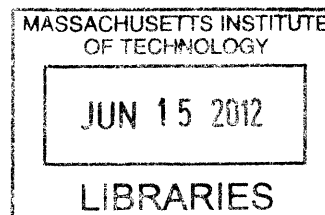


Temperature-Jump 2D IR Spectroscopy to Study Protein Conformational Dynamics

by

Kevin C. Jones

B.S. Chemistry
Yale University, 2005



ARCHIVES

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

© Massachusetts Institute of Technology. All rights reserved.

Signature of Author: _____

Kevin C. Jones
May 8, 2012

Certified by: _____

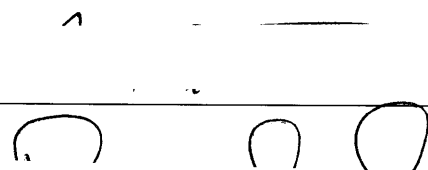
Andrei Tokmakoff
Professor of Chemistry
Thesis Supervisor

Accepted by: _____


Robert W. Field
Haslam and Dewey Professor of Chemistry
Chairman, Departmental Committee on Graduate Studies

This doctoral thesis has been examined by a committee of the Department of Chemistry that included:

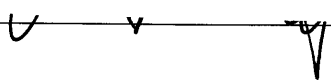
Professor Robert W. Field
Chair

Handwritten signature of Professor Robert W. Field, consisting of several loops and a horizontal line.

Professor Robert G. Griffin

Handwritten signature of Professor Robert G. Griffin, featuring a large initial 'R' and a horizontal line.

Professor Andrei Tokmakoff
Thesis Supervisor

Handwritten signature of Professor Andrei Tokmakoff, with a stylized 'A' and a horizontal line.

Temperature-Jump 2D IR Spectroscopy to Study Protein Conformational Dynamics

by
Kevin C. Jones

Submitted to the Department of Chemistry on May 8, 2012
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Abstract

Temperature-jump (T-jump) two-dimensional infrared spectroscopy (2D IR) is developed, characterized, and applied to the study of protein folding and association. In solution, protein conformational changes span a wide range of timescale from nanoseconds to minutes. Ultrafast 2D IR spectroscopy measures time-dependent structural changes within the protein ensemble by probing the frequency changes associated with amide I backbone vibrations. Combining 2D IR with a perturbing laser-induced T-jump enables the study of conformational dynamics from 5 ns to 50 ms. To access a finer time-sampling of the conformational evolution, a one-dimensional variant of 2D IR, heterodyne-detected dispersed vibrational echo spectroscopy (HDVE), is implemented. The framework for interpreting transient HDVE and 2D IR spectra is developed, and we propose a method to remove the linear absorption distortions along both frequency axes. We first present the T-jump 2D IR spectra of a dipeptide to reveal the general amide I baseline response expected in the absence of conformational change. To facilitate the analysis of T-jump data, singular value decomposition (SVD) is employed for reducing noise, identifying the number of distinguishable states, and separating spectral changes based on shared timescales. Finally, T-jump 2D IR spectroscopy is applied to study the unfolding of ubiquitin, disordering of the 12-residue β -hairpin peptide trpzip2 (TZ2), and the dissociation of insulin dimers to monomers. Experimental results for ubiquitin highlight the importance of linear absorption corrections for interpretation of the data. In response to the T-jump, 2D IR results indicate β -sheet structure melts in ubiquitin with a small amplitude ($\sim 10 \mu\text{s}$) and large amplitude (17 ms) response. Isotope-labeling T-jump experiments on TZ2 allow for the proposal of a free energy surface in which transitions from a native and misfolded state proceed through a disordered hub-like state with a 1-2 μs timescale. Multiple timescales are observed in the T-jump induced dissociation of insulin. Based on their spectral features and concentration dependence, the insulin timescales can be assigned to dissociation, disordering, and oligomerization processes. With these applications, we demonstrate the capability of T-jump 2D IR spectroscopy to reveal detailed molecular dynamics.

Thesis Supervisor: Andrei Tokmakoff
Title: Professor of Chemistry

Acknowledgements

Seven years is a long time. Although certain events stick out, the overpowering memory is a blur of the stereotypical day: bouncing back and forth between lab and the office, eating lunch with labmates, going to seminar, lowering the blinds next to my desk at ~3 PM when the sun pokes up over building 4, and finally removing my coat from the coat tree and leaving the office either to a chorus of "see-you-later" or the click of the light switch.

I spent the majority of my time in the big lab. The green and white tiles, the faint glow of laser light from the OPA and HeNe, the hiss of the dry air, the feel of the stainless steel laser table, the flashing of the labview programs, and the whirring of the T-jump laser are sensations that have become strangely comforting. Working in lab has always been a challenge – any given day can end with the overwhelming despair of failure or the euphoria of getting the experiment to work. Although the disappointing, frustrating days certainly outnumber the others, the feeling of success is rewarding enough that it has begun to dominate my memory. In the end, for me, regardless of the underlying driving science, the laser system and experiment are both challenging and elegant enough to justify themselves.

Andrei has been encouraging through all of the frustrating days, and I thank him for pushing me. It took four years of tinkering to get the T-jump experiment into the state we wanted, and Andrei never gave up on me. His constant pursuit of understanding the science at the heart of problems is an inspiration, as is the trust that he shows for his group.

As I look around the office, I am reminded of the wonderful people with whom I have worked. Matt DeCamp was the first to show me how to align a laser. Despite my tripping on a BNC and wrenching a camera out of place in the middle of a long night of data taking, he was always patient with me. Hoi Sung Chung built the original T-jump spectrometer, and, while working with it, I have been nothing but impressed by his creation. Aside from his excellent work, Hoi Sung was extremely kind and helpful in passing on knowledge. Although I never worked with him in the lab, Joe Loparo was instrumental in keeping spirits up in the office, and I am glad to see the tradition of lightheartedness continue today. Poul Petersen was a great friend who often provided perspective on life as a graduate student. He sets a good balance between personal and professional life, and I fondly remember his good-natured cheerfulness. Although only around for one year, I enjoyed working in lab with Benjamin Dietzek, who pointed out the simple joy of watching the FID oscillations mapped out by labview. With Adam Smith, I learned about the day-to-day operation of the big lab. I have always been impressed with his focus, conceiving a project and making it happen quickly. I also appreciate his music suggestions. Lauren DeFlores was always willing to talk, and from her I learned the importance of beautiful figures and being precise in lab. I marvel at how much work Sean Roberts could do despite never appearing to eat or sleep. Aside from doing difficult simulations and experiments, Sean was always willing to discuss 2D IR theory, video games, recent novels, or music. I spent countless days in lab with Ziad Ganim, and through patiently answering my sometimes half-formed questions, I learned a great deal from him. I appreciate his guidance and advice. I overlapped almost my entire time with Becky Nicodemus, who was always welcoming and inclusive. Aside from making me laugh, she helped show me the way through grad school. Josh Lessing and I

entered and finished at the same time, and I am glad that he has been around to help every step of the way. He has always stood up for grad students' rights and reminded me that there is a world that exists beyond MIT. Krupa Ramasesha is a considerate labmate who has unflinchingly put up with a lot of my nonsense and distractions around the office. She is a great scientist and has a wonderful sense of humor. Recently, I have spent most of my time in lab with Sam Chunte Peng, whose questions have enriched my own understanding. His comments and thoughts have improved all of the work that I have done, and I look forward to the future of the big lab in his hands. Andrew Horning has always been helpful either through his googling skills or computer fixing knowledge. His talents extend far beyond that (he's particularly good at wrapping objects in aluminum foil), and he has a bright and exciting future lying ahead of him. Aritra Mandal brings a great perspective, and I have enjoyed learning about water and cricket from him. Luigi De Marco has borne the brunt of the most recent office shenanigans and returned in kind. I appreciate his laughter. Mike Reppert also has an infectious laugh, and his uncompromising pursuit of understanding the underlying theory inspires me. Carlos Baiz arrived right before I left lab to begin writing up, and his presence has made the transition smooth and easy. He has provided extremely helpful comments on my recent work, and I am excited to see which directions he takes the lab. Finally, in Paul Stevenson I see the excitement (and some disillusionment) of a first year student discovering the Tokmakoff lab and IR spectroscopy for the first time. Although there is a long road ahead, I can tell he is better prepared than I was.

Thankfully, my time at MIT has not been spent entirely in the lab. I would like to express my gratitude for all my friends from the entering class of 2005 (all of whom graduated before me!), the friends who transitioned from Yale to Boston with me, the friends I have met in the basement from the Field and Nelson groups, and, in particular, my roommates from 275 Putnam (Michael, Dan, Scott, Patrick, and Nicole).

Finally, I want to thank Amanda, who has been entirely supportive from the moment we met. With you I have enjoyed running, eating your wonderful cooking, dancing, traveling, exploring, chatting, drinking, watching too much television, and laughing. I do not think I would have made it without you. My younger sister Maddie spent a wonderful summer in Boston, and I cherished the time we had together. I am glad she has put up with my sporadic calling, my stubborn unwillingness to try new foods, and my new found interest in getting medical diagnoses for every affliction, real or imaginary. Carrie has always been the perfect older sister, giving advice and guidance ever since we visited MIT together in 2005. Luckily, she and her husband Brian have been close by the entire time, and their house has always felt like a home-away-from-home for me. Lastly, I want to thank my Mom and Dad (and their dog, Eli). Their unwavering encouragement throughout has meant the world to me. Knowing that I can call them or hop on a bus and see them at any time has sustained me through too many tough weeks. Thank you.

Farewell.

Contents

List of Figures

List of Appendix Figures

List of Tables

1	Introduction	15
1.1	Introduction	15
1.2	Protein Folding.....	16
1.2.1	Thermodynamics.....	17
1.2.2	Folding Pathways.....	19
1.2.3	Relevant Timescales	20
1.3	Infrared Spectroscopy	22
1.3.1	Amide I spectroscopy	22
1.3.2	Isotope labels	24
1.3.3	2D IR spectroscopy.....	25
1.3.4	Transient spectroscopy.....	27
1.3.5	Challenges and advantages	27
1.4	Thesis outline	28
1.5	References	29
2	2D IR Background	31
2.1	Introduction	31
2.2	Linear IR spectroscopy.....	32
2.2.1	Theory	32
2.2.2	Experiment and Detection.....	35
2.2.3	Discussion	37
2.3	2D IR spectroscopy background	38
2.3.1	Theory	38
2.3.2	Specific 2D IR measurements.....	42
2.3.3	Detection	43
2.3.4	Discussion	48
2.3.5	Discussion of amide I spectra	53
2.4	Conclusion.....	53
2.5	References	55
3	2D IR Experiment	56
3.1	Introduction	56
3.2	2D IR equilibrium optical setup	57
3.2.1	800 nm Generation.....	57
3.2.2	6 μm generation	59
3.2.3	HeNe Overlap	60
3.2.4	Interferometer Box.....	61
3.2.5	Sample Box	63
3.2.6	Timing up pulses.....	65
3.2.7	Timing up the LO.....	66
3.2.8	Collecting equilibrium pump-probe spectra.....	66

3.2.9	Collecting equilibrium 2D IR spectra	67
3.2.10	Processing equilibrium 2D IR spectra	68
3.3	T-jump t-2D IR optical setup	70
3.3.1	2 μm generation	71
3.3.2	T-jump profile	73
3.3.3	Data collection	75
3.3.4	T-jump laser induced interferometric fluctuations	76
3.4	Interpreting the T-jump data	78
3.4.1	t-2D IR spectral features	78
3.4.2	Times dependence of the signal	78
3.5	Conclusion.....	87
3.6	References	88
4	Heterodyne-Detected Dispersed Vibrational Echo Spectroscopy	89
4.1	Abstract	89
4.2	Introduction	90
4.3	Signal Detection Methods	91
4.3.1	Heterodyne Detection	91
4.3.2	Separation of amplitude and phase	93
4.4	Experimental Section	96
4.5	Results and Discussion.....	97
4.5.1	Extension to Combination Methods.....	99
4.6	Conclusion.....	103
4.7	Acknowledgment	103
4.8	References	104
5	Transient Two Dimensional Spectroscopy with Linear Absorption Corrections Applied to Temperature Jump 2D IR.....	106
5.1	Abstract	106
5.2	Introduction	107
5.3	Experimental	109
5.3.1	Mid-Infrared Spectroscopy	109
5.3.2	Temperature-Jump	110
5.3.3	Sample Preparation	112
5.4	Transient Data Collection and Processing	112
5.4.1	Theoretical Background.....	112
5.4.2	Data Collection	115
5.4.3	Post Processing	117
5.4.4	t-2D IR	118
5.4.5	t-HDVE.....	119
5.5	Results and Discussion.....	119
5.5.1	Diglycine.....	120
5.5.2	Trpzip2.....	126
5.6	Conclusion.....	130
5.7	Acknowledgment	131
5.8	References	132
	Appendix 5.A : Equilibrium Temperature Dependence.....	135
6	Singular Value Decomposition.....	139

6.1	Introduction	139
6.2	Methods and Discussion.....	140
6.2.1	SVD application and the state and mathematical bases	140
6.2.2	Information content and the intermediate basis	144
6.2.3	Calculating intermediate basis U, V from mathematical basis U, S, V	150
6.2.4	Information content for a two component system	154
6.2.5	Two component re-weighting applied to t-DPP and t-2D IR	156
6.2.6	Hypothesizing the state basis Σ_{ijk} and Z_{ijk} from t-DPP and t-2D IR	165
6.2.7	Comparison of re-weighting to projections	169
6.2.8	Shortcomings of SVD	172
6.3	Conclusion.....	174
6.4	Acknowledgment	175
6.5	References	176
	Appendix 6.A : List of Variables	178
	Appendix 6.B : Special case of $\mathbf{u}_j \propto u_i$	180
	Appendix 6.C : Transformation matrices, W	182
	Appendix 6.D : Re-Weighting applied to structure determination	183
	Appendix 6.E : Temperature dependence of $[I(\tau)]$ and $[I'(\tau)]$	186
	Appendix 6.F : Kinetic Models.....	193
7	Equilibrium and Transient Unfolding of Ubiquitin	195
7.1	Introduction	195
7.1.1	Background.....	196
7.2	Experimental	198
7.3	Results and Discussion.....	199
7.3.1	Equilibrium 2D IR spectra	199
7.3.2	Transient DVE spectra.....	202
7.3.3	Transient DPP spectra.....	204
7.3.4	SVD analysis of t-DPP spectra	206
7.3.5	Transient 2D IR spectra	209
7.3.6	SVD analysis of t-2D IR spectra.....	212
7.3.7	SVD analysis of t-transmission spectra	214
7.3.8	Comparison of results	217
7.4	Conclusion.....	218
7.5	References	221
8	Transient Studies of Isotope-Labeled Trpzip2.....	223
8.1	Introduction	223
8.2	Methods.....	226
8.2.1	TZ2 Samples	226
8.2.2	Optical setup	226
8.2.3	SVD Analysis.....	227
8.3	Results and Interpretation of Experimental Spectra.....	228
8.3.1	Equilibrium 2D IR Spectra	228
8.3.2	Equilibrium structural assignment	231
8.3.3	Transient dispersed pump-probe (t-DPP) Spectra	232

8.3.4	Singular value decomposition (SVD) into spectral components	235
8.3.5	t-2D IR – ns response.....	236
8.3.6	t-2D IR – μ s response	240
8.4	Discussion	242
8.4.1	Assignment of Kinetic States.....	242
8.4.2	Mechanism	246
8.5	Conclusion.....	248
8.6	Acknowledgment	248
8.7	References	249
	Appendix 8.A : Transient 2D IR spectra.....	253
	Appendix 8.B : SVD Analysis and Hypothesized Spectra.....	254
	Appendix 8.C : Transient Absorption	267
9	T-Jump Studies of Insulin	269
9.1	Introduction	269
9.2	Background	271
9.2.1	Kinetics of molecular association	271
9.2.2	Insulin Dimer Formation.....	274
9.3	Methods.....	277
9.4	Results	280
9.4.1	Equilibrium Experiments	280
9.4.2	Transient Experiments	281
9.4.3	Separation of Spectra According to Timescale.....	284
9.4.4	SVD Separation	286
9.4.5	Separated Timescales.....	287
9.4.6	Separated Spectra.....	289
9.5	Discussion	293
9.5.1	Mechanism	296
9.6	Conclusion.....	300
9.7	Acknowledgment	301
9.8	References	302
	Appendix 9.A : Kinetic Models	307
	Appendix 9.B : SVD separation of <i>kinetic phases</i>	312
	Appendix 9.C : Fitting a two-state model	318

Curriculum Vitae

List of Figures

Figure 1.1: Relevant timescales for biomolecular folding and association processes.....	21
Figure 1.2: The amide I vibration	23
Figure 1.3: The crystal structure, FTIR, and 2D IR spectra of selected small proteins...	26
Figure 2.1: Inhomogeneous and homogeneous broadening in FTIR and 2D IR spectra.	35
Figure 2.2: FTIR spectrometer based on a Michelson interferometer.	36
Figure 2.3: The discussed time delays.....	39
Figure 2.4: The boxcar geometry.....	40
Figure 2.5: The S_{2D} , probed energy level structure, and stick diagram for RDC	50
Figure 3.1: The full laser setup	57
Figure 3.2: The triggering timing signals	58
Figure 3.3: The HeNe overlap	60
Figure 3.4: The interferometer.....	62
Figure 3.5: The sample box	64
Figure 3.6: LO time determination through interferometric cross-correlations	67
Figure 3.7: $S_{2D}(\omega_3, \tau_2=150 \text{ fs}, \tau_1, \tau_{LO}=0 \text{ fs})$ slices for 20 mg/mL ubiquitin.....	69
Figure 3.8: Different representations of the 2D surfaces for 20 mg/mL ubiquitin	71
Figure 3.9: The projections of the S_R+S_{NR} 2D surfaces onto the ω_3 axis	72
Figure 3.10: The time delays between pulses	73
Figure 3.11: Parameters from a stretched exponential fit to the transient transmission time traces shown in Figure 3.12	74
Figure 3.12: The transient transmission time traces and selected difference spectra at five different initial temperatures.....	74
Figure 3.13: The transient data collection.....	76
Figure 3.14: Modulations in the interferometrically detected signal	77
Figure 3.15: Expected transient difference spectra.....	79
Figure 3.16: Convolution of exponentials with the temperature profile.....	81
Figure 3.17: Two-state kinetic model signal changes.....	83
Figure 3.18: Signal amplitude dependence on convolution.....	84
Figure 3.19: Mapping of convolutions onto underlying exponentials	85
Figure 3.20: Deconvolution of the signals predicted by the kinetic models.....	87
Figure 4.1: The four methods for calculating the complex heterodyned.....	94
Figure 4.2: The linear absorbance, linear phase, DVE, and DPP spectra.....	98
Figure 4.3: A: The NMA HDVE and DVE peak intensities versus concentration	100
Figure 4.4: Insulin FTSI HDVE amplitude peak ratios versus concentration	101
Figure 4.5: 15 mg/ml ubiquitin single-shot DVE and DPP spectra.....	102
Figure 5.1: The sample box	110
Figure 5.2: Time delays between T-jump and IR pulses	111
Figure 5.3: The absorptive spectra of t-2D IR of diglycine with different corrections .	120
Figure 5.4: Difference linear spectra for diglycine in D_2O and correction factors.....	122
Figure 5.5: 20 mg/mL diglycine DPP slices with different corrections	123
Figure 5.6: DPP diglycine time traces	124
Figure 5.7: TZ2 DPP and DVE transient spectra from t-HDVE measurements	126
Figure 5.8: The TZ2 time dependent trace and exponential fit	127
Figure 5.9: The absorptive TZ2 t-2D IR spectra at selected time delays	128

Figure 5.10: TZ2 uncorrected difference DPP and 2D IR spectra.....	129
Figure 6.1: Depiction of the mathematical basis SVD terms	142
Figure 6.2: Eq. (6.3) is depicted using the SVD mathematical basis	143
Figure 6.3: Two model basis Lorentzian $U_{1,2}$ spectra, $V_{1,2}$ traces, and model S spectra	147
Figure 6.4: The first three mathematical U and V SVD results from Figure 6.3	148
Figure 6.5: Illustration of the 1 st step of transformation shown in eq (6.6).....	152
Figure 6.6: Illustration of the 2 nd step of transformation shown in eq (6.6).....	153
Figure 6.7: The information content of a two-component system.....	155
Figure 6.8: Diagram of the two-component SVD re-weighting	156
Figure 6.9: Exaggerated illustration of the free-energy intra- and inter- well states	159
Figure 6.10: The transformation process and assumptions for a two-state case.....	160
Figure 6.11: The three-state and four-state cases equations	162
Figure 6.12: The three-state T-jump kinetic model results.....	163
Figure 6.13: The four-state T-jump kinetic model results	164
Figure 6.14: The $p_{n,\mu}$ projection results from the S shown in Figure 6.3	171
Figure 6.15: The S peak-shift data and resulting SVD singular values	171
Figure 6.16: Mathematical basis U_{1-4} and V_{1-4} SVD results from the peakshift	172
Figure 6.17: The $S \Pi(\omega)$ data and resulting SVD singular values.....	173
Figure 6.18: Mathematical basis U_{1-4} and V_{1-4} SVD results from the $S \Pi(\omega)$	174
Figure 7.1: The structure of ubiquitin.....	196
Figure 7.2: Cartoon of the downhill unfolding scenario.....	198
Figure 7.3: The ubiquitin equilibrium 2D IR spectra	200
Figure 7.4 2 nd SVD component ubiquitin ‘melting curve’ from 2D IR spectra.....	201
Figure 7.5: Ubiquitin t-DVE spectra.....	202
Figure 7.6: Ubiquitin ω_3 -corrected t-DPP spectra calculated from t-HDVE.....	204
Figure 7.7: $U_{1,2,3}$ and $V_{1,2,3}$ SVD components from the ubiquitin t-DPP.....	206
Figure 7.8: The ubiquitin t-2D IR without absorption correction.....	210
Figure 7.9: The difference t-2D IR spectra after ω_3 -correction	211
Figure 7.10: The ubiquitin SVD results from t-2D IR and t-DPP	213
Figure 7.11: The transient transmission spectra of ubiquitin.....	214
Figure 7.12: SVD mathematical basis results from ubiquitin transient transmission...	215
Figure 7.13: SVD intermediate basis results from ubiquitin transient transmission	216
Figure 8.1: The structure of TZ2 showing native registry of cross-strand H-bonds	225
Figure 8.2: The TZ2 2D IR spectra simulated from select markov states	229
Figure 8.3: Equilibrium FTIR and 2D IR TZ2 spectra	230
Figure 8.4: TZ2 t-DPP spectra at selected T-jump delay times.....	232
Figure 8.5: Time traces from data shown in Figure 8.4.....	233
Figure 8.6: SVD intermediate basis results for TZ2	236
Figure 8.7: Proposed Free Energy Surface	243
Figure 9.1: Different proposed mechanisms for association of proteins A and B.....	270
Figure 9.2: Different insulin structures determined by X-ray crystallography or NMR	274
Figure 9.3: Window mode analysis of insulin's FTIR spectrum	276
Figure 9.4: Sampled temperature and concentration range for insulin.....	277
Figure 9.5: Convolutions of exponentials with the temperature profile	279
Figure 9.6: The temperature dependent K_D and predicted monomer fraction.....	280
Figure 9.7: Equilibrium and difference insulin temperature dependent 2D IR spectra.	281

Figure 9.8: Insulin difference ZZYY t-2D IR spectra	282
Figure 9.9: Insulin difference ZZZZ t-2D IR spectra	282
Figure 9.10: Insulin difference ZZZZ t-DPP spectra.....	284
Figure 9.11: Insulin difference ZZZZ t-DPP traces <i>after</i> subtracting the temperature-tracking response	285
Figure 9.12: SVD analysis of the $T_i = 30^\circ\text{C}$ ZZZZ t-DPP spectra	287
Figure 9.13: The deconvolved exponential time constants for components <i>b</i> , <i>c</i> , and <i>d</i>	288
Figure 9.14: The magnitude of components <i>b</i> and <i>c</i> versus temperature	289
Figure 9.15: SVD analysis of the 10 mg/mL, $T_i = 40^\circ\text{C}$ t-DPP and t-2D IR spectra	290
Figure 9.16: SVD analysis of the 30 mg/mL, $T_i = 35^\circ\text{C}$ t-DPP spectra	291
Figure 9.17: The spectra and time-dependence for 10 mg/mL samples.....	294
Figure 9.18: SVD analysis of the 10 mg/mL, $T_i = 40^\circ\text{C}$, 40% EtOD t-DPP and t-2D IR spectra	295
Figure 9.19: A small subset of simple possible dimerization mechanisms	297

List of Appendix Figures

Figure 5.A.1: SVD results from temperature-dependent diglycine 2D IR spectra.....	135
Figure 5.A.2: t-2D IR spectra of 1% NMA in D ₂ O	136
Figure 5.A.3: SVD results from temperature-dependent NMA 2D IR spectra	137
Figure 6.D.1: SVD applied to a model set of spectra with varying secondary structure content.....	184
Figure 6.E.1: Eq. (6.E.9) is compared to the results from the three-state case kinetic model.....	188
Figure 6.E.2: Eq. (6.E.21) is compared to the results from the three-state case kinetic model.....	191
Figure 8.A.1: The t-2D IR difference data for TZ2 isotopologues.....	253
Figure 8.B.1: Flow chart depicting the SVD process and transformation.....	254
Figure 8.B.2: The first three UL TZ2 t-DPP U , S , V components	255
Figure 8.B.3: The singular values, S , that result from SVD.....	256
Figure 8.B.4: The hypothesized 2D IR $S_{F0/M0}$, S_{F/M_b} and S_D spectra and $P_{F0/M0}$, P_{F/M_b} and P_D population traces.....	263
Figure 8.B.5: Varying c_n and c_μ for hypothesizing spectra	265
Figure 8.C.1: SVD analysis of the TZ2 t-abs. difference data	268
Figure 9.B.1: Separating the spectra and associated timescales	312
Figure 9.B.2: SVD analysis of the $T_i = 30^\circ\text{C}$ ZZZZ t-DPP spectra.....	315
Figure 9.B.3: SVD analysis of the 10 mg/mL insulin, $T_i = 30^\circ\text{C}$ ZZZZ t-DPP spectra	316
Figure 9.C.1: The de-convolved observed rates from the b proces.....	318
Figure 9.C.2: Calculating the activation energy	319
Figure 9.C.3: The calculated time-dependent concentrations, DPP spectra, and 2D IR spectra for the F and F ₂ states	321

List of Tables

Table 8.1: Table of measured rates from indicated TZ2 transient spectra

234

Chapter 1

Introduction

1.1 Introduction

Although the mention of biology, the study of living organisms, usually conjures up memories of experiments aimed at relatively large length-scales – counting cells in microscopes, dissecting worms, modeling predator-prey cycles – the behind-the-scenes reactions that make-up the complex processes collectively referred to as ‘life’ occur at the molecular level on molecularly relevant timescales. While in the gym pumping iron, we tend to focus on the flexing of the biceps rather than the myosin proteins that are actually doing the work of pulling actin filaments in quantized, nanometer-scale contractions¹. Along with often seeing the forest rather than the tree molecules, we also sometimes ascribe desire or free-will to reactions that are governed by nature and physical laws. We can lift weights against gravity not because our consciousness is instructing the myosin proteins to create contractions, but rather because our body makes it favorable for the myosin proteins to react with adenosine triphosphate (ATP) to cause a conformational change that results in contraction. Stochastic diffusion delivers ATP to the myosin protein, which sometimes moves backwards (retraction) despite our desire and the fact that it is energetically better to contract. There is an invisible hand that drives biological reactions in living organisms, but it is the same invisible hand that drives reactions everywhere. By tweaking rates through catalysis or changing concentrations of reactants, our body creates a favorable environment such that the desired reactions proceed in the desired direction on the

correct timescales. Thus, one can argue that the secret of life (and perhaps the universe and everything) is revealed when one understands the thermodynamics and kinetics that determine the favorability and rates of reactions.

The work presented in this thesis is motivated by understanding molecular interactions of proteins in solution on molecularly relevant timescales. Fully characterizing these phenomena requires measuring the time-dependent structures and populations such that we can hypothesize the underlying free energy surface, which provides the answer to why and under which conditions reactions proceed. The described projects are motivated both by fundamental physical chemistry questions related to how protein molecules behave in solution and by biologically relevant questions dealing with how proteins interact and fold.

Measuring protein behavior requires time-dependent sensitivity to structure. By observing the interaction of light with proteins, spectroscopy provides such sensitivity, and it can be used to tease out the molecular level information. Of the different types of spectroscopy, there is a balance between the structural sensitivity and time-resolution. Nuclear magnetic resonance provides atomic resolution, but its time resolution is limited by the frequency of the nuclear spins that it probes. Visible spectroscopy, which tracks excitations of electrons with fast time-resolution, is generally sensitive to a few specific residues within the protein. Although each type of spectroscopy is immensely important for addressing a particular set of problems, we choose multidimensional infrared spectroscopy to study protein structures. Infrared light excites vibrations (and sometimes rotations) within the protein that can be localized to two atoms or extended across the entire molecule when all the atoms are moving in a concerted fashion. The frequency of the vibrations depends on the mass of, and bonds between, atoms. The bond strength is sensitive to the surrounding electric charges, which means that information about the local environment surrounding a particular oscillator is encoded in the frequency of the vibration. Multidimensional infrared spectroscopy, such as two dimensional infrared (2D IR), uses intense pulses to correlate multiple vibrations. Aside from the structural sensitivity, infrared spectroscopy also has the required time resolution, making it an ideal technique for studying the biophysics of protein folding and association.

1.2 Protein Folding

Proteins are involved in almost every biological process from catalyzing reactions as molecular machines and platforms, to signaling the speed up or slow down of pathways, to

serving as the structural infrastructure within cells. Although each protein is created from splicing together amino acid building blocks selected from the same ~20 amino acid library, proteins exist in a wide variety of shapes and sizes. The protein's form is determined by its sequence, and its function is extremely dependent on its form. Questions remain about how proteins reach their final structure and how stable they are once they get there: How do proteins end up in their functional structures despite the staggeringly large set of conformational states accessible to them? How many significantly populated states are there, and what are the timescales for their conversion? What is the mechanism or, stated differently, the pathway of sequential structures through which the protein passes to convert to the native state?

A significant research thrust is focused on answering these questions and characterizing the protein folding processes. Understanding how and why proteins retain their functional forms will allow for better protein engineering and possible therapy for diseases that result from misfolding and aggregation^{2, 3}. Different fields approach the protein folding questions from different directions. Computational experiments interpret molecular dynamics simulations to predict how proteins behave with atomic resolution⁸. Another computational approach is de novo structure prediction, in which our understanding of molecular forces can be refined by attempting to predict the structure of proteins based on their sequence⁹. This is related to the experimental analog, in which de novo design studies attempt to synthesize particular sequences to fold into wanted and predicted structures¹⁰⁻¹². Others perform time-resolved experiments to track the kinetics of protein structural changes following a perturbation¹³. In our approach, time-resolved experiments are performed on fast timescales with structural sensitivity to measure dynamics. We want to distinguish the number of important states, the timescales that separate them, and the mechanism of folding and binding.

1.2.1 Thermodynamics

The relative populations of different structures of a protein are dictated by the delicate balance between entropy, the tendency for the protein to morph into many different configurations, and enthalpy, the energy stabilization created by favorable contacts that trap a protein into a particular structure. Although there are a staggering number of possible conformations, the enthalpy term (and steric hindrance) greatly limits the number of relevant conformations, as is elegantly illustrated through proposed folding funnel energy landscapes^{14, 15}. For many proteins, the balance between large (hundreds of kcal/mol) entropy and enthalpy terms

is close enough (10-70 kcal/mol) that modest changes in temperature or denaturant can cause dramatic changes in the relative populations of states¹⁶⁻¹⁸. In this thesis we rely on the temperature dependence to create measurable changes in these populations.

At the molecular level, the enthalpic term is the result of interactions between the solvent, protein, and ions^{16, 17}. Electrostatics are the dominant long range forces, at distances $> 5 \text{ \AA}$. Side-chains can be charged, as is determined by the pH of the solution, and the repulsive or attractive forces can be shielded by salts and the solvent. Despite earlier hypotheses that over-emphasized intra-protein ion pairing, its scarcity indicates that it is not the dominant stabilizer in proteins. Ion-pairs contribute an estimated 1-3 kcal/mol stabilization per pair^{17, 19}.

At short-range, van der Waals attractions and repulsion and hydrogen bonding are important¹⁷. The prevalent secondary structural forms are stabilized by intra-protein hydrogen bonds (H-bonds), yet it is not clear that protein-protein H-bonds are appreciably energetically stronger than protein-water varieties: H-bonds are on the order of 2-10 kcal/mol^{16, 17}, with protein-protein H-bonds being favored over water-protein H-bonds by 0-2 kcal/mol per residue¹⁸. The similarity in water-protein and protein-protein H-bond energies suggest that although they are important for defining and stabilizing secondary structure, H-bonding may not be the only important stabilizing interaction in proteins.

The hydrophobic effect, the tendency for non-polar side chains to bury themselves within the interior of the protein while polar side chains tend to be solvent exposed, is identified by some as the most dominant protein folding force¹⁷. Polar and non-polar residues have different propensities for being solvent exposed. When a polar residue is solvent exposed, it can H-bond with the solvent, while a non-polar residue cannot. The non-polar residue upsets the natural H-bond network that would persist if the non-polar residue was absent. There are entropic considerations to be made as well – the solvent molecules may make an ordered pocket to surround a non-polar residue which represents a loss in entropy for the solvent. Calculated through transfer studies from organic to water solvents, hydration of side chains is associated with energies that range on the order of 0.1 to -0.4 kcal/g¹⁶.

