in ΔH^0 (see Figure 8-11), which is a steep decrease from 0-20%, followed by a continuing gradual decrease; the 42% drop in surface tension upon adding 20% volume fraction of ethanol corresponds to a 44% drop in ΔH^0 from 135 to 76 kcal/mol. The concomitant changes in ΔG at 22°C are relatively much smaller, 5.4 kcal/mol to 2.9 kcal/mol, which is a demonstration of the enthalpy-entropy compensation widely seen in protein biophysics.

An estimation of the hydrophobic surface area from the molecular dynamics simulations used to calculate the 2D IR spectra reveals a value of 32 nm2 for the dimer and a range from 20-25 nm^2 for the monomer, depending on the ensemble chosen. Using an empirical, linear correlation for the hydration enthalpy of hydrocarbons (data obtained from Gallichio et al.[304]), the 8-18 nm^2 of buried surface area in the dimer corresponds to $\Delta H_{Hydrophobic} \approx$ -22 to -49 kcal/mol in pure water for the hydrophobic contribution to the enthalpy. Such a large hydrophobic stabilization of the dimer rationalizes the experimentally observed change upon adding 20% ethanol, ΔH^0 = -59 kcal/mol (see Figure 8-6). These facts are also consistent with an inspection of the structure (Figure 8-1), all of which indicate that the insulin dimer is largely stabilized by hydrophobic interactions.

Our measured value of K_D from 2D IR experiments, 70 μM for bovine insulin in 0.27 M DCl in D_2O, is consistent with previous measurements in protonated water and similar conditions, which range from 25 μM to 261 μM (literature K_D values are tabulated in the Table 8.1). It has been observed that the addition of ethanol greatly destabilizes dimers relative to monomers.[267,299] This work demonstrates that K_D is increased by ~100x, but no corroborating evidence was seen for the previously observed complete dissociation in 5% ethanol.

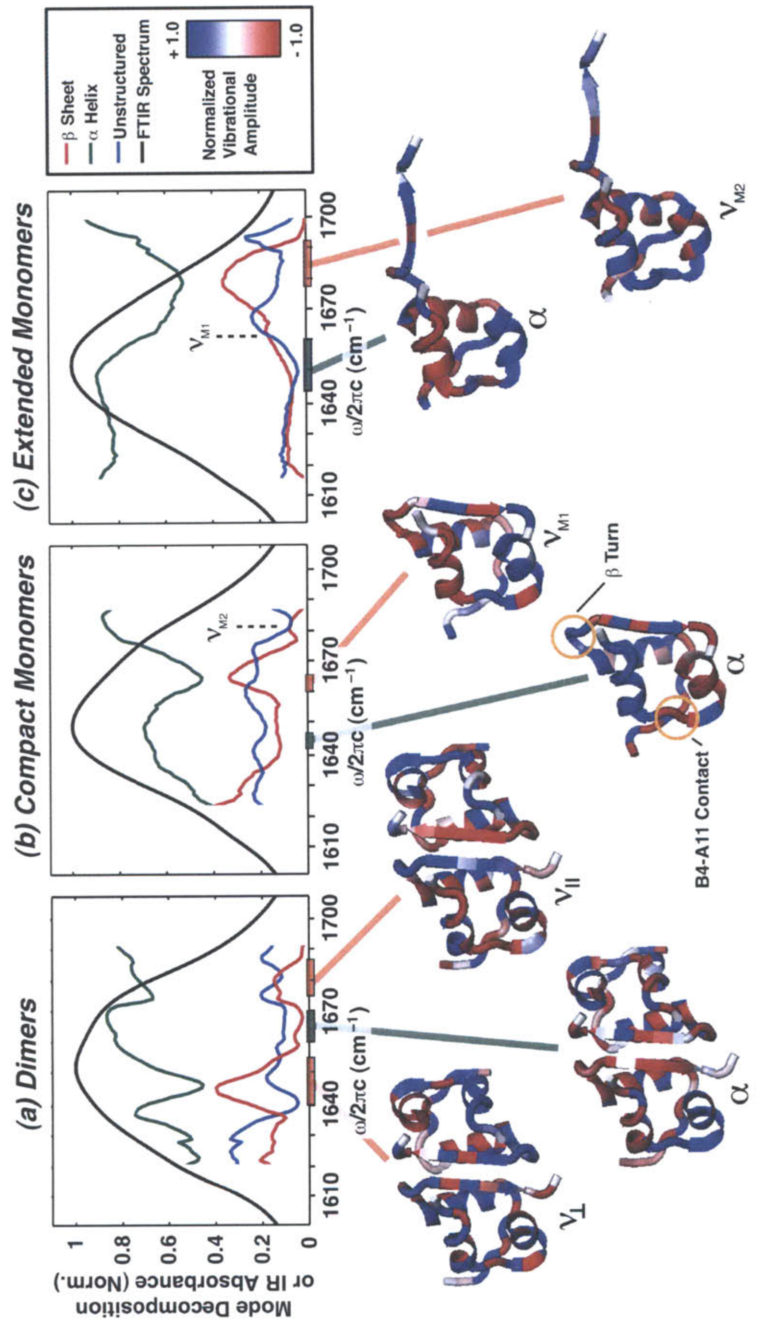


Figure 8-10: Bright state analysis and calculated FT IR spectra. Top: plot of the mode composition from bright state analysis decomposed by secondary structure, and the calculated IR absorption spectrum for the (a) dimer, (b) compact monomer, and (c) extended monomer ensembles. For the mode decomposition, the independent axis gives the beginning of the 5 cm^{-1} window. Below: Visualizations of the bright states for representative spectral regions, prepared using PyMol.(Delano Scientific) The color (red or blue) represents the vibrational phase of backbone amide I oscillators, while the intensity of color reflects the amplitude of vibration. For each bright state, the vibrational amplitudes were normalized to the large magnitude value.

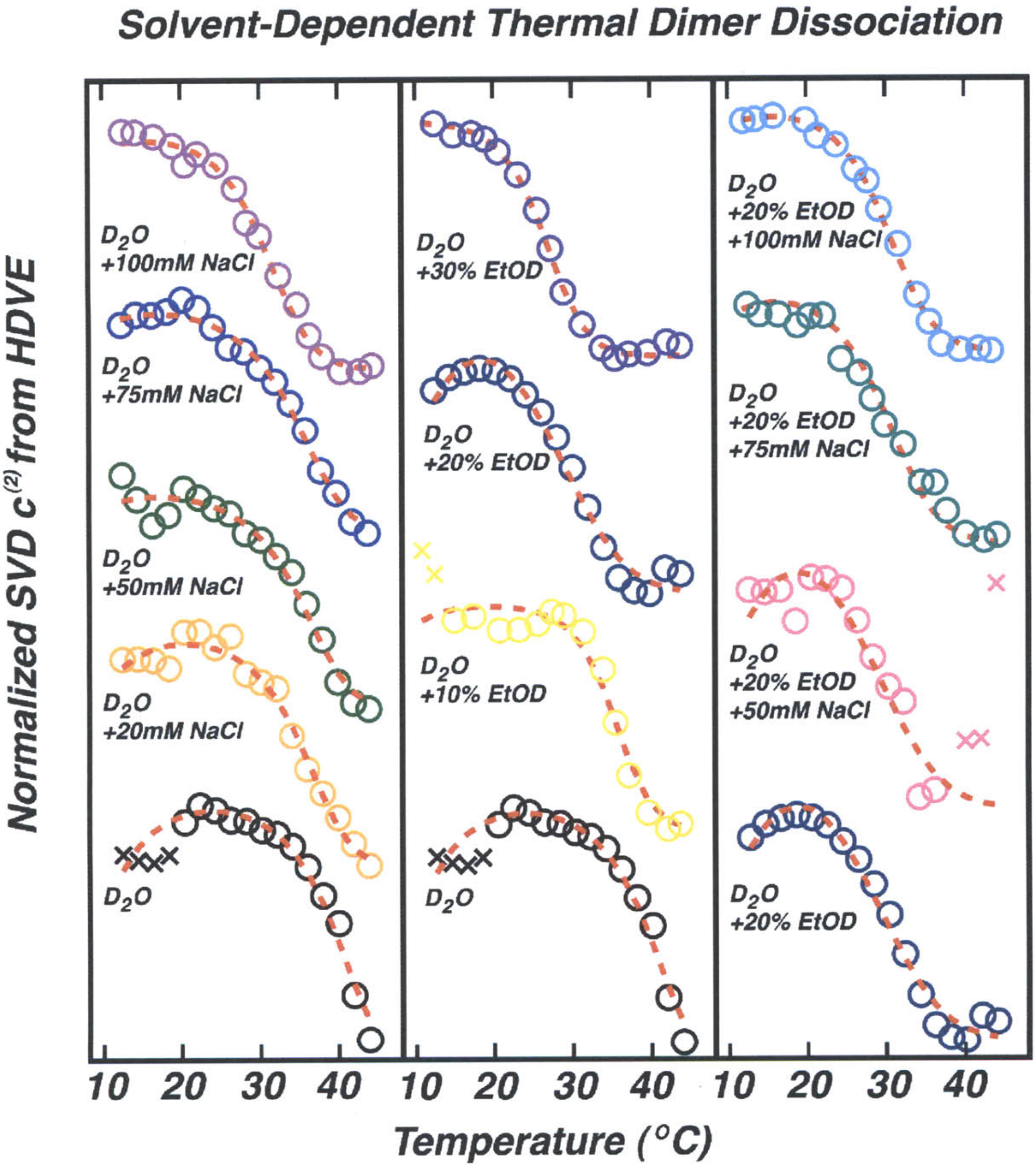


Figure 8-11: Insulin melting and dissociation curves. Melting and dissociation curves in all tested solvents along with two-state model fits, shown in Figure 8-6. The total insulin concentration was 1.7 mM and all solvents included 0.27 M DCl. HDVE experiments were repeated 3-7 times in each solvent, and the sets of spectra chosen for analysis had the least aggregation. For solvents that caused persistent aggregatation, aggregate spectra (points indicated with an X) were excluded from the fit (see Figure 8-7).

Year	Researchers	Technique	pH	Ionic Strength (M)	T(°C)	K_{12} (M^{-1})	$K_D(=1/K_{12})$ in μM
1966	Jeffery, PD et al.[262]	Sedimentation	2	0.05	25	7950	126
			2	0.1	25	10200	98
			2	0.15	25	3830	261
			2	0.2	25	6640	151
1973	Lord, RS et al.[266]	Concentration-Difference UV-Vis Absorption	2	0.1	15	109000	9
			2	0.1	25	40000	25
			2	0.1	30	18000	56
			2	0.1	35	19000	53
			2	0.1	44	6100	164
			3.5	0.1	26	32000	31
			2	0.01	26	11000	91
			3.5	0.01	26	7600	133
1980	Pocker, Y.[270]	CD	2	0.005	25	750000	1.3
1985	Strazza, S et al.[271]		2	0.1	17	27000	37
				0.1	21	16000	63
				0.1	25	12000	83
				0.1	30	10000	100
				0.1	34	6900	144
				0.1	38	5100	196
			7	0.01	25	20000	50
1976	Jeffery, PD et al.[262]	Sedimentation	7	0.2	25	111000	10
1971	Goldman, J et al.[305]	Sedimentation	8	0.1	25	222000	4.5
2000	Nettleton, EJ et al.[265]	Mass Spectrometry	3.3		22	10000	100

Table 8.1: Measures of Insulin Dimerization K_D

Our values for the enthalpy and entropy of dimer dissociation are consistent with previous findings in protonated solvents at room temperature. Investigations using uv-vis absorption at pH 2.0, 25°C, and 0.1 M ionic strength have found ΔH=-17.2 kcal/mol, ΔS=-29 cal/deg mol[266] and ΔH=-12 kcal/mol ΔS=-18 cal/deg mol,[271] which can be compared to our values of ΔH=-9.6 kcal/mol and ΔS=-50 cal/deg mol. Differences in these values may be due to our higher ionic strength, lower pH, and deuterated water. Another difference is that our data was sampled at 2.5 °C, which is finer than some previous work ($\sim$10 °C), and may have neccessitated $\Delta C_P > 0$ when fitting our data. A more recent calorimetry investigation by Dzwolak et al.[272] was analyzed with attention towards cooperativity in the unfolding, and found a transition temperature (not directly comparable to T_M) around 60 °C. However, our value of ΔC_P is much larger (5 kcal/mol vs 0.5 kcal/mol). Dzwolak et al. also found that D_2O lowered the temperature for aggregation by 7 °C, which provides evidence that straightforward comparisons of melting behavior in protonated and deuterated solvents are not possible.

8.5 Conclusions

The dimer dissociation reaction of insulin has been probed with two-dimensional infrared spectroscopy and related nonlinear spectroscopies. These techniques provide a structurally sensitive probe of the monomer-dimer transition that were used to measure the binding constant, K_D, and to parameterize a thermodynamic model for the dimer fraction as a function of temperature and concentration for a range of solvent conditions. The monomer-stabilization effect of 20% ethanol was quantified- at 22 °C, it shifts K_D from 210 μM to 6.9 mM- and rationalized by the increased hydrophobic surface area when the dimer interface is exposed.

Atomistic interpretation of the diagonal and off-diagonal spectral changes were provided by comparison to spectra calculated from atomistic molecular dynamics simulations of insulin, which included the effects of structural disorder and spectral interference. This provided both a point of comparison and a complement to NMR and x-ray crytallography structural models. Upon dimer dissociation, the 1630 cm^{-1} $\nu_\perp$ mode was lost, the 1665 cm^{-1} α-helix mode was redshifted, and the 1690 cm^{-1} $\nu_\|$ mode decreased in intensity. It was shown that the most sensitive marker for insulin dimers, the $\nu_\perp$ mode, derives its sensitivity due to delocalization over both monomer units mediated by the intermolecular anti-parallel β-sheet. Both spectral interference effects and a loss of coupling to the $\nu_\perp$ mode cause the shift in the α-helix mode. The $\nu_\|$ mode undergoes a change in character, from a delocalized β-sheet mode in the dimer to a turn-localized mode that is sensitive to the conformation and disorder in the B-chain C terminus.

From a comparison of the simulated and experimental spectra, it was found that neither structural model entirely reproduces the 2D IR spectrum of the monomer; the experimental monomer spectra show more disorder than the compact monomer model, but less than the extended monomer model. Disorder has been identified as a key variable to describe the monomer ensemble, which suggests that the fly-casting dimerization mechanism should be given further consideration.

One key feature of 2D IR is its intrinsically ultrafast time resolution, which is dictated by the dephasing dynamics of the system and typically extends to a few ps for protein amide modes. While time-resolved techniques are widely applied to study protein folding, they are less commonly used to study the conformational dynamics of protein-protein binding. The combination of time and structural resolution has allowed 2D IR to be used as a transient probe in time-resolved experiments on protein and peptide conformational dynamics and folding, and it will be further exploited to study the monomer/dimer features identified in this investigation. We hope that the background work presented here provides the basis for using insulin dimerization to be studied as a model system for coupled folding/binding protein-protein interactions by a variety of complementary approaches.

8.6 Acknowledgements

This work was supported by a grant from the National Science Foundation (CHE-0616575). I would like to acknowledge fellowship support from the Poitras Foundation. I would also like to thank Benjamin Dietzek for valuable discussions and assistance with data acquisition in the early stages of the experiments.

8.7 Afterword

This chapter was excerpted from "Insulin Dimer Dissociation and Unfolding Revealed by Amide I Two-Dimensional Infrared Spectroscopy," by Ganim et al.[306] I would like to thank Kevin Jones for assistance with data acquisition and stimulating discussions.

Chapter 9

Insulin Monomer and Dimer Equilibrium Fluorescence Experiments

9.1 Motivation

It has been demonstrated in previous chapters that two-dimensional infrared spectra can be analyzed to reveal atomistic information about secondary structure contacts. However, amide I 2D IR spectra are mostly sensitive to secondary structure, and secondary structure is only one of the dimerization coordinates identified in Section 7.3. One key complement would be to measure the interprotein separation without sensitivity to specific contact formation. Because the only intermolecular coordinate identified in the amide I 2D IR spectra in Chapter 8 was the anti-parallel β-sheet contacts, a non-specifically associated state lacking these contacts cannot be discriminated against relative to monomers without additional tools. In this case, a contact is defined by the coupling between amide I oscillators, which changes rapidly with sub-Ångstrom displacements, and drops to nearly zero for distances greater than 4 Å.

Optical fluorophores can possess large transition dipoles that allow Förster resonance energy transfer (FRET) to take place in the 10-100 Å range. A FRET experiment will be most sensitive to distance changes around the Förster radius, R_0, which must be considered based on the desired structural sensitivity. This experiment would require a mixture of proteins labeled with donors and acceptors, which will invariably produce homodimers of donor-labeled and acceptor-labeled proteins that cannot undergo FRET. In addition to the synthetic challenge, it is unknown whether fluorescent labeling will alter the biophysics of association for a small protein such as insulin, and distance ambiguity will always be a concern due to the size of the dye and linker. Despite these challenges, the addition of long-range sensitivity is a valuable complement to 2D IR that is worth pursuing.

Fluorophores with a small Stokes shift are capable of self-quenching. This experiment would eliminate the problem of creating donor-donor and acceptor-acceptor pairs. This chapter describes two types of fluorescence experiments that were used to study the equilibrium between insulin monomers and dimers in preparation for transient temperature-jump experiments. The first approach used insulin labeled with Texas Red fluorophores, and the second approach used intrinsic tyrosine fluorescence of insulin, both of which undergo increased fluorescence quenching in the dimer state.

9.2 Dye-Labeled Insulin Fluorescence

Previous insulin-labeling experiments have used amine-reactive chemistry or site-specific mutagenesis. Insulin self-association was measured with a FRET experiment using 2-Amino benzoyl and nitrotyrosyl groups as the donor/acceptor pair, which were attached at the Lys B29.[307,308] A Trp B30 mutant of insulin has also been synthesized, which was used to measure insulin adsorption onto Teflon particles.[309] Winter et al. synthesized Alexa 488-labeled insulin using an amine-reactive succimidyl ester form of the dye to measure the solvent-dependent rotational correlation time.[267]

For this work, amine-reactive dyes were used to synthesize Texas red-labeled insulin, which is schematically shown in Fig 9-1. In this reaction, the amine-reactive N-Hydroxysuccinimide ester (NHS) dye acts as a stable leaving group to allow the formation of a strong amide bond. Texas Red is especially suited for looking at self-association because it is an excellent self-quencher.[310] Ester hydrolysis is a significant, acid-catalyzed competing reaction that degrades the NHS-dye. Hydrolysis can be minimized in basic conditions, as long as the pH is low enough to produce primary amines on the protein capable of nucleophilic attack.

9.2.1 Synthesis, Purification, and Spectroscopic Methods

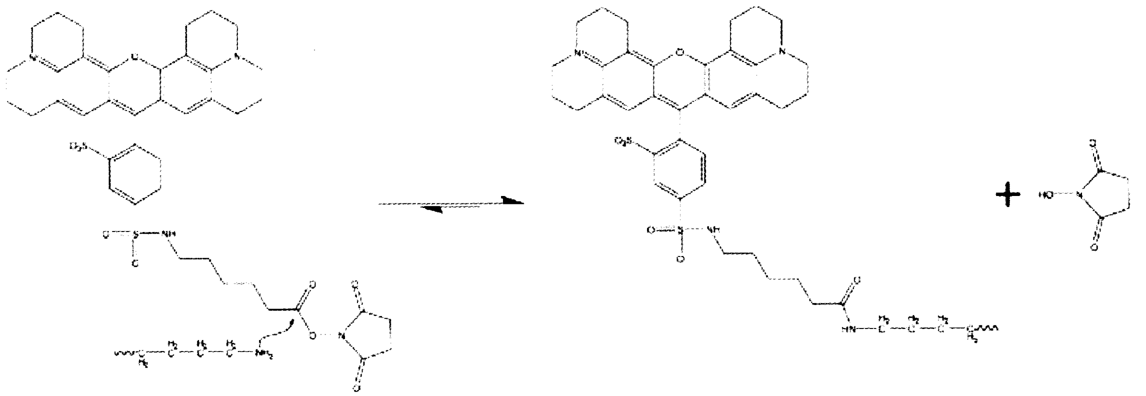


Figure 9-1: *N*-hydroxysuccinimide ester Texas Red reacts with primary amines to form a very stable amide bond.

Insulin was labeled with Texas Red (TR) using procedures adapted from previously published methods.[311] Mixed isomers of Texas Red-X, succinimidyl ester, 6-(tetramethylrhodamine-5-(and-6)-carboxamido)hexanoic acid, succinimidyl ester (Tetramethylrhodamine), and malachite green isothiocyanate were obtained from Molecular Probes (Eugene, OR). The dye-labeling reaction was carried out with 9 mg of insulin per synthesis, dissolved to a concentration of 1 mg/mL in pH 8.1 phosphate buffer at 4 °C. Only Millipore filtered water was used throughout the synthesis. Anhydrous DMSO (80 μL) was added to 1 mg of Texas Red, which was introduced to the insulin solution slowly while stirring. The labeling reaction was allowed to proceed for at least one hour, but could be run overnight without degrading the product.

The reaction mixture was purified using size exclusion chromatography to separate unreacted TR from unlabeled and Texas Red-labeled insulin (TR-insulin). A 19 mm inner diameter x 45.7 cm glass column was loaded with 20 g of hydrated Sephadex G-25 beads (medium size, 50-150 μm when dry), allowed to settle, and equilibrated with two bed volumes of water. Sephadex is a gel filtration medium composed of cross-linked dextran, in which larger biomolecules are adsorbed less and elute quicker. The eluent was allowed to drop to a thin ($<$ 1mm) layer above the beads and the reaction mixture was gently added to the column to maintain a flat band. Care was taken to avoid allowing the beads to dry, which produces channels in the column and diminishes its separation ability. When the reaction mixture had settled into the column, 2 mL of water was added and allowed to settle. At this point, the reaction mixture had begun separating and a reservoir of water could be added to the top of the column without diluting the sample. Insulin and TR-insulin traveled together in a dark purple band that eluted first. The remaining, hydrolysed dye eluted in a slower, ligher pink band. The two fractions were easily distinguishable by eye, as seen in Fig 9-2. The dye-labeled protein band was collected ($\approx$ 15 mL) and lyophilized.

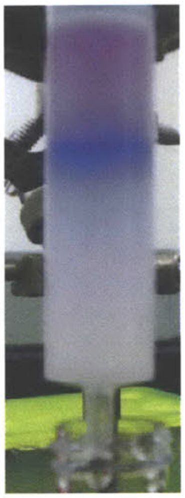


Figure 9-2: A digital camera photograph of the Texas Red-labeled insulin (dark, purple) and free dye bands (light, pink).

For amine-reactive NHS-ester dye labeling, the pH must be chosen to compromise between the rate of acid-catalyzed ester hydrolysis and the concentration of primary amines available for nucleophilic attack on the dye. Figure 9-3 shows the location of primary amines in insulin, which include a lysine residue and an N terminus for each of the A and B chains. Of these three, the N terminus of the B chain is the farthest from the dimer interface, and thus is the desired target for dye labeling. Table 9.2.1 shows the percentage of primary amines (assuming nominal pKa values) and the half life of the NHS ester. A pH of 8.1 was found to maximize the yield of dye-labeled insulin, which also has a >100x molar excess of primary amines from the N termini relative to neutral lysine residues. Based on this reasoning, it is hypothesized that the dye labeling occurs at either Gly A1 or Phe B1, the latter of which appears to be more solvent-accessible based on inspection of the crystal structure.

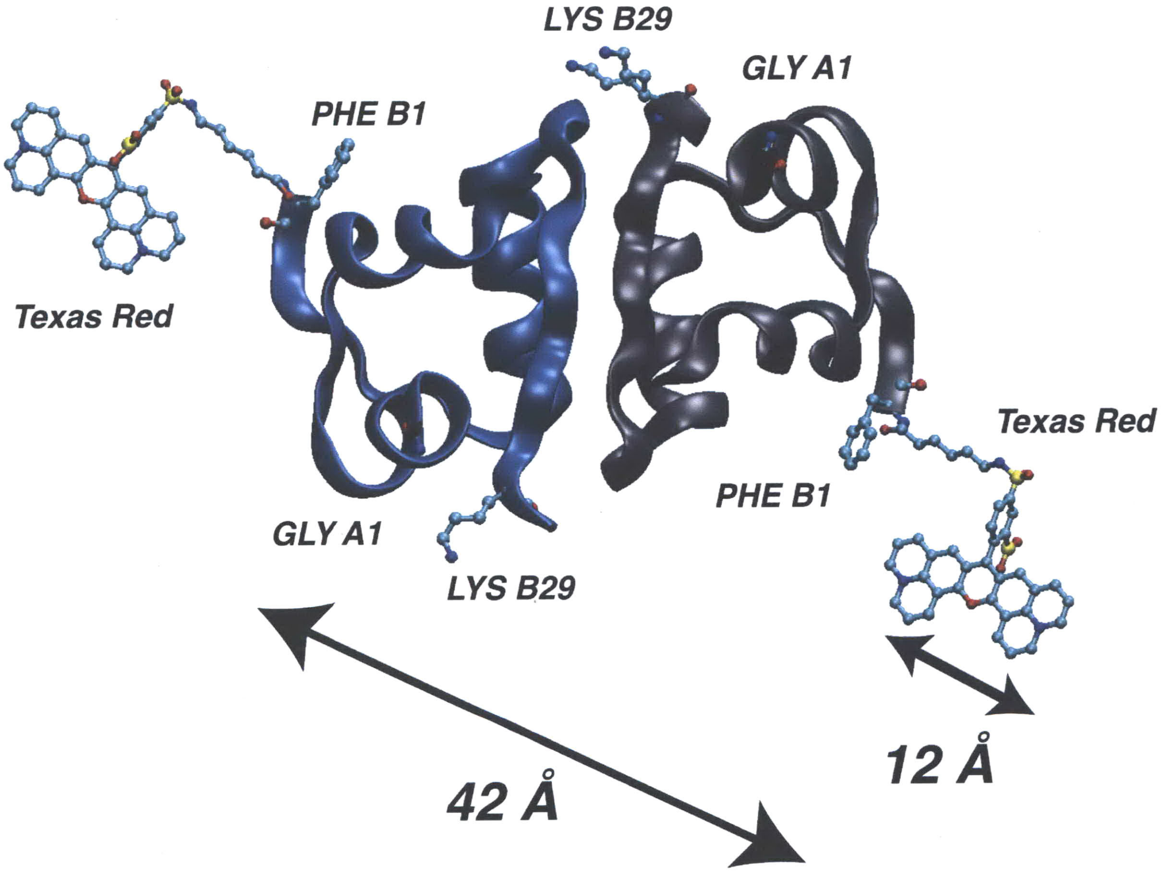


Figure 9-3: Diagram of the primary amines in the insulin dimer crystal structure,[16] with Texas Red conjugated to one of the N-termini, drawn to scale.