These energetic terms can be recast in terms of a Gibbs free energy, which defines the distribution of populations in different conformations. At equilibrium, the ratio of populations between two interconverting states, A and B, where



is given by the equilibrium constant:

$$K = \frac{[B]}{[A]} = \frac{k_1}{k_{-1}}. \quad (1.2)$$

The equilibrium constant depends on the Gibbs Free Energy, ΔG :

$$K = \exp\left[-\frac{\Delta G}{RT}\right]. \quad (1.3)$$

ΔG depends on the enthalpy, ΔH , entropy, ΔS , and the change in the heat capacity at constant pressure, ΔC_p . When A and B are taken to be the folded and unfolded state of a protein, for a cooperative folding transition, assuming a temperature-independent ΔC_p results in:

$$\Delta G = \Delta H^\circ - T\Delta S^\circ + \Delta C_p \left[T - T_M - T \ln\left(\frac{T}{T_M}\right) \right] \quad (1.4)$$

where T_M is defined as the temperature at which $\Delta G = 0$ (at T_M , A and B have equal populations)²⁰. For a classic, general two-state protein with an unfolded and folded state, movement from the folded to unfolded state has a negative change in enthalpy, positive change in entropy, and positive ΔC_p . At high temperatures, ΔG dips below zero due to the second term in eq. (1.4). At low temperatures (often below the freezing point of water) there is also a projected low temperature cold-denaturation point due to the third term in eq. (1.4). Many proteins display this behavior, and in the proteins we study, an increase in temperature generally results in an increase in the unfolded state population.

1.2.2 Folding Pathways

Aside from the ratios of populations, which are defined by the thermodynamics described above (assuming equilibrium has been reached), we also want to understand the folding mechanism. Generally, the concept of traditional folding pathways – the sequential series of structures through which a protein moves to transition from a non-native state to a folded state – are antithetical to the statistical folding funnel view¹⁵, in which populations fold through parallel, heteroneneous routes. Discussing potential paradigmatic mechanisms, however, is useful for specific simple cases (perhaps small peptides) and for making connections to previous work.

Proposed protein folding mechanisms differ by the order in which hierarchical structures

form. In the framework model²¹, local sections of the protein form secondary structural elements prior to the eventual coarse-grained collapse into proper tertiary structure. In the hydrophobic collapse model²², the protein collapses in a non-specific way into a molten globule before re-arrangement occurs to form native structures. In the nucleation-condensation model²³, tertiary-contact-stabilized secondary structures nucleate a diffuse, partially-folded structure in which hierarchical elements fold in parallel.

Similar paradigmatic mechanisms have also been proposed for protein association, where binding must also be considered along with folding. In the lock-and-key mechanism, the two interacting species are rigidly folded into complementary, interlocking shapes. In the conformational selection mechanism²⁴, one or both of the species are interconverting between associable and unassociable structures, and only when they both encounter while in the proper conformation does association occur. This is related to the induced-fit mechanism^{25, 26}, in which a partially folded species is induced to fold into its final, associated structure upon encountering its matching species. In the fold-on-contact model, species can fold after they encounter each other, such as in the specific fly casting method²⁷, in which disordered proteins associate and then fold.

The prevalence of paradigmatic mechanisms in our conception of protein folding and association is partially due to experimental limitations. Individual experimental probes can usually discern just a few states, which means that simple mechanisms are adequate to describe the data. Growing sensitivity in experimental techniques and increased computational power for simulations have begun to provide a more complex view of protein folding, in which tens of structurally distinguishable states are populated with significant populations²⁸. It is this ‘Markov state’ view that is adopted for this thesis, in which structurally and kinetically defined states interconvert with predicted exponential rates dependent on the separating barriers²⁹, as described in Eq. (1.5) below.

1.2.3 Relevant Timescales

An important piece of understanding protein folding pathways and mechanisms is identifying the timescales of interconversion. The relevant timescales for different biomolecular, condensed phase processes are summarized in Figure 1.1. Collective H-bond rearrangement in water sets a good standard for the fastest local motions expected in biomolecules at 10^{-12} s³⁰. Rotational diffusion of small proteins is on the 10^{-9} s timescales, suggesting a rough lower

timescale limit for diffusion of protein subdomains. Actual measured folding rates range from 10^{-7} s out to minutes or hours⁶. The range of timescales indicates that some special, fast folding proteins potentially fold at diffusion-limited timescales, an exciting and unusual phenomenon.

The kinetic rate equation that describes the movement of population over a barrier of height E_a is given by:

$$k = k_0 \exp\left(\frac{-E_a}{RT}\right) \tag{1.5}$$

where k_0 is the prefactor. For the fast-folding proteins and peptides with negligible activation barriers, the equation reduces to the diffusional, ‘downhill’ limit given by the prefactor. This prefactor has been estimated as being the # residues/100 x 1 μ s, although this is a rough estimate⁶. To track changes on these relevant timescales, we require a technique with fast enough

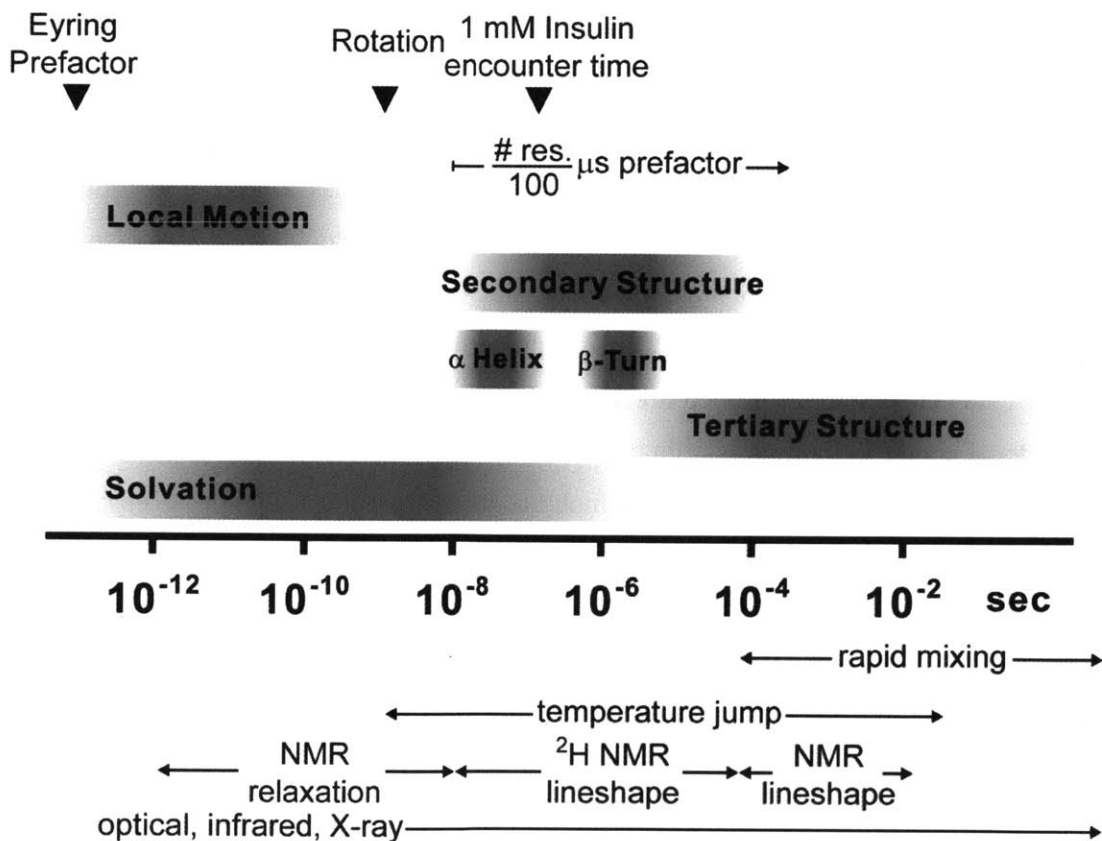


Figure 1.1: Relevant timescales for biomolecular folding and association processes, adapted from Ref. 4. The top row indicates diffusional limits: the Eyring prefactor (gas phase), protein rotation, encounter time for 1mM insulin dimers in solution⁴, and the (# residues)/100 μ s empirical prefactor⁶. Protein structural timescales⁶ are studied with a variety of techniques⁷ (bottom).

time-resolution.

1.3 Infrared Spectroscopy

Infrared spectroscopy provides the time-resolution and structural sensitivity necessary to characterize protein folding processes on the timescales identified in Figure 1.1. Infrared spectroscopy probes inherent vibrations, which makes it non-intrusive. Other techniques, such as fluorescence, often require tagging the protein with dye molecules, which may disrupt the molecule of interest. The local vibrations monitored have periods on the order of 20 fs, which sets a theoretical lower limit on the time resolution. The pulse-width of the interrogating laser pulse raises the lower limit to ~90 fs. For temperature-jump experiments, which are used in this thesis to expand the accessible timescales, the time-resolution is instead defined by the ~1 ps dephasing time of amide I vibrations, which sets the theoretical time-resolution of the T-jump technique, although experimental limitations imposed by the temperature-jump initiation may decrease the time-resolution.

1.3.1 Amide I spectroscopy

Relative to the other inherent protein vibrations, amide I has a large transition dipole moment of $\sim 4 D \text{ \AA}^{-1} \text{ amu}^{-1/2}$ per oscillator³¹, which makes it a popular mode to study with infrared spectroscopy. In the laboratory, however, the weakness of infrared sources and the noisiness of detectors dictates protein concentrations normally on the 5 mg/mL level, which is much higher than concentrations needed for fluorescence studies.

The local amide I vibration is primarily due to C=O stretching and a small amount of N-H in-plane bend, as is shown in Figure 1.2, which occurs in the 1600-1700 cm^{-1} (1667 $\text{cm}^{-1} = 6 \mu\text{m}$) region. The predominant C=O stretch character makes the vibration sensitive to H-bonding. Each H-bond is expected to shift the C=O stretching frequency down by ~16 wavenumbers³². The local site energy of the localized stretch is also affected by surrounding charges, as is indicated by simulations⁵, and it is relatively insensitive to side-chains³³. In a protein with $N + 1$ residues, there are N local amide I vibrations.

The structural sensitivity of the amide I vibration is due to the coupling between nearby oscillators, which creates delocalized vibrations whose frequencies differ based on the secondary structure element in which they reside. This structural sensitivity has been empirically known and has recently been modeled with an excitonic model which reveals the importance of nearest-

neighbor coupling and inter-strand coupling^{31, 33-35}. The empirical frequencies associated with particular secondary structural elements are shown in Figure 1.2.

Within an anti-parallel β -sheet, amide I oscillators are arranged as shown in the bottom left corner of Figure 1.2. Coupling between individual local vibrators creates characteristic delocalized normal modes that are visualized at the bottom right of Figure 1.2, where the phase of the local vibration is either blue (0°) or red (180°), and the participation ratio of the unit oscillation is indicated by the shading. Summing the local dipoles in-phase perpendicular to the

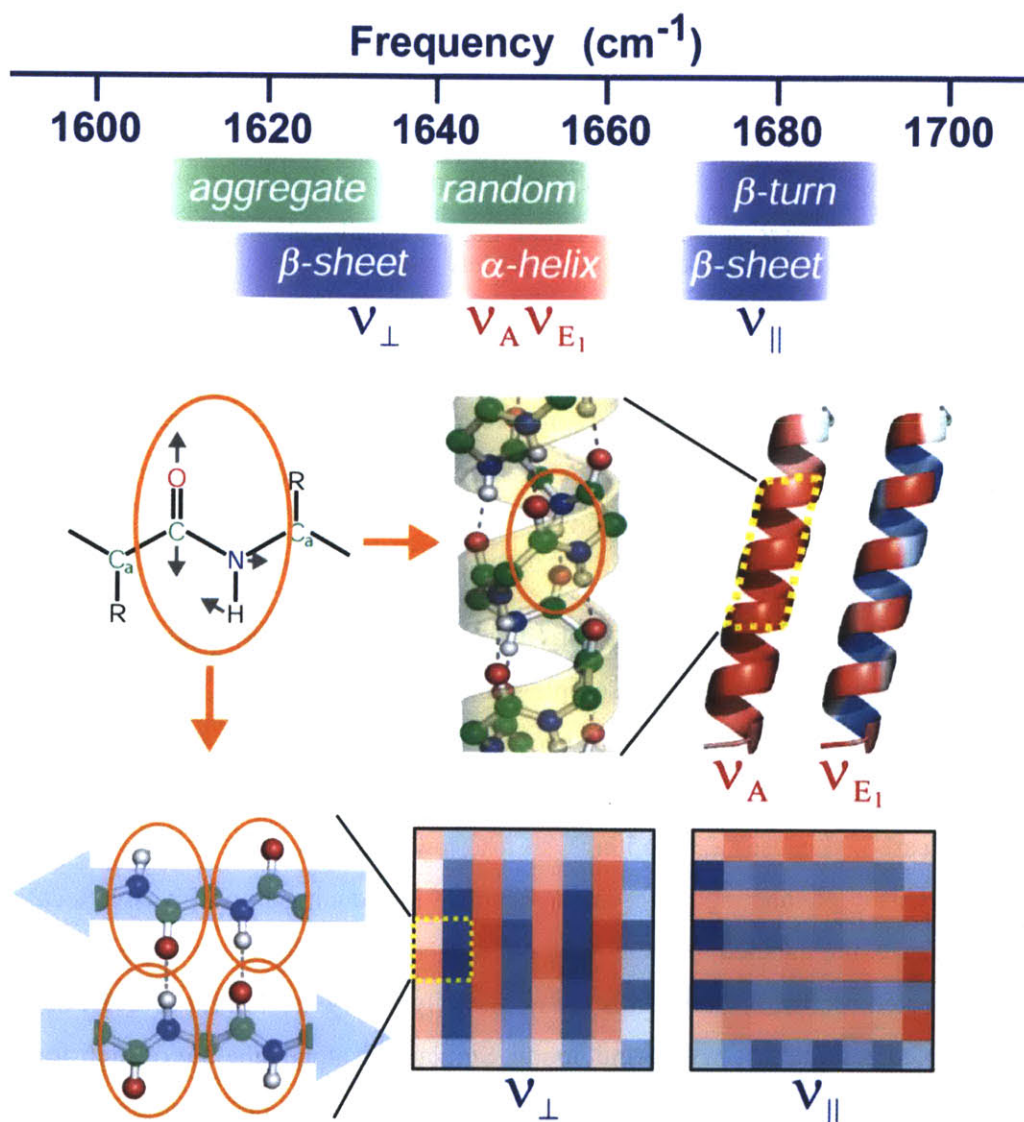


Figure 1.2: The frequencies, local modes, and normal modes associated with the structurally sensitive amide I vibration, reproduced from Ref. 5.

β -strand (each oscillator along the strand is out-of-phase with its neighbor) results in a strong normal mode dipole that is perpendicular to the β -strand, a mode called ν_{\perp} , whose $\sim 1630\text{ cm}^{-1}$ frequency depends on the size of the β -sheet³³⁻³⁵. Summing the local dipoles in-phase parallel to the β -strand results in a weaker normal mode dipole that is parallel to the β -strand, a mode called ν_{\parallel} , whose $\sim 1680\text{ cm}^{-1}$ frequency less-sensitively depends on the size of the β -sheet³⁵. Although the thought experiment of summing of the dipoles within the unit cell is a simplification, it is consistent with excitonic normal mode calculations, and it describes the character of the underlying vibrations. ν_{\perp} and ν_{\parallel} are characteristic normal modes whose presence is indicative of β -sheet secondary structure.

The other characteristic common normal mode vibrations are those for the α -helix, which has a strong A mode vibration, in which each oscillator vibrates in-phase, and a weaker set of degenerate E1 modes⁵. The frequency for these vibrations is $\sim 1650\text{ cm}^{-1}$. Disordered structures lacking canonical secondary structure generally oscillate with a frequency of $1640\text{-}1650\text{ cm}^{-1}$.

Because a vibration of H_2O overlaps with the amide I vibration, solution amide I experiments are usually performed with a D_2O solvent. The N-H proton exchanges with the D_2O solvent, which shifts the amide I' (where the prime refers to deuteration) to a frequency $\sim 10\text{ cm}^{-1}$ lower than amide I. In this thesis, all the spectra are collected with a D_2O solvent, and any reference to amide I is actually referring to amide I'.

1.3.2 Isotope labels

Although the secondary structure sensitivity of amide I is experimentally attractive, more localized probes are also advantageous. Isotopic substitution, also termed isotope-labeling, can provide structurally localized information because it shifts the local amide I vibrational frequency out of the overlapping main amide I $1600\text{-}1700\text{ cm}^{-1}$ region. Placing a ^{13}C and/or an ^{18}O label on the carbonyl causes a $30\text{-}65\text{ cm}^{-1}$ shift to lower frequency (^{13}C , $^{18}\text{O} \sim 40$; $^{13}\text{C} + ^{18}\text{O} \sim 60\text{ cm}^{-1}$). Ideally, isotope labeling represents a means of introducing a non-invasive local probe of structure because the labeled residue will no longer be hidden under overlapping band structure. The labeled oscillator vibration is not necessarily purely localized, however, because it may still be coupled and interact in a normal mode. Similarly, isotope labeling of high frequency residues may cause a shift to frequencies that still overlap the red-edge of the amide I region³².

Upon localization with an isotope label, the amide I vibration reports on the local environment surrounding the carbonyl. The vibrational frequency is particularly sensitive to the

H-bonding state. When the C=O is buried and H-bonded to the N-H of another residue, the carbonyl can accept fewer (generally one) H-bonds than if it is extended out into the solvent, where it can accept up to two H-bonds from water. Thus, the frequency becomes an indicator of the level of solvation: higher frequencies indicate fewer H-bonds and burial while lower frequencies tend to indicate more H-bonds and solvations.

1.3.3 2D IR spectroscopy

Two-dimensional infrared spectroscopy (2D IR) is the specific infrared spectroscopy used throughout this thesis to probe the amide I vibration³⁶. It is a multidimensional spectroscopy whose intensity depends nonlinearly on the input beams' intensities. A common example of nonlinear spectroscopy is pump-probe spectroscopy, in which the probing intensity is modulated by the pumping beam(s). 2D IR can be thought of as a more specific variant of pump-probe spectroscopy, in which the dependence of the probe spectrum modulation is measured as a function of the frequency of the pump beam. Through the 2D IR measurement, the pumping and probing frequencies can be correlated to reveal coupling and broadening mechanisms.

The major advantage of 2D IR spectroscopy is that it spreads the spectral information over two axes rather than one, as is done in linear FTIR spectroscopy. In Figure 1.3, the spectra of eight select proteins and peptides are shown to illustrate the increased information content provided by 2D IR. Shown above each 2D IR spectrum is the corresponding protein structure and FTIR spectrum. The FTIR spectra are remarkably similar, while the 2D IR spectrum reveals secondary structure specific peaks that correspond well with the known structures. Ridges in the 1630 and 1680 cm^{-1} regions (most prominent in Con A) are the result of coupling between the ν_{\perp} and ν_{\parallel} anti-parallel β -sheet modes, which give a characteristic “Z” shaped spectrum. Conversely, the presence of the α -helix mode (most prominent in myoglobin) gives a squashed “figure 8” spectrum.

Another advantage of 2D IR spectroscopy is structure based modeling^{5, 37}. Molecular dynamics simulations can be run to attain atomistically detailed trajectories of protein dynamics. They can be used to identify possible structures, identify the significantly populated structures, and model the effect of temperature or other macroscopic variables. 2D IR spectra can then be simulated based on the structures to either assign ensemble features or characteristics associated with particular conformations. Performing simulations before experiments also allows for more thoughtful experimental design, as certain features can be predicted and the sensitivity can be

assessed. Experiment and simulation can be cross-validated through comparison of respective spectra. Through combined simulation and experiment, 2D IR can gain potential atomic resolution. Also, understanding how to properly model the 2D IR spectra allows us to assess the important factors that affect energies and couplings within the system. Modeling has shown the importance of H-bonding, electrostatic surroundings, and nearest neighbor orientations.

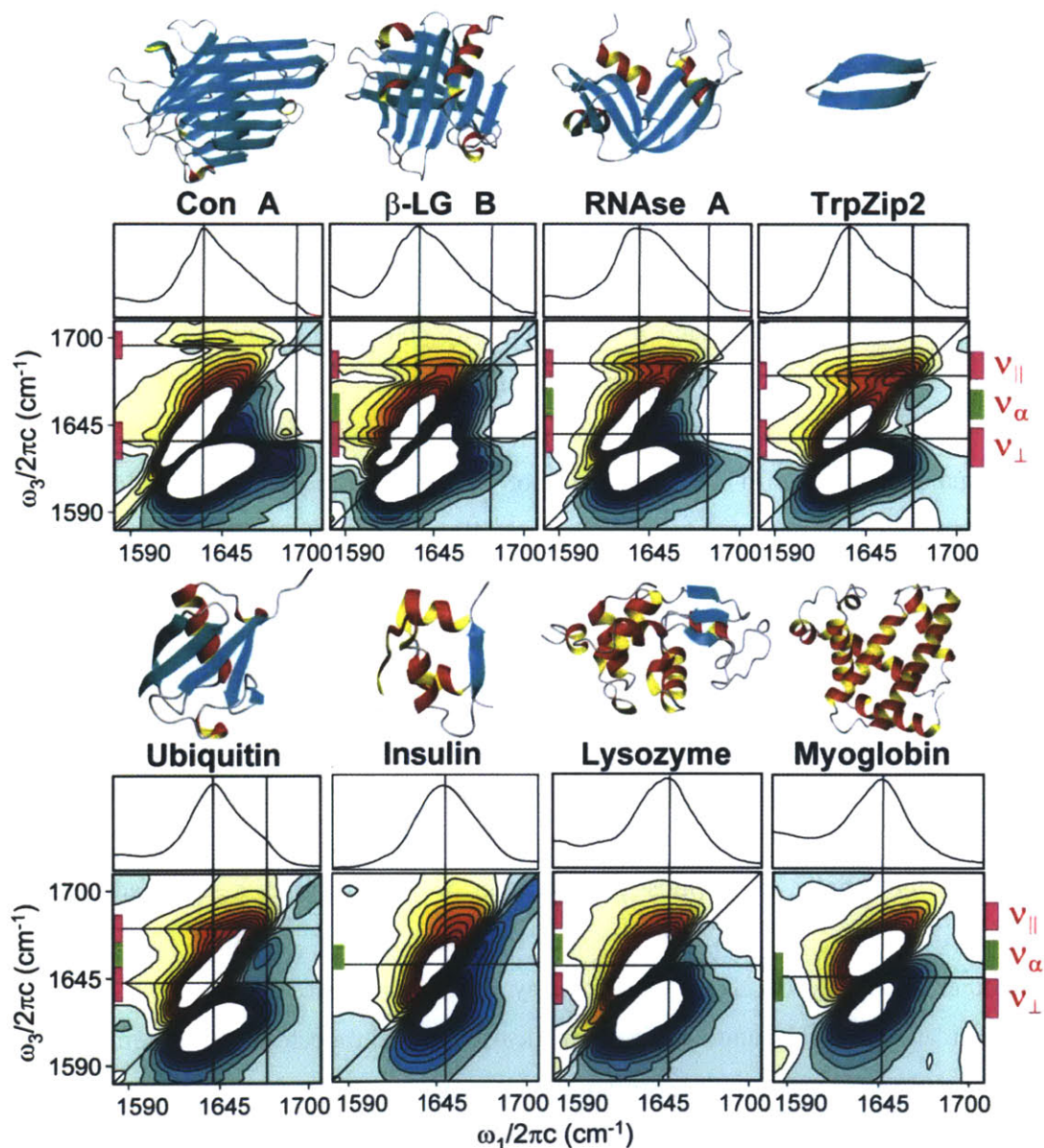


Figure 1.3: The crystal structure, FTIR, and 2D IR spectra of selected small proteins with varying β -sheet and α -helix structure. Adapted from Ref. 5.

More details of analyzing the 2D IR spectrum are given in Chapter 2.

1.3.4 Transient spectroscopy

Although 2D IR spectroscopy has been successful for equilibrium structural determination, pairing it with a transient spectroscopy extends the technique to non-equilibrium experiments by introducing access to another time variable. Conventional pump–probe and 2D IR waiting time experiments, in which the time delay between pumping and probing pulses is varied, are limited by the ~ 1 ps vibrational dephasing of amide I in water. To access longer timescales, we introduce a 6 ns T-jump laser pulse into the setup. Thus, we can collect 2D IR spectra in picosecond snapshots following the temperature rise. For our current setup and sample cell, we can track changes induced by the T-jump from ~ 5 ns to 50 ms, although the temperature profile begins to relax after ~ 100 μ s.

As described in 1.2.1, the rates of transfer between protein states and the relative populations are temperature dependent. If the ratio of populations in different states do not match the equilibrium ratio at the current temperature, the populations will redistribute to reach the equilibrium distribution. In the T-jump experiment, we quickly change the temperature from its initial value to a final value. If the rates of interconversion between states are slower than the 6 ns change in temperature, as is generally the case, then the system will not be at equilibrium: the populations will be distributed according to the initial temperature rather than the current final temperature. In response, the populations will redistribute at a rate determined by the current temperature. Through spectroscopy, we can measure the time-dependence of the redistribution. Thus, by performing the T-jump experiment, we attain the necessary high time resolution afforded by infrared spectroscopy and the desired high structural resolution provided by 2D IR. Ideally, the observed rates can be linked to structurally assigned spectral changes to reveal the kinetics and the important states.

1.3.5 Challenges and advantages

Although they are used extensively, infrared studies of proteins require high concentrations of proteins, which must be dissolved in appropriately transparent solvents, such as D₂O. Special optics must be used that are transparent in the infrared, which limits the quality and accessibility of polarizers. 6 μ m Infrared beams are invisible to humans and conventional IR viewers, which means that it is necessary to overlap the IR beams with visible tracer beams for easier alignment. The diffraction limit dictates a larger focus size than would be attained at

shorter wavelengths. For the 6 μm frequencies used here, fast and sensitive array cameras are only now becoming commercially available at somewhat non-astronomical prices. Detectors with the necessary sensitivity and speed are still expensive and are currently limited to ~ 100 pixels.

Despite the challenges listed above, technological advances in detectors warrant optimism for the future, suggesting a significant technology-driven change. The long wavelengths are also a blessing in multidimensional spectroscopy, where the error bars on stage position and phase stability can be high, unlike in the visible or ultraviolet frequency ranges.

1.4 Thesis outline

This introductory chapter is intended to whet the appetite for the thesis that follows hereafter. Chapter 2 is a general discussion of the background to 2D IR spectroscopy and detection. Chapter 3 follows up by mapping out the experimental setup used to collect equilibrium and transient 2D IR spectra. Chapter 4 describes the characterization of the amplitude and phase of the 2D IR signal. Chapter 5 describes the specific instrumentation for, and improvements to, the transient T-jump 2D IR spectrometer, whose data is adjusted to correct for linear absorption of the incident and signal beams. Example transient T-jump spectra are also described and interpreted as exhibiting general temperature-dependent spectral shift features expected for all infrared T-jump experiments. To aid data interpretation, Chapter 6 discusses the singular value decomposition technique, a general mathematical decomposition used to link observed timescales with spectral changes. Chapter 7 assesses some recent transient T-jump 2D IR spectra taken of ubiquitin and compares it to previous data collected with different data processing techniques. Chapter 8 proposes a free energy surface to describe the isotope-label-informed transient T-jump 2D IR spectra collected for the model β -turn peptide Trpzip2. Through T-jump 2D IR experiments, chapter 9 analyzes the association processes in insulin, which exists in a dimer-monomer equilibrium.

1.5 References

- (1) C. Veigel, J. E. Molloy, S. Schmitz, and J. Kendrick-Jones, "Load-dependent kinetics of force production by smooth muscle myosin measured with optical tweezers," *Nat. Cell Biol.* **5**, 980-986 (2003).
- (2) C. Soto, "Unfolding the role of protein misfolding in neurodegenerative diseases," *Nat. Rev. Neurosci.* **4**, 49-60 (2003).
- (3) F. Chiti and C. M. Dobson, "Protein Misfolding, Functional Amyloid, and Human Disease," *Annu. Rev. Biochem.* **75**, 333-366 (2006).
- (4) Z. Ganim, "2D IR spectroscopy and computational modeling : application to protein folding and binding," (Massachusetts Institute of Technology, Cambridge, 2010).
- (5) Z. Ganim, H. S. Chung, A. W. Smith, L. P. DeFlores, K. C. Jones, and A. Tokmakoff, "Amide I Two-Dimensional Infrared Spectroscopy of Proteins," *Acc. Chem. Res.* **41**, 432-441 (2008).
- (6) J. Kubelka, J. Hofrichter, and W. A. Eaton, "The protein folding 'speed limit'," *Curr. Opin. Struct. Biol.* **14**, 76-88 (2004).
- (7) J. K. Myers and T. G. Oas, "Mechanisms of fast protein folding," *Annu. Rev. Biochem.* **71**, 783-815 (2002).
- (8) J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature* **267**, 585-590 (1977).
- (9) P. Bradley, K. M. S. Misura, and D. Baker, "Toward High-Resolution de Novo Structure Prediction for Small Proteins," *Science* **309**, 1868-1871 (2005).
- (10) L. Regan and W. DeGrado, "Characterization of a helical protein designed from first principles," *Science* **241**, 976 (1988).
- (11) M. D. Struthers, R. P. Cheng, and B. Imperiali, "Design of a Monomeric 23-Residue Polypeptide with Defined Tertiary Structure," *Science* **271**, 342-345 (1996).
- (12) A. G. Cochran, N. J. Skelton, and M. A. Starovasnik, "Tryptophan zippers: Stable, monomeric β -hairpins," *Proc. Natl. Acad. Sci. USA* **99**, 9081-9081 (2001).
- (13) K. W. Plaxco and C. M. Dobson, "Time-resolved biophysical methods in the study of protein folding," *Curr. Opin. Struct. Biol.* **6**, 630-636 (1996).
- (14) H. Frauenfelder, S. Sligar, and P. Wolynes, "The energy landscapes and motions of proteins," *Science* **254**, 1598-1603 (1991).
- (15) K. A. Dill and H. S. Chan, "From Levinthal to pathways to funnels," *Nat. Struct. Biol.* **4**, 10-19 (1997).
- (16) B. Nölting, *Protein Folding Kinetics*, 2nd Ed. ed. (Springer, Berlin, 2006).
- (17) K. A. Dill, "Dominant forces in protein folding," *Biochemistry* **29**, 7133 (1990).
- (18) G. D. Rose, P. J. Fleming, J. R. Banavar, and A. Maritan, "A backbone-based theory of protein folding," *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16623-16633 (2006).
- (19) A. R. Fersht, "Conformational equilibria in α - and δ -chymotrypsin: The energetics and importance of the salt bridge," *J. Mol. Biol.* **64**, 497-509 (1972).
- (20) W. J. Becktel and J. A. Schellman, "Protein stability curves," *Biopolymers* **26**, 1859-1877 (1987).
- (21) O. B. Ptitsyn and A. A. Rashin, "A model of myoglobin self-organization," *Biophys. Chem.* **3**, 1-20 (1975).
- (22) K. A. Dill, "Theory for the folding and stability of globular proteins," *Biochemistry* **24**, 1501-1509 (1985).

- (23) A. R. Fersht, "Nucleation mechanisms in protein folding," *Curr. Opin. Struct. Biol.* **7**, 3 (1997).
- (24) H. R. Bosshard, "Molecular Recognition by Induced Fit: How Fit is the Concept?," *Physiology* **16**, 171-173 (2001).
- (25) K. C. A. Holmes, I.; Kull, F. J.; Jahn, W.; Schoder, R. R., "Electron Cryo-Microscopy Shows How Strong Binding of Myosin to Actin Releases Nucleotide," *Nature* **425**, 423-427 (2003).
- (26) K. Sugase, H. J. Dyson, and P. E. Wright, "Mechanism of coupled folding and binding of an intrinsically disordered protein," *Nature* **447**, 1021-1025 (2007).
- (27) B. A. Shoemaker, J. J. Portman, and P. G. Wolynes, "Speeding molecular recognition by using the folding funnel: the fly-casting mechanism," *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8868-8873 (2000).
- (28) G. R. Bowman, V. A. Voelz, and V. S. Pande, "Taming the complexity of protein folding," *Curr. Opin. Struct. Biol.* **21**, 4-11 (2011).
- (29) N. Singhal, C. D. Snow, and V. S. Pande, "Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin," *J. Chem. Phys.* **121**, 415-425 (2004).
- (30) C. J. Fecko, J. D. Eaves, J. J. Loparo, A. Tokmakoff, and P. L. Geissler, "Ultrafast Hydrogen-Bond Dynamics in the Infrared Spectroscopy of Water," *Science* **301**, 1698-1698 (2003).
- (31) H. Torii and M. Tasumi, "Model Calculations on the amide-I infrared bands of globular proteins," *J. Chem. Phys.* **96**, 3379-3387 (1992).
- (32) A. W. Smith, J. Lessing, Z. Ganim, C. S. Peng, A. Tokmakoff, S. Roy, T. L. C. Jansen, and J. Knoester, "Melting of a β -Hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations," *J. Phys. Chem. B* **114**, 10913-10924 (2010).
- (33) A. Barth and C. Zscherp, "What vibrations tell us about proteins," *Q. Rev. Biophys* **35**, 369-430 (2002).
- (34) C. M. Cheatum, A. Tokmakoff, and J. Knoester, "Signatures of β -sheet secondary structures in linear and two-dimensional infrared spectroscopy," *J. Chem. Phys.* **120**, 8201-8215 (2004).
- (35) N. Demirdöven, C. M. Cheatum, H. S. Chung, M. Khalil, J. Knoester, and A. Tokmakoff, "Two-dimensional infrared spectroscopy of antiparallel β -sheet secondary structure," *J. Am. Chem. Soc.* **126**, 7981-7990 (2004).
- (36) M. Khalil, N. Demirdöven, and A. Tokmakoff, "Coherent 2D IR Spectroscopy: Molecular Structure and Dynamics in Solution," *J. Phys. Chem. A* **107**, 5258-5279 (2003).
- (37) Z. Ganim and A. Tokmakoff, "Spectral Signatures of Heterogeneous Protein Ensembles Revealed by MD Simulations of 2DIR Spectra," *Biophys. J.* **91**, 2636-2646 (2006).

Chapter 2

2D IR Background

2.1 Introduction

Two dimensional infrared spectroscopy (2D IR) is technique in which multiple, ultrashort laser pulses pump and probe vibrations. By varying the time delay between the pulses, coherences can be mapped to determine the frequencies of the excited vibrations. The short time duration of the pulses corresponds to broad spectra, which allow for coincidental excitation of modes of varying frequencies.

In this chapter, the background to 2D IR spectroscopy is discussed with particular emphasis on detection. Although the theory is discussed, it is done briefly and qualitatively to highlight the information content provided by 2D IR and motivate the technical discussion presented in Chapter 3.

First, linear FTIR spectroscopy theory and experiment are discussed to indicate the information content and underlying molecular processes probed, to provide a conceptualization of spectroscopic signal generation, and to discuss the detection method.

The FTIR discussion provides a starting point from which we can begin to discuss the improved information content inherent to 2D IR spectroscopy. The underpinnings of 2D IR theory are discussed. The different detection methods and representations of the 2D IR signal are described.

2.2 Linear IR spectroscopy

2.2.1 Theory

In linear IR spectroscopy, the sample's absorption spectrum is measured. The fraction of incident light that is transmitted through the sample, $T(\omega)$, is usually written using a power of 10 or an exponential term:

$$T(\omega) = \frac{I_0(\omega)10^{-\varepsilon(\omega)l\zeta}}{I_0(\omega)} = \frac{I_0(\omega)\exp(-\alpha(\omega)l)}{I_0(\omega)} \quad (2.1)$$

where ζ is the concentration (M), l , is the length of the sample (cm), $\varepsilon(\omega)$ is the molar decadic extinction coefficient ($M^{-1} \text{ cm}^{-1}$), $\alpha(\omega)$ is the absorption coefficient (cm^{-1}), $I_0(\omega)$ is the incident light intensity spectrum, and the numerator is the detected intensity transmitted through the sample. In order to isolate the exponential argument, which is linearly proportional to concentration, the absorbance spectrum, $S(\omega)$ (in optical density units), is calculated through:

$$S(\omega) = -\log_{10}(T(\omega)) = \varepsilon(\omega)l\zeta = \frac{\alpha(\omega)l}{\ln(10)}. \quad (2.2)$$

where the central two terms are Beers law.

Although on the microscopic level absorption is usually thought of as depletion, we can also conceptualize the absorption process as generation of phase shifted light: the complex electric field of the incident light, $\tilde{E}_0(\omega)$, creates a polarization in the sample which emits a phase-shifted, complex signal electric field, $\tilde{E}_{sg}(\omega)$, which, due to the phase shift, depletes the intensity of light measured after the sample. When the incident field is resonant with the transition, the induced polarization is $\pi/2$ phase-shifted relative to the incident field, and the emitted field is $\pi/2$ phase-shifted relative to the polarization, which sums to a total shift of π between incident and emitted. We use the convention that the electric field in the time domain is real. Therefore, we can express the electric field in the frequency domain as an amplitude, $e(\omega)$, and a phase, $\varphi(\omega)$, which is related to the time domain electric field by a Fourier transform (FT):

$$\begin{aligned} \tilde{E}_j(\omega) &= \text{FT}[E_j(t)] \\ &= \int_{-\infty}^{\infty} E_j(t) \exp(i\omega t) dt. \\ &= e_j(\omega) \exp[i\varphi_j(\omega)] \end{aligned} \quad (2.3)$$

In this 'signal generation' conceptualization, the intensity measured at the detector after passing

through the sample is given by:

$$\left| \tilde{E}_0(\omega) + \tilde{E}_{sig}(\omega) \right|^2 = e_0(\omega)^2 + e_{sig}(\omega)^2 + 2e_0(\omega)e_{sig}(\omega)\cos[\varphi_{sig}(\omega) - \varphi_0(\omega)] \quad (2.4)$$

This is an example of interferometric, heterodyned detection, in which two fields are overlapped for detection. Because this overlap is incidental, it is called self-heterodyned. The two electric fields are seemingly experimentally inseparable, which prohibits measurement of just $\left| \tilde{E}_{sig}(\omega) \right|^2$. As a result of the heterodyning overlap, the cross term is linear in the signal electric field amplitude, which is in turn linear in concentration. The cross term also retains the difference phase information, and it depends on the product of the amplitudes of each input electric field, a trait which can be used to amplify the smaller field. If we assume that the emitted signal is small, or $e_0(\omega) \gg e_{sig}(\omega)$ and $\alpha(\omega) \ll 1$, then equating eq (2.4) to a Taylor expansion of the numerator of eq. (2.1) reveals:

$$\alpha(\omega) \approx -\frac{1}{l} \frac{2e_{sig}(\omega)\cos[\varphi(\omega)_{sig} - \varphi(\omega)_0]}{e_0(\omega)}. \quad (2.5)$$

As discussed, the absorption coefficient depends on the product of the emitted signal amplitude multiplied by the cosine of the difference between the signal and incident phases. $\alpha(\omega)$ should not depend on $e_0(\omega)$, and therefore we expect $e_{sig}(\omega)$ to be proportional to $e_0(\omega)$.

Although this conceptualization of $\tilde{E}_0(\omega)$ generating $\tilde{E}_{sig}(\omega)$ seems unnecessary considering the adequate, useful, and universally accepted absorption process described in eqs. (2.1) and (2.2), we present it here because it is analogous to the convention usually used to describe nonlinear spectroscopies. For nonlinear spectroscopies, this representation is more natural because a signal electric field is often generated in directions that are not co-linear with the incident beams, which means there is no transmitted beam, and no obvious denominator if we try to use an expression similar to eq. (2.1).

Although there are many expressions that describe the molar extinction coefficient, an appropriate expression for infrared vibrational spectroscopy adapted from Ref. 3 is:

$$\varepsilon(\omega) \propto -\frac{4\pi^2}{\hbar c} \omega \sigma(\omega) \quad (2.6)$$

where \hbar is Planck constant, c is the speed of light, we have assumed that the ground state is the only thermally populated state, and $\sigma(\omega)$ is the lineshape function given by the FT of the dipole-

dipole correlation function:

$$\sigma(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp(-i\omega t) \langle \hat{\epsilon} \cdot \bar{\mu}(0) \hat{\epsilon} \cdot \bar{\mu}(t) \rangle \quad (2.7)$$

in which $\hat{\epsilon}$ is a vector indicating the orientation of the incident field polarization vector, $\langle \rangle$ indicates ensemble averaging, and $\bar{\mu}$ is the dipole operator. For vibrational spectroscopy, the dipole operator is expanded along the vibrational coordinates, $q(t)$, and the transition dipole is given by the leading term $\frac{\partial \bar{\mu}}{\partial q}$. $\hat{\epsilon}$ projects onto the dipole and results in a separable term - 1/3 for an isotropic light field. Thus, for vibrational spectroscopy, the $\langle \rangle$ term reduces to the product of: a scalar term of the transition dipole squared, $|\mu|^2$, an orientational term, and the vibrational coordinate correlation $\langle q(0)q(t) \rangle$.

For a given transition between states n and m with energy splitting of $h\omega_{nm}$, this correlation function can be written out phenomenologically to account for different broadening mechanisms³:

$$\langle q(0)q(t) \rangle \approx \exp(i\omega_{nm}t) \exp\left[-t\left(\frac{1}{T_2^*} + \frac{1}{2T_1} + \frac{1}{\tau_{or}}\right)\right] \exp\left(\frac{-\Delta^2 t^2}{2}\right). \quad (2.8)$$

Eq. (2.8) indicates the general features of the time domain electric field that is emitted from the sample in response to incident driving light. The time domain response is easiest to conceptualize in terms of the response to a coherent or infinitely short incident pulse, and we describe it here in these terms. The case of incident incoherent light is harder to imagine, as it involves imagining the response of the sample molecules to a randomly phased continuous light field, but the resulting spectrum is equivalent to that measured using coherent light. Due to the first term of eq. (2.8), we expect to see a field oscillating at the frequency associated with the energy difference between the m and n states. Due to the second and third terms, this ω_{nm} oscillation is damped because of homogeneous and inhomogeneous broadening, respectively,

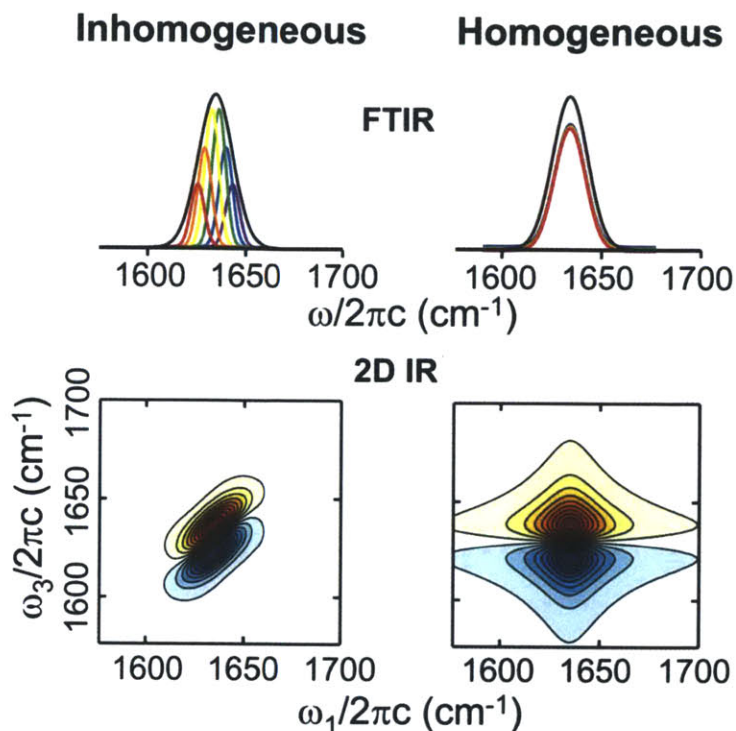


Figure 2.1: A diagram illustrating the manifestation of inhomogeneous and homogeneous broadening in FTIR and 2D IR spectra. Modified from Ref. 1.

which would individually result in Lorentzian and Gaussian frequency domain lineshapes. The $1/2T_1$ accounts for lifetime relaxation, loss of excited population due to spontaneous emission or non-radiative decay. $1/T_2^*$ is due to pure dephasing, the scrambling of phase between different molecules in the ensemble. $1/\tau_{or}$ accounts for orientational relaxation of the dipoles as they tumble in solution. The third exponential describes inhomogeneous broadening due to the molecular environment, which can create a distribution of frequencies associated with the same transition.

For this linear case, these broadening effects cannot be separated, as is illustrated in Figure 2.1, where inhomogeneous and homogeneous broadening result in similar lineshapes. The observed lineshape results from all of them. One of the advantages of 2D IR is that by using many pulses, nonlinear spectroscopies can distinguish different broadening mechanisms.

2.2.2 Experiment and Detection

Linear IR spectroscopy can be collected in the frequency or time domain using incoherent or coherent light. Collection using coherent light is described in Chapter 4, and the

principles and equations are similar, but, based on their prevalence, we focus on discussing the incoherent techniques, where a heated lamp creates continuous black-body radiation in the infrared frequency range.

In the frequency domain collection case, the intensity is measured as a function of frequency after it is dispersed off a grating. Either the entire spectrum is collected on a stripe of pixels or the intensity at each frequency is measured on one pixel as the grating is rotated to sweep out the spectrum.

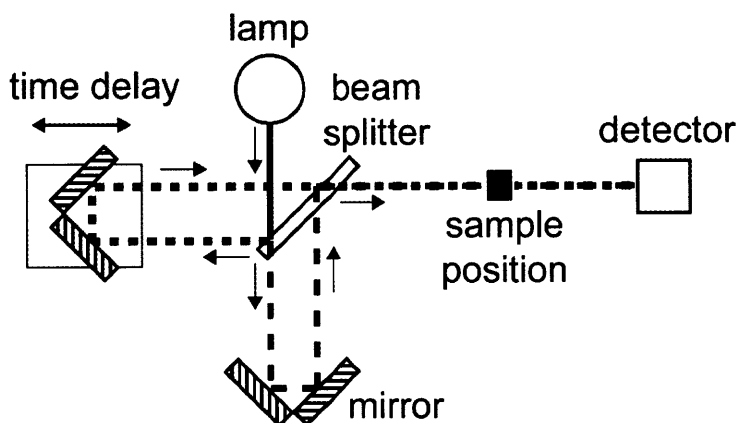


Figure 2.2: An illustration of a FTIR spectrometer based on a Michelson interferometer.

Due to many reasons, including increased frequency resolution, stability, and signal-to-noise, linear IR spectroscopy is usually done in the time domain using Fourier Transform infrared spectroscopy (FTIR). In this technique, a Michelson interferometer is used to split the incident radiation into two beams which are recombined and focused onto a detector. A diagram of the setup is shown in Figure 2.2. Once the two beams are generated, they travel along different arms: one is fixed while a delay stage in the other introduces a time delay between the two beams. Because the two beams both pass through the sample, they are exact copies of each other (within a proportionality factor, c_I , to account for inequalities in loss due to the beam splitter and mirrors) separated by a user defined time delay. If we define $E_{det}(t)$ as the detected electric field of one of the pulses, and the delay between the pulses is given by τ_{del} , the observed intensity will be given by integrating the intensity on the detector over time:

$$\begin{aligned}
I_{obs}(\tau_{del}) &\propto \int_{-\infty}^{\infty} |E_{det}(t) + c_1 E_{det}(t - \tau_{del})|^2 dt \\
&\propto \int_{-\infty}^{\infty} (1 + c_1) [E_{det}(t)]^2 dt + 2c_1 [E_{det} * \bar{E}_{det}](\tau_{del})
\end{aligned} \tag{2.9}$$

or in the frequency domain:

$$\begin{aligned}
I_{obs}(\tau_{del}) &\propto \int_{-\infty}^{\infty} |\tilde{E}_{det}(\omega) + c_1 \tilde{E}_{det}(\omega) \exp(-i\omega\tau_{del})|^2 d\omega \\
&\propto \int_{-\infty}^{\infty} [1 + c_1^2 + 2c_1 \cos(\omega\tau_{del})] e_{det}(\omega)^2 d\omega
\end{aligned} \tag{2.10}$$

\bar{E} is the complex conjugate of E . The first line of eq. (2.10) results from expressing the time domain electric fields in the frequency domain (Parseval's theorem). If the τ_{del} -independent, constant terms are ignored (or subtracted), Fourier transformation of either eq. (2.9) and (2.10) along τ_{del} results in:

$$\begin{aligned}
\text{FT}[I_{obs}(\tau_{del}) - I_{const}] &\propto \int_{-\infty}^{\infty} [E_{det} * \bar{E}_{det}](\tau_{del}) \exp(i\omega\tau_{del}) d\tau_{del} \\
&\propto \text{FT}[E_{det}] \times \text{FT}[\bar{E}_{det}] \\
&\propto e_{det}(\omega)^2
\end{aligned} \tag{2.11}$$

or

$$\begin{aligned}
\text{FT}[I_{obs}(\tau_{del}) - I_{const}] &\propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\cos(\omega'\tau_{del}) e_{det}(\omega')^2 d\omega'] \exp(i\omega\tau_{del}) d\tau_{del} \\
&\propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2} \{ \exp[i\tau_{del}(\omega + \omega')] + \exp[i\tau_{del}(\omega - \omega')] \} e_{det}(\omega')^2 d\omega' d\tau_{del} \\
&\propto \int_{-\infty}^{\infty} \frac{1}{2} \{ \delta(\omega + \omega') + \delta(\omega - \omega') \} e_{det}(\omega')^2 d\omega' \\
&\propto e_{det}(\omega)^2
\end{aligned} \tag{2.12}$$

Thus, after FT, the interferometric measurement results in the same quantity that is measured through frequency domain collection.

2.2.3 Discussion

Linear IR spectroscopy is the linear, simpler analog to 2D IR. We have discussed it here 1) to illustrate the connection between the absorption and signal generation pictures, 2) to relate the observed spectrum to the time domain signal, which can be approximated by multiple

phenomenological broadening mechanisms, and 3) to show the equivalence of time-domain and frequency-domain collection.

In the linear IR absorbance spectrum, positive peaks are attributed to absorbance, and they are proportional to the concentration. If there are multiple species within the sample, linear IR spectroscopy cannot distinguish which peaks can be assigned to each species. Similarly, linear IR spectroscopy cannot differentiate between different broadening mechanisms, all of which contribute to broadening the frequency domain peaks.

As a technical detail related to our specific study of the amide I protein vibration, for protein solution samples dissolved in deuterated water, although D₂O has a much smaller dipole moment than the amide I mode of proteins in the 1550 – 1750 cm⁻¹ region, the D₂O generally has a much higher concentration, which results in similar absorbance optical densities. Thus, for these samples, the solution spectrum must be separately collected for subtraction to separate the solute response.

2.3 2D IR spectroscopy background

2D IR is a nonlinear 3rd order spectroscopy, in which signal is generated as a result of three interactions between the sample and the light field. Although there can be multiple interactions per pulse, we find it most straightforward to discuss the case in which there are three pulses that each interact with the sample once. Although nonlinear signal is generated using incoherent and continuous coherent light, we restrict ourselves to the more intuitive case in which each pulse is an impulsive, infinitely short light pulse. Thus, in our description of the process, three ultrashort, broadband pulses interact with the sample to create the emitted signal. By varying the time delay between each pulse, we can measure the energy level structure of the molecule by pushing quanta of energy through the different vibrational modes. Variation of time delay is a specific case of varying the phase ($\Delta\phi = \Delta t\omega$), and the same spectrum can be acquired by measuring a more general (not linear in time) phase-dependence.

2.3.1 Theory

From an ensemble perspective, the three electric fields generate a third order polarization that depends on the respective wavevectors, \mathbf{k}_j , the time delay between each field, and the inherent response of the sample, $\tilde{\mathbf{R}}^{2,4-6}$:

$$\mathbf{P}^{(3)}(\mathbf{k}_{sig}, t, \tau_2, \tau_1) = \int_0^\infty \int_0^\infty \int_0^\infty \vec{\mathbf{R}}(\tau_3, \tau_2, \tau_1) \cdot \mathbf{E}_3(\mathbf{k}_3, \nu_3, t - \tau_3) \times \mathbf{E}_2(\mathbf{k}_2, \nu_2, t - \tau_3 - \tau_2) \mathbf{E}_1(\mathbf{k}_1, \nu_1, t - \tau_3 - \tau_2 - \tau_1) d\tau_1 d\tau_2 d\tau_3 \quad (2.13)$$

where τ_1 and τ_2 are the time delays between the first/second and second/third pulses, respectively. τ_3 is the time variable associated with emission of the signal electric field, or, equivalently, detection of the electric field. The point at which the absolute time variable is arbitrarily defined to be zero, $t = 0$, is (usually) set when the third pulse interacts with the sample. Using this definition, $\tau_3 = t$. Thus, the left side of eq. (2.13) can be interpreted as follows – the $\mathbf{P}^{(3)}$ polarization emits a proportional signal electric field along the signal wavevector \mathbf{k}_{sig} . The signal electric field, $E_{sig}(\mathbf{k}_{sig}, t = \tau_3, \tau_2, \tau_1)$, is a real time domain field that parametrically depends on the time delays τ_1 and τ_2 . The three electric fields and the time delays between them are shown in Figure 2.3.

Variations in different third order spectroscopies arise based on the number and geometry of the beams, the orientation of the detector, the detection method, and the time delays (τ_1, τ_2, τ_3) or conjugate frequencies ($\omega_1, \omega_2, \omega_3$) measured. Later Chapters 4 and 5 are almost entirely

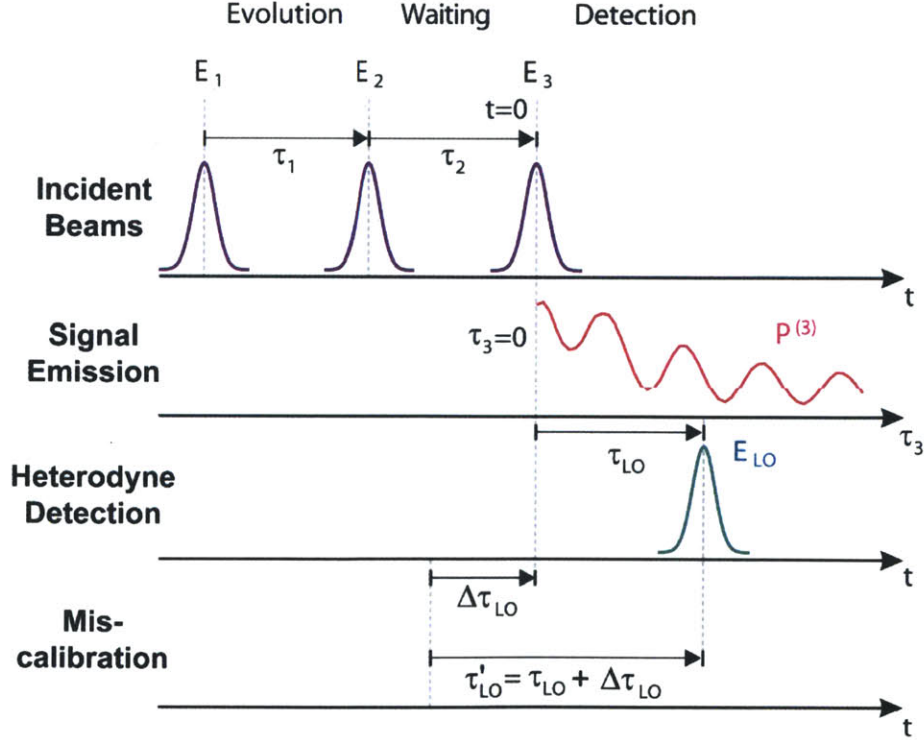


Figure 2.3: A diagram illustrating the discussed time delays. Modified from Ref. 2.

focused on detection of E_{sig} , and we discuss it then as an arbitrary electric field whose detection is completely independent of its origin.

Our goal is to measure the response function of the sample. If we conceptualize the process in terms of photon absorption and emission, at most three photons can be absorbed, one

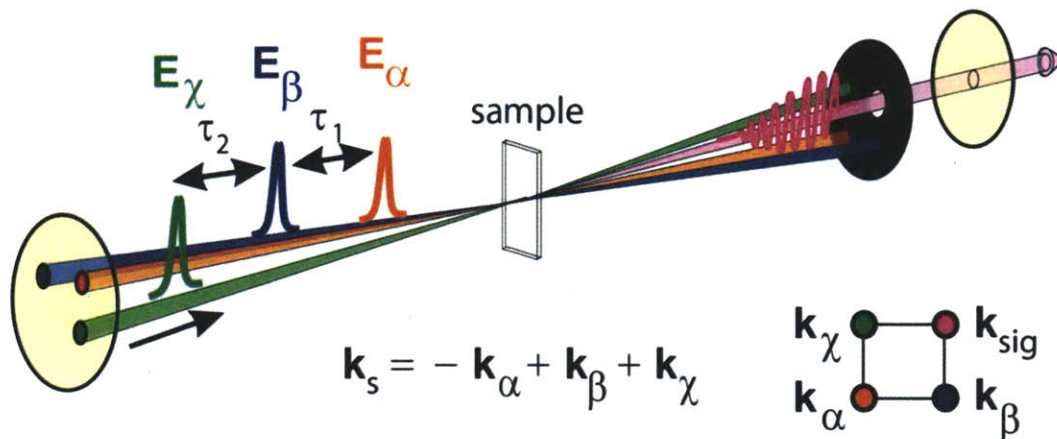


Figure 2.4: A diagram illustrating the boxcar geometry. Reproduced from Ref. 2.

from each beam. Each interaction provides the possibility of photon absorption or emission. That is, each interaction has the potential to either excite the system into a coherent superposition (in which the photon may or may not have been absorbed) or cause emission of a previous potentially absorbed photon. Due to physicality, because each successive interaction acts on the excited system left by the previous (otherwise the signal would not be third order), the three interactions map out a pathway of absorption and emission in the molecule called a Liouville pathway. Each absorbed photon has momentum associated with it, and the direction of the emitted signal depends on which photons are potentially absorbed from which beam. Thus, an experimentalist can measure the signal emitted due to selected sets of pathways by arranging the incident beams with different directions and appropriately placing the detector. Further selectivity is achieved by changing the time ordering between pulses whose wavevectors constrain the type of interaction.

For the 2D IR experiments described here, the three excitation beams are arranged in the boxcar geometry, as shown in Figure 2.4, in which each beam originates from a different corner of a rectangular prism, overlap in the center, and exit through the opposite corners. The detector is placed to detect the signal emitted along the fourth, unused box corner. Beam χ is always the

third beam in the sequence, but the time ordering of the α and β beams is reversed to select for different sets of pathways. Based on the limited bandwidth of the incident pulses, the selection rules for vibrational IR spectroscopy, the geometry of the beams, and the time ordering of the beams, only the 0, 1, and 2 quanta energy levels are probed.

Using perturbation theory and a number of assumptions, the response function can be expressed as the sum of eight (four unique, four conjugate terms) $\tilde{\mathbf{R}}_n$ terms which each represent a subset of Liouville pathways². Each $\tilde{\mathbf{R}}_n$ term can be further broken down into the product of different contributions. To begin, we define an orientational term, Y_n , which describes how the polarizations {X, Y, or Z} of the input beams (I,J,K) and detection polarization (L) map onto the transition dipoles connecting the probed eigenstates a, b, c, d ²:

$$\tilde{\mathbf{R}}_n(\tau_3, \tau_2, \tau_1) = \sum_{IJKL} \sum_{a,b,c,d} (Y_n)_{IJKL}^{a,b,c,d}(\tau_3, \tau_2, \tau_1) R_n^{a,b,c,d}(\tau_3, \tau_2, \tau_1) \quad (2.14)$$

Y_n also takes into consideration the time dependent re-orientation of the molecules as they tumble in solution during the three time periods. R_n can be further separated into the product of multiple terms. For example, in the $n = 2$ case, the three sequential interactions potentially cause consecutive movement between (not necessarily unique) eigenstates a to d , a to b , d to c , c to b , and R is²:

$$R_2^{a,b,c,d}(\tau_3, \tau_2, \tau_1) = P_a \mu^{c,b} \mu^{d,c} \mu^{b,a} \mu^{a,d} \times \exp(-i\omega_{b,c}^0 \tau_3 + i\omega_{d,b}^0 \tau_2 + i\omega_{d,a}^0 \tau_1) F_2^{a,b,c,d}(\tau_3, \tau_2, \tau_1) \quad (2.15)$$

Here we see each interaction and the final emission contributes a $\mu^{p,q}$, the transition dipole between the two eigenstates, and an $\exp(\pm i\omega_{p,q}^0 \tau_m)$ that oscillates at the energy spacing between the eigenstates that are potentially populated (coherent superposition) during the time period m . The average energy splitting between the eigenstates is given by $\omega_{p,q}^0$. The + or - depends on the wavevector of the interacting pulse and whether the interaction causes absorption or emission. P_a is the population of the initial ground state, a . F_2 is the dephasing function that describes the different inhomogeneous and homogeneous broadening mechanisms (except for $1/\tau_{or}$, which is folded into Y_n) described in eq. (2.8) for linear absorption. A more detailed description of F_n is given in Ref. 2, and, as in the linear case, it can be interpreted as a correlation function. These broadening mechanisms occur during every time period. In the impulsive pulse limit, in which the vibrational dynamics are longer than the pulse length, the emitted signal electric field is

proportional to the real part of the response function:

$$E_{sig}(\mathbf{k}_{sig}, \tau_3, \tau_2, \tau_1) \propto \text{Re} \left[\vec{\mathbf{R}}(\tau_3, \tau_2, \tau_1) \right] \equiv S(\tau_3, \tau_2, \tau_1). \quad (2.16)$$

Re-examining eq. (2.15) after equating $\vec{\mathbf{R}}$ to the emitted signal in eq. (2.16) allows us to better understand the signal generation. The first three pulses prepare the system during τ_1 and τ_2 to emit the signal. Due to the τ_1 and τ_2 periods, the amplitude of the emitted electric field is scaled by the accumulated: 1) the population of the ground state (P_a), 2) the magnitude ($\mu^{p,q}$) of the first three transition dipoles, 3) the orientation ($Y_n(\tau_1, \tau_2)$) of the first three dipoles relative to each other and to the polarizations of the incident fields, and 4) the decay mechanisms ($F_n(\tau_1, \tau_2)$) that occurred during τ_1 and τ_2 . As a result of this preparation, immediately following the third pulse, the system is in an excited state that emits radiation along \mathbf{k}_{sig} . This final transition causes the emitted signal electric field to oscillate in time (τ_3) with a frequency determined by the average energy spacing between the final two eigenstates. The amplitude is further affected by the magnitude of the final transition dipole ($\mu^{p,q}$) and the orientation of the final dipole relative to the detection polarization. The oscillation decays in time due to re-orientation ($Y_n(\tau_3)$) and decay mechanisms ($F_n(\tau_1, \tau_2, \tau_3)$, where F_n is often not separable into the product of terms solely dependent on one time delay). The phase of the oscillating signal field results from the accumulated phase during τ_1 and τ_2 .

Thus, as is done in the experiment, by measuring the τ_1 and τ_2 dependence of $E_{sig}(\mathbf{k}_{sig}, t = \tau_3)$, the energy levels and broadening rates and mechanisms can be determined. Because multiple photons are used, the two quanta energy levels are also probed and studied.

2.3.2 Specific 2D IR measurements

Different 3rd order infrared measurements vary based on which time (τ_1 , τ_2 , τ_3) and/or frequencies (ω_1 , ω_2 , ω_3) are plotted, whether the signal is integrated over any of these variables, and which representation of the signal is plotted (real, amplitude, or phase). Spectroscopies measured for this thesis include dispersed pump-probe (DPP), dispersed vibrational echo (DVE), heterodyne-detected dispersed vibrational echo (HDVE), and 2D IR. The explicit equations describing these representations are given below after discussion of the detection. Briefly, all four are collected in ω_3 , which is the origin of the dispersed label. DPP, DVE, and HDVE are one dimensional measurements of the nonlinear signal generated for a given τ_1 and τ_2 delay, where τ_1 is usually 0 (the first and second pulses are time-coincident). DPP, DVE, and HDVE

are all measuring the same signal field. They are differentiated based on whether that signal field is detected interferometrically with a reference beam, and whether that reference beam can be independently delayed. DPP is the real part of the signal, DVE is the amplitude squared of the signal, and HDVE is a measurement from which the complex signal can be calculated. Thus, HDVE can be used to calculate the DPP and the DVE spectrum.

2D IR involves full measurement of the signal as a function of τ_1 . The 2D IR is usually displayed as the real part of the spectrum that results from Fourier transforming along τ_1 to give (ω_1, ω_3) axes. Based on the projection-slice theorem⁵, integration of the complex 2D IR spectrum along ω_1 gives the equivalent complex HDVE spectrum collected at the same $(\tau_1 = 0, \tau_2, \tau_{LO})$ delays. Thus, the 2D IR signal can be related back to the HDVE, DPP, and DVE spectra.

2.3.3 Detection

For our purposes, the signal field can be detected as an integrated quantity in frequency or in time, and in a homodyne or heterodyne measurement. In a frequency domain measurement, the electric field (or fields) being detected is diffracted off of a grating.

In a homodyne measurement, only one electric field is measured to give the power spectrum. This may be done, as in FTIR described above in eq. (2.9), by creating two copies of the same pulse and measuring their square amplitude through interferometric detection.

In a heterodyne measurement, two *different* electric fields are measured. Usually, the phase and amplitude of one ‘reference’ pulse is known such that the phase and the amplitude of the remaining pulse can be determined.

We begin by describing detection generally before describing it specifically to measure $E_{sig}(\mathbf{k}_{sig}, \tau_1, \tau_2, \tau_3)$ in the homodyne and heterodyne cases. Thus, we start by talking about a general detected electric field $E_{det}(t)$ which may be the sum of many fields.