The fluorescence spectra were collected using a simple, home-built fluorimeter. An unfocused, pulsed Nd:YAG laser (532 nm, 5 kHz) was used for excitation. The sample was held in a 1 cm path length, quartz cuvette. Fluorescence was collected

pH	**Neutral N Terminus**	**Neutral Lys**	**NHS-ester Half Life**
6	0.63%	0%	
7	5.9%	0.03%	4-5 h
7.4	14%	0.07%	
8	39%	0.29%	1 hr
8.6	72%	1.1%	10 min
9	86%	2.8%	

Table 9.1: Availability of Primary Amines in Insulin as a Function of pH, and half lives for NHS-ester hydrolysis.[18–20]. N terminus pKa was taken to be 8.2 and Lys pKa was taken to be 10.5.

at 90° relative to the excitation; time-resolved fluorescence was filtered with a 550 nm longpass plastic filter, collected on an Electro-Optics Technology ET 2030A 1.2 GHz amplified silicon photodetector, and acquired with a Tektronics DPO3054 oscilloscope. Fluorescence spectra were acquired with the filter using an Ocean Optics S2000 spectrometer coupled to a personal computer.

9.2.2 Characterization of Labeled Insulin

The most straightforward method for characterizing the labeling reaction is using UV-vis absorption spectroscopy. It is assumed that dye conjugation only weakly perturbs the optical properties of the protein and dye, and thus the optical density allows for the protein and dye concentration to be independently determined. Figure 9-4 shows the absorption and fluorescence spectra of all of the reactants and products in the labeling reactions attempted. Unlabeled insulin has no absorption in the visible range, which provides a clean baseline to measure dye absorption. The measured Texas Red absorption spectrum peaks at 589 nm, and undergoes a slight redshift to 592 nm upon conjugation to insulin. Its fluorescence spectrum peaks at 613 nm and is nearly unaffected by the labeling.

Tetramethylrhodamine (TAMRA) and Malachite Green (MG) conjugated insulin were also synthesized and purified using the methods described for Texas Red. (Malachite Green isothiocyanate reacts with amines to form thiourea, but identical procedures were followed.) TAMRA-insulin is capable of serving as a FRET donor for

TR-insulin acceptors. TR-insulin can also serve as a FRET donor when coupled with MG-insulin as an acceptor. MG-insulin offers an additional benefit of almost entirely nonradiative relaxation, which precludes the need to filter out acceptor fluorescence when measuring donor fluorescence quenching. The absorption of free TAMRA peaks at 552 nm with a peak width standard deviation of 16 nm that redshifts by $\approx$ 10 nm and broadens to 44 nm upon conjugation. This is also reflected in the fluorescence spectrum, whose peak shifts from 578 nm to 585 nm. Malachite Green displays the opposite trend- its absorption peak blueshifts from 642 to 612 nm and becomes narrower. The yield and labeling efficiency for these reactions were substantially lower than TR-ins, which contributed to failure of the FRET experiments. A further discussion is provided in 9.2.5

To characterize the reaction yield, the eluate from the column was collected in 0.5 mL aliquots and UV-vis spectra were collected. Using the molar absorptivity of Texas Red (ϵ(593 nm)=80 mM^{-1} cm^{-1}[311]), the dye concentration was determined, and the dye contribution to the optical density at 280 nm was subtracted (using $\epsilon(280nm)/\epsilon(593nm) = 0.18$ for Texas Red[311]) to determine the protein concentration (ϵ(280 nm)=5.53 mM^{-1} cm^{-1} for insulin[312]). Figure 9-5 shows the results of a clean separation between the insulin and the free dye, as gauged by a drop in absorption at 593 nm to baseline in between the bands at 8 mL and >20 mL. (The additional UV absorption appearing at 17.5 mL may be due to the NHS group, or denatured insulin fragments.) After the sample was lyophilized, the labeling efficiency (concentration dye/concentration protein) was calculated using the same algorithm, and was typically $\approx$ 0.3.

Another characterization of the Texas Red-insulin was using liquid chromatography-mass spectrometry (LC-MS), which was run by the MIT Biopolymers Facility and appears in Appendix B. The most significant protein-containing fractions were fraction 8 (Figure B-3, and expanded in Figure B-4), fraction 9 (Figure B-5), and fraction 10 (Figure B-6). Fractions 8 and 9 contained a mixture of insulin and insulin-TR. Fraction 10, which was much smaller than fractions 8 and 9, was purely insulin-TR, but only measured $\approx$0.1 mg when lyophilized, and did not

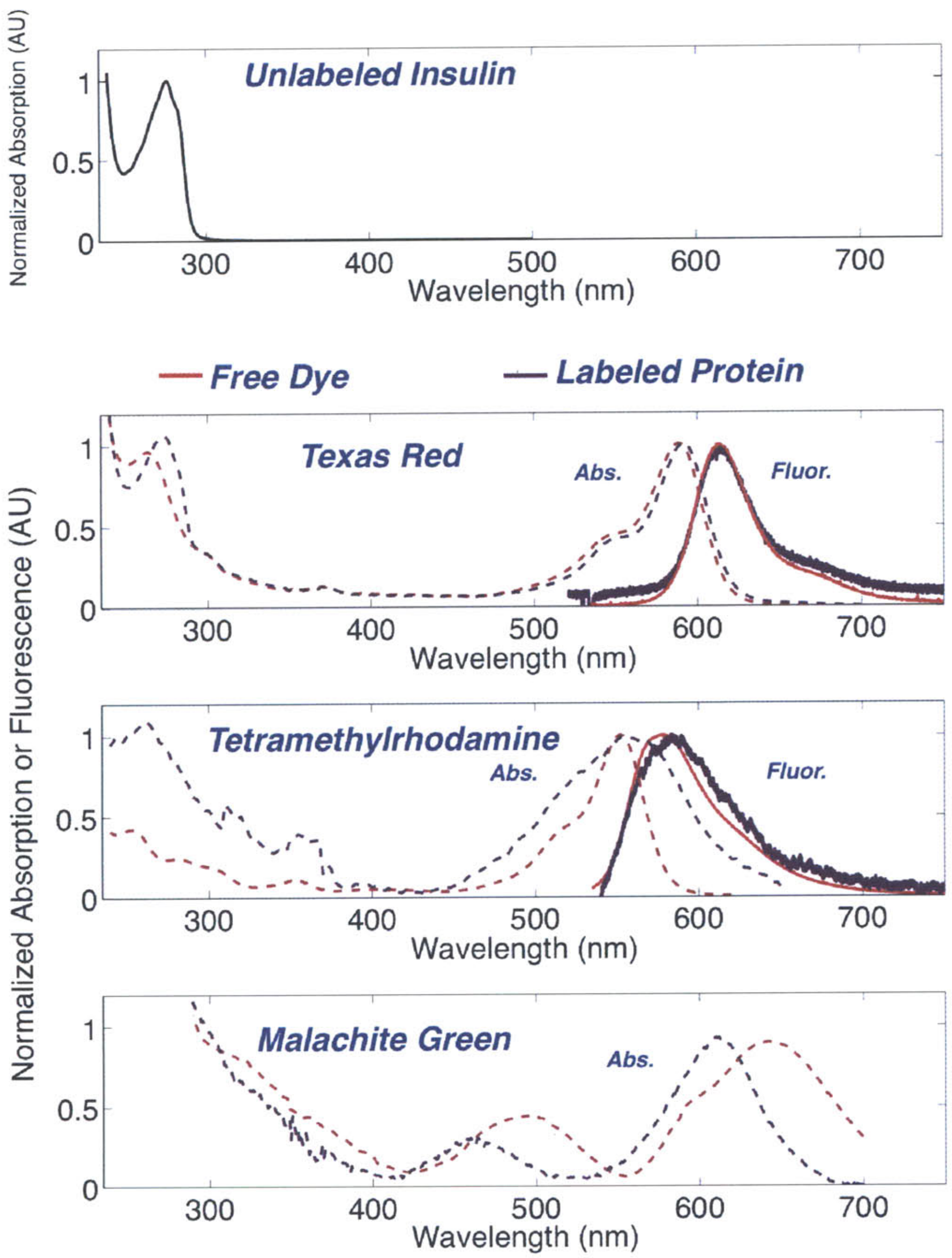


Figure 9-4: Insulin, Free Dye, and Dye-Labeled Insulin Absorption and Fluorescence Spectra. Normalized absorption and fluorescence spectra for all NHS dyes used for labeling. Malachite green is a nonfluorescent dye. All solutes were optically dilute (OD < 0.5) in water. Fluorescence spectra were collected with excitation at 532 nm.

constitute a sufficient amount for any experiments. Fractions 13 (Figure B-7) and fraction 14 (Figure B-8) had masses consistent with a mixture of single, double, and triple-labeled insulin, but the yields were much smaller than fractions 8 and 9. Due

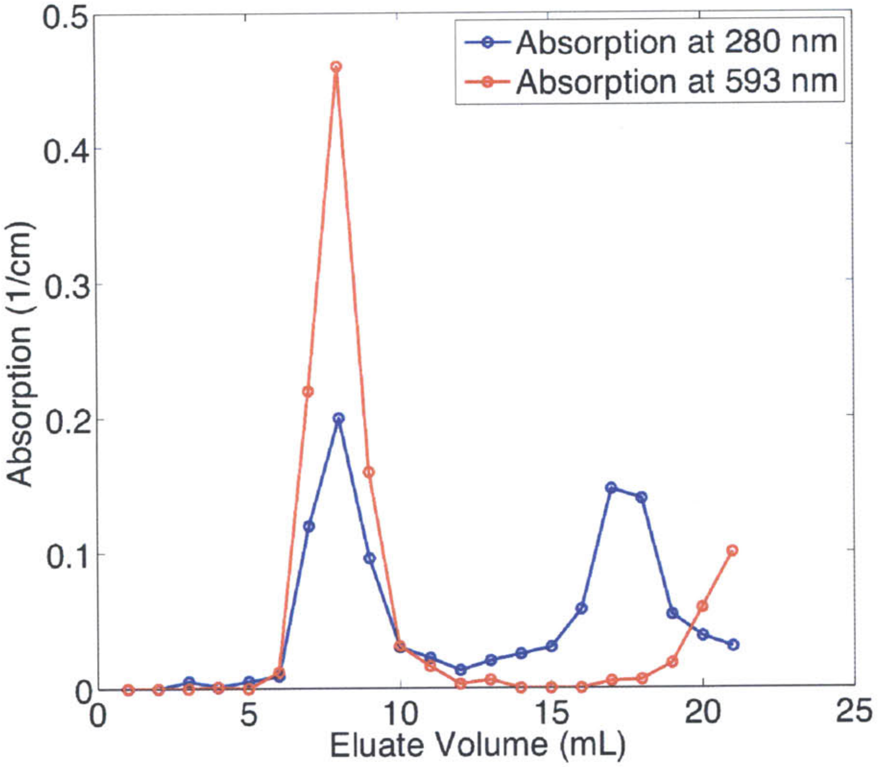


Figure 9-5: Characterizing aliquots from the column-separated TR-ins synthesis using UV and visible absorption. The absorption at 593 nm arises from Texas Red chromophores, which may be covalently bound to insulin. The absorption at 280 nm arises from insulin, Texas Red, and possibly also the NHS leaving group.

to the volume of sample loaded on the column, the Biopolymers Facility indicated that the UV-vis data could not be used for quantitative determination of the amount of protein and dye.

9.2.3 Measuring the Dissociation Constant, K_D

To use the Texas Red-labeled insulin to measure the dissociation constant in a self-quenching experiment, there must be a difference in the quenching rate for monomers and dimers. For example, if FRET was the mechanism for energy transfer, it would be ideal for the interdye separation, R to be $\ll R_0$, the Förster radius in the dimer state and $R \gg R_0$ in the monomer state.

A series of concentration-dependent time-resolved fluorescence emission traces are shown in Figure 9-6. (The data were Fourier filtered to remove high frequency noise due to channel crosstalk in the oscilloscope.) The solvent was 0.27M HCl:H_2O. At the lowest concentrations, the data has a simple exponential form. As the concentration was increased, an additional, faster timescale was observed, which was potentially due to dimer formation-induced quenching. The data set was globally fit to a sum of two exponentials with concentration-dependent amplitudes and fixed timescales,

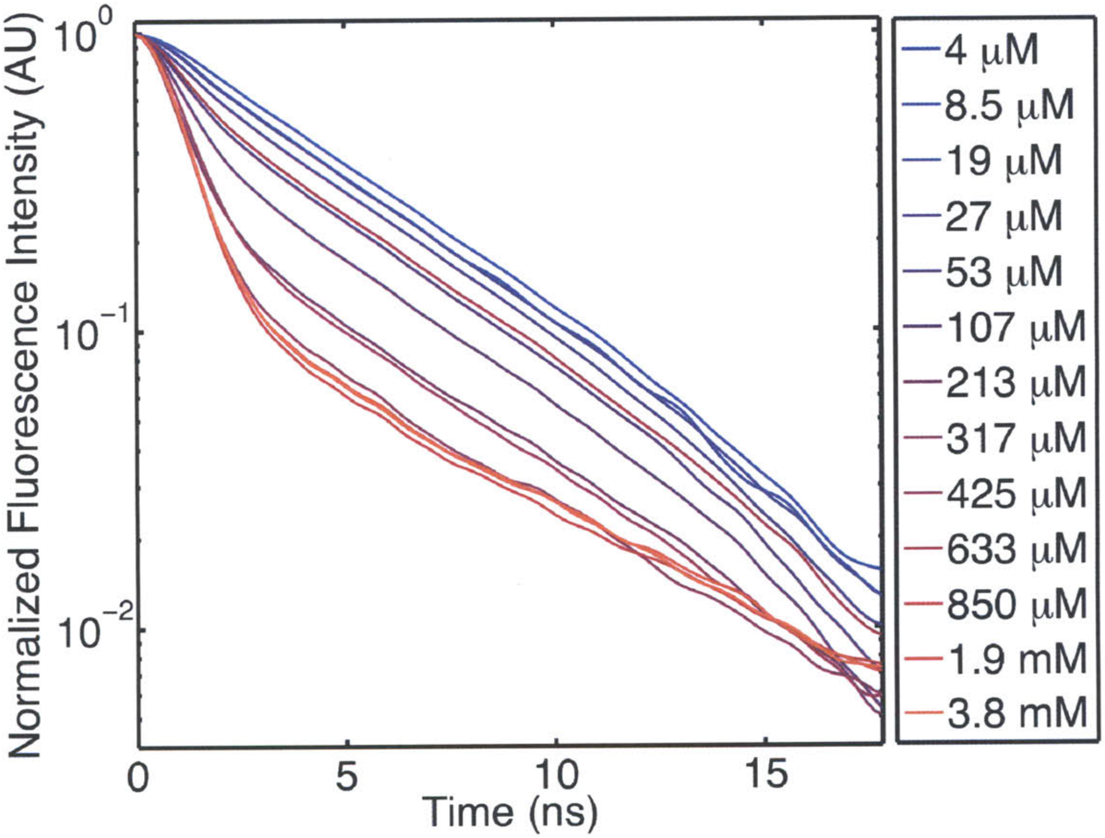


Figure 9-6: Concentration-dependent fluorescence decays of Texas Red-Labeled Insulin.

$$I(t,c) = A_1(c)e^{-t/\tau_1} + A_2(c)e^{-t/\tau_2}. \tag{9.1}$$

The slower timescale, 5.0 ns was assigned to the lifetime of Texas Red in the monomer state, and the faster timescale, 1.4 ns, was assigned to dye quenching, potentially due to insulin dimer formation.

Figure 9-7 shows a plot of the fast amplitude fraction,

$$\theta_A(c) = A_1(c) / \left(A_1(c) + A_2(c)\right). \tag{9.2}$$

For a simple bimolecular equilibrium with total protein concentration c, and concetrations c_M and c_D of monomers and dimers,

$$\begin{aligned} D &\rightleftharpoons 2M & (9.3)\\ 2c_D + c_M &= c, & (9.4) \end{aligned}$$

it would be expected that the dimer fraction, $\theta_D(c)$ provides a measure of the dissociation constant, K_D,

$$\begin{aligned} K_D &= \frac{c_M^2}{c_D} & (9.5)\\ \theta_D(c) &= \frac{2c_D}{2c_D + c_M} & (9.6)\\ &= \frac{4c + K_D - \sqrt{K_D^2 + 8K_D c}}{4c}. & (9.7) \end{aligned}$$

However, the experimentally observed transition is a sharper function of concentration than this form predicts. This may arise if the stoichiometry for dye quenching was different than the stoichiometry for dimer formation; in the nominal, biomolecular association equilibrium, when two insulin monomers associate, two TR dyes would quench. A sharper transition would occur if when two insulin monomers associate, more than two TR dyes quench. A steeper transition such as the one observed can arise if the proteins are multiply labeled, although this scenario is inconsistent with the labeling efficiency. Assuming the statistics of rare events, a Poisson distribution would predict that if the average labeling efficiency is 0.30, the probability of one dye to conjugate is 0.22, and 0.04 for more than one dye to conjugate.

Despite this rationale, there is evidence that this concentration-dependent experiment is sensitive to the monomer-dimer equilibrium. The baselines at high and low

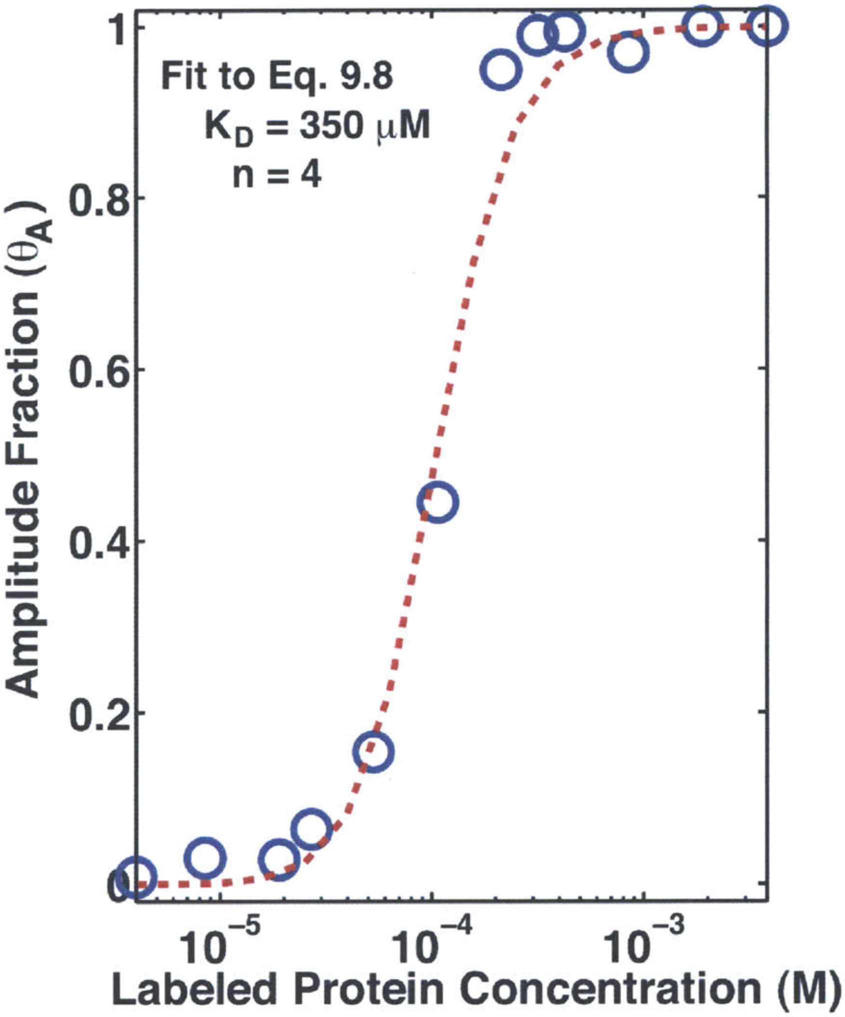


Figure 9-7: Extracting the Dimer Fraction from Fluorescence Decay Amplitudes.

concentrations as well as the midpoint of the transition are all consistent with the expectations of insulin's monomer-dimer equilibrium. At low pH, the equilibrium constant for tetramer formation is expected to be ≈ 100x greater than for dimer formation[265], which would be outside this concentration range. Quenching due to any higher order aggregate formation would result in an even *steeper* slope. The average intermolecular separation is 255 Å at 100 μM, where changes to the lifetime are very clear. Even at the highest concentration (3.8 mM), the average intermolecular separation is 76 Å, which argues against through-space quenching of unassociated proteins across the concentration range. The Förster distance for resonant energy transfer between Texas Red molecules is 58 Å, and the predicted distance between conjugation points in the insulin dimer is 30-42 Å.

The data was fit to the form of the Hill Equation, which describes the binding of multiple ligands, and allows tunability in the midpoint (K_D) and slope (n) of the dissociation curve,

$$\theta_A(c) = \frac{c^n}{K_D^n + c^n}, \tag{9.8}$$

and the best fit was obtained for $n = 4$ and $K_D = 350\mu M$. If it can be assumed that $\theta_A(c) \approx \theta_D(c)$, then this is a measurement of the dissociation constant for insulin. This value can be compared to K_D obtained from 2D IR spectroscopy in Chapter 8 of 70 μM, and previously published values that range from 25 μM to 261 μM.[262, 265, 266, 270, 271, 305]

9.2.4 Thermal Dimer Dissociation

Additional data favoring the interpretation that the quenching assay monitors the monomer-dimer equilibrium of insulin comes from temperature-dependent experiments. Figure 9-8 shows the temperature-dependent fluorescence spectra of TR-insulin. The buffer pH has a dramatic effect on the equilibrium constant. Three insulin dimers can assemble to form a hexamer, which is inhibited at pH ≤ 3.3, because the hexamerization interface contains titratable histidine residues. The series of fluorescence spectra acquired at pH 8.1 show a 15% change across the temperature range, and appear to be a linear function of the temperature. In 0.27 M HCl (pH < 1), 80% changes are observed, with a sigmoidal form; there is a high temperature baseline, an identifiable midpoint (42 °C), and low temperature changes indicative of cold dissociation, which are all consistent with the 2D IR experiments presented in Chapter 8. The changes at pH 8.1 may reflect thermally induced changes to the fluorophore properties without changes to the monomer-dimer equilibrium, or dissociation to a much lesser degree than at low pH.

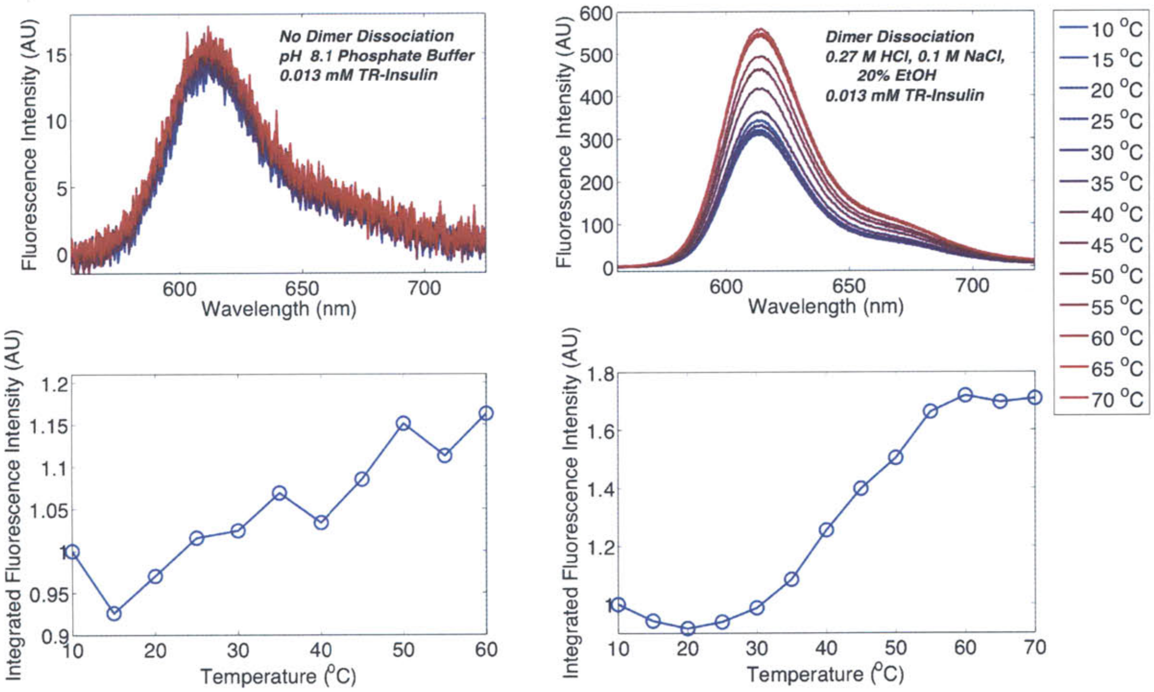


Figure 9-8: Temperature-dependent fluorescence spectra of TR-insulin. Dimer dissociation is observed in acidic conditions as an increase in fluorescence due to the loss of quenching. At pH 8.1, where dissociation is strongly inhibited, a much smaller increase is seen.

9.2.5 Conclusions

In this section, the synthesis and characterization of fluorescent, Texas Red-labeled insulin was described. As a bright dye that efficiently quenches, Texas Red provides sensitivity to the monomer-dimer equilibrium, which was observed using concentration-dependent and temperature-dependent experiments. This fluorescence assay is a great complement to the infrared experiments; Texas Red can be observed at the single molecule level,[313] and it was straightforward to observe the fluorescence from 4 μM TR-insulin, while the lowest concentration probed by the 2D IR experiments was 88 μM. Moreover, fluorescence quenching is expected to only be sensitive to the distance and orientation between fluorophores, and not to any secondary structure. In principle, this fluorescence assay will allow for the discrimination between disordered insulin dimers and insulin monomers. Both of these states would appear similar in

2D IR experiments due to the loss of β-sheet structure, but would display different fluorescence intensities and lifetimes.

The main limitation of this fluorescence assay is the labeling efficiency (≈ 0.3). Unlabeled insulin proteins contribute to aggregation, scattering, and generate fluorescence background, because Texas Red in heterodimers of unlabeled and labeled insulin would be unquenched and display the same optical properties as monomeric insulin. To model this process, consider the conservation equations and laws of mass action, where five species are in equilibrium: labeled monomers (c_M^L), unlabeled monomers (c_M^{UL}), labeled homodimers (c_D^{L-L}), unlabeled homodimers (${c_D}^{UL-UL}$), and heterodimers (c_D^{L-UL}) with labeling efficiency γ assuming that dye labeling does not change K_D,

$$c = c_M^{UL} + c_M^L + 2c_D^{L-L} + 2c_D^{L-UL} + 2c_D^{UL-UL} = c_M + 2c_D \tag{9.9}$$

$$c_M = c_M^L + c_M^{UL} \tag{9.10}$$

$$c_D = c_D^L + c_D^{UL-UL} + c_D^{L-UL} \tag{9.11}$$

$$K_D = \frac{\left(c_M^{UL} + c_M^L\right)^2}{c_D^{L-L} + c_D^{L-UL} + c_D^{UL-UL}} = \frac{c_M^2}{c_D} \tag{9.12}$$

$$\tag{9.13}$$

If two signals can be measured- one proportional to the population of quenched, labeled proteins (I_Q) and one proportional to the concentration of unquenched, labeled proteins (I_{UQ}), these signals will report,

$$I_Q \propto 2c_D^{L-L} = 2\gamma^2 c_D \tag{9.14}$$

$$I_{UQ} \propto c_D^{L-UL} + c_M^L = 2\gamma\left(1-\gamma\right)c_D + \gamma c_M \tag{9.15}$$

Because the proportionality constant is unknown, the ratio is measured, which is proportional to the dimer fraction defined in Eq. (9.6)

$$\begin{aligned}\frac{I_Q}{I_Q+I_{UQ}} &= \frac{2\gamma^2 c_D}{2\gamma^2 c_D + 2\gamma\,(1-\gamma)\,c_D + \gamma c_M} && (9.16)\\ &= \gamma\frac{2c_D}{2c_D+c_M} && (9.17)\\ &= \gamma\,\theta_D, && (9.18)\end{aligned}$$

Thus, this curve always has the same functional form as the dimer fraction, and the difference signal scales with the labeling efficiency, as Fig 9-9 shows.