General

Detection occurs at a pixel made of a semiconductor material whose bandgap is small enough that absorption of an infrared photon causes the generation of an electron, which is then measured electronically as a voltage. In practice, the detection time resolution is limited by the response time of the pixel – although the light is absorbed instantaneously, electrons and current are generated on a slower timescale. For the mercury cadmium telluride (MCT) detectors used here, the response times are $\sim 5 \mu\text{s}$. Thus, because the incident electric field pulse is infinitely short ($< \text{ns}$) relative to the response time, the observed output from the MCT pixel is an

exponential with a $\sim 5 \mu\text{s}$ decay whose amplitude (equivalently the area) is scaled by the integrated electric field intensity at the detector $|E_{\text{det}}|^2$. The observed intensity at pixel l is then:

$$I_{\text{obs}}(l) \propto \int_{-\infty}^{\infty} |E_{\text{det}}(l, t)|^2 dt \quad (2.17)$$

where the limits are not actually infinite, just effectively infinite due to the $< \text{ns}$ length of the signal electric field pulse and the $> \mu\text{s}$ response time of the detector pixel. The proportionality indicates the linearity between the observed electronic voltage and the intensity. Also, only $\sim 1/2$ of the detector decaying exponential amplitude is integrated by the gating electronics to avoid collecting noise, which again results in proportionality.

If all of the detected electric field is focused to a single pixel, then the total integrated intensity is the same whether the electric field is written in the time or frequency domain:

$$I_{\text{obs}} \propto \int_{-\infty}^{\infty} |E_{\text{det}}(t)|^2 dt = \int_{-\infty}^{\infty} |\tilde{E}_{\text{det}}(\omega)|^2 d\omega. \quad (2.18)$$

If the detected electric field is dispersed off a grating and measured on multiple pixels, then each pixel will be measuring a different frequency:

$$I_{\text{obs}}(l = \omega) \propto |\tilde{E}_{\text{det}}(\omega)|^2. \quad (2.19)$$

Here the grating effectively Fourier transforms the detected electric field from the time domain (t) to the frequency domain, ω , where the field is complex. Thus, the detected electric field is split into many different monochromatic fields that still propagate through space until their intensities are measured by the detector.

Based on the above discussion, I_{obs} can be measured in two ways. If all of the detected electric field is focused onto one pixel, then I_{obs} is proportional to the integration of the absolute value square of the time domain detected electric field (eq. (2.17)). If the electric field is split apart into different frequency components, then the square of the amplitude of each frequency component is measured (eqs. (2.19)) and the ϕ_{det} information is lost.

In the following discussion, specific electric fields will be substituted into the above equations for E_{det} .

Homodyne

In the homodyne case, the $E_{\text{det}} = E_{\text{sig}}$. Introducing the parametric τ_1 and τ_2 dependence means that we can write the signal out more explicitly as $E_{\text{sig}}(t = \tau_3, \tau_2, \tau_1)$. If the signal is

measured on a single pixel, then the observed intensity, $I_{\text{obs}}(\tau_2, \tau_1)$, is given by eq. (2.18). If the signal is dispersed and measured, then the observed intensity, $I_{\text{obs}}(\omega_3, \tau_2, \tau_1)$, is given by eq. (2.19). If the signal electric field is split into copies and interferometrically measured, then $I_{\text{obs}}(\tau_{LO}, \tau_2, \tau_1)$ is measured as given in eq. (2.9), which, after FT, becomes $I_{\text{obs}}(\omega_3, \tau_2, \tau_1)$.

Dispersed vibrational echo spectroscopy is a type of homodyne detected nonlinear infrared spectroscopy. The DVE spectrum, $S_{\text{DVE}}(\omega_3, \tau_2, \tau_1)$ is given by $I_{\text{obs}}(\omega_3, \tau_2, \tau_1)$, which equals $[e_{\text{sig}}(\omega_3, \tau_2, \tau_1)]^2$. For experimental convenience, S_{DVE} is collected in the frequency domain. It can also be collected interferometrically as described in the previous paragraph, a method which will take longer to collect, but results in better frequency resolution.

Heterodyne

In the heterodyne case, a second reference pulse of known phase and amplitude, called the local oscillator (LO), is overlapped with the unknown signal electric field, and $E_{\text{det}} = E_{\text{sig}} + E_{LO}$. In time domain detection, the observed intensity is measured as a function of the time delay between the pulses, τ_{LO} , which is shown in relation to the other time delays in Figure 2.3. This τ_{LO} is similar to the τ_{del} from eq. (2.9). The only distinction is that we use τ_{del} for the general case of interferometric homodyne detection (the two beams are the same, but delayed by τ_{del}). τ_{LO} is reserved for the specific interferometric heterodyne detection (the two beams are different and delayed by τ_{LO}).

Defining the $\tau_{LO} = 0$ point is an interesting endeavor if the pulses are arbitrary. How can the center time of an asymmetrical pulse be determined such that a time separation can be defined? For our case, $E_3(t)$ and $E_{LO}(t)$ are copies of each other (within a scaling factor), which means that the $t = \tau_3 = \tau_{LO} = 0$ can be defined based on a chosen point in $E_3(t)$. Our system puts out symmetric pulses, and this zero time is defined as the center of the pulse.

For the following discussion, we describe the general and practical heterodyne case in which the τ_{LO} may be mis-calibrated by a factor $\Delta\tau_{LO}$. In this case, the delay that is measured is $\tau_{LO}' = \tau_{LO} + \Delta\tau_{LO}$, which is shown at the bottom of Figure 2.3. In some cases this $\Delta\tau_{LO}$ is intentionally introduced.

The observed intensity resulting from time domain detection of $E_{\text{det}} = E_{\text{sig}} + E_{LO}$ is given by substituting this relation into eq. (2.18):

$$\begin{aligned}
I_{obs}(\tau'_{LO}, \tau_2, \tau_1) &\propto \int_{-\infty}^{\infty} \left| E_{sig}(t, \tau_2, \tau_1) + E_{LO}(t - \tau'_{LO}) \right|^2 dt \\
&\propto \int_{-\infty}^{\infty} \left| E_{sig}(t, \tau_2, \tau_1) \right|^2 dt + \int_{-\infty}^{\infty} \left| E_{LO}(t) \right|^2 dt \\
&\quad + \left[E_{sig}(\tau_2, \tau_1) * \bar{E}_{LO} \right](\tau'_{LO}) + \left[E_{LO} * \bar{E}_{sig}(\tau_2, \tau_1) \right](\tau'_{LO})
\end{aligned} \tag{2.20}$$

If the electric fields are re-expressed in the frequency domain (Parseval's theorem), we have:

$$\begin{aligned}
I_{obs}(\tau'_{LO}, \tau_2, \tau_1) &\propto \int_{-\infty}^{\infty} \left| \tilde{E}_{sig}(\omega_3, \tau_2, \tau_1) + \tilde{E}_{LO}(\omega_3) \exp(-i\omega_3 \tau'_{LO}) \right|^2 d\omega_3 \\
&\propto \int_{-\infty}^{\infty} \left| e_{sig}(\omega_3, \tau_2, \tau_1) \exp[i\varphi_{sig}(\omega_3, \tau_2, \tau_1)] + e_{LO}(\omega_3) \exp[i\varphi_{LO}(\omega_3) - i\omega_3 \tau'_{LO}] \right|^2 d\omega_3 \\
&\propto \int_{-\infty}^{\infty} \left\{ \left[e_{sig}(\omega_3, \tau_2, \tau_1) \right]^2 + \left[e_{LO}(\omega_3) \right]^2 + \right. \\
&\quad \left. 2e_{LO}(\omega_3) e_{sig}(\omega_3, \tau_2, \tau_1) \times \cos[\varphi_{sig}(\omega_3, \tau_2, \tau_1) - \varphi_{LO}(\omega_3) + \omega_3 \tau'_{LO}] \right\} d\omega_3
\end{aligned} \tag{2.21}$$

As illustrated explicitly in eqs. (2.11) and (2.12), FT of $[I_{obs} - I_{const}]$ for either eq. (2.20) or eq. (2.21) along τ_{LO} results in:

$$\begin{aligned}
\text{FT}[I_{obs}(\tau'_{LO}, \tau_2, \tau_1) - I_{const}(\tau_2, \tau_1)] &\propto 2e_{LO}(\omega_3) e_{sig}(\omega_3, \tau_2, \tau_1) \times \\
&\quad \cos[\varphi_{sig}(\omega_3, \tau_2, \tau_1) - \varphi_{LO}(\omega_3) + \Delta\tau_{LO}\omega_3]
\end{aligned} \tag{2.22}$$

If the heterodyne detection is done in the frequency domain, then the observed intensity is given by eq. (2.19) similar to eq. (2.4):

$$\begin{aligned}
\left| \tilde{E}_{sig}(\omega_3, \tau_2, \tau_1) + \tilde{E}_{LO}(\omega_3, \tau_{LO}) \right|^2 &= e_{LO}(\omega_3)^2 + e_{sig}(\omega_3, \tau_2, \tau_1)^2 \\
&\quad + 2e_{LO}(\omega_3) e_{sig}(\omega_3, \tau_2, \tau_1) \times \\
&\quad \cos[\varphi_{sig}(\omega_3, \tau_2, \tau_1) - \varphi_{LO}(\omega_3) + \tau_{LO}\omega_3]
\end{aligned} \tag{2.23}$$

At this point, the observed intensity has been measured without any attempted correction for the time delay between the signal and LO pulse. Thus, the $\tau_{LO} \omega_3$ phase term results from the true time difference between the pulses. If the experimentalist thinks that the time delay is $\tau'_{LO} = 100$ fs, but it is actually $\tau_{LO} = 90$ fs ($\Delta\tau_{LO} = 10$ fs), then correction of the cross term for τ'_{LO} leads to the same term given on the right in eq. (2.22), in which there is a $\Delta\tau_{LO}\omega_3$ phase term.

Removing the constant terms and assuming that the time delays are properly known allows us to define the desired heterodyne-detected third order spectrum, which parametrically

depends on the τ_2 and τ_1 delays:

$$S(\omega_3, \tau_2, \tau_1) = 2e_{LO}(\omega_3)e_{sig}(\omega_3, \tau_2, \tau_1) \times \cos[\varphi_{sig}(\omega_3, \tau_2, \tau_1) - \varphi_{LO}(\omega_3)] \quad (2.24)$$

As expected, this spectrum is the product of both the local oscillator and signal spectral amplitudes and the observed phase is the difference between the signal and the LO phases. Thus, the heterodyne detection measures the characteristics of the signal electric field relative to the reference LO field. Inspection of S reveals these appealing properties: it is linear in e_{sig} , it is amplified by e_{LO} , and it retains the phase information.

Now that the general heterodyne and homodyne spectrum has been described, we can explicitly write out the equations for specific types of nonlinear spectra, dispersed vibrational echo, dispersed pump-probe, heterodyne-detected dispersed vibrational echo, and 2D IR:

$$S_{DVE}(\omega_3, \tau_2, \tau_1) = [e_{sig}(\omega_3, \tau_2, \tau_1)]^2 \quad (2.25)$$

$$S_{DPP}(\omega_3, \tau_2, \tau_1 \equiv 0) = 2e_{LO}(\omega_3)e_{sig}(\omega_3, \tau_2, \tau_1 = 0) \times \cos[\varphi_{sig}(\omega_3, \tau_2, \tau_1 = 0) - \varphi_{LO}(\omega_3)] \quad (2.26)$$

$$\tilde{S}_{HDVE}(\omega_3, \tau_2, \tau_1, \tau_{LO}) = 2e_{LO}(\omega_3)e_{sig}(\omega_3, \tau_2, \tau_1) \times \left\{ \begin{array}{l} \cos[\varphi_{sig}(\omega_3, \tau_2, \tau_1) - \varphi_{LO}(\omega_3) + \tau_{LO}\omega_3] + \\ i \sin[\varphi_{sig}(\omega_3, \tau_2, \tau_1) - \varphi_{LO}(\omega_3) + \tau_{LO}\omega_3] \end{array} \right\} \quad (2.27)$$

$$\begin{aligned} S_{2D}(\omega_3, \tau_2, \omega_1) &= \text{Re} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(\tau_3, \tau_2, \tau_1) e^{i\omega_1\tau_1} e^{i\omega_3\tau_3} d\tau_1 d\tau_3 \right] \\ &= \text{Re} \left[\int_{-\infty}^{\infty} S(\omega_3, \tau_2, \tau_1) e^{i\omega_1\tau_1} d\tau_1 \right] \\ &= 2e_{LO}(\omega_3)e_{sig}(\omega_3, \tau_2, \omega_1) \times \cos[\varphi_{sig}(\omega_3, \tau_2, \omega_1) - \varphi_{LO}(\omega_3)] \end{aligned} \quad (2.28)$$

The imaginary part, absolute value, and phase 2D spectra can also be calculated by starting with the last equation in eq. (2.28) and replacing the cosine with a sine, taking the amplitude squared part, and plotting the argument of the cosine term, respectively. For the S_{2D} , errors in the timing will result in additional $\Delta\tau_{LO}\omega_3$ and $\Delta\tau_1\omega_1$ phase terms. Further inspection of eq. (2.28) reveals that it can be thought of as a two-dimensional amplitude multiplied by a two dimensional phase.

The DVE, DPP, and HDVE spectra are discussed in more detail in Chapters 3 and 4.

Each can be related to a projection of the 2D spectrum based on the projection-slice theorem⁵. These projection relationships are discussed in Chapters 3, 4, and 5.

2.3.4 Discussion

As an example of heterodyne-detected nonlinear spectroscopy, a 2D IR spectrum of dicarbonyl-acetylacetonato rhodium (RDC) is shown in Figure 2.5 with the corresponding energy level diagram². In RDC, the two carbonyl vibrations are highly coupled, such that the observed one-quantum eigenstates, a and s , are attributed to the asymmetric and symmetric stretches that involve both carbonyl groups. Due to the multiple interactions in nonlinear spectroscopy, the two-quanta states are also probed.

Although for RDC and proteins we prefer to speak in terms of the one-quantum eigenstates (a and s for RDC, amide I and II for proteins), it is more correct to think either in the local mode (carbonyl stretches) or the full eigenstate basis. Diagonalization of the zero, one, and two quantum Hamiltonian will result in a set of zero, one, and two quantum eigenstates that are a mix of the local state modes. The one quantum eigenstate modes are the a,s or Amide I/II modes. The relationship between the two quantum eigenstates and the one quantum eigenstates is unclear. Therefore, referring to the two eigenstate modes as $2a,2s,as$, assumes that the real two quantum eigenstates are close to what we expect for two quantum states of the one quantum eigenstates.

A 2D IR spectrum can be thought of as a spectrally resolved 'pump-probe' experiment, in which photons are 'pumped' into the system at a frequency given by ω_1 , the energy is allowed to redistribute during τ_2 , and then the system is 'probed' (ω_3) to reveal where the input energy has gone. In our adopted convention, there is no true 'probing' beam, because the LO is viewed as a reference beam used to measure the signal electric field.

The 2D IR spectrum contains positive (red) and negative (blue) peaks. As with linear spectroscopy, we can think of the positive and negative peaks as arising due to two physical pictures: 1) as an increase/depletion of photons in a probing beam or 2) as generation of a signal electric field whose phase relative to a reference pulse creates an apparent increase or decrease in the intensity. Our entire discussion of nonlinear spectroscopy has fallen into the realm of 2), but previous historical conceptions operate within 1) and we discuss them both to reconcile the two pictures.

Experimentally, the heterodyne-detected nonlinear signal is determined by modulation.

In the simplest case, the LO intensity, $|E_{LO}|^2$, is subtracted from the signal+LO intensity, $|E_{sig+E_{LO}}|^2$.

If the same pulse is used for the third interaction and the LO (self-heterodyne), then it is natural to think in convention 1), where the beams 1/2 are the ‘pump’ that causes either an increase or decrease in photons in the ‘probing’ pulse, beam 3/LO. Here, the positive peaks result from a phenomenon called stimulated emission while the process producing the negative peaks is called induced absorption. In this formulation, the ‘pump’ beam (generally) excites the system up to a one quantum state. From this one quantum state, the ‘probe’ can either stimulate emission of photons, which will result in an increase of photons in the ‘probe,’ or photons from the ‘probe’ can be absorbed to excite the system into a two-quantum state, which will deplete the ‘probe.’ In the latter process, the ‘pump’ has induced absorption of the ‘probe.’ Thus, in the RDC case, if the ‘probe’ causes emission from a or s , then we observe a positive peak. If the ‘probe’ is absorbed to create a two-quantum state ($2s$, $2a$, or as), then a negative peak is observed. This is consistent with a linear transmittance spectrum, in which a depletion of photons is negative while an increase is positive.

In the convention 2), where the three interactions generate a signal electric field, then the final step is always thought of as emission from a penultimate state. Thus, both the one- and two-quantum states emit photons, but the accumulated phases attribute to the one- and two-quantum emitted photons are 180° different such that interference with a LO results in the appearance of positive and negative features (respectively). As a justification for adopting this interpretation, we can analyze the homodyne detection case. Here, we measure the square amplitude of the signal electric field in the absence of any reference pulse, and the phase information is lost. The S_{DVE} spectrum has peaks at the frequencies associated with the positive and negative peaks in the S_{DPP} , indicating that photons are emitted in each case.

Even within convention 2), we can still think of the independent (ω_1) and dependent (ω_3) axes of a 2D IR spectrum in terms of a map of where energy is ‘pumped’-in and later ‘probed.’ The ω_1 frequency is determined by the frequency of the coherent superposition created by the first pulse, E_1 . The ω_3 frequency is defined by the frequency of the coherent superposition created directly as a result of the third pulse, E_3 , although this state is actually prepared indirectly by the first two pulses as well.

Based on this discussion, we can better interpret the RDC 2D IR spectrum shown in

Figure 2.5. Peaks 1 and 1' are positive, on-diagonal peaks. They are positive because they are due to emission from the one quantum state s and a , respectively. This is also apparent from their frequencies, where (not considering the lineshape width) $\omega_1 = \omega_3$, indicating that, for peak 1, the first and third interaction produce a coherent superposition between the ground state, 0, and s .

In 2D IR spectroscopy, each diagonal peak is paired with an anharmonically shifted

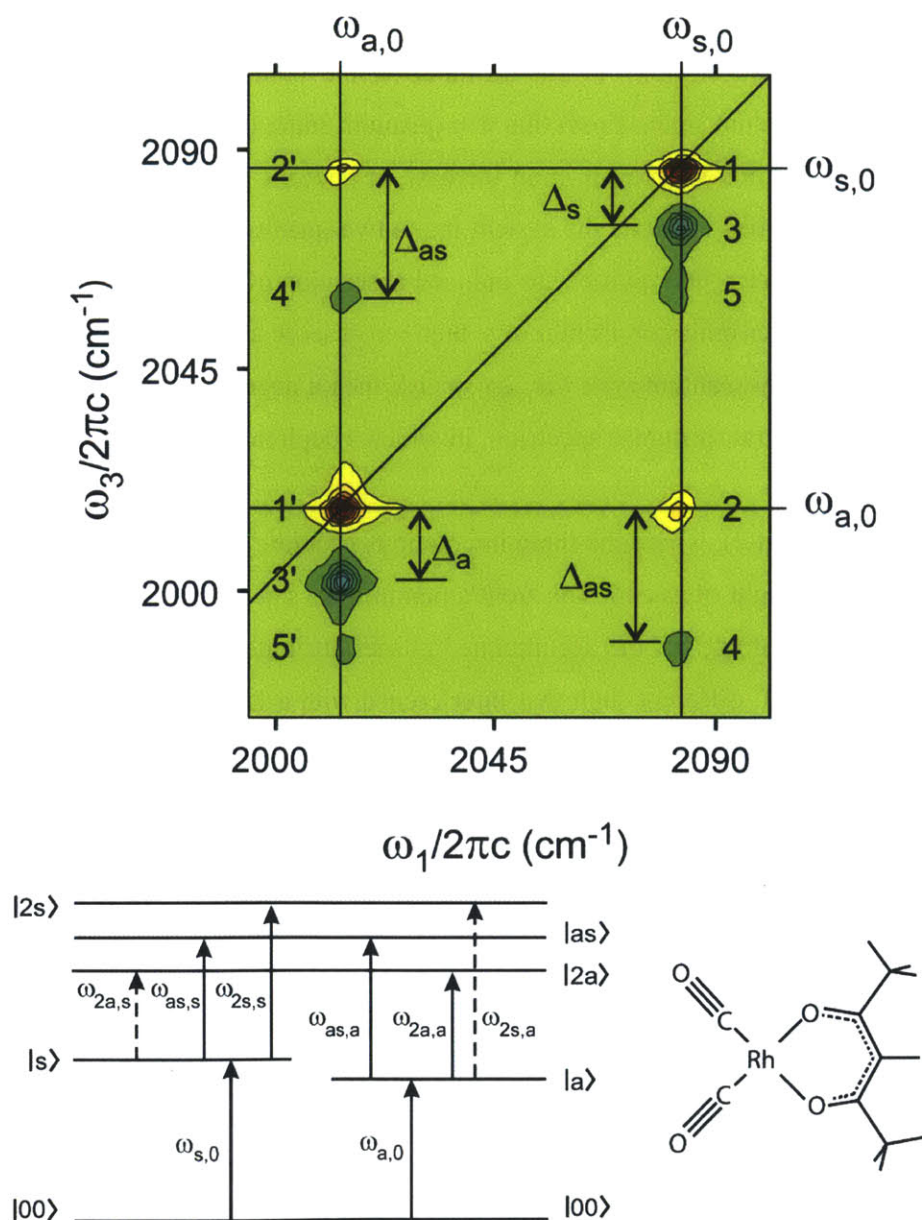


Figure 2.5: The S_{2D} , probed energy level structure, and stick diagram for RDC. Modified from Ref. 2.

negative peak, 3 and 3', that results from probing the two quantum states ($2s$ or $2a$ in this case). For peak 3, $\omega_1 = \omega_{s,0}$, while $\omega_3 = \omega_{2s,s}$ where the subscripts refer to the coherent superposition. Because of the anharmonic nature of vibrations, $E_{2s} = \hbar\omega_{2s}$ is different than twice the energy of the one quantum state: $\omega_{2s} = 2\omega_s - \Delta_s$, where the anharmonicity Δ_s is assumed here to be positive such that the two-quanta energy is less than twice the one-quantum energy. In the limit of no thermally populated one-quantum states ($\hbar\omega \gg k_bT$), if the anharmonicity is 0, then the 1 and 3 peaks will exactly cancel out to give no signal.

The off-diagonal cross-peaks (peaks 2/4 and 2'/4') also occur in complementary pairs in 2D IR. The appearance of these cross-peaks indicates coupling. This coupling is between the local mode two quantum states (which is expected if there is coupling between the local mode one quantum states). Another way to describe these peaks is that they exist because the energy it takes to populate the combination band of two one-quantum eigenstates is less than the sum of the energies of each of the individual states. For RDC, that means that $\omega_{as} = \omega_{s,0} + \omega_{a,0} - \Delta_{as}$. The positive peak 2 occurs due to the first beam exciting into the $\omega_1 = \omega_{s,0}$ superposition and the third interaction exciting into the $\omega_3 = \omega_{a,0}$. Negative peak 4 occurs because of the same initial interaction, $\omega_1 = \omega_{s,0}$, and then an interaction that involves the two quantum state $\omega_3 = \omega_{as,s}$. If the states a and s were uncoupled local modes, then the negative peak 4 would exactly overlap with peak 2 ($\omega_{as} = \omega_s + \omega_a$) and both peaks would cancel. In this way, the existence of these cross-peaks indicates the existence of coupling in the local mode basis. Modeling spectra, in particular the magnitude of Δ_{as} , allows for quantifying the coupling terms. For RDC, identification of the one quantum modes as eigenstates of the two carbonyls ensures the existence of cross-peaks. That is, we know the local modes are highly coupled, and therefore we expect to see cross-peaks. In a system with unassigned peaks, however, the mere presence of cross-peaks indicates that the underlying local modes are coupled and thus the observed one-quantum states are actually eigenstates that result from mixing local modes. The ability to qualitatively determine which modes are connected is a powerful advantage, and it separates 2D IR from linear spectroscopies.

The cross-peaks can also be explained using the 'pump-probe' interpretation – for peak 2 we input a photon at $\omega_1 = \omega_{s,0}$ and we observe output from the $\omega_3 = \omega_{a,0}$ state. This interpretation is particularly helpful for chemical exchange processes, in which varying τ_2 reveals the kinetic conversion of different species. Energy is input into the reactant at a particular frequency ω_1 .

During τ_2 the reaction proceeds, which changes the frequency of the excited vibration. The light is then output at the frequency assigned to the product.

The lineshapes indicate information about the homogeneous and inhomogeneous broadening processes, as shown in Figure 2.3. In the inhomogeneous case, diagonally elongated peaks are observed – light put in at an initial frequency is emitted at the same frequency and $\omega_1 = \omega_3$. If there is homogeneous broadening, then the absorption tagged with the initial interaction can wander to another environment of different frequency before the interaction with the final pulse, and $\omega_1 \neq \omega_3$. These frequency definitions are similar to definitions used to describe diagonal and off diagonal peaks. This raises the interesting question of how to distinguish many overlapping peaks from one inhomogeneously broadened peak. Generally, this distinction is made based on the number of underlying oscillators in the system. Each oscillator is considered to be part of the system, and the number of oscillators per (non-interacting) molecule will result in a single one-quantum peak, where any inhomogeneous broadening observed in this peak is attributed to the bath.

The inhomogeneous broadening occurs along the diagonal, while the homogeneous broadening occurs along both axes. The rounded lineshapes observed for RDC indicates homogeneous broadening. The diagonal width is a measure of the inhomogeneous broadening while the anti-diagonal width is a measure of the homogeneous broadening. Generally, lineshapes change with the τ_2 time. In this regime, at early τ_2 times, the correlation between the ω_1 and ω_3 frequencies is retained and the peaks are inhomogeneously broadened. With increasing τ_2 times, the correlation is blurred, and the peaks gradually shift from inhomogeneously to homogeneously broadened. If the τ_2 (or τ_1 or τ_3) is too long, the correlation between ω_1 and ω_3 will be entirely scrambled and the nonlinear signal will drop to zero. Thus, the decay mechanisms place a limit on the time delays.

Aside from the information determined from the peak positions and lineshapes, the relative magnitude of peaks (particularly their dependence on incident field polarization) and the waiting time dependence all reveal further information about the system. The magnitude of the peaks is related to the interacting dipoles and their projection onto the incident light fields. The τ_2 dependence reveals the processes that occur between the second and third pulses. Polarization and τ_2 dependences, also called anisotropy measurements, can be combined to reveal the re-orientational tumbling timescales.

2.3.5 Discussion of amide I spectra

The 2D IR experiments presented in this thesis are almost exclusively taken on the amide I vibration of proteins and peptides. Although the same principles and equations govern the analysis and interpretation of the 2D IR spectra of both RDC (discussed above) and proteins (discussed in the following chapters and Chapter 1), much of the contained information is inaccessible due to the overlapping nature of protein 2D IR spectra. If the protein has $N+1$ peptide units, then the spectrum will be the result of nonlinear peaks resulting from pathways involving N one-quantum and $(N^2+N)/2$ two-quantum energy levels. Although many of the resulting peaks will have negligible amplitudes, the remainder will combine to create an overlapping mess with cross-peaks showing up as ridges. The homogeneous linewidth of each peak is $\sim 10 \text{ cm}^{-1}$ due to the ~ 1 ps damping exponential. Even though the spectra appear blobby, they contain more information than the analogous FTIR spectra.

Although the macro-peaks observed are the result of many underlying overlapping peaks, structural information can be gained by analyzing a single spectrum. Secondary structures tend to create observed peaks in different frequency ranges.

Due to the overlapping nature of the many peaks, in the absence of isotope-labeling, τ_2 -dependent and quantitative polarization-dependent experiments reveal little valuable information. Thus, 2D IR spectra are usually collected at a single τ_2 delay. Instead of varying the τ_2 or polarization, protein structure is usually tracked by observing changes in the 2D IR spectrum in response to a temperature change, which may be introduced slowly or quickly. When introduced quickly through a temperature jump, the structural changes can be tracked as a function of time following the initiation.

2.4 Conclusion

Linear and nonlinear infrared spectroscopies measure how molecules respond to incident infrared electric fields. Linear absorption measurements reveal vibrational transitions and energies, but different broadening mechanisms are indistinguishable. By using three incident pulses, 2D IR can track multiple successive excitations to reveal coupling between different vibrations and separate homogeneous and inhomogeneous broadening. Although the signal emitted due to linear and nonlinear processes encodes different information about the vibrations within the system, the same equations apply to detection. Delaying a reference beam relative to

the signal beam allows for interferometric measurement of the signal. The technical details regarding collecting the nonlinear spectra are presented in Chapter 3.

2.5 References

- (1) J. J. Loparo, "Ultrafast structural fluctuations and rearrangements of water's hydrogen bonded network," (Massachusetts Institute of Technology, Cambridge, MA, 2007).
- (2) M. Khalil, N. Demirdoven, and A. Tokmakoff, "Coherent 2D IR Spectroscopy: Molecular Structure and Dynamics in Solution," *J. Phys. Chem. A* **107**, 5258-5279 (2003).
- (3) A. Tokmakoff, "OpenCourseWare: 5.74 Introductory Quantum Mechanics II" (2005), retrieved <http://ocw.mit.edu/OcwWeb/Chemistry/5-74Spring-2005/CourseHome/index.htm>.
- (4) J. Sung and R. J. Silbey, "Four wave mixing spectroscopy for a multilevel system," *J. Chem. Phys.* **115**, 9266-9266 (2001).
- (5) D. M. Jonas, "Two-dimensional femtosecond spectroscopy," *Annu. Rev. Phys. Chem.* **54**, 397-424 (2003).
- (6) S. Mukamel, *Principles of Nonlinear Optical Spectroscopy*, Principles of Nonlinear Optical Spectroscopy (Oxford University Press, New York, 1995).

Chapter 3

2D IR Experiment

3.1 Introduction

In this chapter, the 2D IR technique used to collect the data presented in this thesis is discussed. First, the equilibrium 2D IR optical setup, timing, and data processing are described. Particular attention is given to the HeNe overlap, interferometer, and sample box, all of which have been recently modified. In the middle of the chapter, the transient 2D IR optical setup and timing are discussed. Finally, expected spectral and temporal transient 2D IR features are shown and interpreted with particular focus given to the protein time response.

2D IR experiments represent some of the most recent advancements in infrared spectroscopy. By correlating the pumping and probing axes, increased spectral information can be attained relative to FTIR experiments. By using ultrafast lasers with <100 fs pulsewidths, these experiments are conducted with picosecond time resolution.

Transient 2D IR (t-2DIR) experiments integrate another laser system into the mix, which increases the complexity and provides a 4th time dimension that extends our temporal range by many decades beyond the equilibrium 2D IR vibrational lifetime limit of ~3 ps.

T-jump t-2D IR experiments introduce the biologically relevant and powerful variable temperature. By creating abrupt temperature changes, we can measure the time dependent

response of biological systems.

To date, our group is the only laboratory that has successfully conducted T-jump t-2D IR experiments.

3.2 2D IR equilibrium optical setup

A general schematic of the laser system used to collect the spectra presented in this thesis is shown in Figure 3.1. The timing diagram is shown in Figure 3.2. Each step is discussed in more detail below.

3.2.1 800 nm Generation

Briefly, 4 W continuous wave (CW) 532 nm light generated by a diode-pumped Nd:YVO₄ CW Spectra-Physics Millennia (Mountain View, CA) pumps a Ti:Sapphire Spectra-Physics Tsunami (Mountain View, CA) oscillator to generate 82 MHz, 90 fs pulsewidth, ~16 nm FWHM pulses with a center wavelength of 792 nm. ~250 mW (3 nJ per pulse) of the oscillator input is sent into a Spectra-Physics Spitfire (Mountain View, CA) regenerative amplifier that is pumped by a 1 kHz pulsed 527 nm diode-pumped Nd:YLF Spectra-Physics Empower 30 (Mountain View, CA). To reduce the peak pulse power to avoid damage to optics, before entering the amplifier, the oscillator pulses are elongated in time through multiple reflections off of a grating. Retention in the amplifier cavity is controlled by two Pockels cells that flip the

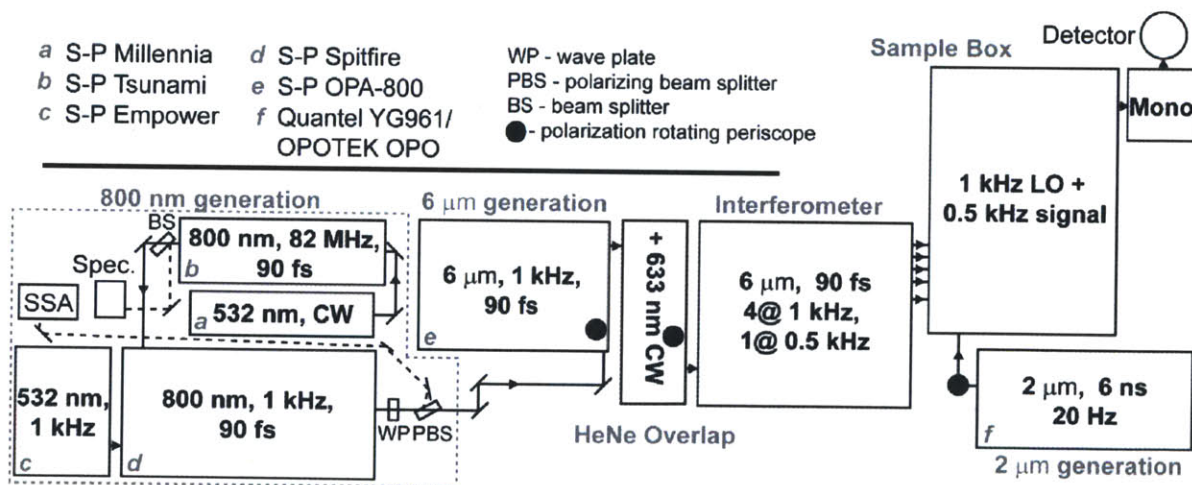


Figure 3.1: A diagram illustrating the full laser setup. Dashed lines indicate the laser light path used for diagnostics. The oscillator output spectrum is measured with an Ocean Optics spectrometer (Spec.), (Dunedin, FL). The amplifier output pulse length is measured with a single shot autocorrelator (SSA). The stretcher and compressor are contained in *d*. The polarization of the beam output of the amplifier and OPA are S and P, respectively.