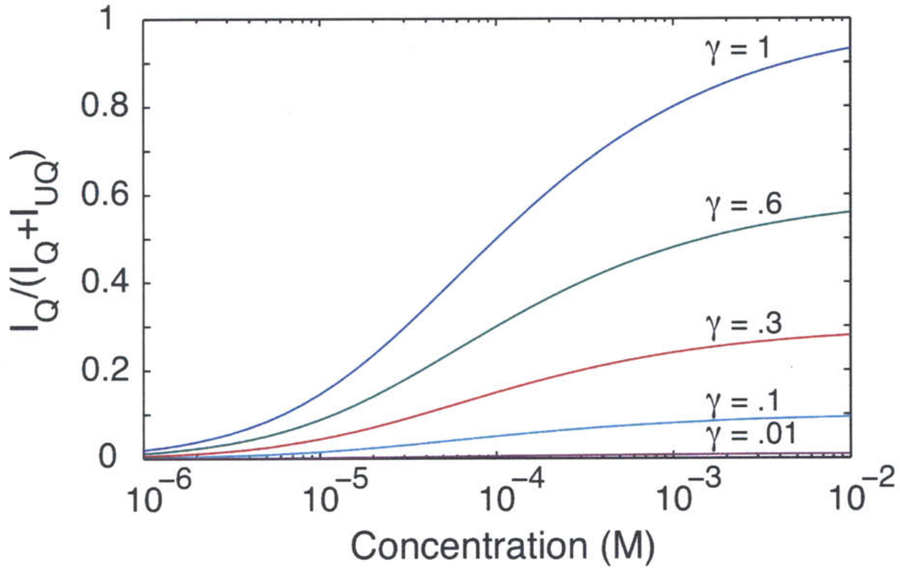


Figure 9-9: Sensitivity to the dimer fraction as a function of concentration and labeling efficiency for $K_D = 100\ \mu M$.

In this section, the synthesis of MG-insulin and TAMRA-insulin were described as potential FRET partners for TR-insulin. An assay of the monomer-dimer equilibrium based on FRET would offer significant advantages over the self-quenching assay described here. The mechanism and distance sensitivity of FRET are well-understood, but self-quenching may occur through-space or when the electronic orbitals of fluorophores overlap to create non-radiative relaxation pathways. A FRET anisotropy experiment would allow for the orientation between the donor and (necessarily fluorescent) acceptor to be measured by controlling the polarization of the excitation relative to the analyzing polarizer for acceptor fluorescence; in a self-quenching experiment, the polarization of the emitted fluorescence only reflects reorientation of the initially excited fluorophore.

Poor labeling efficiency makes FRET experiments nearly impossible because the changes are relative to the *product* of the labeling efficiencies for the donor and acceptor. Furthermore, if the donor undergoes self-quenching, the analysis will be complicated by the several relaxation pathways possible. A molar excess of acceptor-labeled proteins may alleviate the latter problem. For the experiments attempted, an analysis of the labeling efficiencies (≈ 0.3 and ≈ 0.1) predicted that the changes across the concentration range would be $< 1\%$, and were not detectable. This experiment should be attempted again if the labeling efficiencies can be improved.

In any experiment using fluorescent-labeled protein, careful controls must be done to understand the biophysical implications of fluorophore attachment. Consistency in the 2D IR spectra would provide good evidence that the secondary structure was unaffected. However, this diminishes from one of the main advantages in doing fluorescence experiments because infrared spectroscopy still requires higher protein concentrations. Because fluorophores subject to electronic excitation potentially undergo photodamage, fluorescence experiments performed at infrared spectroscopy concentrations would require *more* sample than infrared experiments. In the next section, a solution to these problems is presented by using intrinsic tyrosine fluorescence of insulin to assay the monomer-dimer equilibrium.

9.3 Intrinsic Insulin Fluorescence

9.3.1 Background

An alternative fluorescence experiment for probing the monomer-dimer equilibrium of insulin uses its intrinsic tyrosine fluorescence. Figure 9-10 shows the tyrosine and phenylalanine residues in the insulin dimer. Since insulin has no tryptophan residues, its emission is dominated by tyrosine, which has a larger extinction coefficient (see

Fig. 9-11) and fluorescence quantum yield than phenylalanine (0.14 vs. 0.03[17]). It was demonstrated that tyrosine fluorescence was sensitive to insulin fibrilization state[314] and structural changes due to mutation.[315] Based on the structure, one would expect tyrosine fluorescence to be sensitive to the association state because the cluster of tyrosine and phenylalanine residues at the interface are tightly packed.

Tyrosine fluorescence is ≈ 400 fainter than Texas Red due to its smaller extinction coefficient and fluorescence quantum yield. However, this is naturally mitigated by the goal of directly comparing timescales extracted from temperature-jump fluorescence and 2D IR experiments, both of which require ≈ mM concentrations. For performing such a comparison, the most important advantage of tyrosine fluorescence over TR is that the protein is unmodified, and is unquestionably identical to the one probed by 2D IR. A significant problem that arises is the potential for UV photodamage, which necessitates low excitation powers, continual sample replenishment, and control experiments.

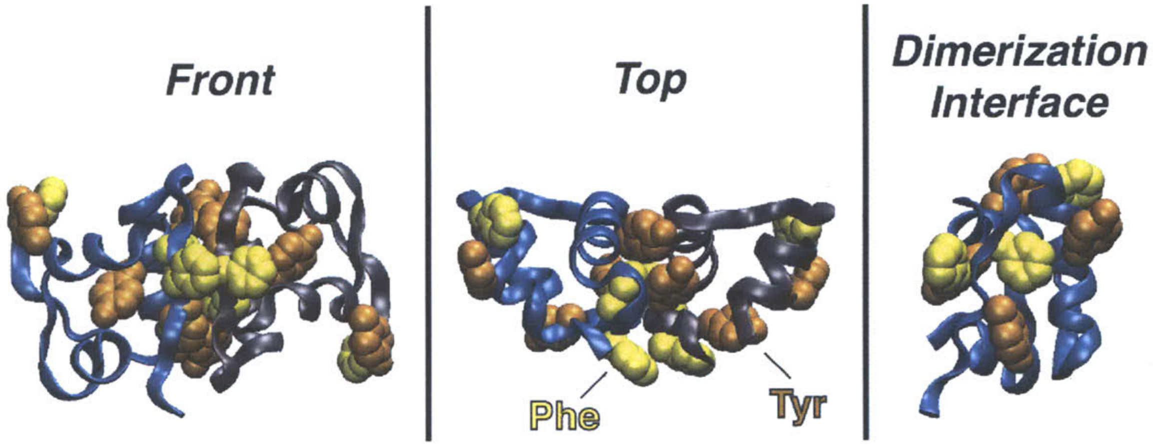


Figure 9-10: Diagram of the six pheynalanine and eight tyrosine residues in the insulin dimer crystal structure.[16]

9.3.2 Experimental Setup

Tyrosine fluorescence is excited with 267 nm, ≈ 100 fs pulses. Figure 9-12 shows the experimental setup for tripling 800 nm light to generate 267 nm. The 800 nm (300

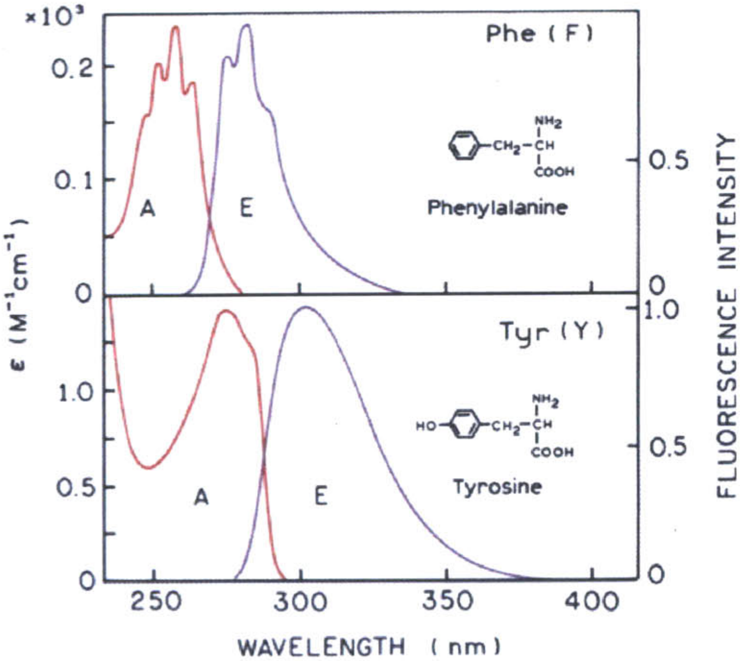


Figure 9-11: Absorption and emission spectra of phenylalanine and tyrosine. Adapted from Lakowicz et al.[17]

mW, 1 kHz, 100 fs) is focused into 0.7 mm thick BBO (θ=29.3°) for type I doubling. The 800 nm and 400 nm are retimed by 2 mm thick BBO crystal (type 1, θ=66°) and rotated into the same polarization by a quartz dual waveplate. The sum frequency at 267 nm is generated by focusing into 0.3 mm thick BBO (type I, θ=44.3°). A prism is used to spatially isolate the 267 nm. Since the fluorescence lifetime of tyrosine is $\approx$ 3 ns, no second prism was added for compression. The power at the sample point was 0.8 μW (with 300 mW input 800 nm) - 4 μ W (with 460 mW input 800 nm). The sample is held in between two 1 mm thick CaF_2, ϕ=1" windows with a 50 or 100 μm Teflon spacer mounted in a brass sample cell, which is temperature controlled with a recirculating water bath. A Schott glass filter is used to reject scattered 400 nm light. The fluorescence is detected using a high speed photodetector, which when sampled using a DPO7104 Tektronics oscilloscope yields an instrument response time of 300 ps. A reflection off the front face of the prism is collected onto a PIN photodetector and used to trigger the oscilloscope.

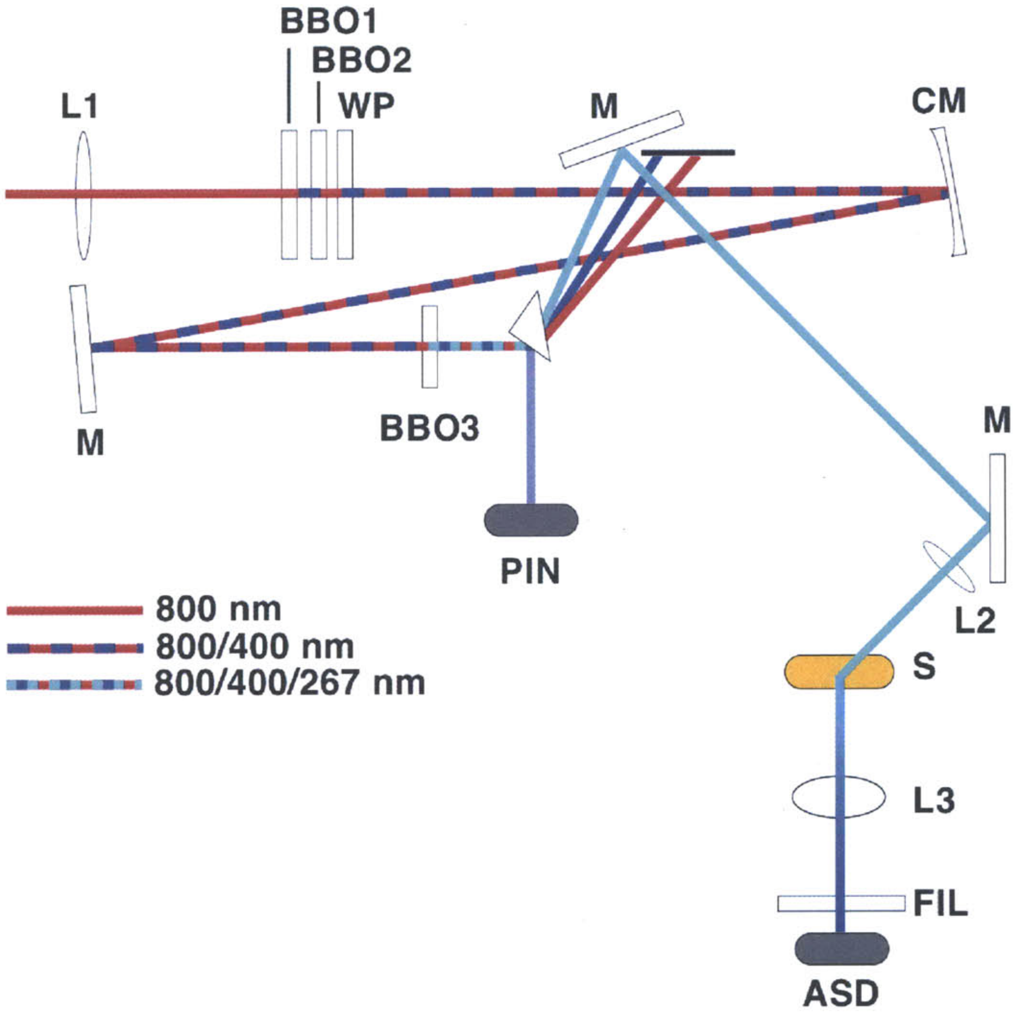


Figure 9-12: Experimental scheme for tyrosine fluorescence experiments. L1 (BK7, f=15 cm), BBO 1-3 (β-BaB_2O_4), WP (quartz waveplate 14 λ at 400 nm, 6.5 λ at 800 nm, 0.584 mm), M (silver mirror), CM (curved aluminum mirror, r=40 cm), PR (fused silica, 1 cm), PIN (Thor Labs DET210 High Speed Silicon Detector), L2 (fused silica f=7.5 cm), L3 (fused silica, f=7.5 cm), S (brass sample cell), FIL (UG 11 Schott Glass Filter), ASD (Electro-Optics Technology ET 2030A 1.2 GHz amplified silicon photodetector)

9.3.3 Results and Discussion for Insulin Fluorescence

Figure 9-13 shows a comparison between free tyrosine fluorescence and intrinsic insulin fluorescence in two solvents. Dilute tyrosine in water has a fluorescence lifetime of 3.1 ns. The emission of insulin predominantly arises from its four tyrosines, but may include a much smaller contribution from its three phenylalanine residues. In 30%

ethanol, where insulin is monomeric, the fluorescence lifetime drops to 2.1 ns. In dimeric insulin, this lifetime further drops to 1.5 ns. These lifetime changes are due to the increasing number of relaxation pathways available upon moving from dilute chromophores, to the monomeric, and dimeric protein.

The lifetime decrease is correlated with an increase in the density of electronic states. Each tyrosine residue has two overlapping electronic transitions (^1L_a, ^1L_b) in the UV. In monomeric insulin, there are four tyrosine and three phenylalanine residues, which all together result in 14 electronic states. The dimer interface has a tight clustering of aromatic amino acids (see Fig. 9-10) where electronic orbital overlap is possible that may create delocalized, excitonic states. Not all of the aromatic residues are in direct contact, but all are within the Förster distance for Tyr-Tyr resonant energy transfer (9-16Å) or Phe-Tyr resonant energy transfer (11.5-13.5Å).[17] The differences in absorption spectra of monomeric and dimeric insulin are on the 0.5% level,[266] and do not provide a straightforward interpretation of the changes in the coupling and electronic structure.

Figure 9-14 shows temperature-dependent insulin fluorescence. With increasing temperature, the fluorescence intensity increases by almost a factor of four, while the lifetime decreases from 1.4 to 1.0 ns. The fluorescence increase is consistent with the dissociation curve characterized using HDVE in Chapter 8. The midpoint for this transition was found to be 30 °C using HDVE, and 33 °C using UV fluorescence (within the ±2°C margin of error for each measurement). Thus, the changes to the fluorescence intensity are correlated with changes to the vibrational spectrum that arise from the loss of vibrational modes delocalized across the dimers. This is strong evidence that the UV fluorescence is sensitive to the monomer-dimer equilibrium.

Based on the results of Fig. 9-13, it would be expected that the fluorescence lifetime should increase in the monomeric state by 0.6 ns relative to the dimeric state. Moreover, conventional wisdom reasons that increased fluorescence intensity can result from a reduction in the number of non-radiative pathways, which would also increase the fluorescence lifetime. However, figure 9-14 shows that temperature causes a decrease in the lifetime. The form of this lifetime change is linear with

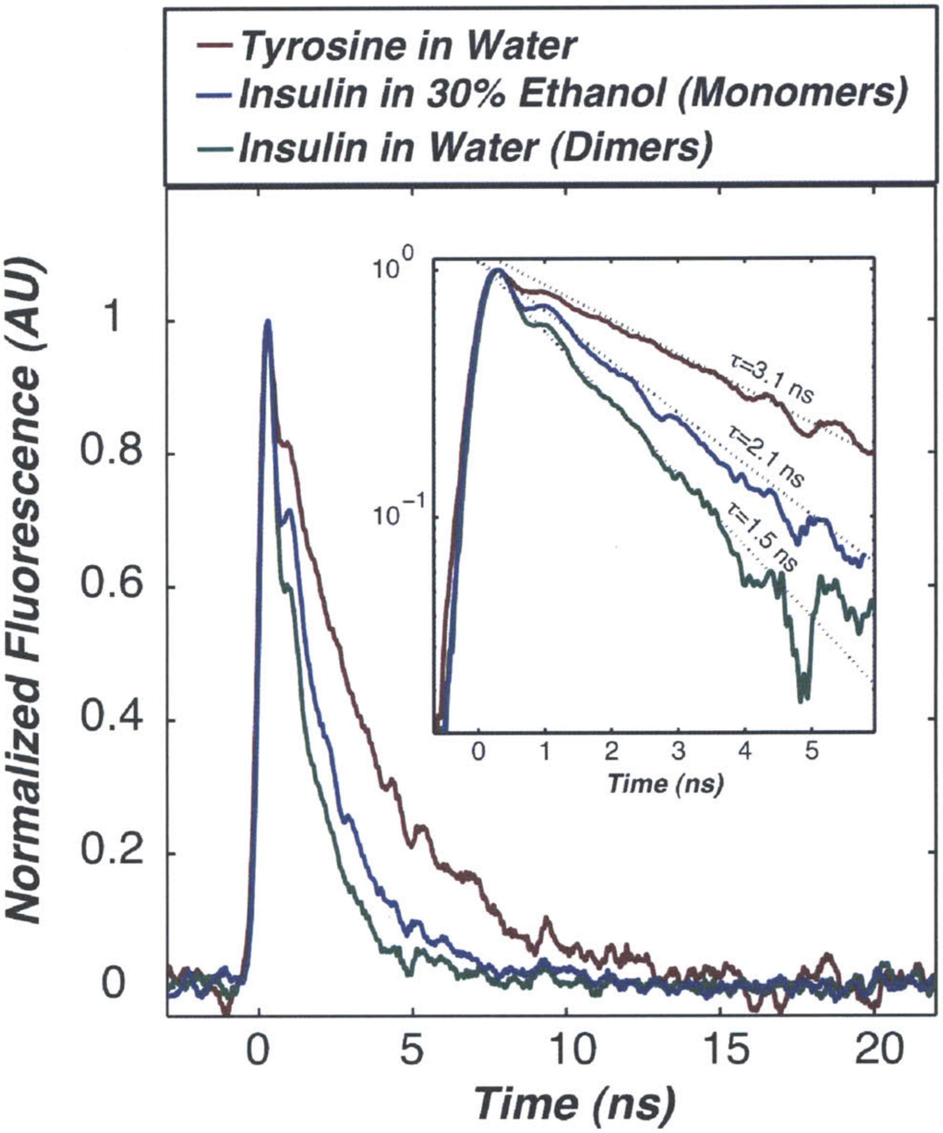


Figure 9-13: Insulin and tyrosine fluorescence.
The free tyrosine concentration was 2.5 mM and the concentration of insulin was 3.4 mM (effectively 14 mM tyrosine).

temperature ($R^2 = 0.97$ for $\tau \propto T$, and $R^2 = 0.98$ for $log\left(\frac{1}{\tau}\right) \propto \frac{1}{T}$), and unlike the sigmoidal shape of the fluorescence intensity or HDVE changes.

The reduction in the fluorescence lifetime of insulin with temperature indicates that it is not controlled by the difference in non-radiative pathways available in the insulin monomer and dimer, but rather an increase in the interconversion rate of tyrosine rotamers. There have been extensive investigations on the fluorescence quenching mechanisms of tyrosine. It is known that tyrosine fluorescence can be quenched by electron transfer to amide groups, and that the quenching rate can vary for different C_α-C_β rotamers.[316–319] It is known that with increasing temperature, the fluorescence quantum yield[320] and lifetime[319] both decrease, which reflects the barrier to exchange between rotamers. An Arrhenius fit to the fluorescence lifetime of insulin yields an activation barrier of 2.3 kcal/mol and a preexponential constant of $(27ps)^{-1}$, which can be compared to 3 kcal/mol and $(31ps)^{-1}$ for N-acetyltyrosineamide.[319] Other electron acceptors can also quench tyrosine, such as the uncharged carboxyl groups available at low pH[321] or molecular oxygen.[322,323]

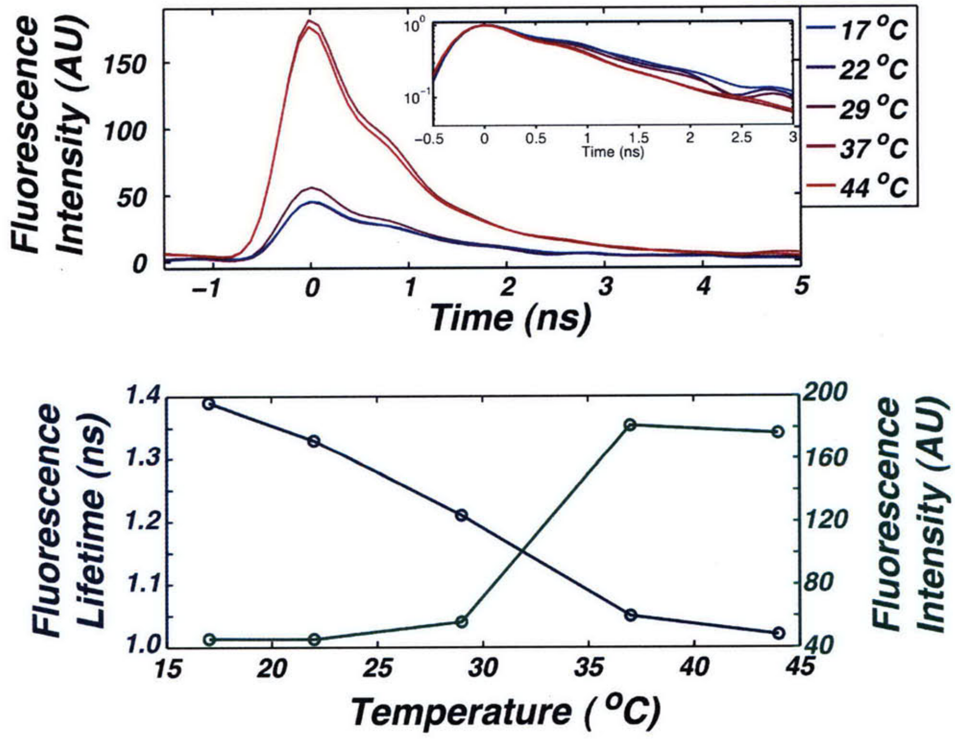


Figure 9-14: Temperature-dependent intrinsic insulin fluorescence. Insulin concentration was 10 mg/ml in 20% EtOD:D_2O, 0.1 M NaCl, 0.27 M DCl.

9.3.4 Conclusions

It has been observed that the intrinsic fluorescence of insulin is sensitive to the monomer-dimer equilibrium. The poor optical properties of tyrosine as a fluorophore are naturally overcome by the concentrations required for comparison to 2D IR experiments. Furthermore, like 2D IR, tyrosine fluorescence is an endogenous protein spectroscopy that requires no sample modification. In comparison to the TR-insulin, the requirement of UV excitation presents practical complications (sample damage), as well as complexity in the data interpretation (photochemistry of tyrosine).

Both TR-insulin fluorescence and intrinsic fluorescence are sensitive to the monomer-dimer equilibrium due to self-quenching. For TR-insulin, it is likely that weak, through-space Förster energy transfer is responsible for the quenching. With tyrosine fluorescence, both through-space and strong coupling due to electronic orbital overlap are possible quenching mechanisms. Furthermore, tyrosine fluorescence has the potential to display intramolecular sensitivity if the quenching can be modulated by monomer unfolding that disrupts the network of aromatic amino acids. For these reasons, it is preferable to probe the monomer-dimer equilibrium using several methods and to develop models that consistently explain changes in the fluorescence and infrared experiments.

9.4 Acknowledgement

I would like to thank Krupa Ramasesha, Chunte Sam Peng, and Katie Barnes for assistance with fluorescence data collection and synthesis.

Appendix A

MATLAB Code for Calculating 2D IR Spectra of Six-Level Systems

```
clear

%%%%1. User Parameters
%%6LS Potential Energy Surface Parameters
%%All in cm^-1
wa=1635;
ws=1680;
dwa=16;
dws=16;
dwas=13;

%%Aliasing frequency (cm^-1)
alias=1550;

%%Polarization
POL='ZZYY'; %%either 'ZZZZ' or 'ZZYY'

%%Transition dipole parameters
ua=1.0;us=.7;theta=70*pi/180;

%%Time axes
dt=10; %time step for t1, t2, t3
t1=0:1:511; %in units of dt
t3=0:1:511; %in units of dt
t2array=[10]; % in units of dt
padlen=2048; %zero padding length

%%Make correlation function parameters visible to other functions
global tau_aa tau_ss tau_as T2aa T2ss T2as dt
global delta_aa delta_ss delta_as delta0_aa delta0_ss corr

%%Correlation function parameters for the generalized Kubo model
%%JR Scmidt, N Sundlass, JL Skinner, CPL 378 (2003) 559-566
%%Modify has, haa, and hss to specify new correlation functions
%%Al times in fs, all deltas in cm^-1
tau_aa=1000;
```

```
tau_ss=500;
tau_as=200;
T2aa=1000;
T2ss=1200;
T2as=900;
delta_aa=15;
delta_ss=7;
delta_as=12;
corr=1; %correlation coefficient between a and s oscillators
delta0_aa=1;
delta0_ss=1;
Dor=0; %orientational diffusion constant 1/fs

%%%%2. Begin Calculations
%%2Q transition dipoles by harmonic scaling
u2a=sqrt(2).*ua;u2s=sqrt(2).*us;
ua_sq=ua.*ua;us_sq=us.*us;
usa=ua;uas=us;
usa_sq=usa.*usa;uas_sq=uas.*uas;
u2a_sq=u2a.*u2a;u2s_sq=u2s.*u2s;

%%Transform from wavenumbers to 1/fs
c=2.99792458e-5;
twopic=2*pi*c;
delta0_aa=delta0_aa*twopic;
delta0_ss=delta0_ss*twopic;
delta_aa=delta_aa*twopic;
delta_ss=delta_ss*twopic;
delta_as=delta_as*twopic;
wa=(wa-alias)*twopic;
ws=(ws-alias)*twopic;
dwa=dwa*twopic;
dws=dws*twopic;
dwas=dwas*twopic;

I=sqrt(-1);
[t3,t2,t1]=ndgrid(t3,t2array,t1);

%%Orientational Reponse Functions
cos2theta=cos(theta)^2; sin2theta=sin(theta)^2;
exp1=(1./9).*exp(-2.*Dor.*(t3+t1).*dt);
exp2=exp(-6.*Dor.*t2.*dt);
if POL=='ZZZZ'
    Yaaaa=exp1.*(1+(4./5).*exp2);
    Yaass=exp1.*(1+(4./5).*exp2.*(cos2theta-0.5.*sin2theta));
    Yasas=exp1.*(cos2theta+0.2.*exp2.*(4.*cos2theta+3.*sin2theta));
elseif POL=='ZZYY'
    Yaaaa=exp1.*(1-(2./5).*exp2);
    Yaass=exp1.*(1-(2./5).*exp2.*(cos2theta-0.5.*sin2theta));
    Yasas=exp1.*(cos2theta-0.1.*exp2.*(4.*cos2theta+3.*sin2theta));
end

%%Oscillatory parts of the response function
wa_dwa = wa-dwa;    wa_dwas = wa-dwas;
ws_dws = ws-dws;    ws_dwas = ws-dwas;
wa_ws = wa-ws;
expi_wa_t3=exp(-I.*wa.*t3.*dt);
expi_wa_t1=exp(-I.*wa.*t1.*dt);
expi_ws_t3=exp(-I.*ws.*t3.*dt);
expi_ws_t1=exp(-I.*ws.*t1.*dt);
expi_wa_dwa_t3=exp(-I.*wa_dwa.*t3.*dt);
```

```
expi_ws_dws_t3=exp(-I.*ws_dws.*t3.*dt);
expi_wa_dwas_t3=exp(-I.*wa_dwas.*t3.*dt);
expi_ws_dwas_t3=exp(-I.*ws_dwas.*t3.*dt);
expi_wa_ws_t2=exp(-I.*wa_ws.*t2.*dt);

%%Rephasing and non-rephasing response functions
S_I=Yaaaa.*...
     (((2.0.*ua_sq.*ua_sq.*expi_wa_t3./expi_wa_t1-ua_sq.*u2a_sq.*expi_wa_dwa_t3./expi_wa_t1).*...
     haa(t2+t3).*haa(t1+t2).*haa(t1).*haa(t3)./(haa(t1+t2+t3).*haa(t2)))+...
     (2.0.*us_sq.*us_sq.*expi_ws_t3./expi_ws_t1-us_sq.*u2s_sq.*expi_ws_dws_t3./expi_ws_t1).*...
     hss(t2+t3).*hss(t1+t2).*hss(t1).*hss(t3)./(hss(t1+t2+t3).*hss(t2)))+...
     Yaass.*...
     (((ua_sq.*us_sq.*expi_ws_t3./expi_wa_t1-ua_sq.*uas_sq.*expi_ws_dwas_t3./expi_wa_t1).*...
     has(t2+t3).*has(t1+t2).*haa(t1).*hss(t3)./(has(t1+t2+t3).*has(t2)))+...
     (ua_sq.*us_sq.*expi_wa_t3./expi_ws_t1-us_sq.*usa_sq.*expi_wa_dwas_t3./expi_ws_t1).*...
     has(t2+t3).*has(t1+t2).*hss(t1).*haa(t3)./(has(t1+t2+t3).*has(t2)))+...
     Yasas.*...
     (((ua_sq.*us_sq.*expi_ws_t3./(expi_wa_t1.*expi_wa_ws_t2)-...
     ua.*us.*usa.*uas.*expi_ws_dwas_t3./expi_wa_ws_t2./expi_wa_t1).*...
     haa(t1+t2).*hss(t2+t3).*has(t1).*has(t3)./(has(t1+t2+t3).*has(t2)))+...
     (ua_sq.*us_sq.*expi_wa_t3.*expi_wa_ws_t2./expi_ws_t1 -...
     ua.*us.*usa.*uas.*expi_wa_dwas_t3./(expi_ws_t1./expi_wa_ws_t2)).*...
     hss(t1+t2).*haa(t2+t3).*has(t1).*has(t3)./(has(t1+t2+t3).*has(t2)));



 S_II=Yaaaa.*...
     (((2.0.*ua_sq.*ua_sq.*expi_wa_t3.*expi_wa_t1-ua_sq.*u2a_sq.*expi_wa_dwa_t3.*expi_wa_t1).*...
     haa(t1+t2+t3).*haa(t2).*haa(t1).*haa(t3)./(haa(t2+t3).*haa(t1+t2)))+...
     ((2.0.*us_sq.*us_sq.*expi_ws_t3.*expi_ws_t1-us_sq.*u2s_sq.*expi_ws_dws_t3.*expi_ws_t1).*...
     hss(t1+t2+t3).*hss(t2).*hss(t1).*hss(t3)./(hss(t2+t3).*hss(t1+t2))))+...
     Yaass.*...
     (((ua_sq.*us_sq.*expi_ws_t3.*expi_wa_t1-ua_sq.*uas_sq.*expi_ws_dwas_t3.*expi_wa_t1).*...
     hss(t3).*haa(t1).*has(t1+t2+t3).*has(t2)./(has(t2+t3).*has(t1+t2)))+...
     ((ua_sq.*us_sq.*expi_wa_t3.*expi_ws_t1-us_sq.*usa_sq.*expi_wa_dwas_t3.*expi_ws_t1).*...
     haa(t3).*hss(t1).*has(t1+t2+t3).*has(t2)./(has(t2+t3).*has(t1+t2))))+...
     Yasas.*...
     (((ua_sq.*us_sq.*expi_wa_t3.*expi_wa_ws_t2.*expi_wa_t1 -...
     ua.*us.*usa.*uas.*expi_wa_dwas_t3.*expi_wa_ws_t2.*expi_wa_t1).*...
     hss(t2).*haa(t1+t2+t3).*has(t1).*has(t3)./(has(t2+t3).*has(t1+t2)))+...
     ((ua_sq.*us_sq.*expi_ws_t3.*expi_ws_t1.*expi_wa_ws_t2 -...
     ua.*us.*usa.*uas.*expi_ws_dwas_t3.*expi_ws_t1.*expi_wa_ws_t2).*...
     haa(t2).*hss(t1+t2+t3).*has(t1).*has(t3)./(has(t2+t3).*has(t1+t2))));

 zerotime=find(t2array==0);
 S_III=zeros(size(S_II));
 if(zerotime > 0)
     tempS3=Yaaaa.*...
          ((ua_sq.*u2a_sq.*(expi_wa_t3.*expi_wa_t1-expi_wa_dwa_t3.*expi_wa_t1).*...
          haa(t3).*haa(t1).*haa(t1+t3).*haa(0)./(haa(t1).*haa(t3)))+...
          (us_sq.*u2s_sq.*(expi_ws_t3.*expi_ws_t1-expi_ws_dws_t3.*expi_ws_t1).*...
          hss(t3).*hss(t1).*hss(t1+t3).*hss(0)./(hss(t1).*hss(t3))))+...
          +Yasas.*...
          (((ua_sq.*uas_sq.*expi_wa_t3.*expi_wa_t1-ua.*uas.*us.*usa.*expi_wa_dwas_t3.*expi_wa_t1).*...
          hss(0).*has(t3).*has(t1).*haa(t1+t3)./(has(t3).*has(t1)))+...
          ((us_sq.*usa_sq.*expi_ws_t3.*expi_ws_t1-us.*usa.*ua.*uas.*expi_ws_dwas_t3.*expi_ws_t1).*...
          haa(0).*has(t3).*has(t1).*hss(t1+t3)./(has(t3).*has(t1)))+...
          ((ua.*uas.*usa.*us.*expi_ws_t3.*expi_wa_t1-ua_sq.*uas_sq.*expi_ws_dwas_t3.*expi_wa_t1).*...
          hss(t3).*has(t1+t3).*has(0).*haa(t1)./(has(t3).*has(t1)))+...
          ((us.*usa.*uas.*ua.*expi_wa_t3.*expi_ws_t1-us_sq.*usa_sq.*expi_wa_dwas_t3.*expi_ws_t1).*...
          haa(t3).*has(t1+t3).*has(0).*hss(t1)./(has(t3).*has(t1))));
     S_III(:,zerotime,:)=tempS3(:,zerotime,:);
```

```
    clear tempS3;
end

%%Create the w axis
waxis=-(padlen/2):1:(padlen/2);
waxis=waxis./(length(waxis)*dt*c);
waxis=waxis(1:end-1);
waxis(find(waxis<0))=waxis(find(waxis<0))-2*alias;
waxis=waxis+alias;
taxis=0:(length(t3)-1);taxis=taxis*dt;

%%Fourier transform and flip the S1 signal to add correctly
S1f=zeros(padlen,length(t2array),padlen);S2f=S1f;S3f=S1f;
for i=1:length(t2array)
    S1=real(squeeze(S_I(:,i,:)));
    S2=real(squeeze(S_II(:,i,:)));
    S3=real(squeeze(S_III(:,i,:)));

    S1f(:,i,:)=circshift(flipud(fftshift(fft2(S1,padlen,padlen))),1);
    S2f(:,i,:)=fftshift(fft2(S2,padlen,padlen));
    S3f(:,i,:)=fftshift(fft2(S3,padlen,padlen));
end

%%%%3. Plotting
%%View some prelims
%figure(1)
%subplot(2,2,1)
%pcolor(taxis,taxis,real(squeeze(S_I(:,1,:))));shading flat
%subplot(2,2,2)
%pcolor(taxis,taxis,real(squeeze(S_II(:,1,:))));shading flat
%subplot(2,2,3)
%contourf(waxis,waxis,squeeze(real(S1f(:,1,:))),15)
%axis([1580 1720 1580 1720])
%subplot(2,2,4)
%contourf(waxis,waxis,squeeze(real(S2f(:,1,:))),15)
%axis([1580 1720 1580 1720])
%return

for i=1:length(t2array)
    %Display the real part of the correlation spectrum
    spec=squeeze(real(S1f(:,i,:)+S2f(:,i,:)+S3f(:,i,:)));

    figure(i);clf
    subplot(5,5,1:20)
    contourf(waxis,waxis,spec,15)
    hold on;
    plot(waxis,waxis,'k');
    axis([1580 1720 1580 1720])
    axis square
    title(['\tau_2␣=␣',num2str(t2array(i)*dt),'␣fs'],'fontsize',30);
    set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
    xlabel('\omega_1/(2␣\pi␣c)')
    ylabel('\omega_3/(2␣\pi␣c)');

    %Show the correlation functions
    subplot(5,5,21:25)
    plot(dt*squeeze(t1(1,1,:)),haa(squeeze(t1(1,1,:))),'r','linewidth',3);hold on;
    plot(dt*squeeze(t1(1,1,:)),hss(squeeze(t1(1,1,:))),'g—','linewidth',3)
    plot(dt*squeeze(t1(1,1,:)),has(squeeze(t1(1,1,:))),'b.','linewidth',3)
    set(gca,'fontsize',20)
    ylabel('C(t)')
    xlabel('t␣(fs)')
    legend('h_{aa}(t)','h_{ss}(t)','h_{as}(t)');
end

figure(1);clf
subplot(3,3,1)
spec=squeeze(real(S1f(:,i,:)+S2f(:,i,:)+S3f(:,i,:)));
contourf(waxis,waxis,spec,15)
hold on;
plot(waxis,waxis,'k');
axis([1580 1720 1580 1720])
axis square
set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
xlabel('\omega_1/(2␣\pi␣c)')
ylabel('\omega_3/(2␣\pi␣c)');

subplot(3,3,2)
spec=squeeze(imag(S1f(:,i,:)+S2f(:,i,:)+S3f(:,i,:)));
contourf(waxis,waxis,spec,15)
hold on;
plot(waxis,waxis,'k');
axis([1580 1720 1580 1720])
axis square
set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
xlabel('\omega_1/(2␣\pi␣c)')
ylabel('\omega_3/(2␣\pi␣c)');

subplot(3,3,3)
spec=squeeze(abs(S1f(:,i,:)+S2f(:,i,:)+S3f(:,i,:)));
spec1=squeeze(real(S1f(:,i,:)+S2f(:,i,:)+S3f(:,i,:)));
spec2=squeeze(imag(S1f(:,i,:)+S2f(:,i,:)+S3f(:,i,:)));
spec=spec1.^2 + spec2.^2;
spec=spec1.^2;
contourf(waxis,waxis,spec,15)
hold on;
plot(waxis,waxis,'k');
axis([1580 1720 1580 1720])
axis square
set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
xlabel('\omega_1/(2␣\pi␣c)')
ylabel('\omega_3/(2␣\pi␣c)');


subplot(3,3,4)
spec=squeeze(real(S1f(:,i,:)));
contourf(waxis,waxis,spec,15)
hold on;
plot(waxis,waxis,'k');
axis([1580 1720 1580 1720])
axis square
set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
xlabel('\omega_1/(2␣\pi␣c)')
ylabel('\omega_3/(2␣\pi␣c)');

subplot(3,3,5)
spec=squeeze(imag(S1f(:,i,:)));
contourf(waxis,waxis,spec,15)
hold on;
plot(waxis,waxis,'k');
axis([1580 1720 1580 1720])
axis square
```

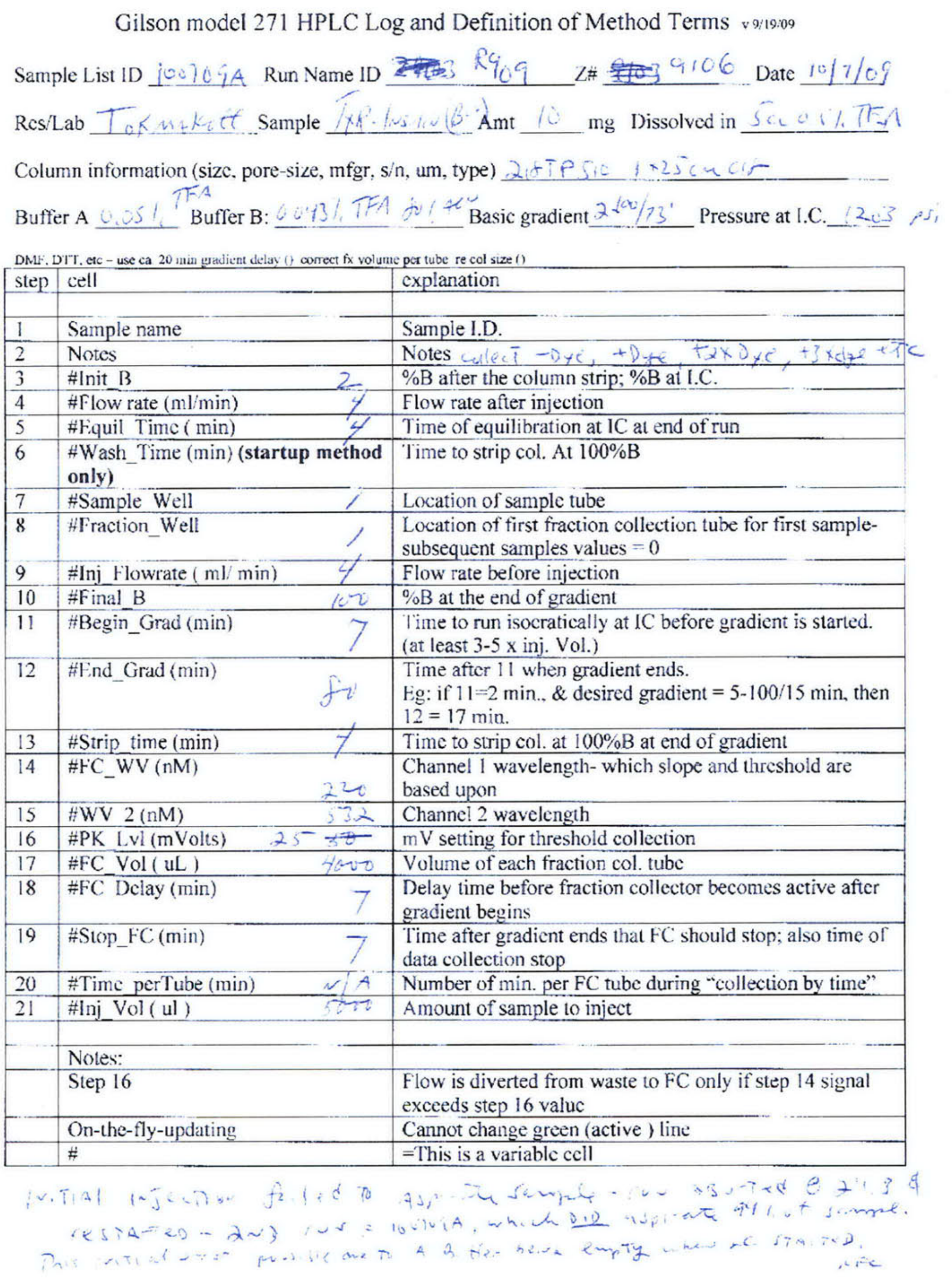
```
set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
xlabel('\omega_1/(2␣\pi␣c)')
ylabel('\omega_3/(2␣\pi␣c)');

subplot(3,3,6)
spec=squeeze(abs(S1f(:,i,:)));
contourf(waxis,waxis,spec,15)
hold on;
plot(waxis,waxis,'k');
axis([1580 1720 1580 1720])
axis square
set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
xlabel('\omega_1/(2␣\pi␣c)')
ylabel('\omega_3/(2␣\pi␣c)');

subplot(3,3,7)
spec=squeeze(real(S2f(:,i,:)));
contourf(waxis,waxis,spec,15)
hold on;
plot(waxis,waxis,'k');
axis([1580 1720 1580 1720])
axis square
set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
xlabel('\omega_1/(2␣\pi␣c)')
ylabel('\omega_3/(2␣\pi␣c)');

subplot(3,3,8)
spec=squeeze(imag(S2f(:,i,:)));
contourf(waxis,waxis,spec,15)
hold on;
plot(waxis,waxis,'k');
axis([1580 1720 1580 1720])
axis square
set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
xlabel('\omega_1/(2␣\pi␣c)')
ylabel('\omega_3/(2␣\pi␣c)');

subplot(3,3,9)
spec=squeeze(abs(S2f(:,i,:)));
contourf(waxis,waxis,spec,15)
hold on;
plot(waxis,waxis,'k');
axis([1580 1720 1580 1720])
axis square
set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
xlabel('\omega_1/(2␣\pi␣c)')
ylabel('\omega_3/(2␣\pi␣c)');

figure(2)
subplot(3,1,1)
spec=atan(squeeze(imag(S1f(:,i,:))./real(S1f(:,i,:))));
contourf(waxis,waxis,spec,15)
hold on;
plot(waxis,waxis,'k');
axis([1580 1720 1580 1720])
axis square
set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
xlabel('\omega_1/(2␣\pi␣c)')
ylabel('\omega_3/(2␣\pi␣c)');

subplot(3,1,2)
```

```
spec=atan(squeeze(imag(S2f(:,i,:))./real(S2f(:,i,:))));
contourf(waxis,waxis,spec,15)
hold on;
plot(waxis,waxis,'k');
axis([1580 1720 1580 1720])
axis square
set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
xlabel('\omega_1/(2␣\pi␣c)')
ylabel('\omega_3/(2␣\pi␣c)');

subplot(3,1,3)
spec=atan(squeeze(imag(S1f(:,i,:)+S2f(:,i,:)+S3f(:,i,:))./real(S1f(:,i,:)+S2f(:,i,:)+S3f(:,i,:))));
contourf(waxis,waxis,spec,15)
hold on;
plot(waxis,waxis,'k');
axis([1580 1720 1580 1720])
axis square
set(gca,'fontsize',25,'TickDir','Out','linewidth',2)
xlabel('\omega_1/(2␣\pi␣c)')
ylabel('\omega_3/(2␣\pi␣c)');


ind1=min(find(waxis>1635));
ind2=min(find(waxis>1680));

figure(2);clf
plot(waxis,450+squeeze(real(S1f(:,i,ind1)+S2f(:,i,ind1)+S3f(:,i,ind1))),'r','linewidth',2);hold on;
plot(waxis,450+squeeze(imag(S1f(:,i,ind1)+S2f(:,i,ind1)+S3f(:,i,ind1))),'g','linewidth',2);
plot(waxis,450+squeeze(abs(S1f(:,i,ind1)+S2f(:,i,ind1)+S3f(:,i,ind1))),'b','linewidth',2);
plot(waxis,200+squeeze(real(S1f(:,i,ind1))),'r—','linewidth',2);
plot(waxis,200+squeeze(imag(S1f(:,i,ind1))),'g—','linewidth',2);
plot(waxis,200+squeeze(abs(S1f(:,i,ind1))),'b—','linewidth',2);
plot(waxis,squeeze(real(S2f(:,i,ind1))),'r-','linewidth',2);
plot(waxis,squeeze(imag(S2f(:,i,ind1))),'g-','linewidth',2);
plot(waxis,squeeze(abs(S2f(:,i,ind1))),'b-','linewidth',2);
plot(waxis,squeeze(real(S2f(:,i,ind1))),'r.','linewidth',2);
plot(waxis,squeeze(imag(S2f(:,i,ind1))),'g.','linewidth',2);
plot(waxis,squeeze(abs(S2f(:,i,ind1))),'b.','linewidth',2);
set(gca,'fontsize',28)
xlabel('\omega_3/2\pi␣c␣(cm^{-1})');
ylabel('Spectral␣Intensity␣(AU)');
axis([1580 1720 -150 650])
```

Appendix B

Texas Red-Insulin Mass Spectra

The nominal masses of bovine insulin, insulin-TR, insulin-2TR, insulin-3TR are 5733.49, 6436.35, 7139.21, and 7842.07 amu. The proteins may also have one or more water molecules associated with them, which will increase these nominal masses by multiples of 18 amu, and may be deprotonated. The mass of Texas Red-NHS is 816.94 amu. Texas Red without the NHS group has a mass of 702.86 amu. Insulin A and B chains, when separated have masses of 2223.53 and 3400.9 amu.



Gilson model 271 HPLC Log and Definition of Method Terms v 9/19/09

Sample List ID 100709A Run Name ID ~~2403~~ R909 Z# ~~9103~~ 9106 Date 10/7/09

Res/Lab Tokmakoff Sample TxR-Insulin (B:) Amt 10 mg Dissolved in 5cc 0.1% TFA

Column information (size, pore-size, mfgr, s/n, um, type) 2.0 TP Sio 1×25cm C18

Buffer A 0.05% TFA Buffer B: 0.043% TFA in 90% ACN Basic gradient 2-100/73' Pressure at I.C. 1203 psi

DMF, DTT, etc – use ca 20 min gradient delay () correct fx volume per tube re col size ()

step	cell		explanation
1	Sample name		Sample I.D.
2	Notes		Notes collect -Dye, +Dye, +2x Dye, +3x dye +TC
3	#Init_B	2	%B after the column strip; %B at I.C.
4	#Flow rate (ml/min)	4	Flow rate after injection
5	#Equil_Time (min)	4	Time of equilibration at IC at end of run
6	#Wash_Time (min) **(startup method only)**		Time to strip col. At 100%B
7	#Sample_Well	1	Location of sample tube
8	#Fraction_Well	1	Location of first fraction collection tube for first sample- subsequent samples values = 0
9	#Inj_Flowrate (ml/ min)	4	Flow rate before injection
10	#Final_B	100	%B at the end of gradient
11	#Begin_Grad (min)	7	Time to run isocratically at IC before gradient is started. (at least 3-5 x inj. Vol.)
12	#End_Grad (min)	80	Time after 11 when gradient ends. Eg: if 11=2 min., & desired gradient = 5-100/15 min, then 12 = 17 min.
13	#Strip_time (min)	7	Time to strip col. at 100%B at end of gradient
14	#FC_WV (nM)	220	Channel 1 wavelength- which slope and threshold are based upon
15	#WV_2 (nM)	532	Channel 2 wavelength
16	#PK_Lvl (mVolts)	25 ~~30~~	mV setting for threshold collection
17	#FC_Vol (uL)	4000	Volume of each fraction col. tube
18	#FC_Delay (min)	7	Delay time before fraction collector becomes active after gradient begins
19	#Stop_FC (min)	7	Time after gradient ends that FC should stop; also time of data collection stop
20	#Time_perTube (min)	N/A	Number of min. per FC tube during "collection by time"
21	#Inj_Vol (ul)	5000	Amount of sample to inject
	Notes:		
	Step 16		Flow is diverted from waste to FC only if step 14 signal exceeds step 16 value
	On-the-fly-updating		Cannot change green (active) line
	#		=This is a variable cell

INITIAL INJECTION failed to aspirate sample - run aborted @ 21.3 [illegible]
RESTARTED - 2nd run = 100709A, which DID aspirate all of sample.
This initial error possible due to A & B [illegible] being empty when run STARTED.
RFC

Figure B-1: Texas Red-Insulin HPLC Log

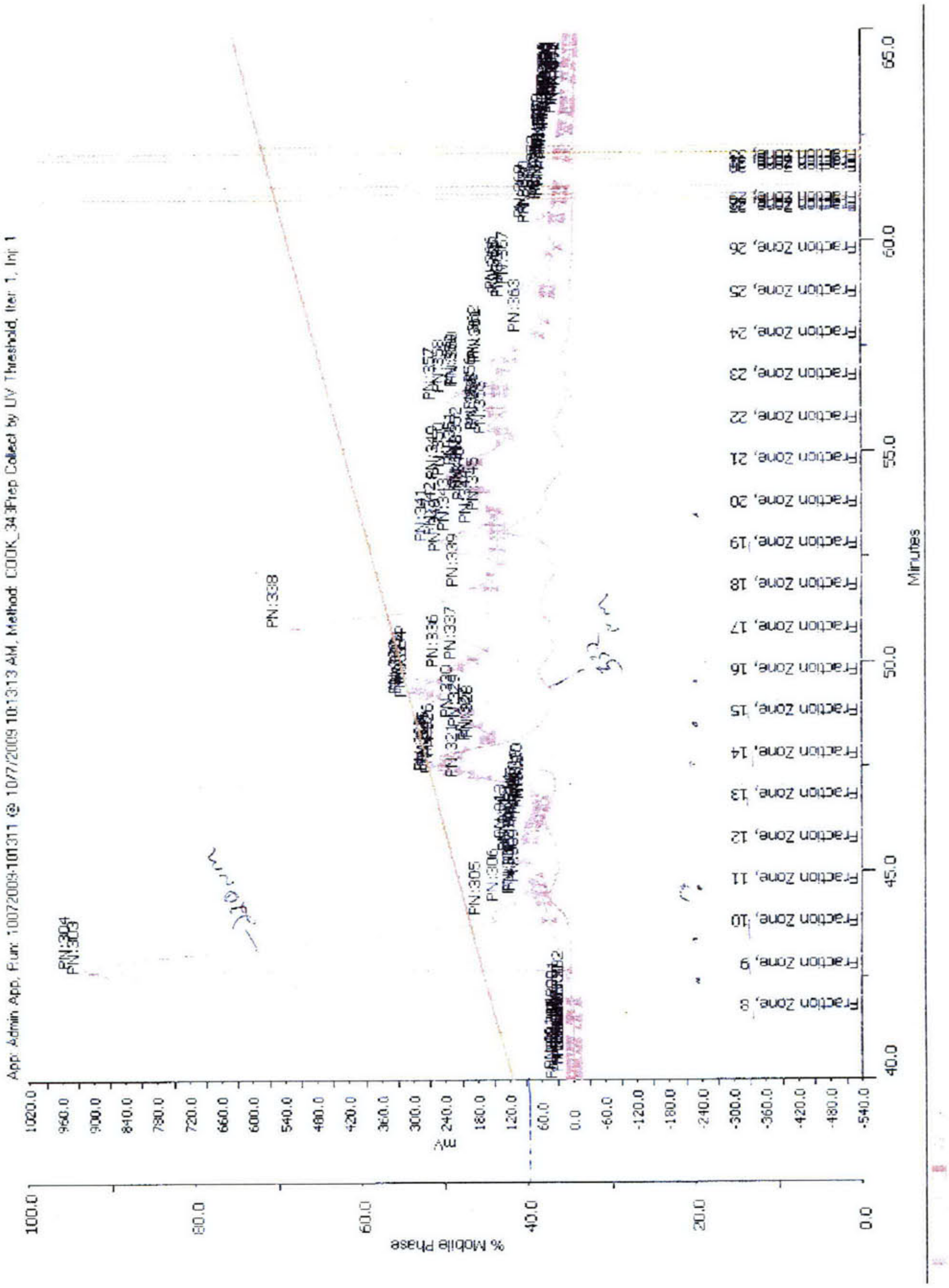


Figure B-2: UV-Vis assay of the LC/MS separation, and demarcation of the fractions.