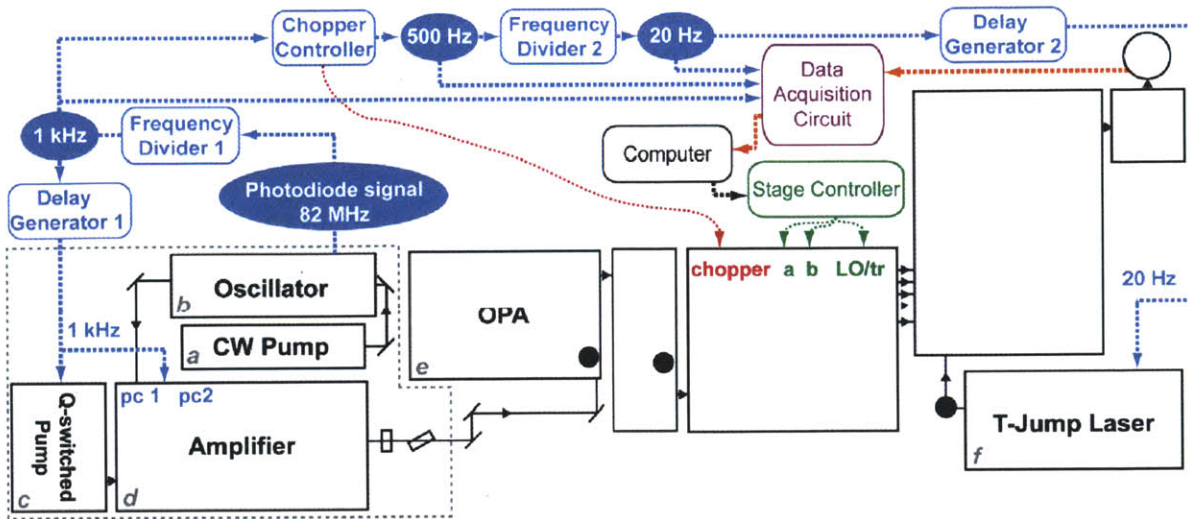


Figure 3.2: A diagram illustrating the triggering timing signals. All timings are derived from the oscillator output. The 82 MHz signal is divided down to 1 kHz (Frequency Divider 1), 500 Hz (Chopper Controller), and 20 Hz (Frequency Divider 2). The Data Acquisition Circuit collects the data and a trigger from each of the 1 kHz, 500 Hz, and 20 Hz triggers. pc1 and pc2 - pockels cells 1 and 2. Compare to Figure 3.1. Modified from Ref. 1.

polarization of selected oscillator pulses so that they are not rejected from the cavity. Each millisecond, the first Pockels cell allows one lucky seed pulse (out of 82,000) from the oscillator to stay in the cavity at a delay optimized to maximize the temporal overlap between the seed pulse and the pumping 532 nm pulse, which dumps its energy into the amplifier Ti:Sapphire crystal. As the lucky, selected seed pulse passes back and forth through the cavity it stimulates emission from the excited Ti:Sapphire crystal, which causes an amplification of power. Once the power is maximized, the amplified pulse is allowed to exit the cavity. The power, stability, and pulse length all depend on the number of passes through the cavity. Too many passes will lengthen the pulse, while too few will lead to instability and low power. Once the buildup produces the proper power and stability, the pulse is allowed to exit the cavity by the second Pockels cell, which sends the pulse to a grating compressor that re-compresses the pulse. The amplified pulse power is ~ 1 mJ (1 W), and its pulse length is ~ 90 fs as monitored by a Positive Light Single Shot Autocorrelator (Los Gatos, CA).

The timing for the 800 nm generation and amplification begins with the oscillator, where the pulsetrain is internally monitored with a diode. The 82 MHz signal is actually measured at half the rate, at 41 MHz, which is passed on to a Medox E/O box (Frequency Divider 1), (now

part of Clark-MXR, Dexter, MI). This Medox box is used to divide the electronic 41 MHz rate down to ~ 1 kHz. This 1 kHz signal is used throughout the laser system. For the 800 nm amplification step, one 1 kHz Medox output is used to trigger a SRS DG535 delay generator (Delay Generator 1),(Sunnyvale, CA), which in turn triggers the Empower firing and the timing of the first and second Pockels cells.

3.2.2 6 μm generation

A more detailed description of the optics and underlying nonlinear processes that result in 6 μm generation is given elsewhere³.

700-800 mW of the 800 nm amplified light is sent to a customized Spectra-Physics OPA-800 optical parametric amplifier (Mountain View, CA) for 6 μm light generation. The beam is initially split into three beams with powers of 1:9:90%. The polarization of the 1% beam is rotated by 90° , and it is focused into a 1 mm sapphire slab to create white light, and then focused into a 3 mm β -Barium Borate (BBO) crystal. The 9% 'first pass' 800 nm light is also focused into the BBO crystal, where it is spatially and temporally overlapped with the white light. The white light acts as a stabilizing seed for the Type II (white light: vertical polarization; 800 nm horizontal polarization) optical parametric amplification process through which the first pass 800 nm photons (ω_3) are split into two photons ($\omega_3 = \omega_2 + \omega_1$), a signal (higher frequency, ω_2) and idler (lower frequency, ω_1). The frequency of the signal (horizontal polarization) and idler (vertical polarization) are determined by the angle of the BBO crystal. Once the signal and idler are generated, the signal is dumped by a dichroic while the idler is transmitted and reflected back into the BBO, where it acts as a seed for the second and final stage of optical parametric amplification. The 90% 'second pass' 800 nm beam is independently focused into the BBO where it is spatially and temporally overlapped with the idler seed pulse. Through the same Type II nonlinear process, ~ 100 mW of signal and idler are produced with 90° rotated relative polarizations. As much of the residual 800 nm light as possible is dumped by dichroics, and the signal and idler are passed on to the final 6 μm generating process, difference frequency generation (DFG).

For our goal of 6 μm generation, each 800 nm (792 nm) photon is split into 1.4 and 1.8 μm photons (the signal and idler from the OPA process). The polarization difference allows for the overlapped beam to be separated by a polarizing beam splitter. The time delay between the two beams is compensated, and they are re-overlapped spatially and focused into a 1mm AgGaS₂

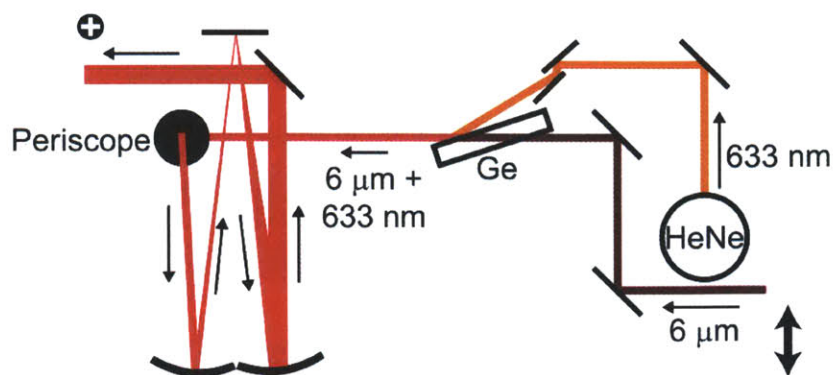


Figure 3.3: A diagram illustrating the HeNe overlap. The $6\ \mu\text{m}$ enters from the OPA on the right, is overlapped with the HeNe beam, and the overlapped pair are sent to the interferometer to the left. Polarization of the $6\ \mu\text{m}$ beam is indicated at bottom right (double-headed arrow, horizontal) and at top left (cross, vertical).

crystal, where a Type II DFG process generates $6\ \mu\text{m}$ photons (horizontal polarization) at the difference frequency between the 1.4 and $1.8\ \mu\text{m}$ frequencies. Following the same convention, for the DFG process, the signal and idler labels are different than in the OPA case: pump = 1.4 , signal = 1.8 , and idler = $6\ \mu\text{m}$. After generation, the $6\ \mu\text{m}$ beam is then passed through a $3\ \text{mm}$ slab of germanium angled at the Brewster's angle to maximize transmission of the $6\ \mu\text{m}$ while absorbing and reflecting residual signal and idler. The germanium also helps correct for the dispersion collected as the beam journeys to the down-table sample. Finally, the $6\ \mu\text{m}$ beam is collimated and sent onto the HeNe overlap box.

The $6\ \mu\text{m}$ beam is absorbed by water vapor lines, and from the DFG until detection, the mid-infrared beam is contained within sealed boxes purged by dried air with a -40°C dew point (Domnick Hunter PNEU DRI MIDiplus DME015, Gateshead, England).

The entire mid-infrared generation produces $\sim 4\ \mu\text{J}$ of $1\ \text{kHz}$, $90\ \text{fs}$ pulsewidth $6\ \mu\text{m}$ light, with an overall efficiency of $\sim 1.4\%$. The $6\ \mu\text{m}$ light that exits the OPA is P (horizontally) polarized.

3.2.3 HeNe Overlap

To make down-table alignment easier, the $6\ \mu\text{m}$ light is overlapped with a visible HeNe beam in the HeNe overlap box. The layout of the HeNe overlap is shown in Figure 3.3. To retain the down-table alignment as much as possible, the $6\ \mu\text{m}$ beam is adjusted to match the HeNe beam, which is immobilized and assumed to be well aligned through the remaining optics. Overlap is maximized by adjusting the $6\ \mu\text{m}$ alignment at the two pre-germanium mirrors to

maximize the signal monitored >5 m down-table after passing through a near- and far-field iris. The HeNe and 6 μm are overlapped on a second 3 mm slab of germanium, which transmits the 6 μm and reflects the HeNe. The HeNe is pre telescoped to match the 6 μm beam size. The polarization of both beams is rotated by 90° to S (vertical) polarization for the 6 μm beam by a periscope, and both beams are expanded to ~ 1 cm diameter and collimated by two spherical focusing mirrors (the beam passes through a focus before diverging and being collimated by the second spherical mirror).

3.2.4 Interferometer Box

The interferometer splits the 6 μm beam (now overlapped and mode-matched with the HeNe 633 nm beam) into 5 beams: 3 are aligned onto three corners of a box for generating the 3rd order nonlinear signal (a,b,c), 1 is aligned on the fourth corner of the box as a tracer beam for alignment (and potentially pump-probe measurements) (tracer), and 1 is used as the local oscillator (LO) to interferometrically detect the generated nonlinear signal. The optical layout for the interferometer box is shown in Figure 3.4.

The five beams are generated by a series of 3 mm thick custom coated Rocky Mountain Instrument Co. ZnSe beam splitters (Lafayette, CO). The anti-reflection (AR) coating is rated at <2.5% average reflection over the 4-8 μm range for S polarization with an angle of incidence of 45° . The 50% coating provides $50\pm 5\%$ reflection of S polarization over the 4-8 μm range with an angle of incidence of 45° .

The first beam splitter (1) is uncoated on the front face (AR on the back), and angled at 56° incident angle to reflect 40% of the power. The transmitted 60% is split by a beam splitter with a 50%/AR front/back face coating to produce the a and b beams with 30% each of the initial 6 μm power.

In the original design, of the transmitted 40% from (1), 25% was intended to be reflected by an uncoated/AR beam splitter (2) angled at 36° incident angle to give a transmitted $40\% \times 75\% = 30\%$ total to create beam c, with the reflected $40\% \times 25\% = 10\%$ total used to create the LO and tracer.

In practice, however, we found that the 10% was unnecessarily too much, and the beam splitter at (2) was replaced with an AR/AR coated beam splitter which theoretically reflects $\sim 2\%$. With this arrangement, the 40% transmitted from (1) is split into a transmitted $98\% \times 40\% = 39\%$ that is used to create the beam c. The reflected $40\% \times 2\% = \sim 1\%$ of the total is further split

into the LO and tracer by a 50%/AR beam splitter.

Ignoring loss due to mirrors and AR coated surfaces, the percentage power is spread across beams a,b,c,tracer,LO with a ratio of 30:30:38:0.5:0.5, which puts the maximum power into beams a, b, and c while still ensuring that the tracer and LO can still saturate the detector. This represents a more ideal arrangement than the previous setup, which used 50%/AR beam splitters to create a ratio of 25:25:25:12.5:12.5, which puts too much power into the LO and tracer at the expense of the other beams. The more ideal arrangement is expected to produce a 50% increase in heterodyned power.

In practice, we find that the intensities of beams a, b, and c are relatively even at the sample, suggesting that reflective losses, differences in focusing, non-ideality in the coatings or incident angles even out the power in each beam. Assessing the overall boost in power is

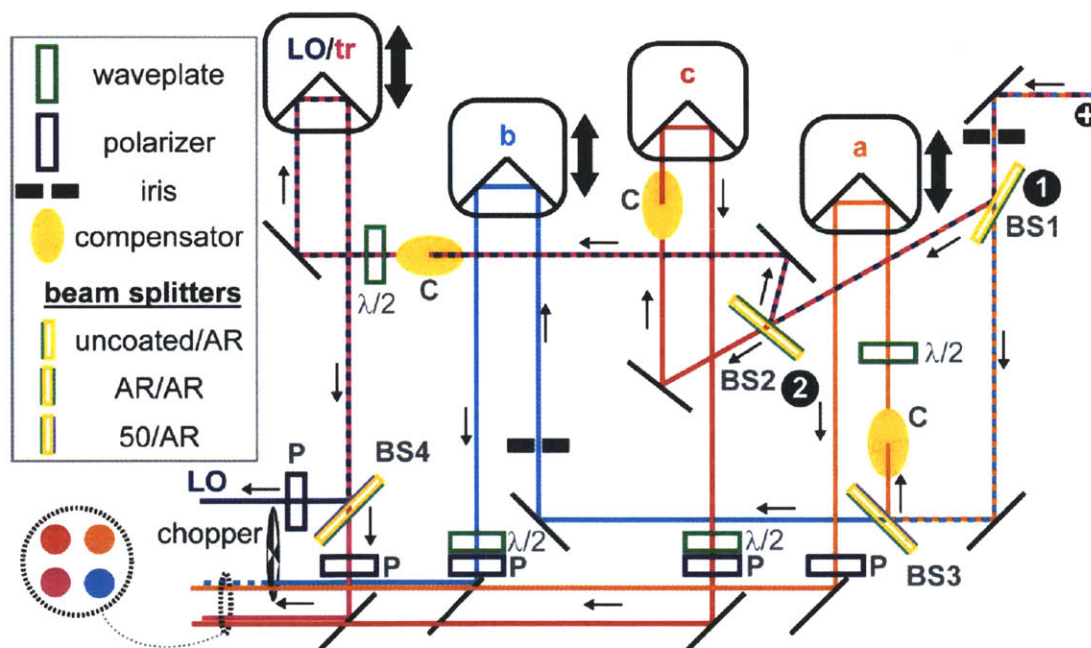


Figure 3.4: A diagram illustrating the interferometer. The 6 μm/HeNe beam entering from the HeNe overlap box at right is split into 5 beams: a, b, c, tr, and LO. At positions 1 and 2 (BS1 and BS2), the front face of the ZnSe beam splitter is uncoated, while the back side is AR coated. At position 2 (BS2), the front and back face of the ZnSe beam splitter are AR coated. The other beam splitters (BS3 and BS4) have a front face coated for 50%/50% reflectance/transmittance, and an AR coated back face. Compensators (C) are AR/AR coated, tilted at a proper angle (for compensation) around the axis parallel to the table. Delays between a, b, c, and LO/tr are created by moving the motorized stages relative to beam c, whose retroreflector is fixed. Vertical displacement of the beams is created by rotating the retroreflectors around the beam axis. The polarization is indicated at top right (vertical, S polarization). λ/2 – half-wave plate; P – polarizer.

difficult given that it is highly dependent on the day-to-day (month-to-month for stable alignments) alignment, but the power increase seems to be in the 25% range.

Beams a, b, c, and LO/tracer undergo an even 2, 2, 2, and 4 reflections, respectively, between the last common mirror and their respective delay stages, which helps to correlate the direction of drifting parallel to the laser table.

AR/AR ZnSe compensators are placed perpendicular to the beam axes (normal incidence) and tilted along the axis parallel to the table to an angle (not the Brewsters angle) that provides the proper amount of material to compensate for the beam splitters. As defined by beam b, which passes through the most ZnSe material, the compensators are tilted such that each beam passes through 10.2 mm of ZnSe. The compensators were tilted towards the laser table to minimize loss (the polarization of the 6 μm beam is 'P' relative to the compensator face). The LO beam only collects its full compensation after passing through the sample box ZnSe 50%/AR beam splitter.

Each of the a, c, and LO/tr beams is independently delayed relative to the fixed beam c. Each beam reflects off of a retroreflector (3" diameter, PLX), which, except for beam c, is mounted on an Aerotech 50-ANT stage (Pittsburgh, PA) with 50 mm stage range, which can produce 333 ps delays. The stages have a 10 nm (0.067 fs) resolution, a 300 nm (2 fs) accuracy, and a 50 nm (0.33 fs) repeatability.

Each beam a, b, c, and LO/tr passes through its own Karl Lambrecht Co. MgF_2 half-wave plate (MWPMF A2-22-6M, Chicago, IL). Beams a, b, c, LO, and tr then pass through their own AR coated 2 mm thick ZnSe wiregrid polarizer (Specac, Inc., Cranston, RI). The half-wave plates are adjusted to maximize the throughput transmitted through the polarizers, which are set to provide specific polarizations at the sample position.

Beams a, b, c, and tr are aligned along the corners of a 1" box. Beam b is chopped at 500 Hz, and its phase is controlled by the chopper controller (New Focus Inc., 3501, Santa Clara, CA) shown in Figure 3.2.

3.2.5 Sample Box

The optical layout for the sample box is shown in Figure 3.5. For all of the spectra collected in this thesis, the internal LO set-up was used. Beams a, b, c are focused into the sample by an unprotected gold 3" diameter, 10 cm effective focal length, 90° off-axis parabolic mirror (Janos Technology, Keene, NH). The beams were consistently focused with 80%

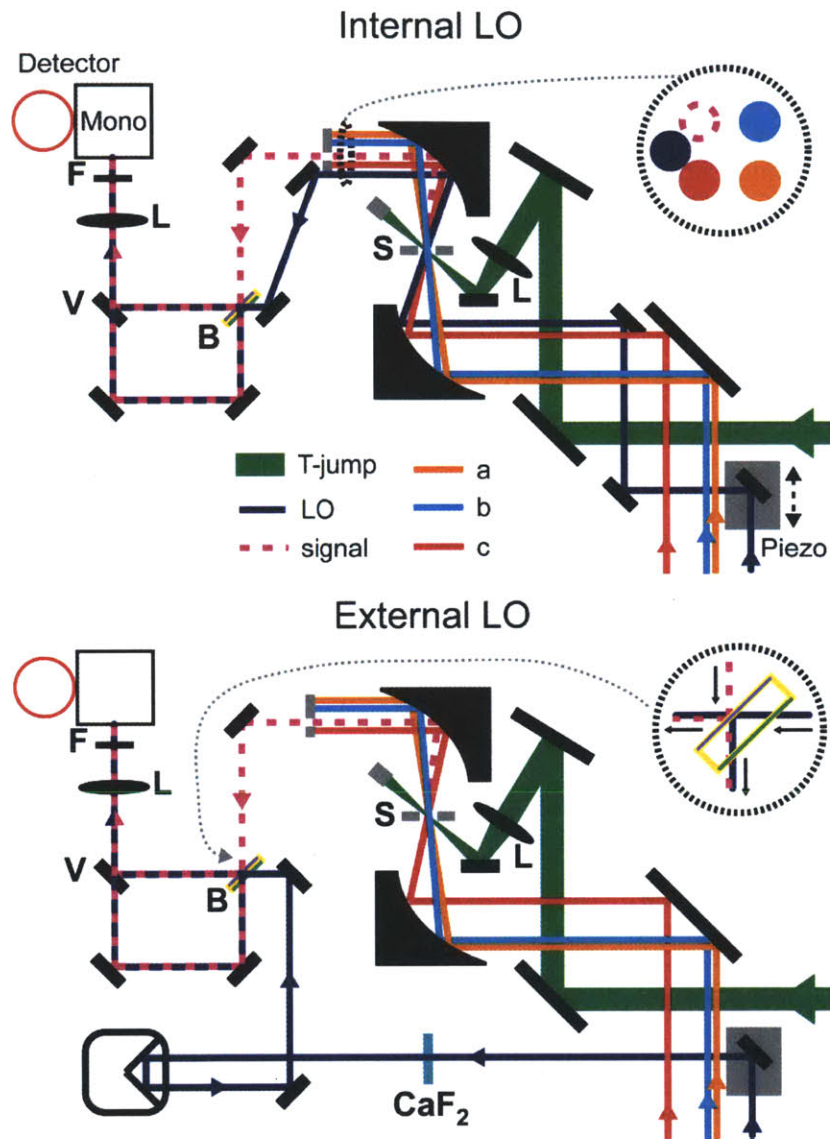


Figure 3.5: A diagram illustrating the sample box. The 6 μm/HeNe beams entering from the interferometer at bottom are focused into the sample, S. *Top* - In the 'Internal LO' case, the local oscillator is passed through the same sample position as the T-jump laser and beams a,b,c, however the LO is delayed by ~25 ps. *Bottom* - In the 'External LO' case, the LO is not passed through the same sample position. The tracer is not shown, although it would emerge from the sample in the same direction as the signal. L - lens, B - 50%/AR beam splitter, V - vertically displaced mirror, F - 3.4 μm low pass filter, CaF₂ - 2 mm Calcium fluoride window to compensate for the transmittance through a 75 μm pinhole. The 3rd order nonlinear signal is emitted along the 4th corner of the box. Beams a, b, and c are blocked, while the signal and LO are transmitted along their respective paths and recombined on a 50%/AR ZnSe beamsplitter, shown in the inset to the

right of the external LO set up in Figure 3.5. The transmitted and reflected beam pairs are vertically separated and focused through the top and bottom of a CaF₂ lens into the 0.25 mm vertical slit entrance of a 190 mm Jobin Yvon Monochromator (Triax 190, Longjumeau, France), diffracted off a 40-lines/mm grating, and collected at 1 kHz on two vertically displaced stripes of 64 Mercury Cadmium Telluride pixels. The pixels are cooled by a liquid nitrogen cooled dewar, and the signal is collected and integrated over a 2.6 μs window by an Infrared Systems Development IR-0144 (Winter Park, FL).

Due to the 180° phase shift due to an external versus internal reflection off of the beam splitter, the signal transmitted/LO reflected and signal reflected/LO transmitted arms produces intensities with oppositely signed cross terms:

$$\begin{aligned}
 I_{obs}^{sig. trans.}(\omega_3) &= \left| \tilde{E}_{sig}(\omega_3) + \tilde{E}_{LO}(\omega_3) \exp(i\pi) \right|^2 \\
 &= \left[e_{sig}(\omega_3) \right]^2 + \left[e_{LO}(\omega_3) \right]^2 - 2e_{LO}(\omega_3)e_{sig}(\omega_3) \times \cos \left[\varphi_{sig}(\omega_3) - \varphi_{LO}(\omega_3) \right] \\
 I_{obs}^{sig. refl.}(\omega_3) &= \left| \tilde{E}_{sig}(\omega_3) + \tilde{E}_{LO}(\omega_3) \right|^2 \\
 &= \left[e_{sig}(\omega_3) \right]^2 + \left[e_{LO}(\omega_3) \right]^2 + 2e_{LO}(\omega_3)e_{sig}(\omega_3) \times \cos \left[\varphi_{sig}(\omega_3) - \varphi_{LO}(\omega_3) \right]
 \end{aligned} \tag{3.1}$$

Subtraction of the two intensities, which are measured on the two stripes of the array, leads to cancelation of the phase independent terms and an increase in the signal to noise¹. This method of detection is called balanced detection.

From its generation until it reaches the sample, the 6 μm beam passes through a KBr lens, 2 x 3 mm Ge windows, 10 mm of ZnSe beam splitters and compensators, a MgF₂ half-wave plate, a 2 mm ZnSe polarizer, and a 1 mm thick CaF₂ sample window. The amount of material passed through is minimized and carefully chosen to minimize and correct for second order dispersion in the pulses⁴.

3.2.6 Timing up pulses

Finding the stage positions that result in time-coincident pulses at the sample is either achieved through 2nd order doubling in a AgGaS₂ crystal positioned at the sample position or through interferometric cross-correlations.

In interferometric cross-correlations, a pinhole (usually with a 75 μm diameter) is positioned at the sample spot, where the beams have been focused and spatially overlapped. Each beam is assumed to be a replica given that they are all generated from the same initial

pulse. The pinhole scatters each incident beam, such that a small portion is directed down the aligned signal path onto the array. Stepping one beam in time relative to another produces an interferometric measurement, whose intensity after integrating over frequency, should be maximized at zero time delay. The equations that describe interferometric detection are described in Chapter 2.

For interferometric cross-correlations, to minimize spatial dependence, an iris previously centered on the signal beam should be closed to only allow a small portion of the scattered light to propagate down the signal arm. Without the spatial masking provided by the iris, the zero times measured on the upper and lower detector stripes may deviate.

3.2.7 Timing up the LO

Finding the proper LO time delay relative to the signal has been an issue in the past. Interferometric cross-correlations allow for the proper LO time delay determination⁵. The time determination is illustrated in Figure 3.6. The scattering pinhole produces intensity of the two incident beams, LO and beam c (the fixed beam, which sets the absolute signal time zero), along both the signal and LO arms. For detecting the time delay between the LO and beam 3 at the sample position, τ_{sample} , the interference between the LO and beam 3 pulses along the signal arm can be measured. To detect the time delay between the LO pulse that travels along the LO arm and the beam 3 pulse that travels along the signal arm, τ_{detector} , we measure the interference between these two pulses traveling different arms.

In the 2D IR experiment, the signal and LO propagate along their respective arms. Determining the time delay between them using an interferometric cross-correlation allows for determination to within +/-10 fs. This is true in principle and in practice.

3.2.8 Collecting equilibrium pump–probe spectra

The pump–probe spectrum is a two beam measurement in which one of the beams must be aligned onto the detector. The other beam is chopped. Based on these requirements, we use beam b as the chopped pump and either the LO or tracer as the probe. If the LO is used, its delay relative to the pump at the sample is set to the required τ_2 delay. It should be noted that the LO is not properly compensated by ZnSe, which means that it will have slightly different dispersion relative to beam b.

For collecting equilibrium 2D and pump–probe spectra in series, it is convenient to use the LO as the probing beam. Delaying the LO and b by many picoseconds relative to beams a

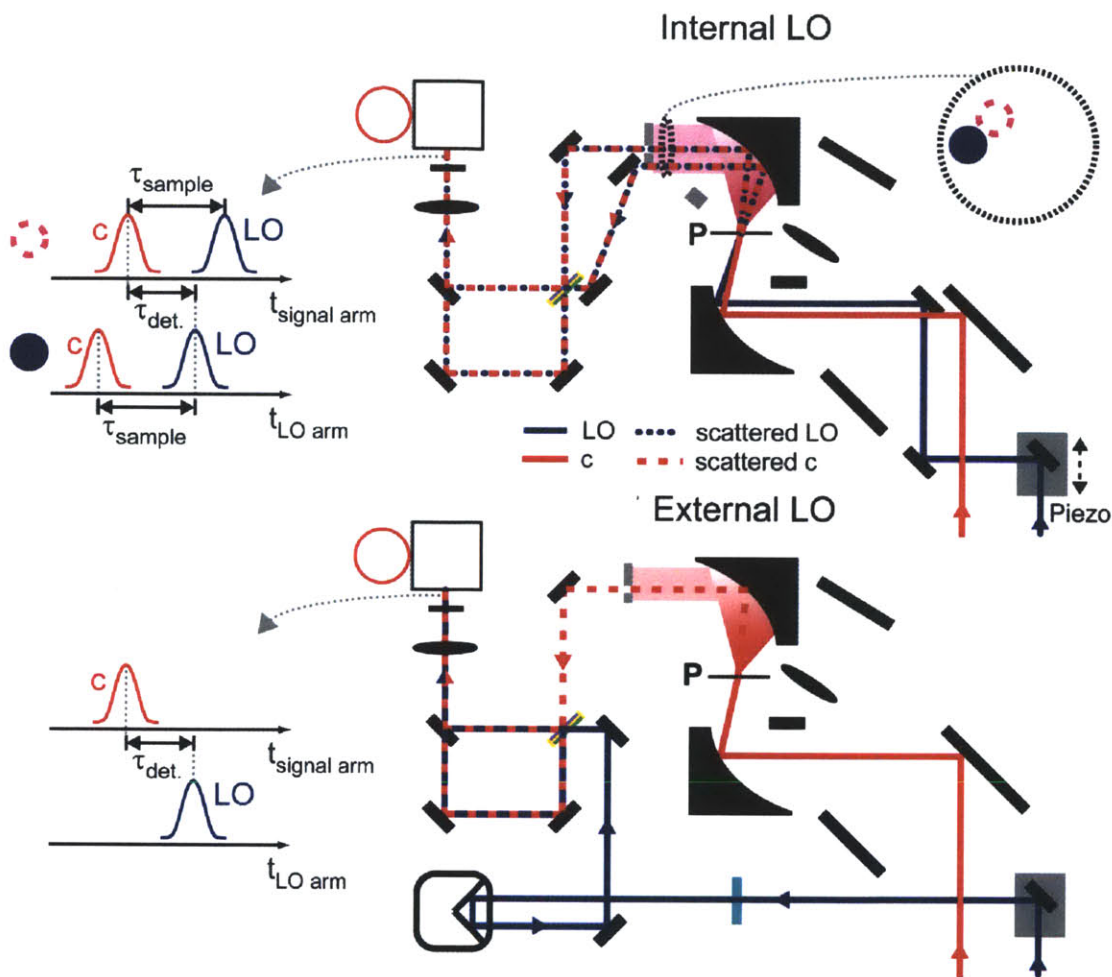


Figure 3.6: A diagram illustrating the LO time determination through interferometric cross-correlations. More detailed labels are given in Figure 3.5. *Internal LO* - Focusing beams c and LO to a pinhole (P) scatters each beam into the LO (blue dotted circle) and signal (red dashed circle) arms, both of which are focused onto the detector. Collecting the time-stepped interference between the c and LO pulse just along the signal arm allows for determining the zero time delay τ_{sample} , which corresponds to time coincidence at the sample. Collecting the time-stepped interference between pulse c propagating along the signal arm and pulse LO propagating along the LO arm allows for determining the zero time delay τ_{det} , which corresponds to time coincidence at the detector (and overlapping beam splitter) between pulses traveling different arms. Although the alignment changes, the LO arm pathlength (measured from sample to detector) is ~ 6 mm (20 ps) shorter than the signal arm.

and c allows for collection of the b/LO pump-probe even with unblocked beams a and c because of the relatively short dephasing time of amide I.

3.2.9 Collecting equilibrium 2D IR spectra

Slight timing errors in the τ_1 and τ_{LO} times creates mixing of the real and imaginary

signals. Before collecting a 2D IR spectrum, the 2-beam pump-probe (b,LO) is collected. The 4-beam (a,b,c,LO) $\text{real}[S_{HDVE}(\omega_3, \tau_2, \tau_1=0, \tau_{LO}=0)]$ is compared to the 2-beam $S_{DPP}(\omega_3, \tau_2, \tau_1=0)$. They should be equal. The $\tau_{LO}=0$ delay is reset such that the 4-beam and 2-beam spectra match.

The (τ_1, ω_3) 2D IR intensity is then collected as a function of τ_1 by stepping either stage a or b. If stage a is stepped, the rephasing surface is collected. If stage b is stepped, then the non-rephasing surface is collected. For the equilibrium spectrum, beam b is chopped at 500 Hz to collect the LO spectrum. Representing the upper(lower) stripe with U(L) and the chopped(unchopped) signal with $\circ(\bullet)$, the measured intensity can be expressed using the convention used in Chapter 2:

$$\begin{aligned} I_{U/L}^\circ &= \left| \tilde{E}_{sig} \pm \tilde{E}_{LO} \right|^2 = e_{sig}^2 + e_{LO}^2 \pm 2e_{sig}e_{LO} \cos(\varphi_{sig} - \varphi_{LO}) \\ I_{U/L}^\bullet &= \left| \tilde{E}_{LO} \right|^2 = e_{LO}^2 \end{aligned} \quad (3.2)$$

where τ_{LO} is assumed to be properly set to 0 (time-coincident with the 3rd pulse). In order to correct for mismatch on the stripes, the unchopped intensity, I° , is subtracted from the chopped intensity, I^\bullet , and the stripes are subtracted to give a slice of the (τ_1, ω_3) 2D IR data $S_{2D}(\omega_3, \tau_2, \tau_1)$:

$$S_{2D} = (I_U^\circ - I_U^\bullet) - (I_L^\circ - I_L^\bullet) \approx 4e_{sig}e_{LO} \cos(\varphi_{sig} - \varphi_{LO}). \quad (3.3)$$

Noise and mismatch between the LO and/or signal intensities on each stripe will lead to incomplete cancelation of the e^2 terms. More complicated post processing to remove the effects of linear absorption is discussed in Chapter 5.