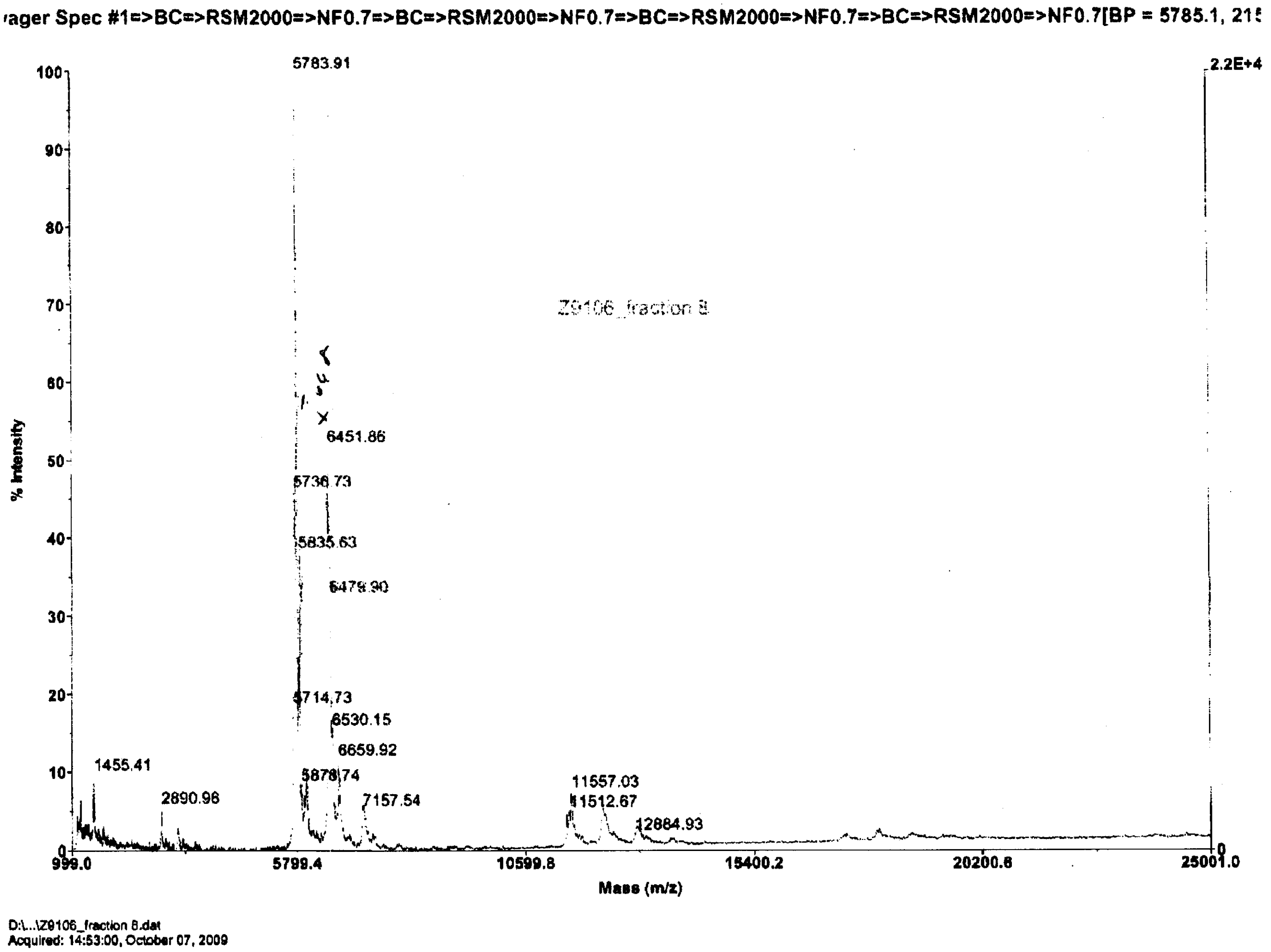


Figure B-3: Mass Spec for Fraction 8
This fraction contains insulin (5783.91) and insulin-TR (6451.86).

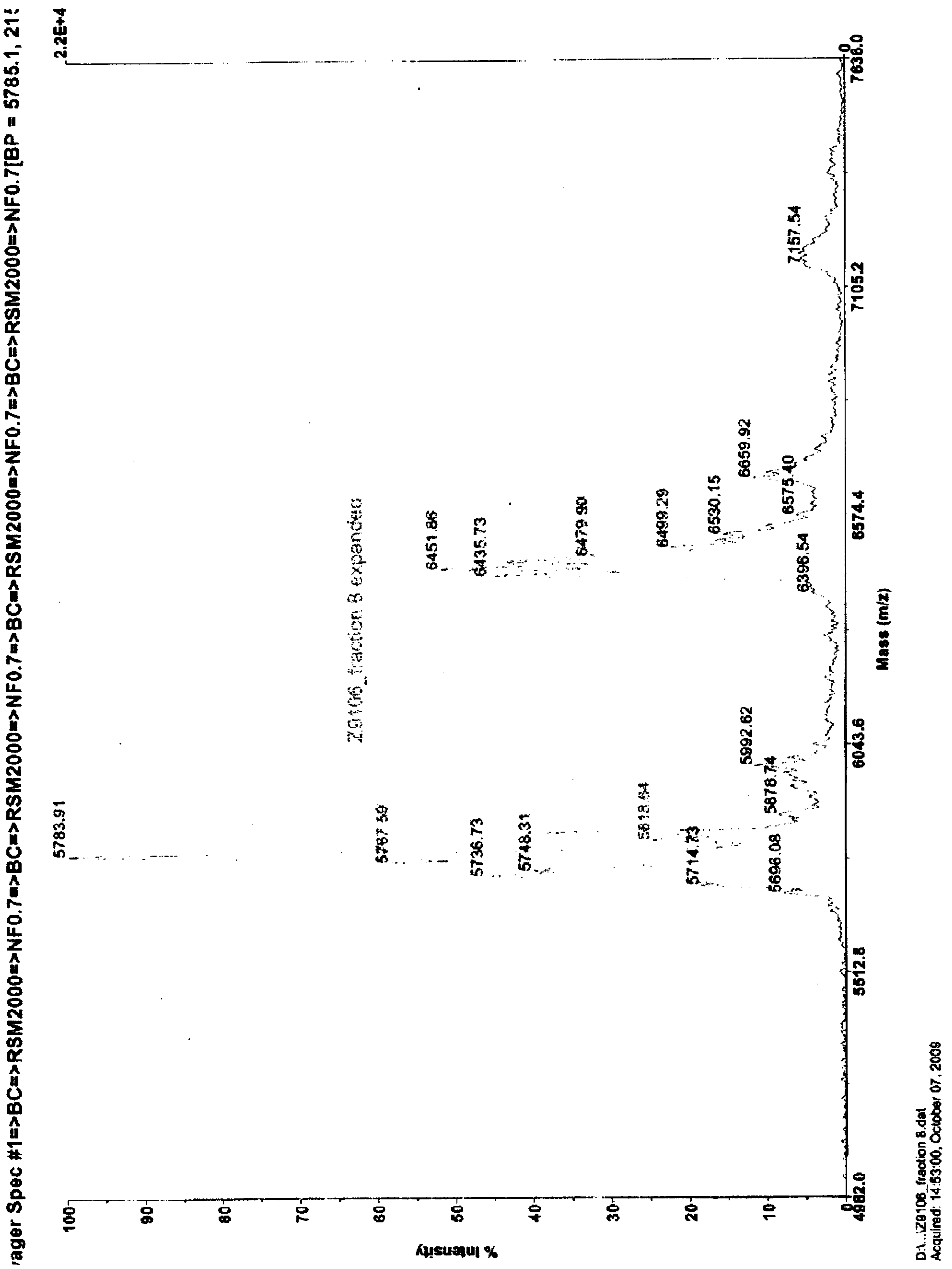


Figure B-4: Mass Spec for Fraction 8, expanded
This fraction contains insulin (5783.91) and insulin-TR (6451.86).

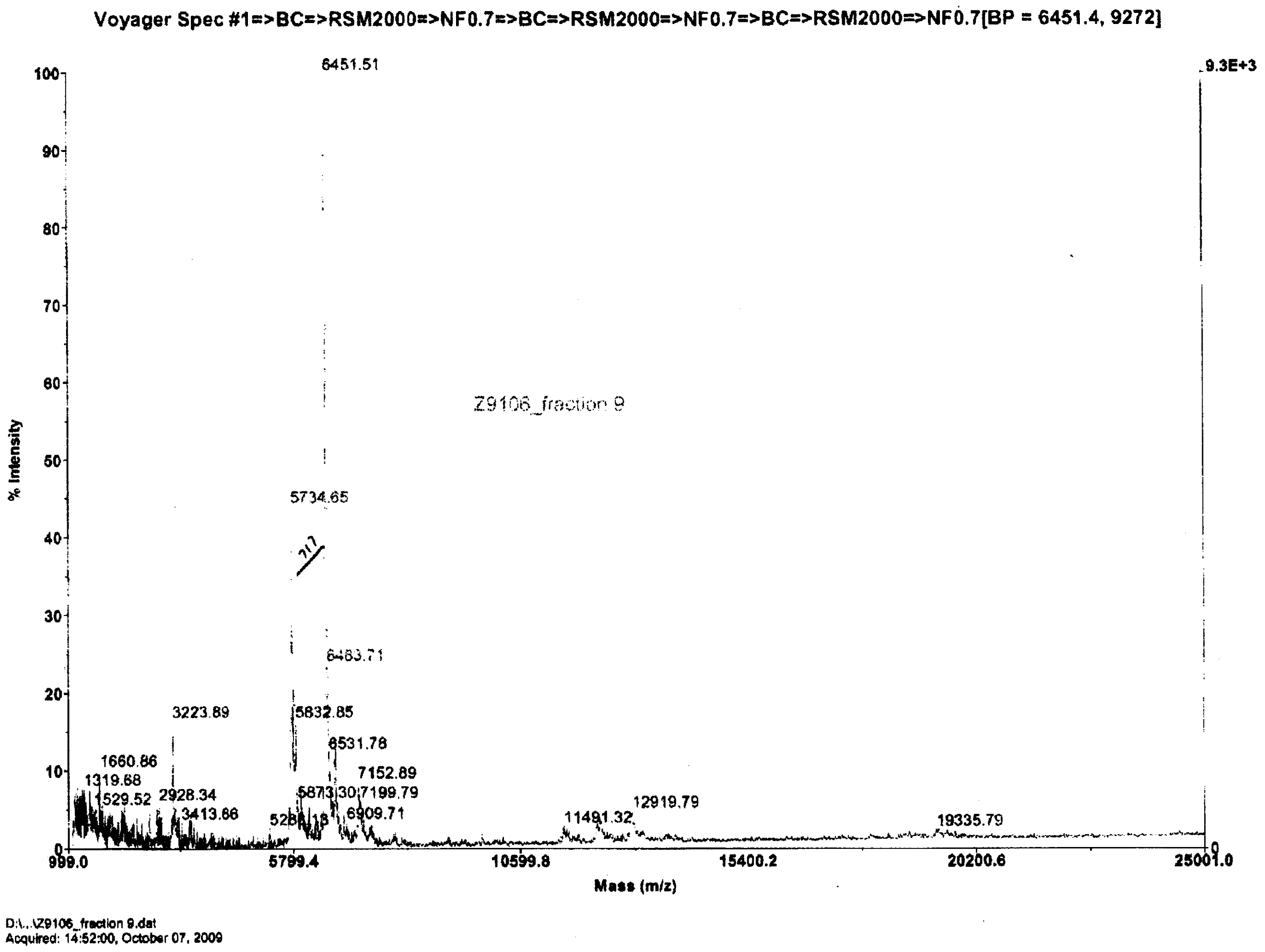
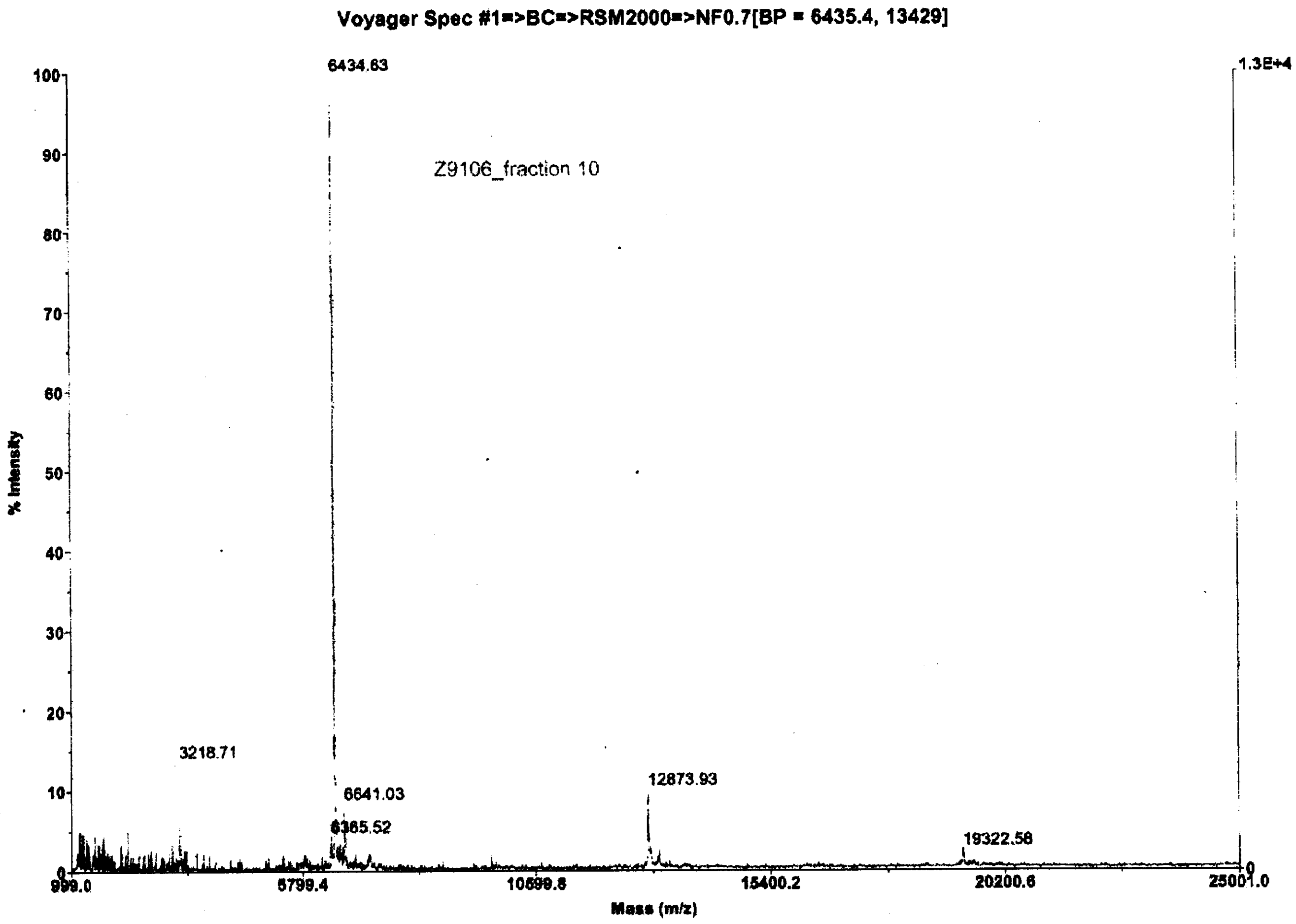


Figure B-5: Mass Spec for Fraction 9
This fraction contains insulin (5734.65) and insulin-TR (6451.51).



D:\...\Z9106_fraction 10.dat
Acquired: 14:54:00, October 07, 2009

Figure B-6: Mass Spec for Fraction 10
This fraction contains insulin-TR (6434.63).

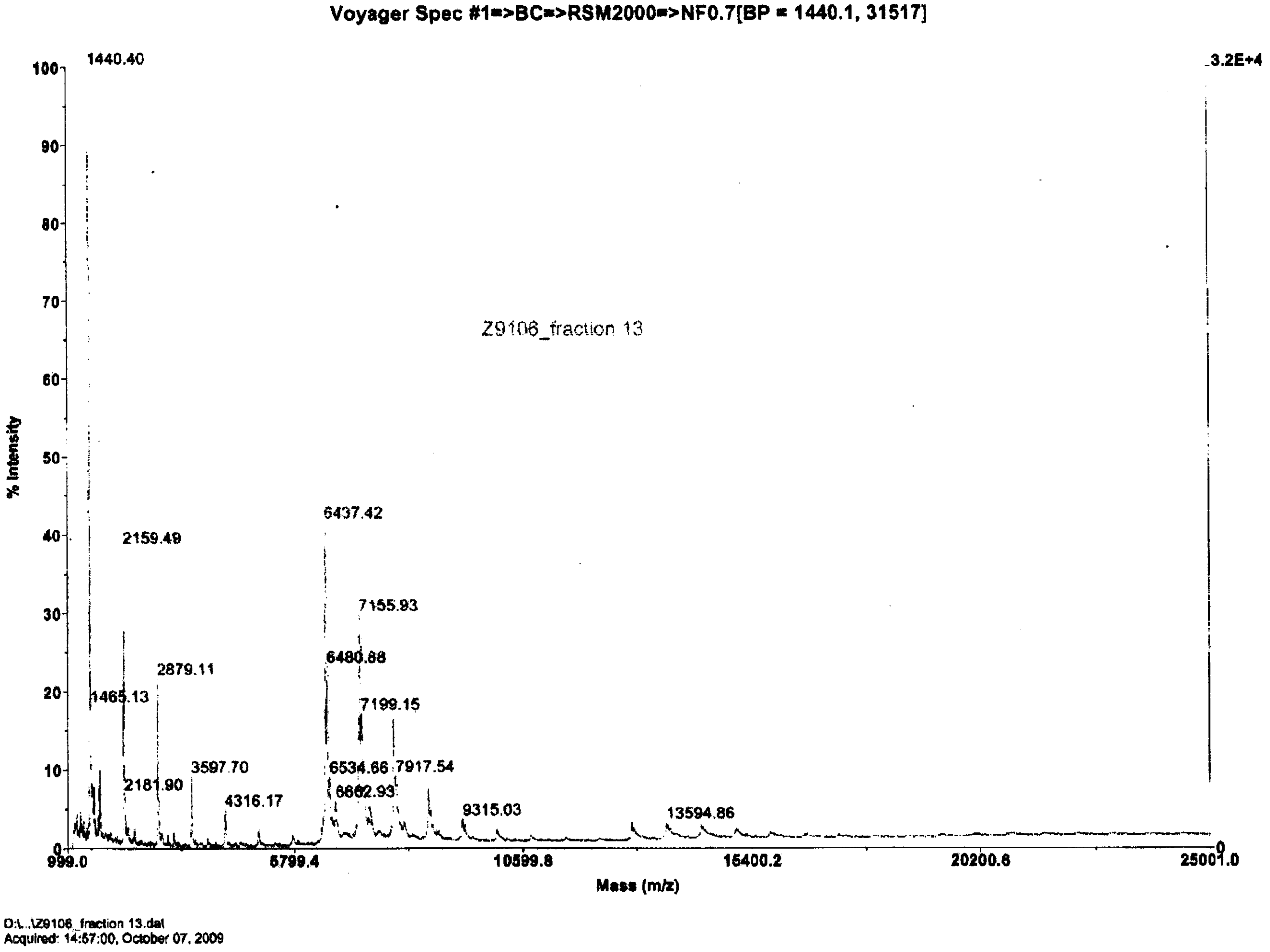
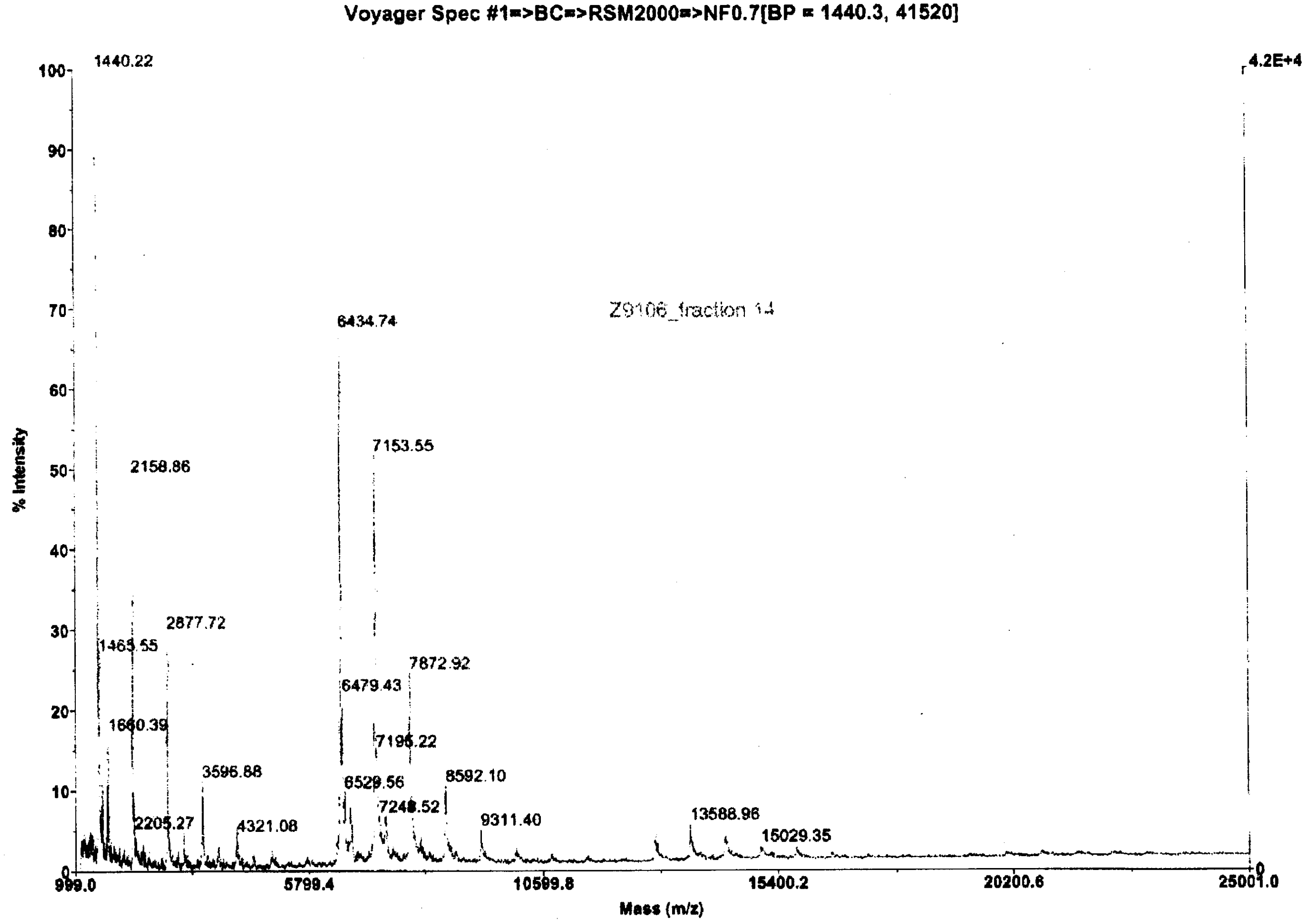


Figure B-7: Mass Spec for Fraction 13
This fraction contains insulin-TR (6437.42) and insulin-2TR (7155.93, 7199.15).



D:\...\Z9106_fraction 14.dat
Acquired: 16:00:00, October 07, 2009

Figure B-8: Mass Spec for Fraction 14

This fraction contains insulin-TR (6434.74), insulin-2TR (7153.55), and insulin-3TR (7872.92).

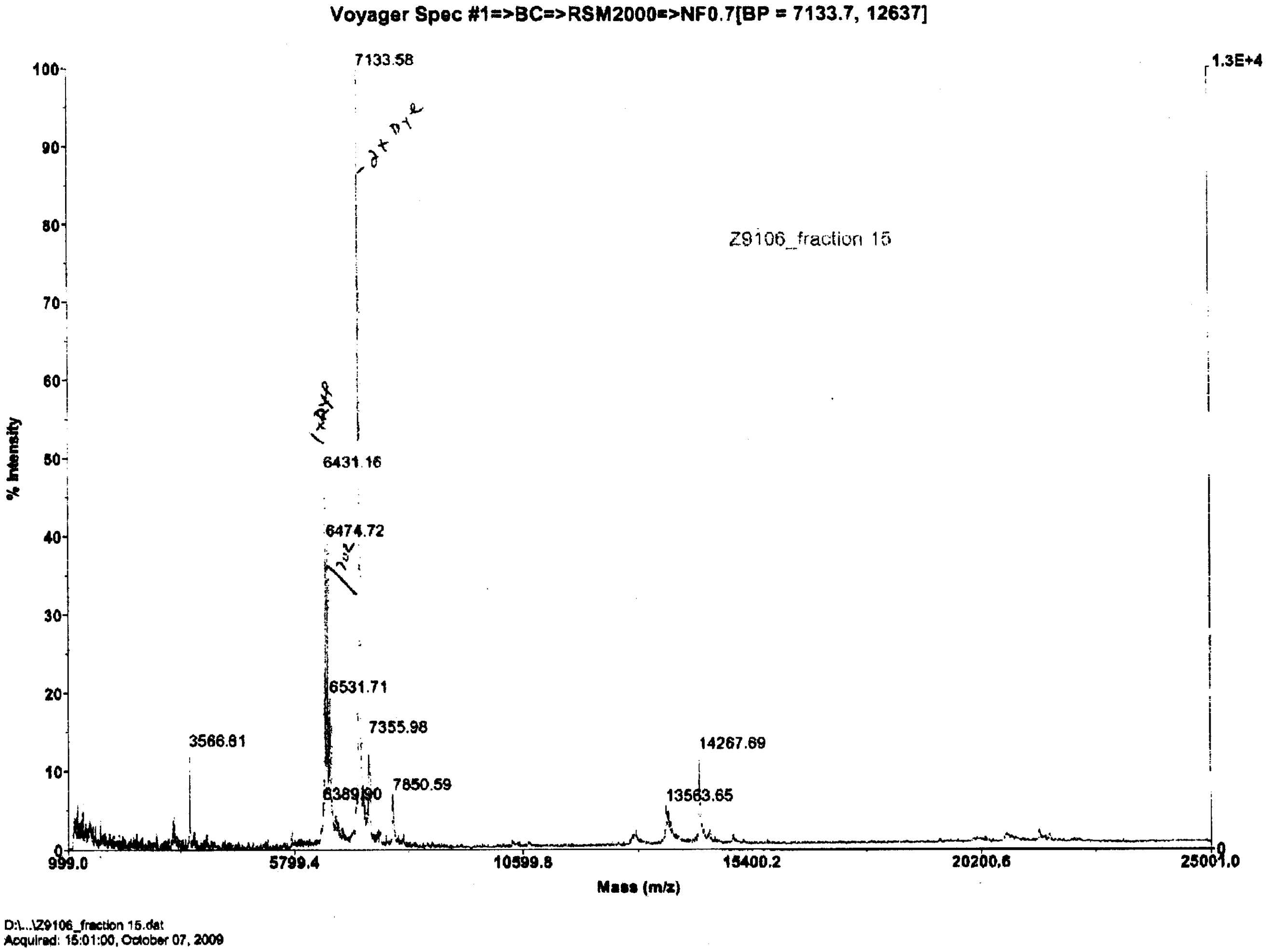


Figure B-9: Mass Spec for Fraction 15

This fraction contains insulin-TR (6431.16) and insulin-2TR (7133.58).

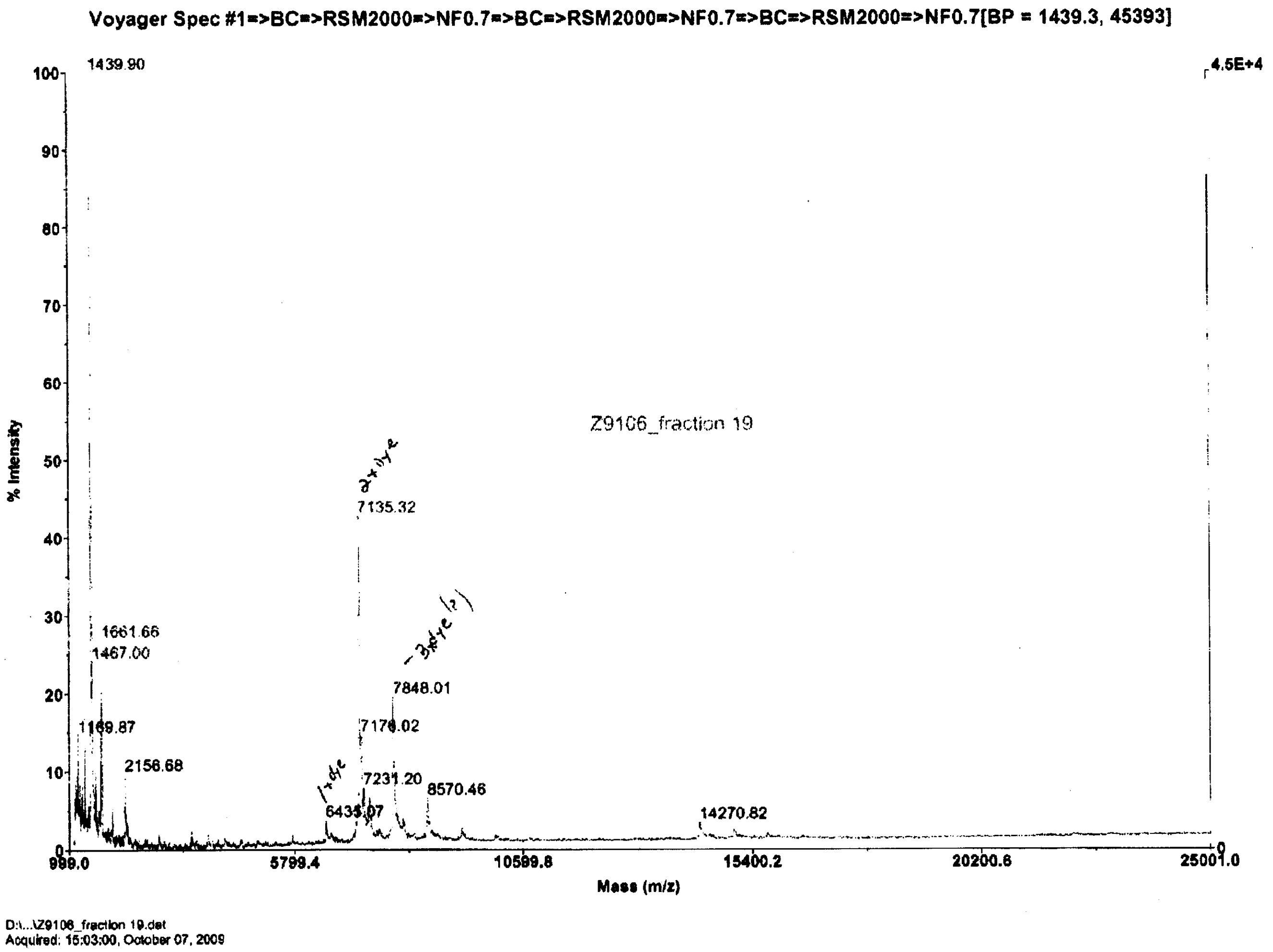


Figure B-10: Mass Spec for Fraction 19
This fraction contains insulin-2TR (7135.32) and insulin-3TR (7848.01).

Bibliography

[1] Ziad Ganim, Hoi Sung Chung, Adam W Smith, Lauren P Deflores, Kevin C Jones, and Andrei Tokmakoff. Amide I two-dimensional infrared spectroscopy of proteins. *Accounts Chem Res*, 41(3):432–441, Jan 2008.

[2] Nurettin Demirdöven, Christopher M Cheatum, Hoi Sung Chung, Munira Khalil, Jasper Knoester, and Andrei Tokmakoff. Two-dimensional infrared spectroscopy of antiparallel beta-sheet secondary structure. *J Am Chem Soc*, 126(25):7981–90, Jun 2004.

[3] HS Chung, M Khalil, and A Tokmakoff. Nonlinear infrared spectroscopy of protein conformational change during thermal unfolding. *J Phys Chem B*, 108(39):15332–15342, Jan 2004.

[4] Adam W. Smith, Hoi Sung Chung, Ziad Ganim, and Andrei Tokmakoff. Residual native structure in a thermally denatured β-hairpin. *J. Phys. Chem. B*, 109:17025–17027, 2005.

[5] Munira Khalil, N Demirdoven, and Andrei Tokmakoff. Coherent 2D IR spectroscopy: Molecular structure and dynamics in solution. *J Phys Chem A*, 107(27):5258–5279, Jan 2003.

[6] Lauren P Deflores, Ziad Ganim, Sarah F Ackley, Hoi Sung Chung, and Andrei Tokmakoff. The anharmonic vibrational potential and relaxation pathways of the amide I and II modes of N-methylacetamide. *J Phys Chem B*, 110(38):18973–18980, Jan 2006.

[7] S. Parkin, B. Rupp, and H. Hope. Structure of bovine pancreatic trypsin inhibitor at 125 K definition of carboxyl-terminal residues Gly57 and Ala58. *Acta Cryst.*, D52:18–29, 1996.

[8] Ziad Ganim and Andrei Tokmakoff. Spectral signatures of heterogeneous protein ensembles revealed by MD simulations of 2DIR spectra. *Biophysical Journal*, 91(7):2636–2646, Jan 2006.

[9] S Vijaykumar, CE Bugg, and WJ Cook. Structure of ubiquitin refined at 1.8 a resolution. *Journal of Molecular Biology*, 194(3):531–544, 1987.

[10] R Koradi, M Billeter, and K Wuthrich. MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graphics*, 14(1):51–&, Jan 1996.

[11] DOV Alonso and V Daggett. Molecular-dynamics simulations of protein unfolding and limited refolding - characterization of partially unfolded states of ubiquitin in 60-percent methanol and in water. *Journal of Molecular Biology*, 247(3):501–520, Jan 1995.

[12] Hoi Sung Chung and Andrei Tokmakoff. Visualization and characterization of the infrared active amide I vibrations of proteins. *J Phys Chem B*, 110(6):2888–2898, Jan 2006.

[13] DS Goodsell and AJ Olson. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.*, 29:105–153, Jan 2000.

[14] AH Elcock, D Sept, and JA McCammon. Computer simulation of protein-protein interactions. *J Phys Chem B*, 105(8):1504–1518, Jan 2001.

[15] W Humphrey, A Dalke, and K Schulten. VMD: Visual molecular dynamics. *J Mol Graphics*, 14(1):33–&, Jan 1996.

[16] E N Baker, T L Blundell, J F Cutfield, S M Cutfield, E J Dodson, G G Dodson, D M Hodgkin, R E Hubbard, N W Isaacs, and C D Reynolds. The structure of 2Zn pig insulin crystals at 1.5 A resolution. *Philos Trans R Soc Lond, B, Biol Sci*, 319(1195):369–456, Jul 1988. RCSB 4INS hexamer.

[17] Joseph R. Lakowicz. *Principles of Fluorescence Spectroscopy, 3rd Ed.* Springer, New York, 2006.

[18] AL Epstein, MM Mizokami, JL Li, PS Hu, and LA Khawli. Identification of a protein fragment of interleukin 2 responsible for vasopermeability. *J Natl Cancer I*, 95(10):741–749, Jan 2003.

[19] K Boels, G Glassmeier, D Herrmann, IB Riedel, W Hampe, I Kojima, JR Schwarz, and HC Schaller. The neuropeptide head activator induces activation and translocation of the growth-factor-regulated Ca2+-permeable channel GRC. *J Cell Sci*, 114(20):3599–3606, Jan 2001.

[20] CH Byun, JY Park, T Akamizu, and CB Chae. Identification of the peptides that inhibit the function of human monoclonal thyroid-stimulating antibodies from phage-displayed peptide library. *J Clin Endocr Metab*, 86(7):3311–3318, Jan 2001.

[21] WY Yang and M Gruebele. Folding at the speed limit. *Nature*, 423(6936):193–197, Jan 2003.

[22] J Sabelko, J Ervin, and M Gruebele. Observation of strange kinetics in protein folding. *P Natl Acad Sci Usa*, 96(11):6031–6036, Jan 1999.

[23] DT Leeson, F Gai, HM Rodriguez, LM Gregoret, and RB Dyer. Protein folding and unfolding on a complex energy landscape. *P Natl Acad Sci Usa*, 97(6):2527–2532, Jan 2000.

[24] CY Huang, Z Getahun, YJ Zhu, JW Klemke, WF DeGrado, and F Gai. Helix formation via conformation diffusion search. *P Natl Acad Sci Usa*, 99(5):2788–2793, Jan 2002.

[25] Hoi Sung Chung, Munira Khalil, Adam W Smith, Ziad Ganim, and Andrei Tokmakoff. Conformational changes during the nanosecond-to-millisecond unfolding of ubiquitin. *P Natl Acad Sci Usa*, 102(3):612–617, Jan 2005.

[26] J Balbach, V Forge, WS Lau, NAJ vanNuland, K Brew, and CM Dobson. Protein folding monitored at individual residues during a two-dimensional NMR experiment. *Science*, 274(5290):1161–1163, Jan 1996.

[27] Colin A Thomas, Erach R Talatya, and James G Bann. 3S-Fluoroproline as a probe to monitor proline isomerization during protein folding by F-19-NMR. *Chem Commun*, 1(23):3366–3368, Jan 2009.

[28] V Schmidt, R Pahl, V Srajer, S Anderson, Z Ren, H Ihee, S Rajagopal, and K Moffat. Protein kinetics: Structures of intermediates and reaction mechanism from time-resolved x-ray data. *P Natl Acad Sci Usa*, 101(14):4799–4804, Jan 2004.

[29] D Bourgeois, B Vallone, F Schotte, A Arcovito, AE Miele, G Sciara, M Wulff, P Anfinrud, and M Brunori. Complex landscape of protein structural dynamics unveiled by nanosecond Laue crystallography. *P Natl Acad Sci Usa*, 100(15):8704–8709, Jan 2003.

[30] Hyun Sun Cho, Naranbaatar Dashdorj, Friedrich Schotte, Timothy Graber, Robert Henning, and Philip Anfinrud. Protein structural dynamics in solution unveiled via 100-ps time-resolved x-ray scattering. *P Natl Acad Sci Usa*, 107(16):7281–7286, Jan 2010.

[31] S Doniach. Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem Rev*, 101(6):1763–1778, Jan 2001.

[32] RB Dyer, F Gai, and WH Woodruff. Infrared studies of fast events in protein folding. *Accounts Chem Res*, 31(11):709–716, Jan 1998.

[33] J Bredenbeck, J Helbing, A Sieg, T Schrader, W Zinth, C Renner, R Behrendt, L Moroder, J Wachtveitl, and P Hamm. Picosecond conformational transition and equilibration of a cyclic peptide. *P Natl Acad Sci Usa*, 100(11):6452–6457, Jan 2003.

[34] K Yamamoto, Y Mizutani, and T Kitagawa. Nanosecond temperature jump and time-resolved raman study of thermal unfolding of ribonuclease A. *Biophysical Journal*, 79(1):485–495, Jan 2000.

[35] Gurusamy Balakrishnan, Ying Hu, Martin A Case, and Thomas G Spiro. Microsecond melting of a folding intermediate in a coiled-coil peptide, monitored

by T-jump/UV raman spectroscopy. *J Phys Chem B*, 110(40):19877–19883, Jan 2006.

[36] XL Xie and JD Simon. Picosecond time-resolved circular-dichroism spectroscopy - experimental details and applications. *Rev. Sci. Instrum.*, 60(8):2614–2627, Jan 1989.

[37] RA Goldbeck, YG Thomas, EF Chen, RM Esquerra, and DS Kliger. Multiple pathways on a protein-folding energy landscape: Kinetic evidence. *P Natl Acad Sci Usa*, 96(6):2782–2787, Jan 1999.

[38] Lorna Dougan, Ainavarapu Sri Rama Koti, Georgi Genchev, Hui Lu, and Julio M Fernandez. A single-molecule perspective on the role of solvent hydrogen bonds in protein folding and chemical reactions. *ChemPhysChem*, 9(18):2836–2847, Jan 2008.

[39] Kambiz M Hamadani and Shimon Weiss. Nonequilibrium single molecule protein folding in a coaxial mixer. *Biophysical Journal*, 95(1):352–365, Jan 2008.

[40] Yann Gambin, Claire Simonnet, Virginia VanDelinder, Ashok Deniz, and Alex Groisman. Ultrafast microfluidic mixer with three-dimensional flow focusing for studies of biochemical kinetics. *Lab Chip*, 10(5):598–609, Jan 2010.

[41] JB Knight, A Vishwanath, JP Brody, and RH Austin. Hydrodynamic focusing on a silicon chip: Mixing nanoliters in microseconds. *Phys Rev Lett*, 80(17):3863–3866, Jan 1998.

[42] CM Phillips, Y Mizutani, and RM Hochstrasser. Ultrafast thermally-induced unfolding of rnase-A. *P Natl Acad Sci Usa*, 92(16):7292–7296, Jan 1995.

[43] Jan Kubelka. Time-resolved methods in biophysics. 9. laser temperature-jump methods for investigating biomolecular dynamics. *Photoch Photobio Sci*, 8(4):499–512, Jan 2009.

[44] Omar F Mohammed, Gouri S Jas, Milo M Lin, and Ahmed H Zewail. Primary peptide folding dynamics observed with ultrafast temperature jump. *Angew Chem Int Edit*, 48(31):5628–5632, Jan 2009.

[45] Kresimir Franjic and R. J. Dwayne Miller. Vibrationally excited ultrafast thermodynamic phase transitions at the water/air interface. *Phys Chem Chem Phys*, pages 1–31, Jan 2010.

[46] Oliver Link, Esteban Voehringer-Martinez, Eugen Lugovoj, Yaxing Liu, Katrin Siefermann, Manfred Faubel, Helmut Grubmueller, R. Benny Gerber, Yifat Miller, and Bernd Abel. Ultrafast phase transitions in metastable water near liquid interfaces. *Faraday Discuss*, 141:67–79, Jan 2009.

[47] Charles Dumont, Tryggvi Emilsson, and Martin Gruebele. Reaching the protein folding speed limit with large, sub-microsecond pressure jumps. *Nat Methods*, 6(7):515–U70, Jan 2009.

[48] A Tokmakoff, MJ Lang, DS Larsen, GR Fleming, V Chernyak, and S Mukamel. Two-dimensional Raman spectroscopy of vibrational interactions in liquids. *Phys Rev Lett*, 79(14):2702–2705, Jan 1997.

[49] JD Hybl, AW Albrecht, SMG Faeder, and DM Jonas. Two-dimensional electronic spectroscopy. *Chem Phys Lett*, 297(3-4):307–313, Jan 1998.

[50] DF Underwood and DA Blank. Measuring the change in the intermolecular raman spectrum during dipolar solvation. *J Phys Chem A*, 109(15):3295–3306, Jan 2005.

[51] Mark A Berg. Hilbert-space treatment of incoherent, time-resolved spectroscopy. I. formalism, a tensorial classification of high-order orientational gratings and generalized MUPPETS "echoes". *J Chem Phys*, 132(14):144105, Jan 2010.

[52] Jens Bredenbeck, Avishek Ghosh, Han-Kwang Nienhuys, and Mischa Bonn. Interface-specific ultrafast two-dimensional vibrational spectroscopy. *Accounts Chem Res*, 42(9):1332–1342, Jan 2009.

[53] Dick T Co, Jenny V Lockard, David W McCamant, and Michael R Wasielewski. Narrow-bandwidth tunable picosecond pulses in the visible produced by noncollinear optical parametric amplification with a chirped blue pump. *Appl Optics*, 49(10):1880–1885, Jan 2010.

[54] Haruhiko Kohno, Klaus-Juergen Bathe, and John C Wright. A finite element procedure for multiscale wave equations with application to plasma waves. *Comput Struct*, 88(1-2):87–94, Jan 2010.

[55] Wei Zhuang, Darius Abramavicius, and Shaul Mukamel. Two-dimensional vibrational optical probes for peptide fast folding investigation. *P Natl Acad Sci Usa*, 103(50):18934–18938, Jan 2006.

[56] S Woutersen and P Hamm. Nonlinear two-dimensional vibrational spectroscopy of peptides. *J Phys-Condens Mat*, 14(39):R1035–R1062, Jan 2002.

[57] CM Cheatum, A Tokmakoff, and J Knoester. Signatures of beta-sheet secondary structures in linear and two-dimensional infrared spectroscopy. *J Chem Phys*, 120(17):8201–8215, Jan 2004.

[58] Amber T Krummel and Martin T Zanni. DNA vibrational coupling revealed with two-dimensional infrared spectroscopy: Insight into why vibrational spectroscopy is sensitive to DNA structure. *J Phys Chem B*, 110(28):13991–14000, Jan 2006.

[59] YS Kim and RM Hochstrasser. Chemical exchange 2D IR of hydrogen-bond making and breaking. *P Natl Acad Sci Usa*, 102(32):11185–11190, Jan 2005.

[60] JR Zheng, K Kwak, J Asbury, X Chen, IR Piletic, and MD Fayer. Ultrafast dynamics of solute-solvent complexation observed at thermal equilibrium in real time. *Science*, 309(5739):1338–1343, Jan 2005.

[61] JB Asbury, T Steinel, C Stromberg, SA Corcelli, CP Lawrence, JL Skinner, and MD Fayer. Water dynamics: Vibrational echo correlation spectroscopy and comparison to molecular dynamics simulations. *J Phys Chem A*, 108(7):1107–1119, Jan 2004.

[62] JD Eaves, JJ Loparo, CJ Fecko, ST Roberts, A Tokmakoff, and PL Geissler. Hydrogen bonds in liquid water are broken only fleetingly. *P Natl Acad Sci Usa*, 102(37):13019–13022, Jan 2005.

[63] Hoi Sung Chung, Ziad Ganim, Kevin C Jones, and Andrei Tokmakoff. Transient 2D IR spectroscopy of ubiquitin unfolding dynamics. *P Natl Acad Sci Usa*, 104(36):14237–14242, Jan 2007.

[64] J Bredenbeck, J Helbing, and P Hamm. Labeling vibrations by light: Ultrafast transient 2D-IR spectroscopy tracks vibrational modes during photoinduced charge transfer. *J Am Chem Soc*, 126(4):990–991, Jan 2004.

[65] Christoph Kolano, Jan Helbing, Mariusz Kozinski, Wolfram Sander, and Peter Hamm. Watching hydrogen-bond dynamics in a beta-turn by transient two-dimensional infrared spectroscopy. *Nature*, 444(7118):469–472, Jan 2006.

[66] N Demirdöven, M Khalil, and A Tokmakoff. Correlated vibrational dynamics revealed by two-dimensional infrared spectroscopy. *Phys Rev Lett*, 89(23):237401, Jan 2002.

[67] A Barth and C Zscherp. What vibrations tell us about proteins. *Q Rev Biophys*, 35(4):369–430, Jan 2002.

[68] Hajime Torii and Mitsuo Tasumi. Three-dimensional doorway-state theory for analyses of absorption bands of many-oscillator systems. *J Chem Phys*, 97(1):86–91, Jan 1992.

[69] JH Choi, S Ham, and M Cho. Inter-peptide interaction and delocalization of amide I vibrational excitons in myoglobin and flavodoxin. *J Chem Phys*, 117(14):6821–6832, Jan 2002.

[70] Christopher M. Cheatum, Andrei Tokmakoff, and Jasper. Knoester. Signatures of β-sheet secondary structures in linear and two-dimensional infrared spectroscopy. *J. Chem. Phys.*, 120:8201–8215, 2004.

[71] Hiroaki Maekawa, Claudio Toniolo, Quirinus B Broxterman, and Nien-Hui Ge. Two-dimensional infrared spectral signatures of 3(10)- and alpha-helical peptides. *J Phys Chem B*, 111(12):3222–3235, Jan 2007.

[72] Kevin C Jones, Ziad Ganim, and Andrei Tokmakoff. Heterodyne-detected dispersed vibrational echo spectroscopy. *J Phys Chem A*, 113(51):14060–14066, Jan 2009.

[73] P Hamm, MH Lim, and RM Hochstrasser. Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J Phys Chem B*, 102(31):6123–6138, Jan 1998.

[74] SA Passino, Y Nagasawa, T Joo, and GR Fleming. Three-pulse echo peak shift studies of polar solvation dynamics. *J Phys Chem A*, 101(4):725–731, Jan 1997.

[75] Kees Lazonder and Maxim S Pshenichnikov. Solvent dynamics in a glass-forming liquid from 300 K to 3 K: What photon echoes can teach us. *Chem Phys*, 341(1-3):123–142, Jan 2007.

[76] CJ Fecko, JD Eaves, JJ Loparo, A Tokmakoff, and PL Geissler. Ultrafast hydrogen-bond dynamics in the infrared spectroscopy of water. *Science*, 301(5640):1698–1702, Jan 2003.

[77] P Hamm, M Lim, and RM Hochstrasser. Non-Markovian dynamics of the vibrations of ions in water from femtosecond infrared three-pulse photon echoes. *Phys Rev Lett*, 81(24):5326–5329, Jan 1998.

[78] KF Everitt, E Geva, and JL Skinner. Determining the solvation correlation function from three-pulse photon echoes in liquids. *J Chem Phys*, 114(3):1326–1335, Jan 2001.

[79] Shaul Mukamel. *Principles of Nonlinear Optical Spectroscopy*, page 116. Oxford University Press, New York, New York, 1995.

[80] V Cervetto, J Helbing, J Bredenbeck, and P Hamm. Double-resonance versus pulsed fourier transform two-dimensional infrared spectroscopy: An experimental and theoretical comparison. *J Chem Phys*, 121(12):5935–5942, Jan 2004.

[81] Lauren P Deflores, Rebecca A Nicodemus, and Andrei Tokmakoff. Two dimensonal fourier transform spectroscopy in the pump-probe geometry. *Opt Lett*, 32(20):2966–2968, Jan 2007.

[82] Adam W. Smith. *Observing the Unfolding Transition of Beta-Hairpin Peptides with Nonlinear Infrared Spectroscopy.* Massachusetts Institute of Technology, 2008.

[83] Ronald Bracewell. *The Fourier Transform and Its Applications, 3rd ed.* McGraw-Hill Science/Engineering/Math, New York, 1999.

[84] M Khalil, N Demirdoven, and A Tokmakoff. Obtaining absorptive line shapes in two-dimensional infrared vibrational correlation spectra. *Phys Rev Lett*, 90(4):047401, Jan 2003.

[85] M. D Fayer. Dynamics of liquids, molecules, and proteins measured with ultrafast 2D IR vibrational echo chemical exchange spectroscopy. *Annu Rev Phys Chem*, 60:21–38, Jan 2009.

[86] Adam W Smith, Hoi Sung Chung, Ziad Ganim, and Andrei Tokmakoff. Residual native structure in a thermally denatured beta-hairpin. *J Phys Chem B*, 109(36):17025–17027, Jan 2005.

[87] Hoi Sung Chung, Ali Shandiz, Tobin R Sosnick, and Andrei Tokmakoff. Probing the folding transition state of ubiquitin mutants by temperature-jump-induced downhill unfolding. *Biochemistry-Us*, 47(52):13870–7, Dec 2008.

[88] Hoi Sung Chung and Andrei Tokmakoff. Temperature-dependent downhill unfolding of ubiquitin. I. nanosecond-to-millisecond resolved nonlinear infrared spectroscopy. *Proteins*, 72(1):474–87, Jul 2008.

[89] Adam W Smith and Andrei Tokmakoff. Probing local structural events in beta-hairpin unfolding with transient nonlinear infrared spectroscopy. *Angew Chem Int Ed Engl*, 46(42):7984–7, Jan 2007.