Representative $S_{2D}(\omega_3, \tau_2=150 \text{ fs}, \tau_1)$ rephasing and non-rephasing data sets for 20 mg/mL ubiquitin dissolved in pH*1 solution of DCl in D₂O are shown in Figure 3.7. The τ_1 delays are generally stepped at 4 fs ($\sim 2\pi/5$) steps out to 2.5(1.5) ps for rephasing(non-rephasing) surfaces. Because the phases are accumulated with opposite signs during the τ_1 and τ_3 periods for the rephasing data, the signal echoes, causing intensity to persist longer in the rephasing data, which requires a longer τ_1 scanning range than the non-rephasing case.

3.2.10 Processing equilibrium 2D IR spectra

Timing errors in τ_1 or τ_{LO} cause mixing of the real and imaginary signals, as is described above in the discussion of the LO timing and in Chapter 2. In the post-processing, the timing errors are assumed to be small (the τ_{LO} is assumed to be zero, τ_1 is assumed to be $< \pi$), and they are corrected by adjusting τ_1 independently for the rephasing and non-rephasing data either in the

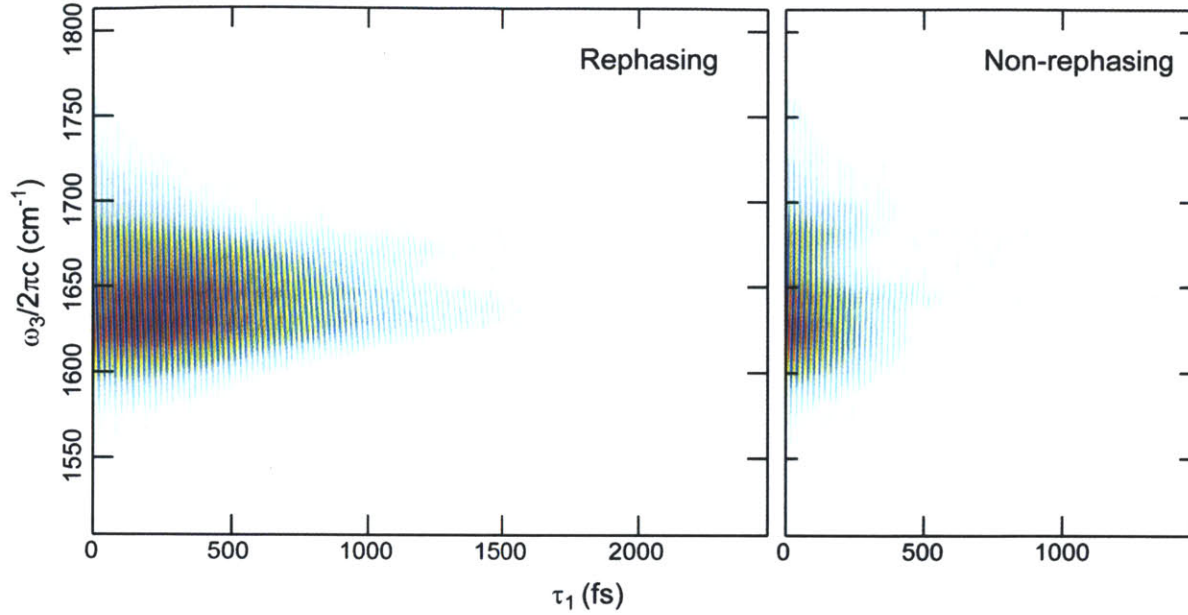


Figure 3.7: $S_{2D}(\omega_3, \tau_2=150 \text{ fs}, \tau_1, \tau_{LO}=0 \text{ fs})$ slices for 20 mg/mL ubiquitin in a pH*1 DCI/D₂O solution. For the rephasing data, beam a precedes beam b and c. For the non-rephasing data, beam b precedes beam a and c.

time or frequency domain. If the τ_1 timing is off by $\Delta\tau_1$, then the FT of either the rephasing or non-rephasing data leads to

$$\begin{aligned}
 \tilde{S}_{2D}(\omega_3, \tau_2, \omega_1) &= \int_{-\infty}^{\infty} S(\omega_3, \tau_2, \tau_1 - \Delta\tau_1) e^{i\omega_1 \tau_1} d\tau_1 \\
 &= 2e_{LO}(\omega_3) e_{sig}(\omega_3, \tau_2, \omega_1) \times \\
 &\quad \exp[i\varphi_{sig}(\omega_3, \tau_2, \omega_1) - i\varphi_{LO}(\omega_3) + i\Delta\tau_1 \omega_1]
 \end{aligned} \tag{3.4}$$

The $\Delta\tau_1$ can either be fixed in the time domain by shifting the data set along τ_1 , or after FT, the complex signal can be multiplied by a factor of $\exp(-i\Delta\tau_1 \omega_1)$. In order to determine $\Delta\tau_1$, we rely on the equivalence of the 2-beam dispersed pump-probe, the 4-beam $\tau_1=0$ 2D/HDVE slice, and the projection of the 2D rephasing and non-rephasing surfaces:

$$\begin{aligned}
 S_{DPP}(\omega_3, \tau_2) &= S_{2D}(\omega_3, \tau_2, \tau_1 = 0) \\
 &= \text{Re} \left[\tilde{S}_{HDVE}(\omega_3, \tau_2, \tau_1 = 0, \tau_{LO} = 0) \right] \\
 &= \int_{-\infty}^{\infty} \text{Re} \left[S_{2D}^r(\omega_3, \tau_2, \omega_1) \right] d\omega_1 \\
 &= \int_{-\infty}^{\infty} \text{Re} \left[S_{2D}^{nr}(\omega_3, \tau_2, \omega_1) \right] d\omega_1
 \end{aligned} \tag{3.5}$$

The $\Delta\tau_1$ correction is usually made in the time domain before FT – the rephasing and non-

rephasing data are independently interpolated along τ_1 and shifted by $\Delta\tau_1$ such that the zero time slice matches the previously collected dispersed pump–probe.

FT of the rephasing and non-rephasing surfaces produces complex 2D spectra. The complex spectra can be displayed individually or as a sum, called the correlation spectrum. The individual or summed spectra can be displayed as real and imaginary or as absolute value (amplitude) or phase.

FT of the non-rephasing surface as shown in Figure 3.7 (with positive τ_1 values) results in the complex conjugate of the desired surface because of the difference in sign in the phase acquired during the τ_1 period. This sign error is corrected in post-processing.

The 2D spectra and their different representations are shown in Figure 3.8. The rephasing and non-rephasing surfaces have dispersive wings along different diagonal and anti-diagonal directions, which is shown more clearly in the phase spectra where constant phase lines are roughly aligned along the diagonal and anti-diagonal directions, respectively. Purely absorptive lineshapes are recovered by summing the rephasing and non-rephasing real spectra⁶.

The 1D equivalent spectra are calculated by projecting the complex 2D spectra (either the rephasing, non-rephasing, or the sum) onto the ω_3 axis (integrating over ω_1). These projections are shown in Figure 3.9. These spectra are displayed to relate and differentiate them. The DPP spectrum is equivalent to the real part of the 2D surface projection. The DVE spectrum is equivalent to the square of the absolute value. The HDVE spectrum is the complex slice at a given τ_1 , τ_2 , and τ_{LO} . HDVE relies on Kramers-Kronig relations or pseudo-phase modulation for complex determination, which differentiates it from the DPP, which is real. Calculating the ‘imaginary’ DPP would otherwise require collecting the full 2D complex spectrum and projecting.

3.3 T-jump t-2D IR optical setup

To collect transient nonlinear 2D IR spectra following a fast change in temperature, a temperature-jump (T-jump)-producing laser is incorporated into the 2D IR spectrometer. Incorporation require spatial alignment, time synching, and proper data collection. The specifics of data collection and processing are described in more detail in Chapter 5.

The T-jump is created by exciting the $\nu=2\leftarrow 0$ OD stretch of D_2O with a ~ 6 ns 1.98 μm pulse at 20 Hz generated as described below.

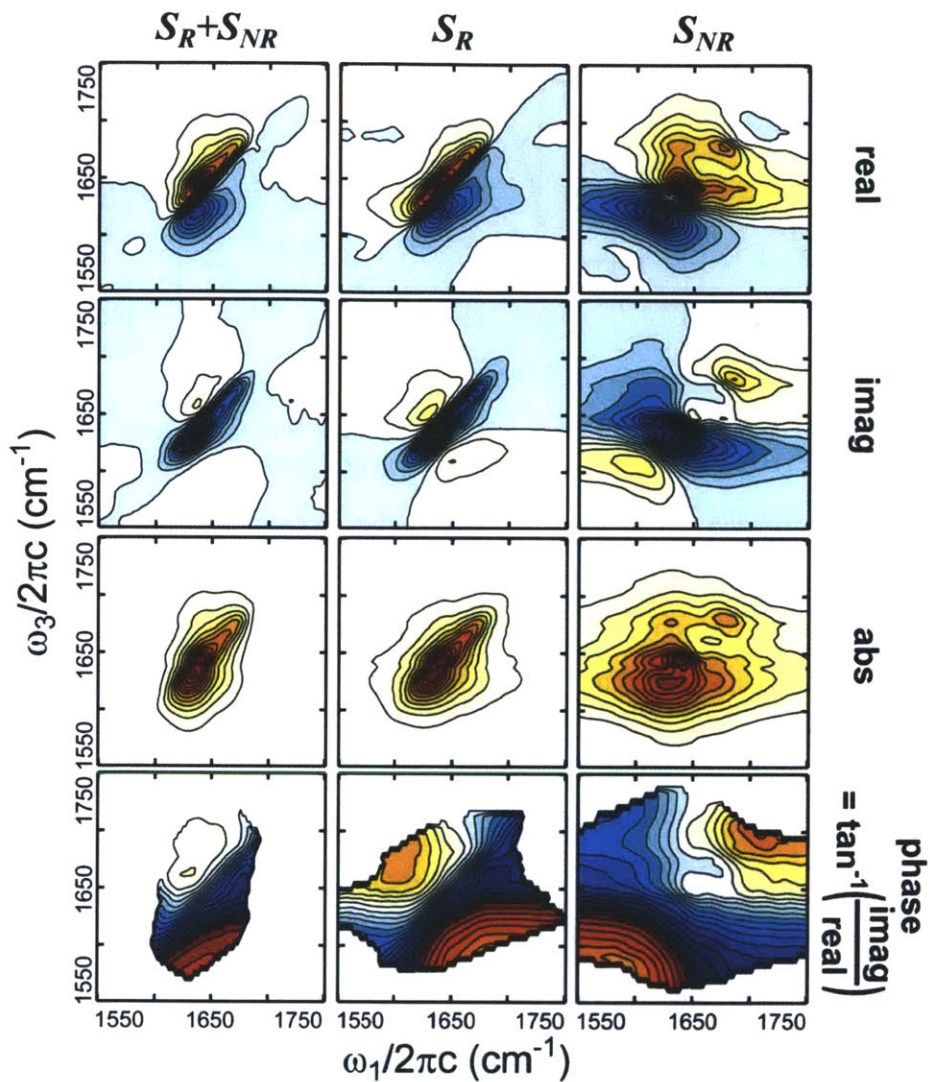


Figure 3.8: Different representations of the rephasing, non-rephasing, and summed 2D surfaces for 20 mg/mL ubiquitin in a pH*1 DCI/D₂O solution (FT from the data shown in Figure 3.7). 25 equally spaced contours are plotted. White contours indicate 0. In the phase spectra the abrupt transition from blue to red indicates the wrapped transition from $-\pi$ to π .

3.3.1 2 μm generation

The 2 μm pulse is generated by a pulsed flashlamp pumped Nd:YAG Q-switched laser (Quantel YG981C, Newbury, UK) coupled to an optical parametric amplifier (OPO) (OPOTEK, Carlsbad, CA). The flashlamps pump an oscillator cavity, whose output is released at 20 Hz by a triggered Q-switch. The released oscillator pulse is amplified by a second flashlamp pumped rod, and the 1064 nm output is frequency doubled and used to pump a BBO based OPO, which

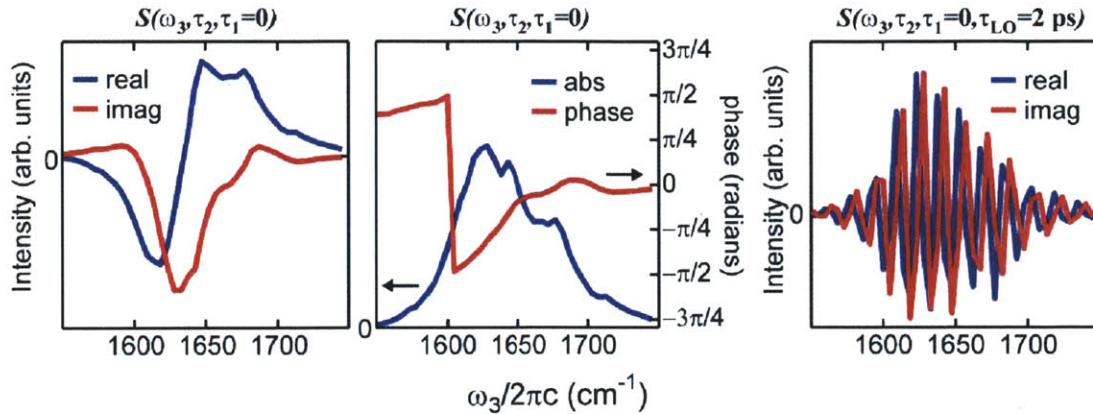


Figure 3.9: The projections of the S_R+S_{NR} 2D surfaces onto the ω_3 axis. To show what the HDVE data would look like at $\tau_{LO} = 2000$ fs, a $\omega_3\tau_{LO}$ linear phase term was added.

produces a signal and idler. The $1.98 \mu\text{m}$ idler pulse is emitted, expanded by a telescope, propagated to the sample, and focused into the sample as shown in Figure 3.5 to a $500 \mu\text{m}$ diameter spot (90% transmission). 6 mJ of power reach the sample, but only $\sim 5\%$ is absorbed by the $50 \mu\text{m}$ long pathlength of the sample. The larger size diameter of the T-jump pulse relative to the $6 \mu\text{m}$ pulses and the low extinction coefficient of D_2O at $2 \mu\text{m}$ ensures a fairly homogeneous heating of the sample volume.

The 20 Hz triggering signal is divided down from the same 82 MHz oscillator pulse train signal that is used to trigger the 1 kHz 800 nm and $6 \mu\text{m}$ laser pulses. The triggering signals are illustrated in Figure 3.2 and in more detail in Figure 3.10. The 500 Hz triggering signal from the chopper controller is divided down by a factor of 25 to produce a 20 Hz triggering signal by frequency divider 2 in Figure 3.2 (PRL260NT, Pulse Research Lab, Torrance, CA). The 20 Hz trigger is sent to delay generator 2 (SRS DG535 delay generator, Sunny Vale, CA). The delay generator triggers the firing of the flashlamps, which, after a $300 \mu\text{s}$ delay, is then followed by the triggering of the Q-switch. The delay between the 1 kHz $6 \mu\text{m}$ beam set (a,b,c,LO,tracer) and the T-jump laser pulse at the sample, τ , is electronically set relative to the *previous* 1 kHz $6 \mu\text{m}$ pulse with a ± 2 ns timing jitter. That is, if the 0^{th} $6 \mu\text{m}$ pulse set arrives at a τ delay following the T-jump, then the 20 Hz delay generator was triggered by a 1 kHz electrical signal associated with the -1^{st} $6 \mu\text{m}$ pulse set. This arrangement sets an upper limit on the τ delay of $\sim 660 \mu\text{s}$, which represents the time necessary for the initial $6 \mu\text{m}$ pulse to set the T-jump pulse generation in motion.

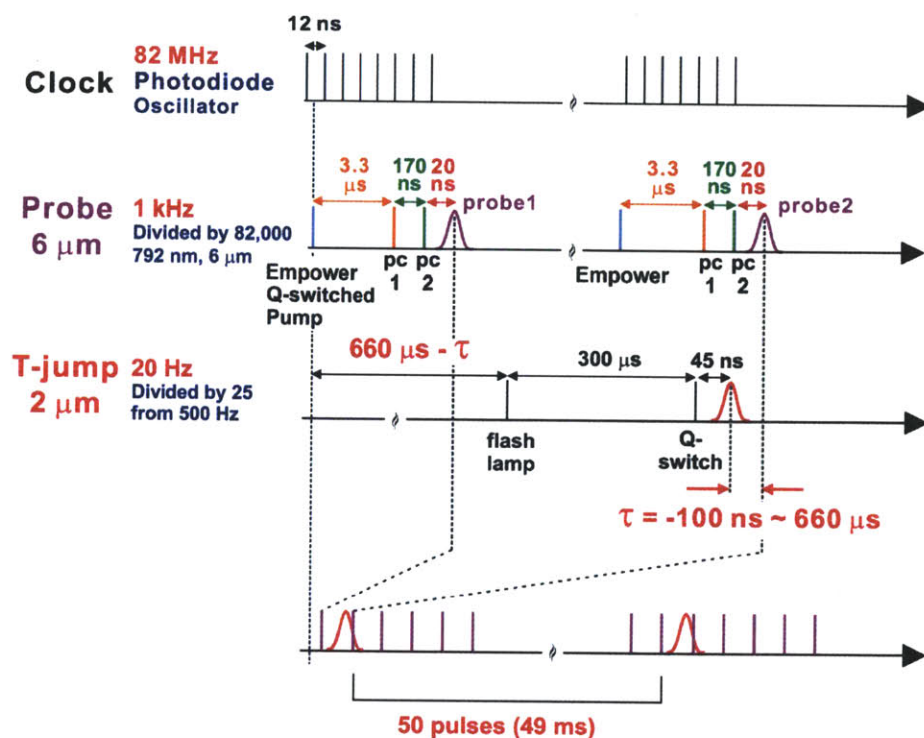


Figure 3.10: The time delays between pulses are illustrated. Modified slightly from Ref 1.

The spatial and timing overlap between the 2 and 6 μm pulses is achieved by monitoring the transient transmission signal (internal beam LO generally) from the D_2O solvent. A 10°C T-jump creates a $\sim 4\%$ increase in transmission in the 6 μm range. The $\tau = 0$ time is set at the delay which produces $1/2$ of the maximum transient transmission change.

Following the first 6 μm pulse set at a delay τ following the T-jump pulse, there are 49 ms pulses at delay $1 \text{ ms} + \tau$, $2 \text{ ms} + \tau \dots$ that are collected to map out the sample response.

3.3.2 T-jump profile

Characteristic transient transmission time dependencies and difference spectra are shown in Figure 3.12. These spectra and time traces represent the change in temperature of the solvent in response to the T-jump pulse. The 5-6 mJ T-jump pulse increases the temperature by $\sim 10^\circ\text{C}$ on the -5 – 5 ns timescale. A slight sloping rise in temperature is observed on the 5 – 200 ns timescale, as is shown in the inset in Figure 3.12. This rise is attributed to cavitation or shockwaves in the sample¹, and its magnitude relative to the total temperature change increases with increasing initial temperature, which suggests that it is not electronic noise. Its magnitude also increases with increasing temperature jump size (not shown).

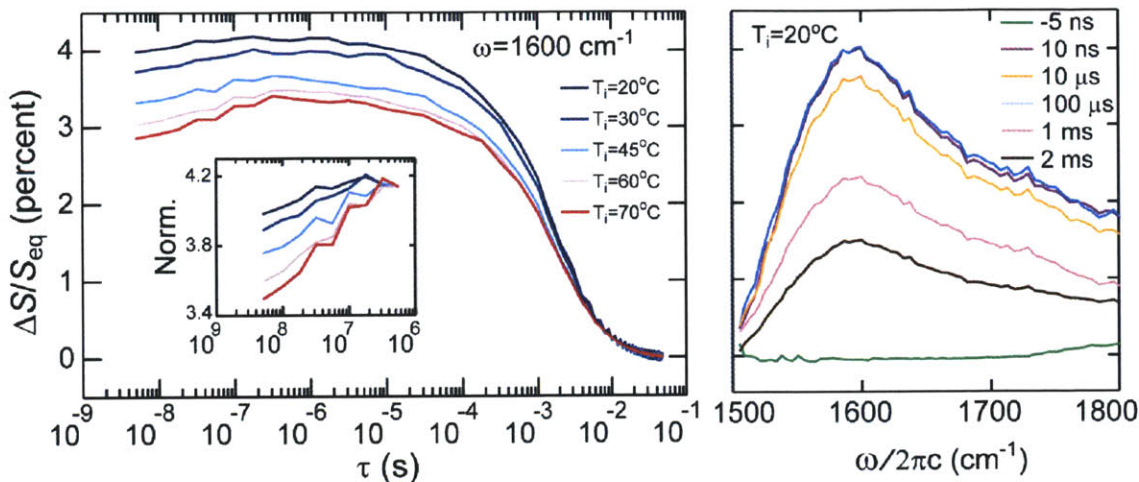


Figure 3.12: The transient transmission time traces at five different initial temperatures (left) and selected difference spectra at $T_i=20^\circ\text{C}$ (right).

From 200 ns until 100 μs , the temperature is constant. After 100 μs , re-equilibration occurs. The >200 ns profile is well fit by a stretched exponential, as is described in Figure 3.11. The shape of the profile is due to two heat diffusion processes: radially out from the beam axis into the solvent and forward and backward along the beam axis into the CaF_2 windows that sandwich the sample. The thermal conductivity is a factor of ~ 10 higher for CaF_2 than for D_2O , which suggests that heat is primarily lost through the windows rather than into the solvent, although both effects will contribute¹. Changing the pathlength or window material will change

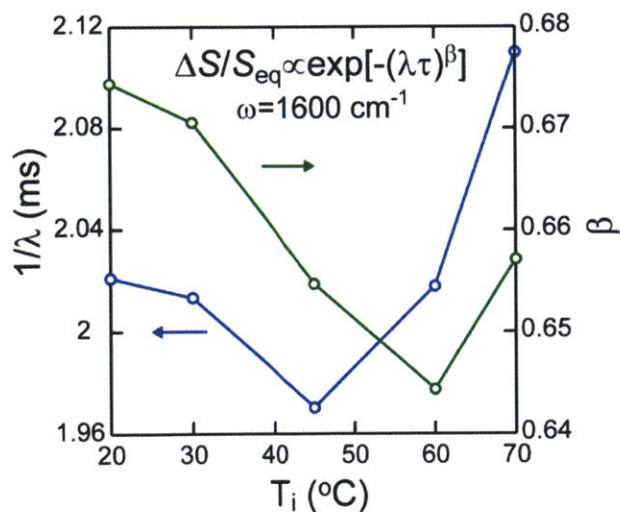


Figure 3.11: Parameters from a stretched exponential fit to the transient transmission time traces shown in Figure 3.12 at delays >500 ns.

the thermal profile. The ideal profile would provide a long constant high temperature period and enough time for re-equilibration before the next 20 Hz T-jump pulse.

3.3.3 Data collection

Both the t-2D IR and t-HDVE measurements are taken following the T-jump. To separate the desired signals, the data must be collected at variable chopper phase (0, 180°), variable T-jump delay (τ) and at variable τ_{LO} delay (for t-HDVE) or at variable τ_1 delay (for t-2D IR). Slices of ω_3 data are collected as a function of adjusted parameters ($\tau; \tau_1$; chopper phase) and ($\tau; \tau_{LO}$; chopper phase) for the t-2D IR and t-HDVE cases, respectively, where ordering of variables indicate the parametric dependence from right to left variables. For example, in the t-2D IR case, first data is collected at one chopper phase. The chopper phase is switched 180° to the other chopper phase and data is collected again. Then, the τ_1 is stepped, and at each step, data at the two chopper phases are collected. Finally, the τ is changed, and the τ_1 and chopper phase stepping begins again. Stepping the T-jump delay τ at each τ_1 (switching the data collection nested loop order) means that each individual surface is collected over the span of the entire collection process, which might create more consistency between different τ spectra. This is how the data was collected for earlier experiments⁷. It has been observed that slow time drift in the stages creates distortions in the Fourier transformed spectra if the time elapsed between the first and last τ_1 data collection is greater than $\sim 1.5 - 2$ h. This distortion may be due to a specific bum stage that might be easily corrected.

At each discrete point in the ($\tau_2, \tau_1, \tau_{LO}, \tau, \text{chopper phase}$) space, intensity is collected with and without the signal (beam b) blocked:

$$I(\omega_3, \tau_2, \tau_1, \tau_{LO}, \tau)^\circ = 2e_{LO}(\omega_3, \tau)e_{sig}(\omega_3, \tau_2, \tau_1, \tau_{LO}, \tau) \times \exp[i\phi_{sig}(\omega_3, \tau_2, \tau_1, \tau) - i\phi_{LO}(\omega_3, \tau) + i\tau_{LO}\omega_3]. \quad (3.6)$$

$$I(\omega_3)^\circ = [e_{LO}(\omega_3, \tau)]^2$$

If the chopper phase is 0°, the even 6 μm shots (of which there are 50 following each T-jump pulse) are open ($I_{j=0,2,\dots,48} = I^\circ, I_{j=1,3,\dots,49} = I$). If the chopper phase is 180°, the odd shots are open ($I_{j=1,3,\dots,49} = I^\circ, I_{j=0,2,\dots,48} = I$). Thus, for each ($\tau, \tau_1, \text{chopper phase}$) or ($\tau, \tau_{LO}, \text{chopper phase}$) setting, two stripes of 64 pixel data is collected for each of the 50 6 μm pulse sets. Usually 140-200 averages (7000-10000 total shots) of this data is collected before changing one of the variables.

The process is illustrated in Figure 3.13. The four 6 μm pulses measure snapshot

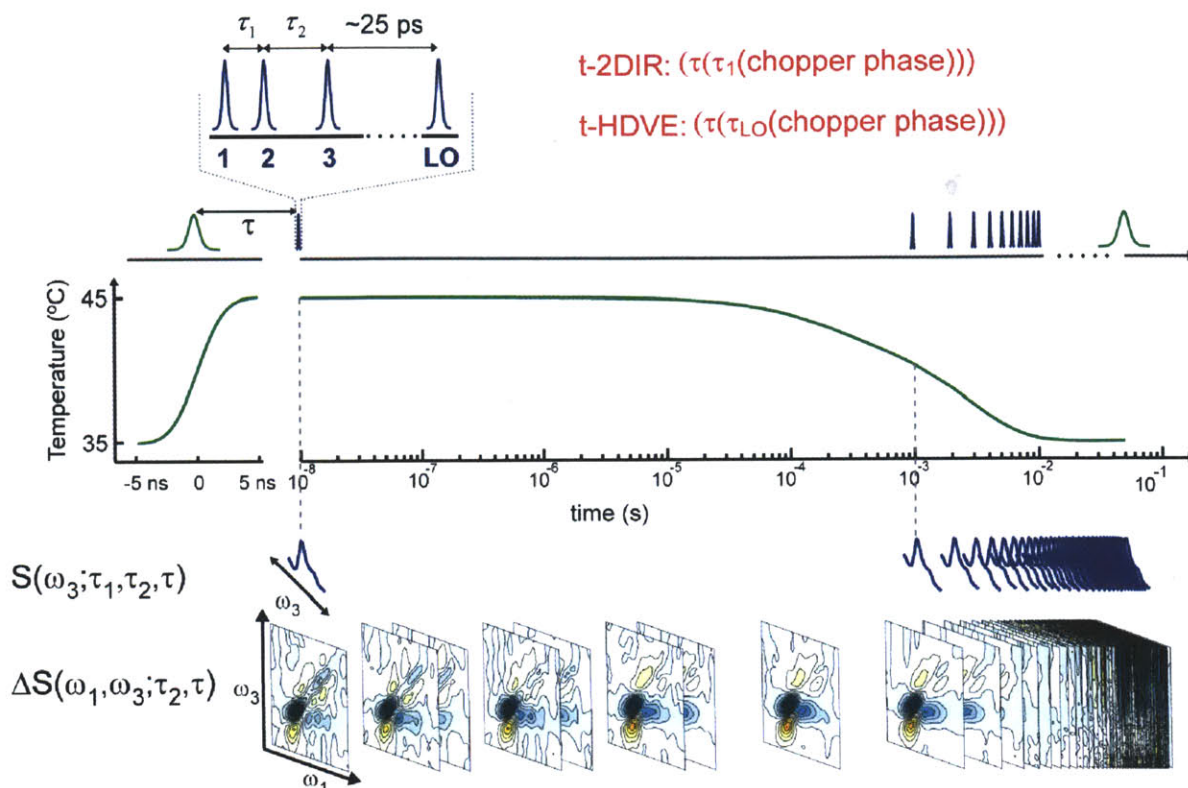


Figure 3.13: The transient data collection is illustrated. The set of four 6 μm pulses (blue) is delayed from the T-jump pulse (green) by τ . The T-jump pulse creates a T-jump in the sample with a profile shown as a function of time at center. The four 6 μm pulses generate a slice of data at τ and at following ms increments. The slices of data are either collected as a function of τ_1 and FT to give a 2D surface at the stepped τ delays (shown at bottom), or the τ_{LO} is stepped to provide the complex t-HDVE spectra (not show).

nonlinear spectra at the delay τ following the T-jump pulse. Parametrically collecting these slice spectra as a function of τ_1 allows for FT to give the t-2D IR. Parametrically collecting these slice spectra as a function of τ_{LO} allows for the complex spectra to be measured to give the t-HDVE.

3.3.4 T-jump laser induced interferometric fluctuations

When placed in direct contact with the laser table, operation of the T-jump laser creates reproducible fluctuations in the interferometric measurements. This suggests that the flashlamp firing creates low-frequency vibrations in the laser table, which in turn create movement in the optics, which causes unwanted modulation of the 6 μm beam time delays. This modulation manifests itself by mixing real and imaginary parts of the measured spectrum, which appears as amplitude changes if one pixel is monitored. The DPP amplitude and calculated pulse delay fluctuations are shown in Figure 3.14, and they correspond to a ~ 0.05 fs slip in the time delay

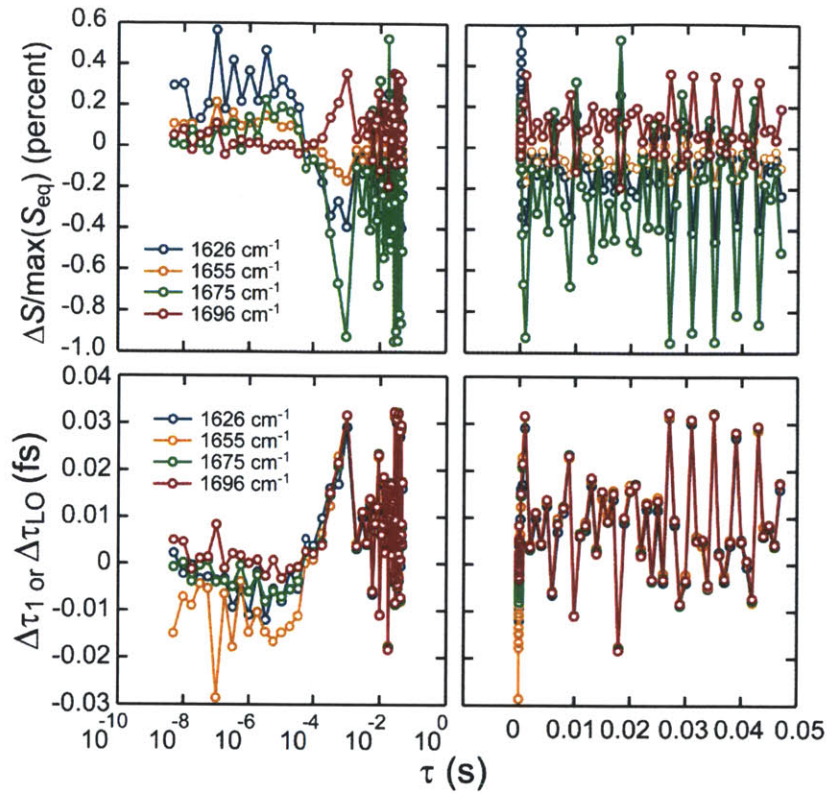


Figure 3.14: Modulations in the interferometrically detected signal. At top, the amplitude of the DPP ($\text{Re}[S_{\text{HDVE}}(\omega_3, \tau_2 = 150 \text{ fs}, \tau_1 = 0 \text{ fs}, \tau)]$) spectrum of diglycine due to the T-jump laser induced fluctuations is shown. At bottom, from the same data set, is shown the change in time $\Delta\tau = \varphi/\omega_3$ calculated from the complex HDVE spectrum. The T-jump laser pulse was blocked during the collection of this data. Modulations are therefore the result of low frequency vibrations created by the firing of the T-jump laser flashlamps, which changes the time delay between the 6 μm pulses.

between the 6 μm pulses – either between the LO, beam a, or beam b. Note that frequencies above and below the equilibrium peak maximum have anti-correlated amplitude changes, which is consistent with a time slip phenomenon. If the vibrations were moving the beams off of the detector, we would observe a monotonic drop in DPP amplitude. If the absolute value is calculated (the $\text{DVE}^{1/2}$), which is insensitive to time slip, the fluctuations are suppressed.

Physically moving the T-jump laser to a different table or turning off the flashlamps removes the fluctuations.

To resolve the unwanted fluctuations, the T-jump laser was vibrationally isolated from the laser table by placing it on foam. Over time this foam compresses, causing the gradual return of fluctuations.