[90] M Khalil, N Demirdoven, and A Tokmakoff. Vibrational coherence transfer characterized with fourier-transform 2D IR spectroscopy. *J Chem Phys*, 121(1):362–373, Jan 2004.

[91] Joseph J Loparo, Sean T Roberts, and Andrei Tokmakoff. Multidimensional infrared spectroscopy of water. II. hydrogen bond switching dynamics. *J Chem Phys*, 125(19):194522, Jan 2006.

[92] SM Gallagher, AW Albrecht, TD Hybl, BL Landin, B Rajaram, and DM Jonas. Heterodyne detection of the complete electric field of femtosecond four-wave mixing signals. *J Opt Soc Am B*, 15(8):2338–2345, Jan 1998.

[93] Matthew F DeCamp, Lauren P Deflores, Kevin C Jones, and Andrei Tokmakoff. Single-shot two-dimensional infrared spectroscopy. *Opt Express*, 15(1):233–241, Jan 2007.

[94] JY Sung and RJ Silbey. Four wave mixing spectroscopy for a multilevel system. *J Chem Phys*, 115(20):9266–9287, Jan 2001.

[95]

[96] JR Schmidt, SA Corcelli, and JL Skinner. Pronounced non-Condon effects in the ultrafast infrared spectroscopy of water. *J Chem Phys*, 123(4):044513, Jan 2005.

[97] J. R Schmidt, S. T Roberts, J. J Loparo, Andrei Tokmakoff, M. D Fayer, and J. L Skinner. Are water simulation models consistent with steady-state and ultrafast vibrational spectroscopy experiments? *Chem Phys*, 341(1-3):143–157, Jan 2007.

[98] O Golonzka and Andrei Tokmakoff. Polarization-selective third-order spectroscopy of coupled vibronic states. *J Chem Phys*, 115(1):297–309, Jan 2001.

[99] Thomas la Cour Jansen and Wioletta M Ruszel. Motional narrowing in the time-averaging approximation for simulating two-dimensional nonlinear infrared spectra. *J. Chem. Phys.*, 128(21):214501, Jun 2008.

[100] B. M Auer and J. L Skinner. Dynamical effects in line shapes for coupled chromophores: Time-averaging approximation. *J Chem Phys*, 127(10):104105, Jan 2007.

[101] JR Schmidt, N Sundlass, and JL Skinner. Line shapes and photon echoes within a generalized kubo model. *Chem Phys Lett*, 378(5-6):559–566, Jan 2003.

[102] Arend G Dijkstra, Thomas la Cour Jansen, Robbert Bloem, and Jasper Knoester. Vibrational relaxation in simulated two-dimensional infrared spectra of two amide modes in solution. *J Chem Phys*, 127(19):194505, Jan 2007.

[103] S Krimm and Y Abe. Intermolecular interaction effects in amide i vibrations of beta polypeptides. *P Natl Acad Sci Usa*, 69(10):2792–&, Jan 1972.

[104] Lauren P Deflores, Ziad Ganim, Rebecca A Nicodemus, and Andrei Tokmakoff. Amide i'-ii' 2d ir spectroscopy provides enhanced protein secondary structural sensitivity. *J Am Chem Soc*, 131(9):3385–91, Mar 2009.

[105] Hajime Torii and Mitsuo Tasumi. Model calculations on the amide-I infrared bands of globular proteins. *J. Chem. Phys.*, 96(5):3379–3387, 1992.

[106] P. Hamm, M.H. Lim, and R. M. Hochstrasser. Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem. B*, 102(31):6123–6138, 1998.

[107] M. F. DeCamp, L. DeFlores, J. M. McCracken, A. Tokmakoff, K. Kwac, and M. Cho. Amide I vibrational dynamics of N-methylacetamide in polar solvents: The role of electrostatic interactions. *J. Phys. Chem. B*, 109:11016–11026, 2005.

[108] Reinhard Schweitzer-Stenner, Fatma Eker, Alejandro Perez, Kai Griebenow, Xiaolin Cao, and Laurence A. Nafie. The structure of tri-proline in water probed by polarized Raman, Fourier transform infrared, vibrational circular dichroism, and electric ultraviolet circular dichroism spectroscopy. *Biopolymers*, 71:558–568, 2003.

[109] Erik Goormaghtigh, Jean-Marie Ruysschaert, and Vincent Raussens. Evaluation of the information content in infrared spectra for protein secondary structure determination. *Biophys. J.*, 90:2946–2957, 2006.

[110] Susana Navea, Romà Taeler, and Anna de Juan. Application of the local regression method interval partial least-squares to the elucidation of protein secondary structure. *Anal. Biochem.*, 336:231–242, 2005.

[111] Hirotsugu Hiramatsu, Yuji Goto, Hironobu Naiki, and Teizo Kitagawa. Core structure of amyloid fibril proposed from IR-microscope linear dichroism. *J. Am. Chem. Soc.*, 126:3008–3009, 2004.

[112] A Elliot and E. J. Ambrose. Structure of synthetic polypeptides. *Nature*, 165:921–922, 1950.

[113] Samuel Krimm. Frequency shift of the CO stretching band in polypeptides and proteins. *J. Chem. Phys.*, 23:1371–1372, 1955.

[114] Tatsuo Miyazawa. Perturbation treatment of the characteristic vibrations of polypeptide chains in various configutations. *J. Chem. Phys.*, 32(6):1647–1652, 1960.

[115] Samuel Krimm and Yasuaki Abe. Intermolecular interaction effects in the amide I vibrations of β polypeptides. *Proc. Natl. Acad. Sci. USA*, 69:2788–2792, 1972.

[116] S. Häseler, A. Poglitsch, G. Bechtold, and W. L. Peticolas. Influence of water on the infrared spectrum of crystalline N-methylacetamide. *J. Chem. Phys.*, 84:1046–1047, 1986.

[117] Noemi G. Mirkin and Samuel Krimm. Ab initio vibrational analysis of hydrogen-bonded *trans*- and *cis*-N-methylacetamide. *J. Am. Chem. Soc.*, 113:9742–9747, 1991.

[118] Noemi G. Mirkin and Samuel Krimm. Ab initio vibrational analysis of isotopic derivatives of aqueous hydrogen-bonded *trans*-N-methylacetamide. *J. Mol. Struct.*, 377:219–234, 1996.

[119] Hajime Torii, Tomoaki Tatsumi, and Mitsuo Tasumi. Effects of hydration on the structure, vibrational wavenumbers, vibrational force field and resonance Raman intensities of N-methylacetamide. *J. Raman Spectrosc.*, 29:537–546, 1998.

[120] Hajime Torii, Tomoaki Tatsumi, Takanori Kanazawa, and Mitsuo Tasumi. Effects of intermolecular hydrogen-bonding interactions on the amide I mode of N-methylacetamide: Matrix-isolation infrared studies and ab initio molecular orbital calculations. *J. Phys. Chem. B*, 102:309–314, 1998.

[121] S Ham and M Cho. Amide I modes in the N-methylacetamide dimer and glycine dipeptide analog: Diagonal force constants. *J. Chem. Phys.*, 118(15):6915–6922, 2003.

[122] Sihyun Ham, Joo-Hee Kim, Hochan Lee, and Minhaeng Cho. Correlation between electronic and molecular structure distortions and vibrational properties. II. Amide I modes of NMA-nD$_2$O complexes. *J. Chem. Phys.*, 118(8):3491–3498, 2003.

[123] P Bour and TA Keiderling. Empirical modeling of the peptide amide I band IR intensity in water solution. *J Chem Phys*, 119(21):11253–11262, Jan 2003.

[124] J. R. Schmidt, S. A. Corcelli, and J. L. Skinner. Ultrafast vibrational spectroscopy of water and aqueous N-methylacetamide: Comparison of different electronic structure/molecular dynamics approaches. *J. Chem. Phys.*, 121(18):8887–8896, 2004.

[125] TL Jansen and J Knoester. A transferable electrostatic map for solvation effects on amide i vibrations and its application to linear and two-dimensional spectroscopy. *J Chem Phys*, 124(4):044502, Jan 2006.

[126] N. Demirdöven, M. Khalil, O. Golonzka, and A.. Tokmakoff. Correlation effects in the two-dimensional vibrational spectroscopy of coupled vibrations. *J. Phys. Chem. A*, 105:8025–8030, 2001.

[127] W Zhuang, D Abramavicius, T Hayashi, and S Mukamel. Simulation protocols for coherent femtosecond vibrational spectra of peptides. *J Phys Chem B*, 110(7):3362–3374, Jan 2006.

[128] S Hahn, S Ham, and M Cho. Simulation studies of amide I IR absorption and two-dimensional IR spectra of beta hairpins in liquid water. *J Phys Chem B*, 109(23):11789–11801, Jan 2005.

[129] Petr Bour and Thomas A. Keiderling. Vibrational spectral simulation for peptides of mixed secondary structure: Method comparisons with the trpzip model hairpin. *J. Phys. Chem. B*, 109(0):23687–23697, 2005.

[130] C. B. Anfinsen, editor. *Advances in Protein Chemistry*, volume 38. Academic Press, Inc., 1986.

[131] Samuel Krimm and William C. Reisdorf, Jr. Understanding normal modes of proteins. *Faraday Discussions*, 99:181–97, 1994.

[132] Hajime Torii and Mitsuo Tasumi. *Ab initio* molecular orbital study of the amide I vibrational interactions between the peptide groups in di- and tripeptides and considerations on the conformation of the extended helix. *J. Raman Spectrosc.*, 29:81–86, 1998.

[133] Sander Woutersen and Peter Hamm. Nonlinear two-dimensional vibrational spectroscopy of peptides. *J. Phys.: Condens. Matter*, 14:R1035–R1062, 2002.

[134] Roman D. Gorbunov, Daniil S. Kosov, and Gerhard Stock. *Ab initio*-based exciton model of amide i vibrations in peptides: Definition, conformational dependence, and transferability. *J. Chem. Phys.*, 122:224904, 2005.

[135] Steven S. Andrews and Steven G. Boxer. Vibrational Stark effects of nitriles II. Physical origins of Stark effects from experiment and perturbation models. *J. Phys. Chem. A*, 106:469–477, 2002.

[136] Ian T. Suydam and Steven G. Boxer. Vibrational Stark effects calibrate the sensitivity of vibrational probes for electric fields in proteins. *Biochemistry*, 42:12050–12055, 2003.

[137] Joel D. Eaves, Andrei Tokmakoff, and Phillip L. Geissler. Electric field fluctuations drive vibrational dephasing in water. *J. Phys. Chem. A*, 109:9424–9436, 2005.

[138] P Hamm and S Woutersen. Coupling of the amide I modes of the glycine dipeptide. *Bull. Chem. Soc. Jpn.*, 75(5):985–988, 2002.

[139] S Ham, S Cha, JH Choi, and M Cho. Amide I modes of tripeptides: Hessian matrix reconstruction and isotope effects. *J. Chem. Phys.*, 119(3):1451–1461, 2003.

[140] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, , and J. A. Pople. Gaussian 98, revision a.9. Gaussian, Inc., Pittsburgh PA, 1998, 1998.

[141] Victor Mũnoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton. Folding dynamics and mechanism of β-hairpin formation. *Nature*, 390:196–199, 1997.

[142] A. R. Dinner, T. Lazaridis, and M. Karplus. Understanding β-hairpin formation. *Proc. Natl. Acad. Sci. USA*, 96:9068–9073, 1999.

[143] Andrea G. Cochran, Nicholas K. Skelton, and Melissa A. Starovasnik. Tryptophan zippers: Stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci. USA*, 98:5578–5583, 2001.

[144] Guanghong Wei, Phillippe Derreumaux, and Normand Mousseau. Sampling the complex energy landscape of a simple β-hairpin. *J. Chem. Phys.*, 119:6403–6406, 2003.

[145] R. Brian Dyer, Shelia J. Maness, Eric S. Peterson, Stefan Franzen, R. Matthew Fesinmeyer, and Niels H. Andersen. The mechanism of β-hairpin formation. *Biochemistry*, 43:11560–11566, 2004.

[146] Jed W. Pitera, Imran Haque, and William C. Swope. Absence of reptation in the high-temperature folding of the trpzip2 β-hairpin peptide. *J. Chem. Phys.*, 124:141102, 2006.

[147] C. Y. Huang, J. W. Klemke, Z. Getahun, W. F. DeGrado, and F. Gai. Temperature-dependent helix-coil transition of an alanine based peptide. *J. Am. Chem. Soc.*, 123:9235–9238, 2001.

[148] Susan Marqusee, Virginia H. Robbins, and Robert L. Baldwin. Unusually stable helix formation in short alanine-based peptides. *Proc. Natl. Acad. Sci. USA*, 86:5286–5290, 1989.

[149] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[150] DOV Alonso and V Daggett. Molecular dynamics simulations of hydrophobic collapse of ubiquitin. *Protein Sci*, 7(4):860–874, Jan 1998.

[151] Arend G. Dijkstra and Jasper Knoester. Collective oscillations and the linear and two-dimensional infrared spectra of inhomogeneous β-sheets. *J. Phys. Chem. B*, 109:9787–9798, 2005.

[152] Oren M. Becker, Alexander D. MacKerell, Jr., Benoît Roux, and Masakatsu Watanabe, editors. *Computational Biochemistry and Biophysics.* Marcel Dekker, Inc., 2001.

[153] Alexander D. MacKerell, Jr. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.*, 25:1584–1604, 2004.

[154] Jian Zhang, Meng Qin, and Wei Wang. Folding mechanism of β-hairpins studied by replica exchange molecular simulations. *Proteins*, 62:672–685, 2006.

[155] Peter G. Bolhuis. Kinetic pathways of β-hairpin (un)folding in explicit solvent. *Biophysical Journal*, 88:50–61, 2005.

[156] J. Tsai and M. Levitt. Evidence of turn and salt bridge contributions to β-hairpin stability: MD simulations of C-terminal fragment from the B1 domain of protein G. *Biophys. Chem.*, 101-102:187–201, 2002.

[157] Pieter Rein ten Wolde and David Chandler. Drying-induced hydrophobic polymer collapse. *Proc. Natl. Acad. Sci. USA*, 99:6539–6543, 2002.

[158] Margaret S. Cheung, Angel E. García, and José N. Onuchic. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. USA*, 99:685–690, 2002.

[159] Young Min Rhee, Eric J. Sorin, Guha Jayachandran, Erik Lindahl, and Vijay S. Pande. Simulations of the role of water in the protein-folding mechanism. *Proc. Natl. Acad. Sci. USA*, 101:6456–6461, 2004.

[160] Michael R. Shirts and Vijay S. Pande. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.*, 122:134508, 2005.

[161] Hui Lu and Klaus Schulten. The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophys. J.*, 79:51–65, 2000.

[162] R. B. Gerber, G. M. Chaban, S. K. Gregurick, and B. Brauer. Vibrational spectroscopy and the development of new force fields for biological molecules. *Biopolymers*, 68:370–382, 2003.

[163] Thomas la Cour Jansen, Wei Zhuang, and Shaul Mukamel. Stochastic Liouville equation simulation of multidimensional vibrational line shapes of trialanine. *J. Chem. Phys.*, 121(21):10577–10598, December 2004.

[164] Tomoyuki Hayashi, Wei Zhuang, and Shaul Mukamel. Electrostatic DFT map for the complete vibrational amide band of NMA. *J. Phys. Chem. A*, 109:9747–9759, 2005.

[165] Tim M. Watson and Jonathan D. Hirst. Theoretical studies of the amide I vibrational frequencies of [Leu]-enkephalin. *Mol. Phys.*, 103:1531–1546, 2005.

[166] AG Cochran, NJ Skelton, and MA Starovasnik. Tryptophan zippers: Stable, monomeric beta-hairpins. *P Natl Acad Sci Usa*, 98(10):5578–5583, Jan 2001.

[167] Adam W. Smith, Joshua Lessing, Ziad Ganim, Chunte Sam Peng, Andrei Tokmakoff, Santanu Roy, Thomas L. C. Jansen, and Jasper Knoester. Melting of a β-hairpin peptide using isotope-edited 2D IR spectroscopy and simulations. *Submitted*, 2010.

[168] Matthew F DeCamp, Lauren P Deflores, JM McCracken, Andrei Tokmakoff, K Kwac, and M Cho. Amide I vibrational dynamics of N-methylacetamide in polar solvents: The role of electrostatic interactions. *J Phys Chem B*, 109(21):11016–11026, Jan 2005.

[169] JW Pitera, I Haque, and WC Swope. Absence of reptation in the high-temperature folding of the trpzip2 beta-hairpin peptide. *J Chem Phys*, 124(14):141102, Jan 2006.

[170] John D Chodera, Nina Singhal, Vijay S Pande, Ken A Dill, and William C Swope. Automatic discovery of metastable states for the construction of markov models of macromolecular conformational dynamics. *J Chem Phys*, 126(15):155101, Jan 2007.

[171] Thomas la Cour Jansen, Arend G Dijkstra, Tim M Watson, Jonathan D Hirst, and Jasper Knoester. Modeling the amide I bands of small peptides. *J Chem Phys*, 125(4):044312, Jan 2006.

[172] Thomas L. C Jansen and Jasper Knoester. Waiting time dynamics in two-dimensional infrared spectroscopy. *Accounts Chem Res*, 42(9):1405–1411, Jan 2009.

[173] Thomas la Cour Jansen and Jasper Knoester. Nonadiabatic effects in the two-dimensional infrared spectra of peptides: Application to alanine dipeptide. *J Phys Chem B*, 110(45):22910–22916, Jan 2006.

[174] S Ham, JH Kim, H Lee, and MH Cho. Correlation between electronic and molecular structure distortions and vibrational properties. II. amide I modes of NMA-nD_2O complexes. *J Chem Phys*, 118(8):3491–3498, Jan 2003.

[175] CM Dobson, A Sali, and M Karplus. Protein folding: A perspective from theory and experiment. *Angew Chem Int Edit*, 37(7):868–893, Jan 1998.

[176] WA Eaton, V Munoz, SJ Hagen, GS Jas, LJ Lapidus, ER Henry, and J Hofrichter. Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.*, 29:327–359, Jan 2000.

[177] M Gruebele. Protein folding: the free energy surface. *Curr Opin Struc Biol*, 12(2):161–168, Jan 2002.

[178] JE Shea and CL Brooks. From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annu Rev Phys Chem*, 52:499–535, Jan 2001.

[179] CD Snow, N Nguyen, VS Pande, and M Gruebele. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature*, 420(6911):102–106, Jan 2002.

[180] AR Fersht and V Daggett. Protein folding and unfolding at atomic resolution. *Cell*, 108(4):573–582, Jan 2002.

[181] MT Zanni and RM Hochstrasser. Two-dimensional infrared spectroscopy: a promising new method for the time resolution of structures. *Curr Opin Struc Biol*, 11(5):516–522, Jan 2001.

[182] P Mukherjee, I Kass, I Arkin, and MT Zanni. Picosecond dynamics of a membrane protein revealed by 2D IR. *P Natl Acad Sci Usa*, 103(10):3528–3533, Jan 2006.

[183] S Woutersen, YG Mu, G Stock, and P Hamm. Subpicosecond conformational dynamics of small peptides probed by two-dimensional vibrational spectroscopy. *P Natl Acad Sci Usa*, 98(20):11254–11258, Jan 2001.

[184] RH Callender, RB Dyer, R Gilmanshin, and WH Woodruff. Fast events in protein folding: The time evolution of primary processes. *Annu Rev Phys Chem*, 49:173–202, Jan 1998.

[185] M Gruebele. The fast protein folding problem. *Annu Rev Phys Chem*, 50:485–516, Jan 1999.

[186] H Roder and MCR Shastry. Methods for exploring early events in protein folding. *Curr Opin Struc Biol*, 9(5):620–626, Jan 1999.

[187] WA Eaton. Searching for "downhill scenarios" in protein folding. *P Natl Acad Sci Usa*, 96(11):5897–5899, Jan 1999.

[188] EC Fulmer, P Mukherjee, AT Krummel, and MT Zanni. A pulse sequence for directly measuring the anharmonicities of coupled vibrations: Two-quantum two-dimensional infrared spectroscopy. *J Chem Phys*, 120(17):8067–8078, Jan 2004.

[189] Joseph J Loparo, Sean T Roberts, and Andrei Tokmakoff. Multidimensional infrared spectroscopy of water. I. vibrational dynamics in two-dimensional IR line shapes. *J Chem Phys*, 125(19):194521, Jan 2006.

[190] Hoi Sung Chung, Munira Khalil, Adam W Smith, and Andrei Tokmakoff. Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics. *Rev. Sci. Instrum.*, 78(6):063101, Jan 2007.

[191] A Tokmakoff. Two-dimensional line shapes derived from coherent third-order nonlinear spectroscopy. *J Phys Chem A*, 104(18):4247–4255, Jan 2000.

[192] MS Briggs and H Roder. Early hydrogen-bonding events in the folding reaction of ubiquitin. *P Natl Acad Sci Usa*, 89(6):2017–2021, Jan 1992.

[193] E Larios, JS Li, K Schulten, H Kihara, and M Gruebele. Multiple probes reveal a native-like intermediate during low-temperature refolding of ubiquitin. *Journal of Molecular Biology*, 340(1):115–125, Jan 2004.

[194] NJ Marianayagam and SE Jackson. The folding pathway of ubiquitin from all-atom molecular dynamics simulations. *Biophys Chem*, 111(2):159–171, Jan 2004.

[195] HM Went and SE Jackson. Ubiquitin folds through a highly polarized transition state. *Protein Eng Des Sel*, 18(5):229–237, Jan 2005.

[196] TR Sosnick, RS Dothager, and BA Krantz. Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. *P Natl Acad Sci Usa*, 101(50):17377–17382, Jan 2004.

[197] M Jourdan and MS Searle. Cooperative assembly of a nativelike ubiquitin structure through peptide fragment complexation: Energetics of peptide association and folding. *Biochemistry-Us*, 39(40):12355–12364, Jan 2000.

[198] JPL Cox, PA Evans, LC Packman, DH Williams, and DN Woolfson. Dissecting the structure of a partially folded protein - circular-dichroism and nuclear-magnetic-resonance studies of peptides from ubiquitin. *Journal of Molecular Biology*, 234(2):483–492, Jan 1993.

[199] R Zerella, PA Evans, JMC Ionides, LC Packman, BW Trotter, JP Mackay, and DH Williams. Autonomous folding of a peptide corresponding to the N-terminal beta-hairpin from ubiquitin. *Protein Sci*, 8(6):1320–1331, Jan 1999.

[200] BJ Stockman, A Euvrard, and TA Scahill. Heteronuclear 3-dimensional NMR-spectroscopy of a partially denatured protein - the A-state of human ubiquitin. *J Biomol NMR*, 3(3):285–296, Jan 1993.

[201] L. E. Riechl. *A Moern Course in Statistical Physics*, section S5.C., pages 266–276. John Wiley & Sons, Inc., New York, New York, second edition, 1998.

[202] G. Stokes. On the effect of the internal friction of fluids on the motion of pendulums. *Trans. Cambridge Philos. Soc.*, 9:8–106, 1856.

[203] A. Einstein. On the theory of the brownian movement. *Ann. Physik (Leipzig).*, 19(4):371–381, 1906.

[204] P. Debye. *Polar molecules.* Dover, New York, 1929.

[205] M. V. Smoluchowski. Mathematical theory for the coagulation kinetics of colloidal solutions. *Z. Phys. Chem.*, 92:129–68, 1917.

[206] CJ Camacho and S Vajda. Protein-protein association kinetics and protein docking. *Curr Opin Struc Biol*, 12(1):36–40, Jan 2002.

[207] FB Sheinerman, R Norel, and B Honig. Electrostatic aspects of protein-protein interactions. *Curr Opin Struc Biol*, 10(2):153–159, Jan 2000.

[208] HS Chan and KA Dill. Protein folding in the landscape perspective: Chevron plots and non-arrhenius kinetics. *Proteins*, 30(1):2–33, Jan 1998.

[209] JD Bryngelson and PG Wolynes. Spin-glasses and the statistical-mechanics of protein folding. *P Natl Acad Sci Usa*, 84(21):7524–7528, Jan 1987.

[210] J Janin and Cyrus Chothia. The structure of protein-protein recognition sites. *Journal of Biological Chemistry*, 265(27):16027–16030, Jan 1990.

[211] C Chothia and J Janin. Principles of protein-protein recognition. *Nature*, 256(5520):705–8, Aug 1975.

[212] S Jones and J M Thornton. Principles of protein-protein interactions. *P Natl Acad Sci Usa*, 93(1):13–20, Jan 1996.

[213] S Jones and JM Thornton. Analysis of protein-protein interaction sites using surface patches. *Journal of Molecular Biology*, 272(1):121–132, Jan 1997.

[214] RP Bahadur, P Chakrabarti, F Rodier, and J Janin. A dissection of specific and non-specific protein - protein interfaces. *Journal of Molecular Biology*, 336(4):943–955, Jan 2004.

[215] C Chothia. Structural invariants in protein folding. *Nature*, 254(5498):304–308, Jan 1975.

[216] F Rodier, RP Bahadur, P Chakrabarti, and J Janin. Hydration of protein-protein interfaces. *Proteins*, 60(1):36–45, Jan 2005.

[217] L Lo Conte, C Chothia, and J Janin. The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology*, 285(5):2177–2198, Jan 1999.

[218] Irina S Moreira, Pedro A Fernandes, and Maria J Ramos. Hot spots-a review of the protein-protein interface determinant amino-acid residues. *Proteins*, 68(4):803–812, Jan 2007.

[219] AA Bogan and KS Thorn. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*, 280(1):1–9, Jan 1998.

[220] R Karlsson and A Falt. Experimental design for kinetic analysis of protein-protein interactions with surface plasmon resonance biosensors. *J Immunol Methods*, 200(1-2):121–133, Jan 1997.

[221] A Szabo, L Stolz, and R Granzow. Surface-plasmon resonance and its use in biomolecular interaction analysis (bia). *Curr Opin Struc Biol*, 5(5):699–705, Jan 1995.

[222] AK Kenworthy. Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy. *Methods*, 24(3):289–296, Jan 2001.

[223] Chun Tang, Junji Iwahara, and G. Marius Clore. Visualization of transient encounter complexes in protein-protein association. *Nature*, 444(7117):383–386, Jan 2006.

[224] PB Crowley and M Ubbink. Close encounters of the transient kind: Protein interactions in the photosynthetic redox chain investigated by NMR spectroscopy. *Accounts Chem Res*, 36(10):723–730, Jan 2003.

[225] R Ishima and DA Torchia. Protein dynamics from NMR. *Nat Struct Biol*, 7(9):740–743, Jan 2000.

[226] FHC Crick. The packing of alpha-helices - simple coiled-coils. *Acta Crystallogr*, 6(8-9):689–697, Jan 1953.

[227] I Halperin, BY Ma, H Wolfson, and R Nussinov. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, 47(4):409–443, Jan 2002.

[228] Nelly Andrusier, Efrat Mashiach, Ruth Nussinov, and Haim J Wolfson. Principles of flexible protein-protein docking. *Proteins*, 73(2):271–289, Jan 2008.

[229] Sergey Lyskov and Jeffrey J Gray. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res*, 36:W233–W238, Jan 2008.