3.4 Interpreting the T-jump data

The transient T-jump experiment collects picosecond snapshot nonlinear infrared spectra that report on the ensemble's response to the T-jump.

A protein system is in equilibrium if its states' populations match those expected based on the free energy at the current temperature. If the protein response to the temperature profile is slower than ~ 10 ns, then the protein's population will be in a non-equilibrium state for at least a portion of the observed experimental time range. This is the ultimate goal of the T-jump experiment: we hope to raise the temperature faster than the protein can fully react such that we may observe subsequent equilibration. The time-dependent spectral changes are interpreted to provide a physical picture of the protein response.

3.4.1 t-2D IR spectral features

The observed t-2D IR changes are often small, on the order of $\sim 5\%$. To highlight the changes, the spectra are presented as difference spectra relative to the equilibrium spectrum. The difference spectral features are often difficult to interpret because positive and negative peaks can represent gain or loss depending on the sign of the corresponding equilibrium peak.

Simulated t-2D IR difference correlation (real [rephasing+non-rephasing]) spectra are shown in Figure 3.15 to illustrate the features associated with gain, loss, frequency shifting, and homogeneous broadening. As is shown in the homogeneous loss case, loss of a peak causes the opposite signed feature in the difference spectrum. Blue shifting corresponds to loss of one peak and the appearance of a higher frequency peak. For the homogeneous blue-shifting, loss and gain features merge in between to create 3 apparent features (red, blue, red moving up the diagonal). Because the homogeneous peak is broader than the inhomogeneous, a shift from inhomogeneous to homogeneous will result in loss along the diagonal (loss of the inhomogeneous intensity) and gain in the off-diagonal regions (gain of the broader homogeneous peak).

3.4.2 Times dependence of the signal

The T-jump pulse generates a temperature change in the solvent that has a profile as described above and shown in Figure 3.12. The protein responds to this temperature profile.

The observed signal, which is (for 2D IR, DPP, and HDVE but not DVE or amplitude) linearly proportional to different populations within the sample, can be difficult to interpret given

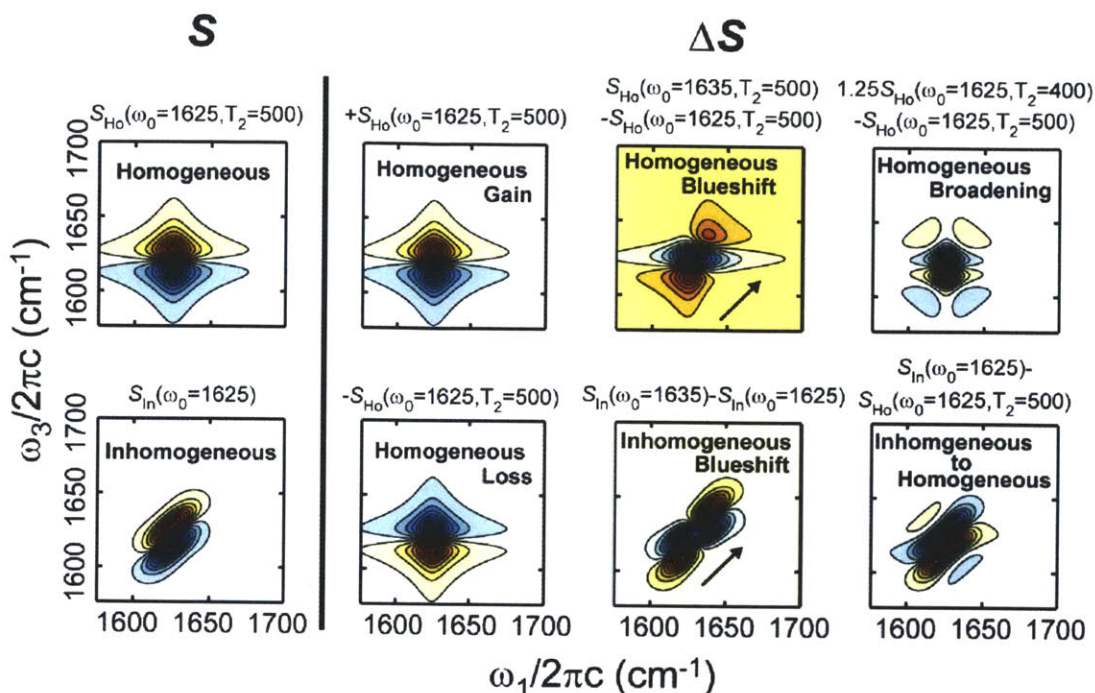


Figure 3.15: A diagram illustrating the difference spectra associated with gain, loss, shifting (loss paired with gain at a different frequency), and broadening. The peaks were generated using code provided by Chunte Sam Peng, described in Ref. 2. Homogeneous T_2 dephasing times are given in fs. A $T_2=500$ fs corresponds to ~ 22 cm^{-1} FWHM. Blueshifting corresponds to loss and gain of two peaks, the second of which is blueshifted relative to the lost, original peak.

that the timescales observed are due to the response of the protein to the temperature profile, which has stretched exponential behavior; the time-dependence of the observed signal is some complicated mixed version of the protein kinetic rates and the temperature profile.

Immediately following the T-jump, the protein responds as populations re-equilibrate to the higher, T-jump-accessed temperature. Once the temperature begins to drop back to the equilibrium value, the protein populations also re-equilibrate at rates determined by the instantaneous solvent temperature. If the reaction rates are much slower than the temperature re-equilibration, the protein populations will not be at equilibrium as the temperature is dropping.

Convolution of ΔT with $\exp(-\lambda\tau)$

To get a sense for how the observed signal corresponds to the protein response, we can approximate the protein as a linear time-invariant (LTI) system. For LTI systems, the observed signal, S , is the convolution of the input (the temperature profile, ΔT) and the protein's response (to an impulse function), R :

$$\begin{aligned}
S(\tau) &= [R * \Delta T](\tau) \\
&= \int_{-\infty}^{\infty} R(\tau - t) \Delta T(t) dt \\
&= \text{FT}^{-1} \{ \text{FT}[R(\tau)] \times \text{FT}[\Delta T(\tau)] \}
\end{aligned} \tag{3.7}$$

For a two-state protein, we can assume that a temperature change will create a mono-exponential kinetic response whose observed rate is equivalent to the forward and backward reaction rates. A true protein is not a LTI system because the rate is temperature dependent. If we temporarily ignore this, we can predict the observed signal if we assume that: 1) the protein response to an impulsive temperature profile is a mono-exponential relaxation whose amplitude is proportional to the magnitude of the impulsive input and 2) the protein response rate is independent of the size of the temperature jump. Under these assumptions, convolving mono-exponentials (the protein response), $R = \exp(-\lambda\tau)$, with the stretched exponential ΔT results in the ‘predicted’ signals S shown at the top right of Figure 3.16.

An equivalent convolution can be done for arbitrary time behavior (stretched exponentials, power law) to predict the signal that would be observed for these cases. Here, we limit ourselves to mono-exponentials.

Each S trace has a rise and a fall. The rise is due to the initial response to the temperature profile, while the fall can be interpreted as the protein’s response to the drop in the temperature profile. For a true protein, the interpretation is generally thought of differently – the drop at long times is due to re-equilibration of population to the equilibrium, initial temperature. For this LTI protein there is no re-equilibration phenomenon because population is always moving to the final state when the temperature is above the equilibrium temperature – there is not return.

When $1/\lambda_{\text{input}} > 500 \mu\text{s}$, the S trace can be fit reasonably well to a bi-exponential with oppositely signed amplitudes (one for the rise, one for the fall). When $1/\lambda_{\text{input}} < 500 \mu\text{s}$, the rise can be well fit to a mono-exponential, but the drop is close to the temperature profile stretched-exponential drop. The plot of the $1/\lambda_{\text{input}}$ versus the rise and fall $1/\lambda_{\text{meas.}}$ is shown in the bottom left of Figure 3.16.

Inspection of this plot leads to some intuitive conclusions. When the protein responds quickly ($1/\lambda_{\text{input}} < 50 \mu\text{s}$), the measured rise rate matches the input rate while the measured drop rate matches the temperature profile drop – the protein rate is fast enough that it responds while the temperature is roughly constant and the temperature-profile drop is relatively slow such that

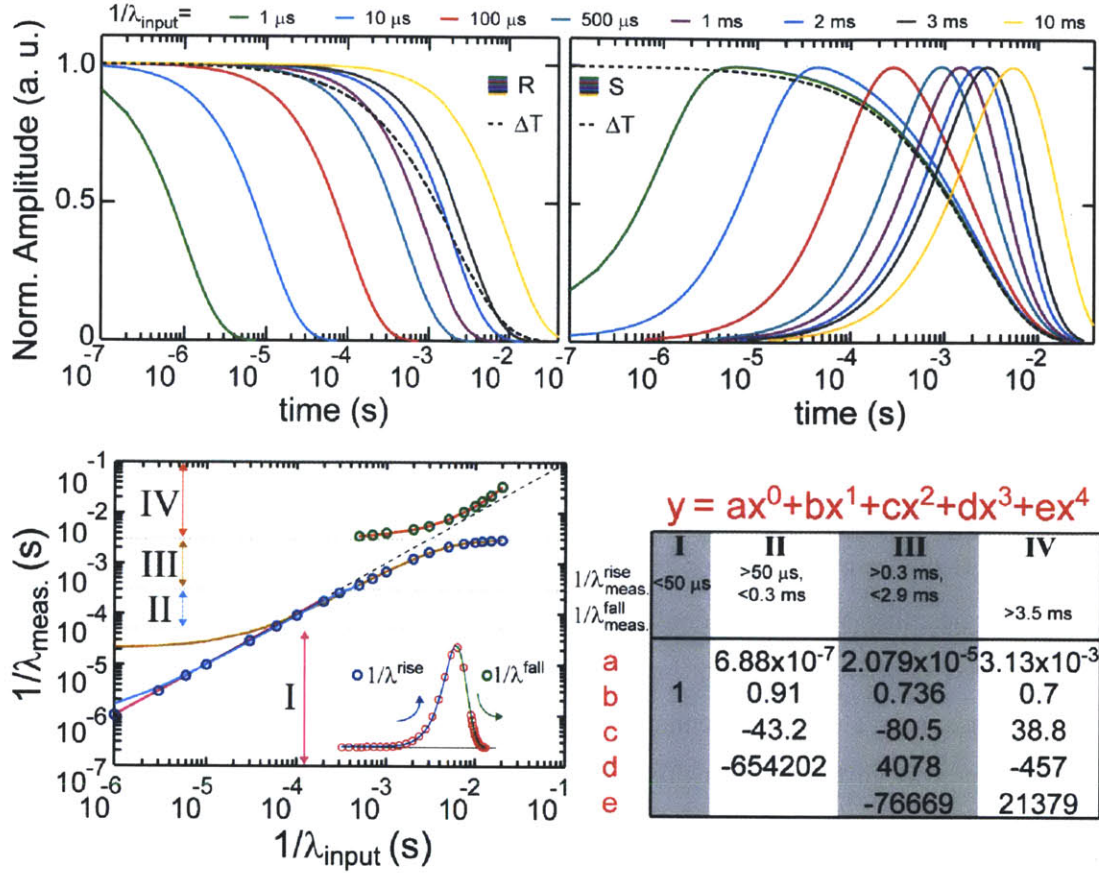


Figure 3.16: *Top left* – The protein response mono-exponentials $R = \exp(-\lambda_{\text{input}}\tau)$ and temperature trace $\Delta T = \exp[-(\tau/2.0544 \text{ ms})^{0.689766}]$ are shown. *Top right* – The observed normalized signal is a convolution of $S = [R * \Delta T]$. *Bottom left* – The observed signal, S , that results from convolution is fit to a bi-exponential with oppositely signed amplitudes. The measured rise and fall rates are plotted as a function of the λ_{input} . Three $\leq 4^{\text{th}}$ order polynomial fits are used to map the measured rise times as a function of the λ_{input} over three different time ranges. One 4^{th} order polynomial fit maps the measured fall times as a function of λ_{input} . *Bottom right* – The $1/\lambda_{\text{meas.}}$ ranges and corresponding coefficients for the mapping fits are shown. The dashed black line is along the diagonal. $x = 1/\lambda_{\text{input}}$ (s), $y = 1/\lambda_{\text{meas.}}$ (s)

the protein responds proportionally at longer times. When the protein response rate is intermediate ($50 \mu\text{s} < 1/\lambda_{\text{input}} < 3 \text{ ms}$), the measured rise rate is faster than expected due to its mixing with the stretched exponential temperature profile. The rise rate and fall rate asymptote towards a $1/\lambda_{\text{meas.}} = \sim 3 \text{ ms}$, which is presumably the point where the temperature profile drop best matches the protein response mono-exponential rate. At slower rates ($3 \text{ ms} < 1/\lambda_{\text{input}}$), the temperature profile drop is faster than the protein response – the rise is defined by the temperature profile while the fall rate is defined by the protein response.

In this convolution picture, multi-exponential protein responses are assumed to result in a convolution that is the sum of their respective, scaled mono-exponential convolutions.

Mapping measured rates to convolutions of input rates

Based on the $1/\lambda_{\text{meas.}}(1/\lambda_{\text{input}})$ plot shown at the bottom left in Figure 3.16, a crude, empirical mapping can be done to relate the measured and underlying, input rates using 4th order polynomials in different measured rate ranges. For example, for measured exponential time constants $1/\lambda_{\text{input}} < 50 \mu\text{s}$, a linear mapping exists. The mapping polynomials are plotted on the bottom left while the coefficients are shown on the bottom right.

Simple modeling of signals

Our interpretation and intuition for a real protein differs from that provided by the LTI picture. The protein system is expected to react with some kinetic response that depends on the temperature. If the temperature drops back to equilibrium much more quickly than the protein kinetic timescales, then we expect a much lower amplitude response as the protein has not had a chance to react. Based on this intuition, the observed signal amplitude should depend on the protein response rates, a dependence that is not included in the LTI.

Another way to think about this amplitude is based on the rate constant. The ratio of populations depends on the Gibb’s free energy differences between states, which are related to the ratio of forward and reverse rate constants. If the rate constants are slower than the temperature relaxation, the system will never reach equilibrium at the T-jump-accessed higher temperature, and the observed population changes will be lower than the changes predicted for an equilibrium system.

To better understand the real phenomenon, the amplitude dependence on rates, how the measured rates correspond to underlying rates, and the limitations of the simple, LTI picture, we performed temperature-dependent kinetic simulations. The simulations are described more fully in Chapter 6 Appendix 6.F. Briefly, the change of population A is monitored in the reaction:



Where the temperature dependence of the forward and reverse reaction rates is given by the Eyring equation with a simple form of the Gibb’s free energy: $\Delta G = \Delta H - T\Delta S$. The inputs into the model are the initial (low temperature), equilibrium forward and reverse rate constants at T_i and the final, high temperature forward and reverse rate constants at $T_f = T_i + \max(\Delta T)$. The

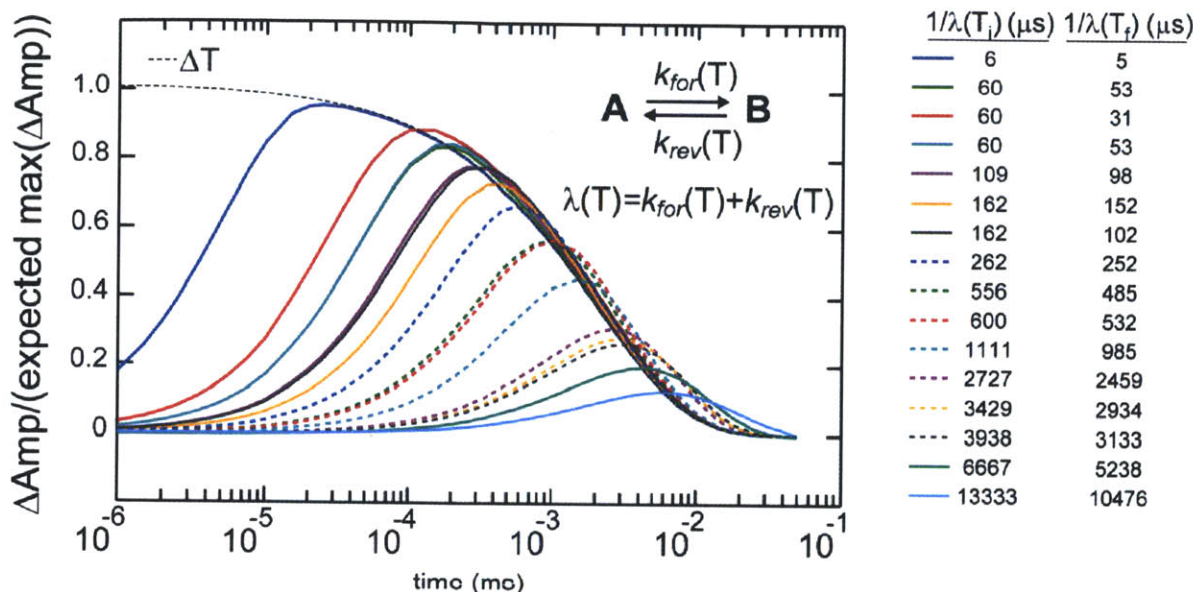


Figure 3.17: The signal changes observed based on a two-state kinetic model responding to the same temperature profile described in Figure 3.16. The initial, low temperature is $T_i=308\text{K}$. The maximum, final temperature reached by the T-jump is $T_f=320\text{K}$. The signal is scaled as $\{[A(t)]-[A(T_i)]\}/\{[A(T_f)]-[A(T_i)]\}$, where $[A(t)]$ is the simulated possibly non-equilibrated concentration of A and $[A(T_k)]$ is the equilibrated concentration of A at temperature k .

expected theoretically observed rate, λ , is the sum of the forward and reverse rates at T, $\lambda(T) = k_{for}(T) + k_{rev}(T)$.

The signals that result from simulations with different $\lambda(T_i)$ and $\lambda(T_f)$ are shown in Figure 3.17. The signal amplitude is normalized by the change in population amplitude expected for an equilibrium change from T_i to T_f . As expected, as the rates become slower, the system has less time to reach equilibrium (reach completion), and the observed amplitude decreases relative to the expected amplitude change. The amplitude's dependence is plotted versus $\lambda(T_f)$ and fit to a mapping bi-exponential in Figure 3.18.

Comparison of convolution-predicted and modeling-predicted signals

Both the convolution study of LTI proteins and the simple kinetic models predict similar signals that grow in and decay in response to the temperature profile. The simple kinetic models predict that the signal level amplitude should decrease (relative to the expected changes) with slower and slower rates.

In Figure 3.16 a map was proposed based on the convolution predictions for mapping the measured exponential time constants onto the underlying, input time constants. In order to test

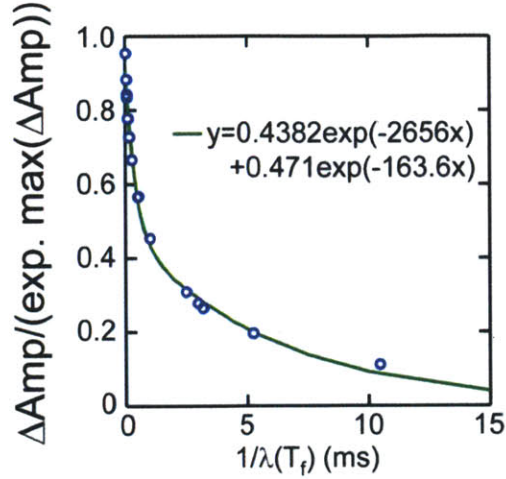


Figure 3.18: The ratio of actual to expected (equilibrium) amplitude change from Figure 3.17 is plotted versus $1/\lambda(T_f)$. The plot is fit to a mapping bi-exponential. $x = 1/\lambda(T_f)$ (s), $y = \Delta\text{Amp}/\text{expected}(\Delta\text{Amp})$.

this map, the measured rates from the simple kinetic models are plotted against that same map in Figure 3.19. For the simple kinetic model case, there are actually two rates fed into the model, $\lambda(T_i)$ and $\lambda(T_f)$, and the temperature dependent rate $\lambda(T)$ is expected to slide between them in response to the temperature profile. We plot the measured rates versus both of these kinetic model rates.

Interestingly, the mapping correlates the measured rise exponential times well with the $\lambda(T_f)$ at $1/\lambda_{\text{meas.}} < \sim 1.5$ ms. In this time range, the $\lambda(T_f)$ is faster than the temperature profile stretched exponential relaxation, which means that the changes at early times are primarily the unadulterated protein response. Because the $\lambda(T_i)$ does not affect the relaxation until the temperature begins to re-equilibrate, there is minimal correlation between the rising $1/\lambda_{\text{meas}}$ and the $\lambda(T_i)$ in this $< \sim 1.5$ ms time constant range.

At long enough rates that the signal fall is not dominated by the stretched exponential temperature profile relaxation, the falling $1/\lambda_{\text{meas.}}$ maps well onto the $1/\lambda(T_i)$ for $1/\lambda_{\text{meas.}} < \sim 8-9$ ms. The mapping is poorer for the $1/\lambda(T_f)$ because the long time re-equilibration (drop in signal) is expected to be the result of the slower rates $\lambda(T_i)$; the temperature has returned to T_i at these long times.

The rates predicted by the LTI protein convolutions and the kinetic models deviate slightly at long $1/\lambda(T_i)$ and $1/\lambda(T_f)$, which may be the result of limited temporal range that leads

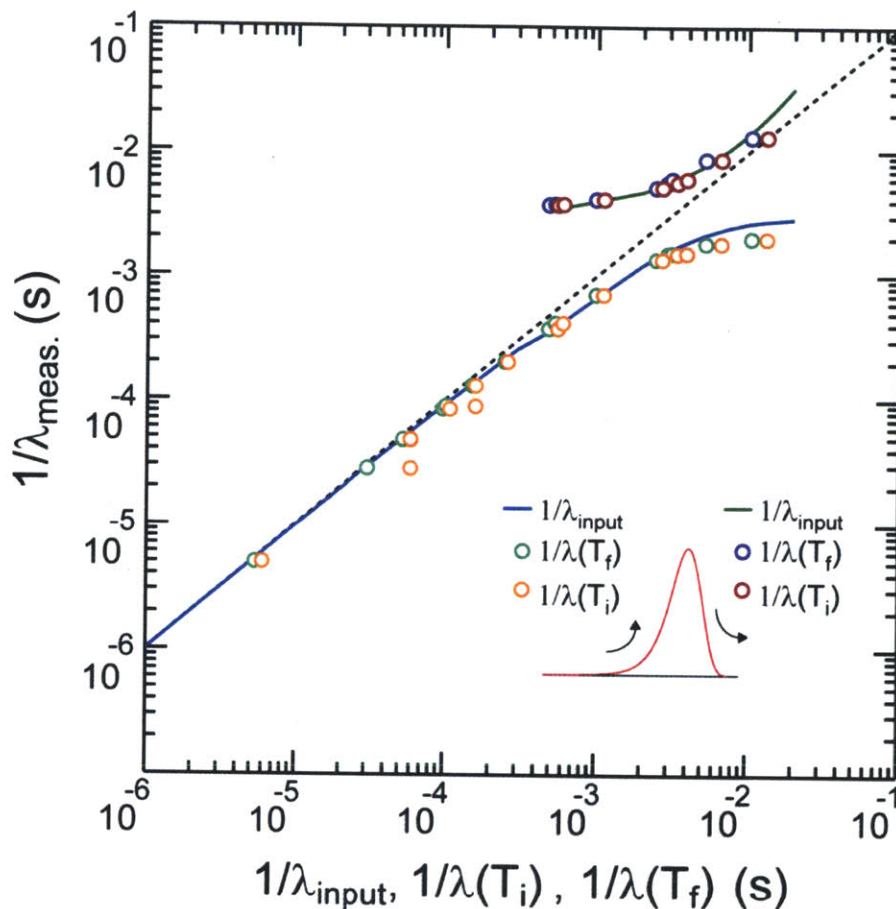


Figure 3.19: The modeled signals shown in Figure 3.17 are fit to bi-exponentials (one for the rise, one for the fall). The measured rising exponential time constants are plotted versus the $1/\lambda(T_f)$ and $1/\lambda(T_i)$ time constants (green and orange circles, respectively). The measured falling exponential time constants are plotted versus the $1/\lambda(T_f)$ and $1/\lambda(T_i)$ time constants (purple and dark red circles, respectively). The relationship between input rate and observed convolution timescale shown in Figure 3.16 are also plotted in blue and to bad fits or bad FT domain convolutions. For both simulations, the time range only extends to 40 ms.

For our kinetic modeling, increases in temperature cause population to flow to state B, and the high temperature rates were always assumed to be faster than the low temperature rates. The simulation details, which may be unrealistic, seem to correspond intuitively to the results, suggesting that conclusions we make are fairly general and not simulation specific. For example, one of the questions is how does the temperature dependence of the rates affect the measured timescales. Based on the modeling, which uses a variety of initial and final rates, the measured time constants seem to either depend on the initial or the final rate – they seem to be roughly

independent of the intermediate or average rate.

If we measure a rising $1/\lambda_{\text{meas}} < 500 \mu\text{s}$, we can map it directly to find $\lambda(T_i)$. If we measure a $500 \mu\text{s} < \text{rising } 1/\lambda_{\text{meas}} < 1.5 \text{ ms}$ and a $500 \mu\text{s} < \text{falling } 1/\lambda_{\text{meas}} < 8\text{-}9 \text{ ms}$, we can map it directly to find $\lambda(T_i)$ and $\lambda(T_f)$. At falling $1/\lambda_{\text{meas}} > 8\text{-}9 \text{ ms}$, we expect that the $1/\lambda_{\text{meas}} \cong \lambda(T_f)$, as predicted by our intuition and the kinetic modeling.

For the data shown here, there is no included noise, which means that the drop in signal amplitude at slower rates does not affect the quality of the exponential fits. In a real protein system, the slower the rates, the less amplitude is expected, and the smaller the signal-to-noise levels will be.

For a multi-exponential protein response we expect that the signal will be the sum of the signals of the individual multi-exponentials. With this modeling we assume exponential responses and a temperature profile given by the equation in Figure 3.16. Changes in either require further modeling. Perhaps the mapping routine can predict a starting timescale which can be fine-tuned through an iterative procedure of convolving it with the temperature profile can comparing the convolution with the experimentally observed signal.

Based on the similar mapping observed for the convolution-predicted and modeling-predicted signals, we conclude that the LTI convolution provides an adequate prediction of the expected time behavior associated with different exponential protein responses. Said in reverse, we predict that the measured signal can be compared to predicted convolutions to back out the protein response. The LTI fails, however, to predict the amplitude dependence on the rate, and it only considers one rate, whereas in reality we should consider the $\lambda(T_i)$ and $\lambda(T_f)$ rates.

Deconvolution

As was shown above, convolving the protein response with the temperature trace predicts a signal trace similar to what is predicted from kinetic modeling, which suggests that the convolution is a fairly good description of the system. Rather than relying on the mapping described above, we can also attempt to deconvolve the observed signal with the temperature profile to directly calculate the protein response:

$$R(\tau) = \text{FT}^{-1} \left\{ \frac{\text{FT}[S(\tau)]}{\text{FT}[\Delta T(\tau)]} \right\}. \quad (3.9)$$

The results of applying this to the signals simulated from the kinetic models (Figure 3.17) is shown in Figure 3.20. Inspection shows that the deconvolution accurately recovers the mono-

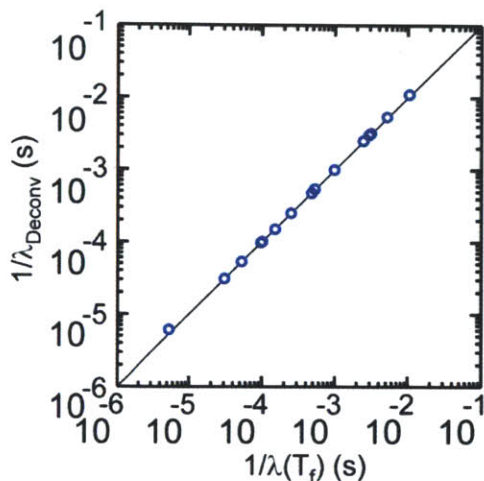


Figure 3.20: Deconvolution of the signals predicted by the kinetic models (Figure 3.17) leads to mono-exponentials that are well fit with a time constant $1/\lambda_{\text{Deconv}}$. The recovered time constants compare well to the $1/\lambda(T_f)$ input into the initial model.

exponential protein response.

Although deconvolution is a valid method of determining the timescales, we expect it to be less successful with noisy data. The deconvolution will also not accurately predict the amplitude, although further study will likely provide the proper corrections. Also, the FT algorithm can be computationally expensive, as the time domain signal usually spans many decades in time and the time spacing must be uniform.

3.5 Conclusion

The optical layout specific to the collection of the equilibrium and transient 2D IR spectra presented in this thesis is presented above. Further data processing details are presented for HDVE in Chapter 4 and for transient T-jump experiments in Chapter 5. The expected spectral and temporal features are also described.

3.6 References

- (1) H. S. Chung, M. Khalil, A. W. Smith, and A. Tokmakoff, "Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics," *Rev. Sci. Instrum.* **78**(2007).
- (2) C. S. Peng, K. C. Jones, and A. Tokmakoff, "Anharmonic Vibrational Modes of Nucleic Acid Bases Revealed by 2D IR Spectroscopy," *J. Am. Chem. Soc.* **133**, 15650-15660 (2011).
- (3) R. A. Nicodemus, "Hydrogen Bond Reorganization and Vibrational Relaxation in Water Studied with Ultrafast Infrared Spectroscopy," (Massachusetts Institute of Technology, Cambridge, 2011).
- (4) N. Demirdöven, M. Khalil, O. Golonzka, and A. Tokmakoff, "Dispersion compensation with optical materials for compression of intense sub-100-fs mid infrared pulses," *Opt. Lett.* **27**, 433-435 (2002).
- (5) E. H. G. Backus, S. Garrett-Roe, and P. Hamm, "Phasing problem of heterodyne-detected two-dimensional infrared spectroscopy," *Opt. Lett.* **33**, 2665-2667 (2008).
- (6) M. Khalil, N. Demirdöven, and A. Tokmakoff, "Obtaining Absorptive Line Shapes in Two-Dimensional Infrared Vibrational Correlation Spectra," *Phys. Rev. Lett.* **90**, 47401-47404 (2003).
- (7) H. S. Chung, Z. Ganim, K. C. Jones, and A. Tokmakoff, "Transient 2D IR spectroscopy of ubiquitin unfolding dynamics," *Proc. Natl. Acad. Sci. USA*, 14237-14242 (2007).

Chapter 4

Heterodyne-detected dispersed vibrational echo spectroscopy

The work presented in this chapter has been published and is reprinted with permission from:

- K. C. Jones, Z. Ganim, and A. Tokmakoff, "Heterodyne-Detected Dispersed Vibrational Echo Spectroscopy," *J. Phys. Chem. A* **113**, 14060-14066 (2009)

Copyright 2009 American Chemical Society

4.1 Abstract

We develop heterodyned dispersed vibrational echo spectroscopy (HDVE) and demonstrate the new capabilities in biophysical applications. HDVE is a robust ultrafast technique that provides a characterization of the real and imaginary components of third-order nonlinear signals with high sensitivity and single-laser-shot capability, and can be used to extract dispersed pump-probe and dispersed vibrational echo spectra. Four methods for acquiring HDVE phase and amplitude spectra were compared: Fourier transform spectral interferometry, a new phase modulation spectral interferometry technique, and combination schemes. These extraction techniques were demonstrated in the context of protein amide I spectroscopy. Experimental HDVE and heterodyned free induction decay amide I spectra were explicitly compared to conventional dispersed pump-probe, dispersed vibrational echo, and absorption

spectra. The new capabilities of HDVE were demonstrated by acquiring single-shot spectra and melting curves of ubiquitin and concentration-dependent spectra of insulin suitable for extracting the binding constant for dimerization. The introduced techniques will prove particularly useful in transient experiments, studying irreversible reactions, and μM concentration studies of small proteins.

4.2 Introduction

Due to their structural sensitivity and ultrafast time resolution, femtosecond nonlinear infrared (IR) spectroscopies are increasingly being used in studies of chemical dynamics in solution and biophysical assays of proteins^{1, 2}, peptides³, and nucleic acids⁴ to reveal structure and time dependent changes. Third-order nonlinear vibrational spectroscopy is particularly powerful because the use of multiple pulses maps out vibrational couplings, measures anharmonicities, and can separate homogeneous and inhomogeneous contributions to absorption lineshapes. Each of these features ultimately increases structural sensitivity and aids in the ability to resolve overlapping peaks⁵. Two-dimensional infrared (2D IR) spectroscopy is the most complete characterization of the third-order vibrational response, but methods that retain nonlinear information and can be acquired and processed quickly or which can be used in single-laser-shot mode are in demand for probing transient and/or irreversible chemical and biophysical dynamics. Two frequency-dispersed probing techniques, dispersed pump-probe (DPP) and dispersed vibrational echo (DVE) meet these requirements yet they have certain disadvantages.⁶ DPP spectra scale linearly with population and are related to changes in sample absorption, but signal levels are small and phase extraction requires certain assumptions. DVE spectra are background-free, but, as amplitude-squared measurements, have signal cross terms that make proportionality to concentration or structure ambiguous. Also, since DVE scales as electric field squared, its intensity scales unfavorably for samples in dilute solutions or with weak transitions.