[230] SR Comeau, DW Gatchell, S Vajda, and CJ Camacho. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res*, 32:W96–W99, Jan 2004.

[231] Andrey Tovchigrechko and Ilya A Vakser. GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res*, 34:W310–W314, Jan 2006.

[232] D Schneidman-Duhovny, Y Inbar, R Nussinov, and HJ Wolfson. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res*, 33:W363–W367, Jan 2005.

[233] CJ Camacho and S Vajda. Protein docking along smooth association pathways. *P Natl Acad Sci Usa*, 98(19):10636–10641, Jan 2001.

[234] O Lichtarge and ME Sowa. Evolutionary predictions of binding surfaces and interactions. *Curr Opin Struc Biol*, 12(1):21–27, Jan 2002.

[235] ZJ Hu, BY Ma, H Wolfson, and R Nussinov. Conservation of polar residues as hot spots at protein interfaces. *Proteins*, 39(4):331–342, Jan 2000.

[236] AS Aytuna, A Gursoy, and O Keskin. Prediction of protein-protein interactions by combining structure and sequence conservation in protein interfaces. *Bioinformatics*, 21(12):2850–2855, Jan 2005.

[237] Anna C. F Lewis, Ramazan Saeed, and Charlotte M Deane. Predicting protein-protein interactions in the context of protein evolution. *Mol. BioSyst.*, 6(1):55–64, Jan 2010.

[238] Sjoerd J de Vries and Alexandre M. J. J Bonvin. How proteins get in touch: Interface prediction in the study of biomolecular complexes. *Curr Protein Pept Sc*, 9(4):394–406, Jan 2008.

[239] J Janin, K Henrick, J Moult, L Ten Eyck, MJE Sternberg, S Vajda, I Vasker, and SJ Wodak. CAPRI: A critical assessment of predicted interactions. *Proteins*, 52(1):2–9, Jan 2003.

[240] Huan-Xiang Zhou and Sanbo Qin. Interaction-site prediction for protein complexes: a critical assessment. *Bioinformatics*, 23(17):2203–2209, Jan 2007.

[241] Lakes Ezkurdia, Lisa Bartoli, Piero Fariselli, Rita Casadio, Alfonso Valencia, and Michael L Tress. Progress and challenges in predicting protein-protein interaction sites. *Briefings in Bioinformatics*, 10(3):233–246, Jan 2009.

[242] Irina S Moreira, Pedro A Fernandes, and Maria J Ramos. Protein-protein docking dealing with the unknown. *J. Comput. Chem.*, 31(2):317–342, Jan 2010.

[243] Carles Pons, Solene Grosdidier, Albert Solernou, Laura Perez-Cano, and Juan Fernandez-Recio. Present and future challenges and limitations in protein-protein docking. *Proteins*, 78(1):95–108, Jan 2010.

[244] Kevin Wiehe, Matthew W Peterson, Brian Pierce, Julian Mintseris, and Zhiping Weng. Protein-protein docking: overview and performance analysis. *Methods Mol Biol*, 413:283–314, Jan 2008.

[245] S Vajda and CJ Camacho. Protein-protein docking: is the glass half-full or half-empty? *Trends Biotechnol*, 22(3):110–116, Jan 2004.

[246] CJ Camacho, DW Gatchell, SR Kimura, and S Vajda. Scoring docked conformations generated by rigid-body protein-protein docking. *Proteins*, 40(3):525–537, Jan 2000.

[247] JD Bryngelson, JN Onuchic, ND Socci, and PG Wolynes. Funnels, pathways, and the energy landscape of protein-folding - a synthesis. *Proteins*, 21(3):167–195, Jan 1995.

[248] Alexander B Sigalov. Protein intrinsic disorder and oligomericity in cell signaling. *Mol. BioSyst.*, 6(3):451–461, Jan 2010.

[249] B A Shoemaker, J J Portman, and P G Wolynes. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *P Natl Acad Sci Usa*, 97(16):8868–73, Aug 2000.

[250] R Koren and G G Hammes. A kinetic study of protein-protein interactions. *Biochemistry-Us*, 15(5):1165–71, Mar 1976.

[251] Michael A Weiss, DT Nguyen, I Khait, K Inouye, BH Frank, M Beckage, E Oshea, SE Shoelson, Martin Karplus, and LJ Neuringer. Two-dimensional NMR and photo-CIDNP studies of the insulin monomer: Assignment of aromatic resonances with application to protein folding, structure, and dynamics. *Biochemistry-Us*, 28(25):9855–9873, Jan 1989.

[252] Q X Hua, M Kochoyan, and M A Weiss. Structure and dynamics of des-pentapeptide-insulin in solution: the molten-globule hypothesis. *P Natl Acad Sci Usa*, 89(6):2379–83, Mar 1992. RCSB 1HIS monomer.

[253] D Keller, R Clausen, K Josefsen, and J J Led. Flexibility and bioactivity of insulin: an nmr investigation of the solution structure and folding of an unusually flexible human insulin mutant with increased biological activity. *Biochemistry-Us*, 40(35):10732–40, Sep 2001. RCSB 1JCO monomer.

[254] Vincent Zoete, Markus Meuwly, and Martin Karplus. A comparison of the dynamic behavior of monomeric and dimeric insulin shows structural rearrangements in the active monomer. *Journal of Molecular Biology*, 342(3):913–29, Sep 2004.

[255] A Budi, S Legge, H Treutlein, and I Yarovsky. Effect of external stresses on protein conformation: a computer modelling study. *Eur Biophys J*, 33(2):121–9, Apr 2004.

[256] Wojciech Bocian, Jerzy Sitkowski, Elzbieta Bednarek, Anna Tarnowska, Robert Kawecki, and Lech Kozerski. Structure of human insulin monomer in water/acetonitrile solution. *J Biomol NMR*, 40(1):55–64, Jan 2008.

[257] Qing xin Hua and Michael A Weiss. Mechanism of insulin fibrillation: the structure of insulin under amyloidogenic conditions resembles a protein-folding intermediate. *J Biol Chem*, 279(20):21449–60, May 2004. RCSB 1SF1 monomer.

[258] Q X Hua, S N Gozani, R E Chance, J A Hoffmann, B H Frank, and M A Weiss. Structure of a protein in a kinetic trap. *Nat Struct Biol*, 2(2):129–38, Feb 1995. RCSB 2HIU (formerly 1HIU) 1XGL monomer.

[259] J Janin. The kinetics of protein-protein recognition. *Proteins*, 28(2):153–61, Jun 1997.

[260] H A Gabb, R M Jackson, and M J Sternberg. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology*, 272(1):106–20, Sep 1997.

[261] JM Creeth. Sedimentation and diffusion studies on insulin: the maximum molecular weight. *Biochem J*, 53(1):41, 1953.

[262] Peter D Jeffrey and JH Coates. An equilibrium ultracentrifuge study of the effect of ionic strength on the self-association of bovine insulin. *Biochemistry-Us*, 5(12):3820–3824, 1966.

[263] AH Pekar and BH Frank. Conformation of proinsulin. comparison of insulin and proinsulin self-association at neutral pH. *Biochemistry-Us*, 11(22):4013–4016, 1972.

[264] CM Yu, CY Chin, EI Franses, and NHL Wang. In situ probing of insulin aggregation in chromatography effluents with spectroturbidimetry. *Journal of colloid and interface science*, 299(2):733–739, 2006.

[265] E J Nettleton, P Tito, M Sunde, M Bouchard, C M Dobson, and C V Robinson. Characterization of the oligomeric states of insulin in self-assembly and amyloid fibril formation by mass spectrometry. *Biophysical Journal*, 79(2):1053–65, Aug 2000.

[266] R S Lord, F Gubensek, and J A Rupley. Insulin self-association. spectrum changes and thermodynamics. *Biochemistry-Us*, 12(22):4385–4392, Jan 1973.

[267] Stefan Grudzielanek, Ralf Jansen, and Roland Winter. Solvational tuning of the unfolding, aggregation and amyloidogenesis of insulin. *Journal of Molecular Biology*, 351(4):879–94, Aug 2005.

[268] Wojciech Bocian, Jerzy Sitkowski, Anna Tarnowska, Elzbieta Bednarek, Robert Kawecki, Wiktor Kozminski, and Lech Kozerski. Direct insight into insulin aggregation by 2D NMR complemented by PFGSE NMR. *Proteins*, 71(3):1057–1065, May 2008.

[269] U Hassiepen, M Federwisch, T Mülders, V J Lenz, H G Gattner, P Krüger, and A Wollmer. Analysis of protein self-association at constant concentration by fluorescence-energy transfer. *Eur J Biochem*, 255(3):580–7, Aug 1998.

[270] Y Pocker and Subhasis B Biswas. Conformational dynamics of insulin in solution. circular dichroic studies. *Biochemistry-Us*, 19(22):5043–5049, Jan 1980.

[271] Steve Strazza, Roger Hunter, Elbert Walker, and Dennis W Darnall. The thermodynamics of bovine and porcine insulin and proinsulin association determined by concentration difference spectroscopy. *Arch Biochem Biophys*, 238(1):30–42, Jan 1985.

[272] Wojciech Dzwolak, R Ravindra, J Lendermann, and Roland Winter. Aggregation of bovine insulin probed by DSC/PPC calorimetry and FTIR spectroscopy. *Biochemistry-Us*, 42(38):11347–11355, Jan 2003.

[273] M Bouchard, J Zurdo, EJ Nettleton, CM Dobson, and CV Robinson. Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Sci*, 9(10):1960–1967, Jan 2000.

[274] U Derewenda, Z Derewenda, E J Dodson, G G Dodson, C D Reynolds, G D Smith, C Sparks, and D Swenson. Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer. *Nature*, 338(6216):594–6, Apr 1989.

[275] G David Smith, Walter A Pangborn, and Robert H Blessing. The structure of T6 human insulin at 1.0 A resolution. *Acta Crystallogr D Biol Crystallogr*, 59(Pt 3):474–82, Mar 2003. RCSB 1MSO hexamer.

[276] M Roy, M L Brader, R W Lee, N C Kaarsholm, J F Hansen, and M F Dunn. Spectroscopic signatures of the T to R conformational transition in the insulin hexamer. *J Biol Chem*, 264(32):19081–5, Nov 1989.

[277] X Chang, A M Jorgensen, P Bardrum, and J J Led. Solution structures of the R6 human insulin hexamer. *Biochemistry-Us*, 36(31):9409–22, Aug 1997. RCSB 1AI0 hexamer.

[278] Qingxin Hua and Michael A Weiss. Towards the solution structure of human insulin: Sequential 2D 1H NMR assignment of a des-pentapeptide analogue and comparison with crystal structure. *Biochemistry-Us*, 29(46):10545–10555, Jan 1990.

[279] Anne Jørgensen, Søren Kristensen, Jens Led, and Per Balschmidt. Three-dimensional solution structure of an insulin dimer : A study of the B9(Asp) mutant of human insulin using nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *Journal of Molecular Biology*, 227(4):1146–1163, Oct 1992. RCSB 1MHI dimer.

[280] S Ludvigsen, M Roy, H Thøgersen, and N C Kaarsholm. High-resolution structure of an engineered biologically potent insulin monomer, B16 Tyr–¿His, as determined by nuclear magnetic resonance spectroscopy. *Biochemistry-Us*, 33(26):7998–8006, Jul 1994. RCSB 1HLS monomer.

[281] Q X Hua, S Q Hu, B H Frank, W Jia, Y C Chu, S H Wang, G T Burke, P G Katsoyannis, and M A Weiss. Mapping the functional surface of insulin by design: structure and function of a novel A-chain analogue. *Journal of Molecular Biology*, 264(2):390–403, Nov 1996. RCSB 1VKT monomer.

[282] E Lindahl, B Hess, and D van der Spoel. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model*, 7(8):306–317, Jan 2001.

[283] H J C Berendsen, D van der Spoel, and R van Drunen. GROMACS - a message-passing parallel molecular-dynamics implementation. *Comput Phys Commun*, 91(1-3):43–56, Jan 1995.

[284] U Essmann, L Perera, M L Berkowitz, T Darden, H Lee, and L G Pedersen. A smooth particle mesh ewald method. *J Chem Phys*, 103(19):8577–8593, Jan 1995.

[285] S Miyamoto and PA Kollman. SETTLE - an analytical version of the shake and rattle algorithm for rigid water models. *J. Comput. Chem.*, 13(8):952–962, Jan 1992.

[286] HJC Berendsen, JPM Postma, WF VanGunsteren, A Dinola, and JR Haak. Molecular-dynamics with coupling to an external bath. *J Chem Phys*, 81(8):3684–3690, Jan 1984.

[287] B Hess, H Bekker, HJC Berendsen, and JGEM Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, Jan 1997.

[288] F Eisenhaber, P Lijnzaad, P Argos, C Sander, and M Scharf. The double cubic lattice method - efficient approaches to numerical-integration of surface-area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.*, 16(3):273–284, Jan 1995.

[289] WL Jorgensen, DS Maxwell, and J TiradoRives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc*, 118(45):11225–11236, Jan 1996.

[290] WL Jorgensen and NA McDonald. Development of an all-atom force field for heterocycles. properties of liquid pyridine and diazenes. *Theochem-J Mol Struc*, 424(1-2):145–155, Jan 1998.

[291] NA McDonald and WL Jorgensen. Development of an all-atom force field for heterocycles. properties of liquid pyrrole, furan, diazoles, and oxazoles. *J Phys Chem B*, 102(41):8049–8059, Jan 1998.

[292] RC Rizzo and WL Jorgensen. OPLS all-atom model for amines: Resolution of the amine hydration problem. *J Am Chem Soc*, 121(20):4827–4836, Jan 1999.

[293] MLP Price, D Ostrovsky, and WL Jorgensen. Gas-phase and liquid-state properties of esters, nitriles, and nitro compounds with the OPLS-AA force field. *J. Comput. Chem.*, 22(13):1340–1352, Jan 2001.

[294] EK Watkins and WL Jorgensen. Perfluoroalkanes: Conformational analysis and liquid-state properties from ab initio and monte carlo calculations. *J Phys Chem A*, 105(16):4118–4125, Jan 2001.

[295] GA Kaminski, RA Friesner, J Tirado-Rives, and WL Jorgensen. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B*, 105(28):6474–6487, Jan 2001.

[296] D Eisenberg and AD McLachlan. Solvation energy in protein folding and binding. *Nature*, 319(6050):199–203, Jan 1986.

[297] HJC Berendsen, JR Grigera, and TP Straatsma. The missing term in effective pair potentials. *J Phys Chem-Us*, 91(24):6269–6271, Jan 1987.

[298] Hajime Torii and Mitsuo Tasumi. Application of the three-dimensional doorway-state theory to analyses of the amide-I infrared bands of globular proteins. *J Chem Phys*, 97(1):92–98, Jan 1992.

[299] Wojciech Dzwolak, Stefan Grudzielanek, V Smirnovas, R Ravindra, C Nicolini, Ralf Jansen, A Loksztejn, S Porowski, and Roland Winter. Ethanol-perturbed amyloidogenic self-assembly of insulin: Looking for origins of amyloid strains. *Biochemistry-Us*, 44(25):8948–8958, Jan 2005.

[300] W J Becktel and J A Schellman. Protein stability curves. *Biopolymers*, 26(11):1859–77, Nov 1987.

[301] Qing xin Hua and Michael A Weiss. Comparative 2D NMR studies of human insulin and des-pentapeptide insulin: Sequential resonance assignment and implications for protein dynamics and receptor recognition. *Biochemistry-Us*, 30(22):5505–5515, Jan 1991.

[302] K Huus, S Havelund, HB Olsen, M van de Weert, and S Frokjaer. Thermal dissociation and unfolding of insulin. *Biochemistry-Us*, 44(33):11171–11177, Jan 2005.

[303] Noel T Southall, Ken A Dill, and A D J Haymet. A view of the hydrophobic effect. *J. Phys Chem. B*, 106(3):521–533, Sep 2002.

[304] E Gallicchio, MM Kubo, and RM Levy. Enthalpy-entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation. *J Phys Chem B*, 104(26):6271–6285, Jan 2000.

[305] José Goldman and Frederick H Carpenter. Zinc binding, circular dichroism, and equilibrium sedimentation studies on insulin (bovine) and several of its derivatives. *Biochemistry-Us*, 13(22):4566–4574, Jan 1974.

[306] Ziad Ganim, Kevin C Jones, and Andrei Tokmakoff. Insulin dimer dissociation and unfolding revealed by amide i two-dimensional infrared spectroscopy. *Phys Chem Chem Phys*, 12(14):3579–3588, Jan 2010.

[307] VJ Lenz, HG Gattner, M Leithauser, D Brandenburg, A Wollmer, and H Hocker. Proteolyses of a fluorogenic insulin derivative and native insulin in reversed micelles monitored by fluorescence emission, reversed-phase high-performance liquid-chromatography, and capillary zone electrophoresis. *Anal Biochem*, 221(1):85–93, Jan 1994.

[308] U Hassiepen, M Federwisch, T Mulders, VL Lenz, HG Gattner, P Kruger, and A Wollmer. Analysis of protein self-association at constant concentration by fluorescence-energy transfer. *Eur J Biochem*, 255(3):580–587, Jan 1998.

[309] Susanne H Mollmann, Lene Jorgensen, Jens T Bukrinsky, Ulla Elofsson, Willem Norde, and Sven Frokjaer. Interfacial adsorption of insulin conformational changes and reversibility of adsorption. *European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences*, 27(2-3):194–204, Feb 2006.

[310] Joseph R. Lakowicz. *Principles of Fluorescence Spectroscopy, 3rd Ed.* Springer, New York, 2006.

[311] Jeff Bigham. Fluorophores and their amine-reactive derivatives, May 2010. http://www.invitrogen.com.

[312] V Sluzky, JA Tamada, AM Klibanov, and R Langer. Kinetics of insulin aggregation in aqueous-solutions upon agitation in the presence of hydrophobic surfaces. *P Natl Acad Sci Usa*, 88(21):9377–9381, Jan 1991.

[313] T Ha, T Enderle, DF Ogletree, DS Chemla, PR Selvin, and S Weiss. Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *P Natl Acad Sci Usa*, 93(13):6264–6268, Jan 1996.

[314] Innocent B Bekard and Dave E Dunstan. Tyrosine autofluorescence as a measure of bovine insulin fibrillation. *Biophysical Journal*, 97(9):2521–31, Nov 2009.

[315] L Nielsen, S Frokjaer, Jens Brange, VN Uversky, and AL Fink. Probing the mechanism of insulin fibril formation with insulin mutants. *Biochemistry-Us*, 40(28):8397–8409, Jan 2001.

[316] W Wiczk, A Rzeska, J Lukomska, K Stachowiak, J Karolczak, J Malicka, and L Lankiewicz. Mechanism of fluorescence quenching of tyrosine derivatives by amide group. *Chem Phys Lett*, 341(1-2):99–106, Jan 2001.

[317] M Noronha, JC Lima, P Lamosa, H Santos, C Maycock, R Ventura, and AL Macanita. Intramolecular fluorescence quenching of tyrosine by the peptide alpha-carbonyl group revisited. *J Phys Chem A*, 108(12):2155–2166, Jan 2004.

[318] Greg S Harms, Steve W Pauls, John F Hedstrom, and Carey K Johnson. Tyrosyl fluorescence decays and rotational dynamics in tyrosine monomers and in dipeptides. *Journal of Fluorescence*, 7(4):273–282, Feb 1997.

[319] Jay R Unruh, Mangala Roshan Liyanage, and Carey K Johnson. Tyrosyl rotamer interconversion rates and the fluorescence decays of N-acetyltyrosinamide and short tyrosyl peptides. *J Phys Chem B*, 111(19):5494–5502, Jan 2007.

[320] G Grabner, G Kohler, J Zechner, and N Getoff. Temperature-dependence of photoprocesses in aqueous phenol. *J Phys Chem-Us*, 84(23):3000–3004, Jan 1980.

[321] Joseph R. Lakowicz. *Principles of Fluorescence Spectroscopy, 3rd Ed.* Springer, New York, 2006.

[322] JR Lakowicz and G Weber. Quenching of fluorescence by oxygen - probe for structural fluctuations in macromolecules. *Biochemistry-Us*, 12(21):4161–4170, Jan 1973.

[323] JR Lakowicz and BP Maliwal. Oxygen quenching and fluorescence depolarization of tyrosine residues in proteins. *Journal of Biological Chemistry*, 258(8):4794–4801, Jan 1983.

Ziad Ganim - Curriculum Vitae

CONTACT INFORMATION	Department of Chemistry Massachusetts Institute of Technology Cambridge, MA Email: ziadg@mit.edu
NATIONALITY	Born January 26, 1981, Stony Brook, New York
EDUCATION	University of California, Berkeley B.S., Chemistry (2003, High Honors) Massachsetts Institute of Technology Ph.D., Physical Chemistry (2010)
AWARDS AND HONORS	UCB Departmental Citation in Chemistry, Spring 2003 Outstanding Teaching Assistant, Spring 2004 (MIT 5.33 and 5.62) Vivian A. and E. Emerson Morse Travel Grant, Fall 2006 Lester Wolfe Fellowship, Spring 2007 Daniel S. Kemp Fellowship, Summer 2008 Poitras Predoctoral Fellowship, Fall 2008 - Fall 2009
INVITED SEMINARS	University of Groningen, Materials Science Centre, May 2006 Massachusetts Institute of Technology, Department of Chemistry, April 2009
CONTRIBUTED SEMINARS AND POSTERS AT NATIONAL MEETINGS	American Chemical Society Undergraduate Research Meeting, May 2002 Western Spectroscopy Association Conference, January 2003 Vibrational Spectroscopy Gordon Research Conference, July 2004 American Physical Society March Meeting, March 2006 The 3rd Int'l Conference on Coherent Multidimensional Spectroscopy, May 2006 American Physical Society March Meeting, March 2007 The Protein Society 21st Annual Symposium, July 2007 Biophysical Society Annual Meeting, February 2008 Biophysical Society Annual Meeting, February 2009 The 14th Int'l Conference on Time-Resolved Vibrational Spectroscopy, May 2009
TEACHING EXPERIENCE	UCB 137A Quantum Mechanics I, Summer 2002 (Grading and Solutions) MIT 5.33 Advanced Chem. Experimentation and Instrumentation Lab, Fall 2003 (TA) MIT 5.62 Physical Chem. II, Spring 2004 (TA, Notes transcribed for OpenCourseWare) MIT 5.74 Introductory Quantum Mechanics II, Spring 2005 (Grading and Solutions) MIT 5.33 Advanced Chem. Experimentation and Instrumentation Lab, Fall 2005 (TA) MIT 5.73 Introductory Quantum Mechanics I, Fall 2005 (Grading and Solutions, Transcribed for OCW) MIT 5.38 Module 11 Time Resolved Molecular Spectroscopy (Laboratory development and TA training)

Contributed Posters at Campus Meetings

MIT Spectroscopy Laboratory Poster Session, January 2004
MIT Spectroscopy Laboratory Poster Session, January 2005
MIT Computational and Systems Biology (CSBi) Oktoberfest, October 2005
MIT Spectroscopy Laboratory Poster Session, January 2006
MIT Computational and Systems Biology (CSBi) Oktoberfest, October 2006
MIT Spectroscopy Laboratory Poster Session, January 2007 (Won Best Science Award)
MIT Spectroscopy Laboratory Poster Session, January 2008
MIT Spectroscopy Laboratory Poster Session, January 2009 (Won Best Science Award)

Professional Memberships

Biophysical Society (2006-), American Physical Society (2006-), American Chemical Society (2007-), Protein Society (2007-), $AX\Sigma$ (2001-)

Peer Reviewed Publications

1. Time-resolved resonance Raman analysis of chromophore structural changes in the formation and decay of rhodopsin's BSI intermediate, D. Pan, Z. Ganim, J. E. Kim, M. A. Verhoeven, J. Lugtenburg, R. A. Mathies, Journal of the American Chemical Society, 124, 17 (2002).

2. Retinal counterion switch in the photoactivation of the G protein-coupled receptor rhodopsin, E. C. Y. Yan, M. A. Kazmi, Z. Ganim, J. M. Hou, D. H. Pan, B. S. W. Chang, T. P. Sakmar, R. A. Mathies, Proceedings of the National Academy of Sciences of the United States of America, 100, 16 (2003).

3. Resonance Raman analysis of the mechanism of energy storage and chromophore distortion in the primary visual photoproduct, E. C. Y. Yan, Z. Ganim, M. A. Kazmi, B. S. W. Chang, T. P. Sakmar, R. A. Mathies, Biochemistry, 43, 34 (2004).

4. Conformational changes during the nanosecond-to-millisecond unfolding of ubiquitin, H. S. Chung, M. Khalil, A. W. Smith, Z. Ganim, A. Tokmakoff, Proceedings of the National Academy of Sciences of the United States of America, 102, 3 (2005).

5. Residual native structure in a thermally denatured β-hairpin, A. W. Smith, H. S. Chung, Z. Ganim, A. Tokmakoff, Journal of Physical Chemistry B, 109, 36 (2005).

6. The anharmonic vibrational potential and relaxation pathways of the amide I and II modes of N-methylacetamide, L. P. DeFlores, Z. Ganim, S. F. Ackley, H. S. Chung, A. Tokmakoff, Journal of Physical Chemistry B, 110, 38 (2006).

7. Spectral Signatures of Hetereogeneous Protein Ensembles Revealed by MD Simulations of 2DIR Spectra, Z. Ganim, A. Tokmakoff, Biophysical Journal, 91, 7 (2006).

8. Transient 2DIR spectroscopy of ubiquitin unfolding dynamics, H. S. Chung*, Z. Ganim*, K. Jones, A. Tokmakoff, Proceedings of the National Academy of Sciences of the United States of America, 104, 36 (2007).

9. Amide I 2D IR Spectroscopy of Proteins, Z. Ganim, H. S. Chung, A. W. Smith, L. P. DeFlores, K. C. Jones, A. Tokmakoff, Acc. Chem. Res., 41, 432 (2008).

10. Amide I-II 2D IR spectroscopy provides enhanced protein secondary structural sensitivity, L. P. DeFlores, Z. Ganim, R. A. Nicodemus, A. Tokmakoff, J. Am. Chem. Soc., 131, 3385 (2009).

11. Heterodyne-Detected Dispersed Vibrational Echo Spectroscopy, K. C. Jones, Z. Ganim, A. Tokmakoff, J. Phys. Chem. A., 113, 14060 (2010).

12. Insulin Dimer Dissociation and Unfolding Revealed by Amide I Two-Dimensional Infrared Spectroscopy, Z. Ganim, K. C. Jones, A. Tokmakoff, Phys. Chem. Chem. Phys., 2010, 12, 3579 (2010).

*Both authors contributed equally to this work.

CONFERENCE PROCEEDINGS

1. Time-resolved resonance Raman microchip measurements of chromophore structural changes in the Batho-to-BSI and BSI-to-Lumi transitions, D. H. Pan, Z. Ganim, J. E. Kim, R. A. Mathies, Abstracts of Papers of the Americal Chemical Society, 223, 375 (2002).
2. Transient 2D IR spectroscopy of ubiquitin unfolding dynamics, H. S. Chung, Z. Ganim, A. W. Smith, K. C. Jones, A. Tokmakoff, Biophysical Journal Suppl. 207A, S (2007).