In this work, we report on the development and characterization of methods that provide linear and phase-sensitive detection of the third-order nonlinear response. Heterodyne detection⁷ has been an invaluable method for amplifying and characterizing the weak signal electric field in third-order measurements,⁸ but 2D IR relies on a Fourier transform (FT) of a set of time domain data to separate phase and amplitude information. Applying a number of alternate phase-sensitive detection protocols, we have developed heterodyned dispersed vibrational echo (HDVE) spectroscopy, a single-dimension frequency domain measurement of the signal

generated by three time-coincident femtosecond infrared pulses. HDVE is related to the projection of the 2D IR spectrum onto the detection axis, and therefore, like DPP and DVE, carries the cross-peak information of the 2D IR spectrum, but may be collected ~250-500 times faster than a 2D IR dataset. In addition, the HDVE measurement can be used to reconstruct both DVE and DPP spectra. Additionally, we show that the complex linear response can be extracted from an analogous broadband heterodyne-detected measurement of the free induction decay (FID), similar to methods that have been used in the optical and terahertz regimes.^{9, 10} The new capabilities are demonstrated using protein amide I HDVE spectroscopy to measure the melting curve of ubiquitin, the dissociation constant of insulin, and to extract the DPP and DVE spectra of ubiquitin from a single laser shot.

4.3 Signal Detection Methods

4.3.1 Heterodyne Detection

In both linear and nonlinear ultrafast coherent spectroscopy, the signal electric field is radiated as a result of the polarization induced in the sample by incident laser pulses. This real, time-domain field, $E'(t)$, can be expressed through a FT as a complex spectral field, $\tilde{E}(\omega)$, with amplitude, e , and phase, ϕ :

$$\tilde{E}(\omega) = e(\omega) \exp(i\phi(\omega)) \quad (4.1)$$

Heterodyne detection of this field is accomplished by collecting an interferogram after mixing with a local oscillator field (LO), \tilde{E}_{LO} , separated by a time delay τ_{LO} . For spectral interferometry, the measured intensity in the frequency domain $I(\omega; \tau_{LO})$ is

$$\begin{aligned} I(\omega; \tau_{LO}) &= \left| \tilde{E}_{sig}(\omega) + \tilde{E}_{LO}(\omega) \exp(-i\omega\tau_{LO}) \right|^2 \\ &= e_{LO}^2(\omega) + e_{sig}^2(\omega) + 2e_{sig}(\omega)e_{LO}(\omega) \cos[\Delta\phi(\omega) + \tau_{LO}\omega] \end{aligned} \quad (4.2)$$

Through phase or amplitude modulation, the spectral interferogram can be obtained by subtracting the square terms to leave the cross term,

$$S'(\omega; \tau_{LO}) = 2e_{sig}(\omega)e_{LO}(\omega) \cos[\Delta\phi(\omega) + \tau_{LO}\omega] \quad (4.3)$$

The cross-term amplifies the signal intensity and creates linear proportionality between the measured intensity change and electric field amplitude. The cross-term is also sensitive to the phase difference between the signal and LO

$$\Delta\phi(\omega) = \phi_{sig}(\omega) - \phi_{LO}(\omega) \quad (4.4)$$

In order to extract the phase difference, the quadrature spectral interferogram component, S'' , which may be calculated through Kramers-Kronig relations or phase modulation, is used:

$$S''(\omega; \tau_{LO}) = 2e_{sig}(\omega)e_{LO}(\omega)\sin[\Delta\phi(\omega) + \tau_{LO}\omega]. \quad (4.5)$$

In the text, S' denotes the in-phase, real, or cosine component of the spectral interferogram while S'' denotes the in-quadrature, imaginary, or sine component. We also define a complex spectral interferogram $\tilde{S} \equiv S' + iS''$.

Heterodyned Ultrafast Free Induction Decay

The complex linear spectrum may be measured through heterodyne detection of an ultrafast broadband pulse transmitted through the sample. In this case, the transmitted electric field is

$$\tilde{E}_{sig}^{(1)}(\omega) = \tilde{E}_0(\omega)\exp[i\tilde{n}(\omega)\omega l / c] \quad (4.6)$$

where $\tilde{E}_0(\omega)$ is the beam that enters the sample, l is the pathlength, c is the speed of light, and $\tilde{n}(\omega)$ is the complex refractive index, $\tilde{n}(\omega) = n(\omega) + i\kappa(\omega)$, that has dispersive, $n(\omega)$, and absorptive, $\kappa(\omega)$, components⁹. From eq. (4.3), the measured heterodyned free induction decay (HFID) can be expressed as:

$$S'_{HFID}(\omega; \tau_{LO}) = 2e_0(\omega)e_{LO}(\omega)\exp[-\kappa(\omega)\omega l / c]\cos[\Delta\phi_{0,LO}(\omega) + n(\omega)\omega l / c + \tau_{LO}\omega]. \quad (4.7)$$

Here $\Delta\phi_{0,LO}$ is the phase difference between the incident excitation field and local oscillator. For comparison, the intensity of the transmitted infrared beam, which is used in conventional absorption spectroscopy, is given by

$$I_{tr}(\omega) = \left| \tilde{E}_0(\omega)\exp[i\tilde{n}(\omega)\omega l / c] \right|^2 = e_0(\omega)^2 \exp[-2\kappa(\omega)\omega l / c] \quad (4.8)$$

Conventional absorption measurements only measure the change in the absorptive component of a sample relative to a reference, $\Delta\kappa(\omega) = \kappa_s(\omega) - \kappa_{ref}(\omega)$, whereas the HFID can measure both $\Delta\kappa(\omega)$ and the real index change $\Delta n(\omega) = n_s(\omega) - n_{ref}(\omega)$. For either conventional absorption (eq. (4.8)) or HFID spectral interferogram (eq. (4.7)), the $\Delta\kappa(\omega)$ difference is calculated from the intensity

$$\Delta\kappa(\omega) = -\frac{\ln(10)c}{2\omega l} \log \frac{I_{IR}^s(\omega)}{I_{IR}^{ref}(\omega)} = -\frac{\ln(10)c}{2\omega l} \log \frac{|\tilde{S}_{HFID}^s(\omega; \tau_{LO})|^2}{|\tilde{S}_{HFID}^{ref}(\omega; \tau_{LO})|^2}. \quad (4.9)$$

$\Delta n(\omega)$ can be extracted by subtracting the sample and reference phases from eq. (4.7); the remaining $\tau_{LO}\omega$ and $\Delta\phi_{0,LO}(\omega)$ terms cancel without needing to know their absolute values⁹.

Heterodyned Ultrafast Nonlinear Spectroscopy

In this work, nonlinear signal characterization is described for three third-order spectroscopies: HDVE, DPP and DVE. Each of these intensities is obtained from a sequence of three resonant excitation pulses, separated by time intervals τ_1 and τ_2 . The signal is radiated during τ_3 . DVE spectroscopy is a homodyne technique that measures the signal field squared. DPP is a self-heterodyned technique whose intrinsic LO is always time-coincident with the signal, but may be affected by transmission through the sample. The three third-order spectra can be described using signal and LO electric field amplitude and phases as

$$S'_{HDVE}(\tau_1 = 0, \tau_2 = 0, \omega_3; \tau_{LO}) = 2e_{sig}(\omega_3)e_{LO}(\omega_3)\cos[\Delta\phi(\omega_3) + \tau_{LO}\omega_3] \quad (4.10)$$

$$I_{DVE}(\tau_1 = 0, \tau_2 = 0, \omega_3) = e_{sig}(\omega_3)^2 \quad (4.11)$$

$$S'_{DPP}(\tau_1 = 0, \tau_2 = 0, \omega_3; \tau_{LO} = 0) = 2e_{sig}(\omega_3)e_{LO}(\omega_3)\cos[\Delta\phi(\omega_3)], \quad (4.12)$$

where the cross-terms have been isolated for HDVE and DPP. As is shown in eq. (4.3), heterodyne detection produces a cosine term that contains phase and amplitude information. Interfering with an LO provides an amplified signal that scales linearly with e_{LO} , e_{sig} , and the chromophore concentration. Additionally, DVE and DPP spectra can be extracted from the HDVE spectrum; once the amplitude $e_{LO}e_{sig}$, phase $\Delta\phi(\omega)$, and LO delay $\tau_{LO}\omega$ have been extracted, the DVE spectrum is recreated from $(e_{LO}e_{sig})^2$ and the DPP spectrum is recreated from $e_{sig}e_{LO}\cos(\Delta\phi(\omega))$.

4.3.2 Separation of amplitude and phase

The most challenging part of full signal characterization is the separation of the amplitude and phase components of the measured intensity cross term. We use two distinct methods for extracting these components from the measured HDVE and HFID: Fourier transform spectral interferometry (FTSI)^{9, 11-14} and phase modulation spectral interferometry (PMSI). Each method extracts the signal field by acquisition of a signal interferogram after

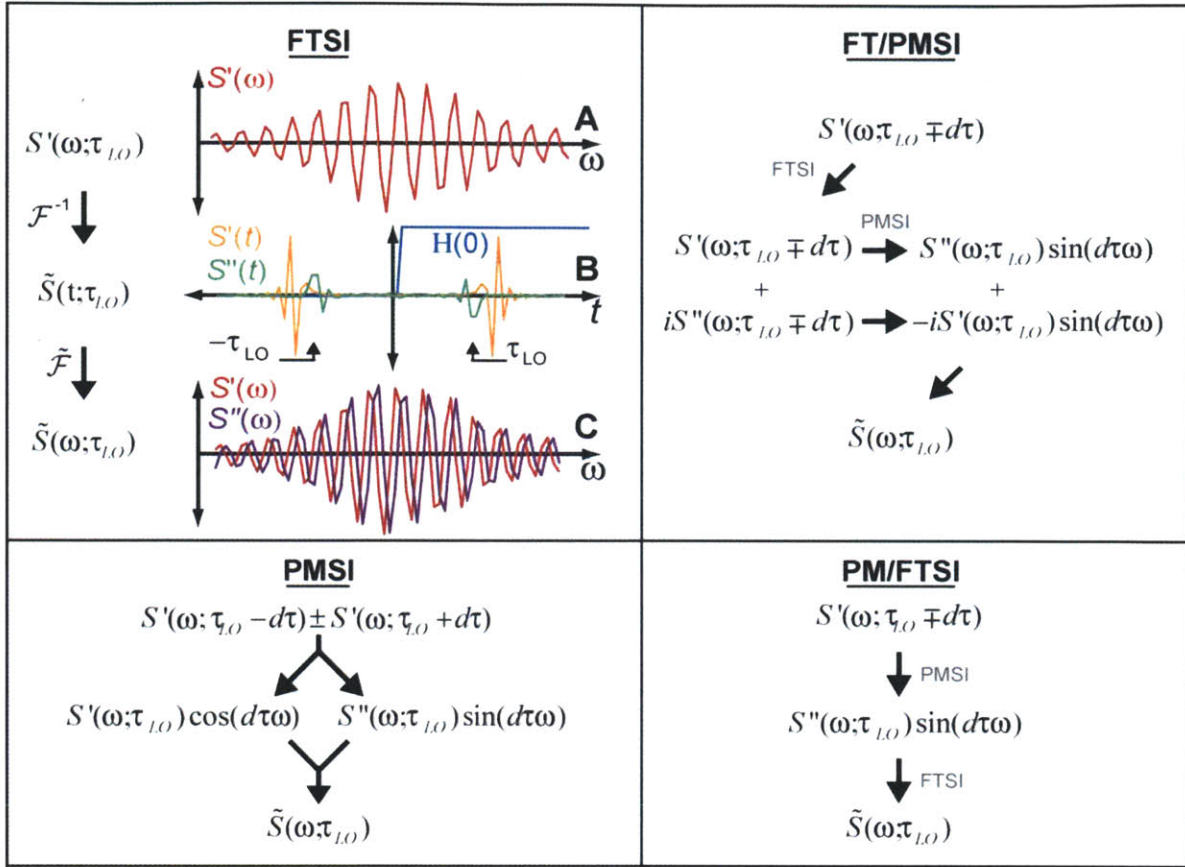


Figure 4.1: The four methods for calculating the complex heterodyned intensity, \tilde{S} , from the measured, real projection, S' , are described with equations and diagrams. **FTSI:** The real, asymmetric interferogram (A) is inversely Fourier Transformed into the time domain (B) to give a complex time domain signal in which $\mathcal{F}^{-1}[\tilde{E}_{sig} \tilde{E}_{LO}^* e^{i\tau_{LO}\omega}]$ and $\mathcal{F}^{-1}[\tilde{E}_{sig}^* \tilde{E}_{LO} e^{-i\tau_{LO}\omega}]$ components are separated onto the positive and negative time axes. After applying a Heaviside window function to isolate one of these components, Fourier Transformation back into the frequency domain gives the complex intensity (C). Actual, 64-pixel *N*-methylacetamide data is shown to the right at each stage of the calculation (A-C) **PMSI:** The difference and sum of the time stepped measured interferograms are calculated to give the real and imaginary components multiplied by $\cos(d\tau\omega)$ and $\sin(d\tau\omega)$, respectively. Division by these known sinusoidal functions allows for calculation of \tilde{S} . **FT/PMSI:** FTFSI is used to calculate $\tilde{S}(\omega; \tau_{LO} - d\tau)$ and $\tilde{S}(\omega; \tau_{LO} + d\tau)$. Real components are subtracted to give $S''(\omega; \tau_{LO}) \sin(d\tau\omega)$. Imaginary components are subtracted to give $S'(\omega; \tau_{LO}) \sin(d\tau\omega)$. $\tilde{S}(\omega; \tau_{LO})$ is then calculated. **PM/FTSI:** PMSI is used to calculate the $S''(\omega; \tau_{LO}) \sin(d\tau\omega)$ from the measured $d\tau$ shifted interferograms. After division by $\sin(d\tau\omega)$, FTFSI allows for the calculation of \tilde{S} .

introducing a known, frequency-dependent phase shift, $\tau_{LO}\omega$. In FTSI, the spectral interferogram is Fourier transformed back into the time domain, where Kramers-Kronig relations allow for calculation of the imaginary component, S'' . In PMSI, the local oscillator time delay, and therefore phase, is sinusoidally cycled to extract the imaginary component. Both of these techniques (illustrated in Figure 4.1) are advantageous because of their simplicity. Each only relies on moving one delay and acquiring the spectral interferogram with an array detector. Other techniques, such as dual-quadrature spectral interferometry and phase-cycling require frequency independent phase control, which is achieved with waveplates¹² or by pulse shaping¹⁵, respectively.

Fourier Transform Spectral Interferometry

In FTSI, we use a Kramers-Kronig relationship to relate real and imaginary components of the spectral interferogram. For discussion, we define the double-sided FT, $\tilde{\mathcal{F}}$,

$$\tilde{\mathcal{F}}[Z(t)] = \int_{-\infty}^{\infty} dt e^{i\omega t} Z(t) \equiv Z(\omega) \quad (4.13)$$

and its conjugate single-sided inverse FT, \mathcal{F}^{-1} ,

$$\mathcal{F}^{-1}[Z(\omega)] = \frac{1}{\pi} \int_0^{\infty} d\omega e^{-i\omega t} Z(\omega) \equiv Z(t) \quad (4.14)$$

where Z may be real or complex. The Kramers-Kronig relationship essentially states the equivalence of the single and double-sided Fourier transforms. Therefore, starting with a real spectral interferogram $S'(\omega; \tau_{LO})$, which is windowed by our detection range to a finite, positive spectrum that spans only the transition of interest, $\mathcal{F}^{-1}[S'(\omega; \tau_{LO})]$ gives a complex time-domain signal $\tilde{S}(t)$ with the property $\tilde{S}(t) = \tilde{S}(-t)^*$. $\tilde{S}(t)$ contains negative and positive time delay components that are symmetrically shifted by τ_{LO} as a result of $\mathcal{F}^{-1}[Z(\omega)\exp[\pm i\tau_{LO}\omega]] = \tilde{Z}(t \pm \tau_{LO})$. By windowing off one component, the symmetry is broken and the full complex spectrum \tilde{S} is recovered after FT back to the frequency domain:

$$\tilde{\mathcal{F}}[W(t)\mathcal{F}^{-1}[S'(\omega; \tau_{LO})]] = \tilde{S}(\omega; \tau_{LO}) = \tilde{E}_{sig}(\omega)\tilde{E}_{LO}^*(\omega)\exp[i\tau_{LO}\omega]. \quad (4.15)$$

$W(t)$ is the windowing function. In our case, this is chosen to be the Heaviside step function, $H(0)$, which removes both DC noise and the negative time signal.

Phase Modulation Spectral Interferometry

The signal amplitude and phase may also be extracted through phase modulation spectral interferometry (PMSI). In this technique, a small time step, $2 d\tau$, modulates τ_{LO} to create a frequency-dependent phase shift that allows for the separation of the real and imaginary parts of the heterodyned intensity based on trigonometric relationships:

$$S'(\omega; \tau_{LO}) = \frac{S'(\omega; \tau_{LO} - d\tau) + S'(\omega; \tau_{LO} + d\tau)}{2 \cos(d\tau\omega)} \quad (4.16)$$

$$S''(\omega; \tau_{LO}) = \frac{S'(\omega; \tau_{LO} - d\tau) - S'(\omega; \tau_{LO} + d\tau)}{2 \sin(d\tau\omega)} \quad (4.17)$$

After taking the sum and difference, the $\sin(d\tau\omega)$ and $\cos(d\tau\omega)$ are divided out to leave the desired complex HDVE or HFID intensities. Noise between $S'(\omega; \tau_{LO} + d\tau)$ and $S'(\omega; \tau_{LO} - d\tau)$ and inaccuracies in $d\tau$ timing may result in high frequency oscillations in the extracted amplitude and phase.

4.4 Experimental Section

The heterodyne techniques were tested with the acquisition of a series of small protein (ribonuclease A, concanavalin A, ubiquitin, insulin) and peptide (*N*-methylacetamide, diglycine) spectra in the amide I region using 1 kHz repetition rate, nearly transform-limited IR pulses centered at $\sim 1650 \text{ cm}^{-1}$ with FWHM bandwidths of 160 cm^{-1} produced as previously described⁵. Signals were detected with ~ 2 or 4.7 cm^{-1} resolution on a 2x64 element mercury cadmium telluride array (IR-0144, Infrared Systems Development) after diffraction off of a 90 or 40 lines/mm grating mounted in a 190 mm monochromator (Triax 190, Jobin Yvon).

The DVE signals were generated by three time-coincident pulses in a boxcar geometry. Unless noted HDVE spectra were acquired by heterodyne detecting the DVE signal by overlapping with an external LO. The DPP was collected with two time-coincident pulses in a noncollinear geometry. The DPP signal is intrinsically self-heterodyned without an external LO. The HFID signal was generated by a single pulse transmitted through the sample and overlapped with an external LO. Pulses were delayed using retro-reflectors mounted on translation stages (ANT-50L, Aerotech) with accuracies of $\pm 0.7 \text{ fs}$. The HFID and HDVE spectra were collected at $\tau_{LO} = -4.000 \text{ ps}$, which provides FTSI time-domain separation for our detection resolution, and at PMSI $d\tau = 2 \text{ fs}$, which introduces $\sim \pi/2$ phase shift. The DVE spectrum is background free, but scatter was removed by subtracting the background generated by chopping consecutive shots of an excitation beam. The $\tau_{LO} = 0$ was initially set at the time where the HDVE matched the

independently measured 2-beam DPP.

The spectral interferogram, S' (eq 2.3), was isolated from the measured intensity, I (eq 2.2), after subtracting the squared terms. For HFID, HDVE, and DPP, the e_{LO}^2 homodyne term (eq 2.2) was measured by chopping an excitation beam to remove the signal in consecutive shots. For HDVE and DPP, the e_{sig}^2 term was ignored because $e_{sig}^2 \ll S'$. For HFID, e_{sig}^2 was subtracted through balanced detection.^{8, 16} For the single-shot HDVE measurements, the e_{LO}^2 and e_{sig}^2 terms were subtracted using a reference spectrum measured in advance.

The in-phase and in-quadrature components of the heterodyned intensity were used to extract the phase and amplitude. The amplitude was calculated from the modulus squared of the complex spectrum. The phase was calculated using

$$\Delta\phi(\omega) + \tau_{LO}\omega = \text{atan}(S''(\omega) / S'(\omega)). \quad (4.18)$$

Due to the cosine nature of the measured heterodyned intensity, only $|\text{mod}(\text{phase}, 2\pi)|$ is measured and an unwrapping procedure was applied to obtain the relative offset. The τ_{LO} was corrected in data processing by up to 4 periods (80 fs) to maximize agreement between the 2-beam DPP and the DPP extracted from the HDVE. Relative times were determined within 0.7 fs (200 nm, $\lambda/30$) from the stage encoder. Slight differences between the directly collected DVE and DPP and their HDVE-extracted counterparts can be attributed to phase and amplitude differences between the external LO and the self-heterodyning beam, which passes through the sample. The LO shaping effects can be corrected through characterization of the HFID phase,¹⁴ but we do not because the unchirped external LO allows for a more faithful measurement of the signal phase.

4.5 Results and Discussion

To validate the FTSI and PMSI methods, we measured complex HDVE and HFID spectra and compared the extracted results with experimental DVE, DPP, IR, and FTIR spectra. The results are shown in Figure 4.2. Agreement between the conventionally collected and heterodyne-extracted spectra indicated that the DVE and linear IR spectra were readily obtained and faithfully reproduced from the HDVE and HFID spectra, respectively. The DPP may also be recreated with knowledge of τ_{LO} . Although determination of τ_{LO} was based on matching the HDVE spectral interferogram to the 2-beam DPP, other promising techniques, such as interferometric autocorrelation between the LO and a pinhole-scattered beam at the sample

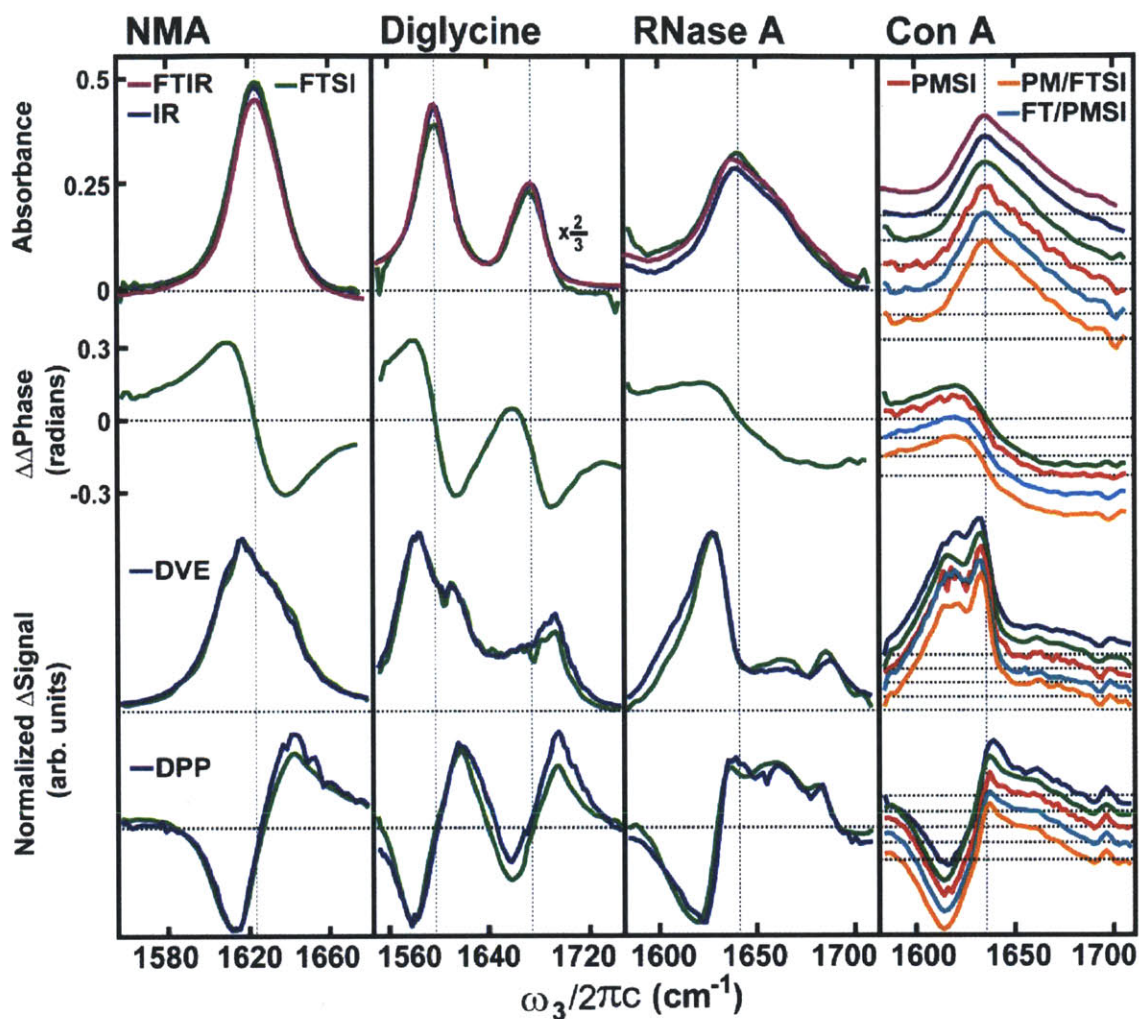


Figure 4.2: The linear absorbance, linear phase, DVE, and DPP spectra (rows, respectively) are plotted for *N*-methylacetamide, diglycine, ribonuclease A, and concanavalin A samples (columns, respectively). FTIR (purple, row 1), IR (blue, row 1), DVE (blue, row 3), and DPP (blue, row 4) spectra are compared to their FTSI (green), PMSI (red, column 4), FT/PMSI (teal, column 4), and PM/FTSI (orange, column 4) heterodyne extracted counterparts. For the $\Delta\Delta$ phase plots (row 2), the absolute zero is chosen arbitrarily. The DVE and DPP spectra were each averaged for 5 minutes. The HFID protein and solvent measurements were each averaged for 15-20 minutes.

position^{17, 18}, may be used to determine τ_{LO} with <1 fs accuracy without the 2-beam DPP spectrum. Slight inconsistencies between the spectra are attributed to slight phase and amplitude differences between the internal and external LO.

For FTSI, the upper and lower bounds of the $|\tau_{LO}|$ timing are limited by the ability to properly separate and sample both wings of the Fourier transformed interferogram. Large values of $|\tau_{LO}|$ induce rapid spectral oscillations that cannot be sampled. Thus, the upper bound is

dictated by the Nyquist period, $(2\Delta\omega c)^{-1}$, where $\Delta\omega$ is the frequency difference between consecutive pixels, and c is the speed of light. The lower bound on $|\tau_{LO}|$ is set by the pulse widths and the range of sampled frequencies; at small, negative LO timings, the positive and negative time components of $\tilde{S}(t; \tau_{LO})$ mix, preventing correct recovery of the complex frequency spectra. At $\tau_{LO} = 0$, finite pulse effects and DC contributions can also lead to mixing of the conjugate signals and distortions that affect the recovered frequency spectra. Therefore, a τ_{LO} of at least a few pulsewidths is ideal.

For PMSI, the maximum $|\tau_{LO}|$ is limited to the Nyquist period by undersampling, but there are no restrictions on the minimum. The time step, $2 d\tau$, should be chosen so that the center frequency is shifted by $\sim \pi/2 + n\pi$, where n is an integer, and the phase shifts across the spectrum are within $\pm\pi/4$ of the central frequency shift to avoid zero crossings in $\sin(d\tau\omega)$ and $\cos(d\tau\omega)$. Oscillations in our extracted PMSI spectra indicate noise and errors in timing. To correct for inaccuracies in measured time delays, the HFID timings $d\tau$ were adjusted in data processing by up to 0.5 fs to remove oscillations from the PMSI extracted absorbance and phase spectra.

4.5.1 Extension to Combination Methods

Combinations of the FTSI and PMSI methods described above can be used to eliminate noise and subtract away homodyne components. This, in some cases, precludes the need for a chopping scheme. FT/PMSI and PM/FTSI use FTSI followed by PMSI, or vice versa, to calculate the complex spectrum from $S'(\omega; \tau_{LO} + d\tau)$ and $S'(\omega; \tau_{LO} - d\tau)$. Both methods are shown schematically in Figure 4.1. The combination methods are advantageous because they rely on subtraction, which allows for removal of noise and homodyne components. When combined with balanced detection, which measures the 0 and π phase components of the signal, the FT/PMSI calculates and appropriately subtracts the $\pi/2$ and $3\pi/2$ components in a method analogous to NMR quadrature detection with a similar resulting reduction in noise. Through subtraction, the PM/FTSI method removes the e_{LO}^2 and e_{sig}^2 terms from the measured intensity, which allows for the extraction of the amplitude and phase without chopping and without balanced detection. The FT/PMSI and PM/FTSI methods are used to extract the DPP, DVE, and IR spectra of concanavalin A as shown in Figure 4.2.

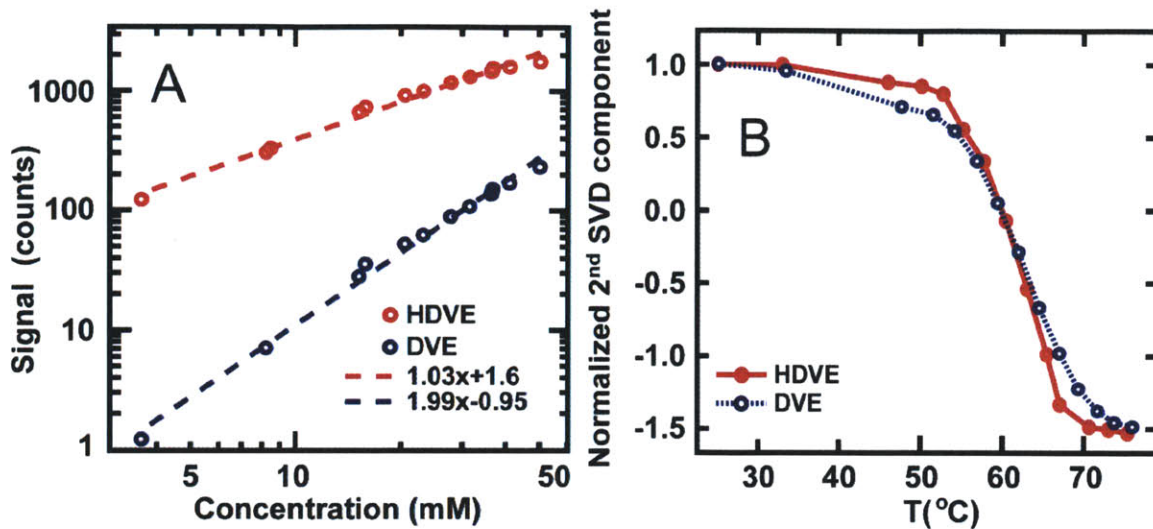


Figure 4.3: **A:** The *N*-methylacetamide 1619 cm^{-1} HDVE(red) and DVE(blue) peak intensities are plotted versus concentration. The log/log plot lines of best fit have slopes of 1.03 ± 0.03 (HDVE) and 1.99 ± 0.05 (DVE). Each point is the result of 5 minutes of averaging. **B:** The normalized 2nd SVD components derived from temperature-dependent HDVE(red) and DVE(blue) spectra of ubiquitin are plotted versus temperature. Each point is the result of 10 minutes of averaging.

As our results show, the FTSI reproduces the spectra with the greatest fidelity. FTSI can be done with a single laser shot, but applying eq. (4.15) to truncated spectra adds noise and ripples. PMSI does not rely on Kramers-Kronig relations, and truncation does not affect the characterization, but inaccuracies in timing and noise between time-stepped spectra leads to ripples in the recovered phase and amplitude. Improvements in the accuracy of time-stepping beyond the $\lambda/30c$ resolution here, in addition to shot-to-shot τ_{LO} modulation, can be implemented with piezoelectric actuators. Given ideal collection schemes with improved τ_{LO} stepping and frequency detection that spans the transition of interest, the FT/PMSI and PM/FTSI extracted spectra should best match the conventional spectra and have the least noise.

A further characterization of the detection linearity is shown in Figure 4.3. The linear concentration dependence of the HDVE amplitude and square concentration dependence of the DVE amplitude are verified by plotting the measured DVE intensity and FTSI-extracted $e_{\text{sig}}e_{LO}$ intensity at varying concentrations of NMA, as shown as a log-log plot in Figure 4.3A. The linear fits indicate the expected linear and squared dependencies for the HDVE and DVE intensities. For our laser stability and 15 bit detection depth, heterodyne detection of the DVE

increases the measured intensity by a factor of 5, which not only allows for simultaneous extraction of DPP and DVE spectra of μM protein samples, but also indicates that equivalent signal to noise measurements can be made in less than twice the time of current DVE and DPP collection techniques.

As an example of an HDVE application, Figure 4.3B shows a comparison of the melting curve for thermal unfolding of ubiquitin extracted using singular value decomposition analysis on a set of DVE and HDVE spectra acquired at temperatures spanning the unfolding temperature. This measurement is desirable for obtaining a measure of the folded population at different temperatures. The importance of the linear detection is illustrated by the sharper transition observed by HDVE. This can be rationalized considering that the DVE is proportional to the amplitude squared, while the HDVE is linearly proportional to the amplitude.

The sensitivity of HDVE for protein applications is illustrated through a characterization of the equilibrium constant for the monomer-to-dimer transition of insulin (Figure 4.4). Formation of the dimer of the largely helical protein leads to the folding of an intermolecular β

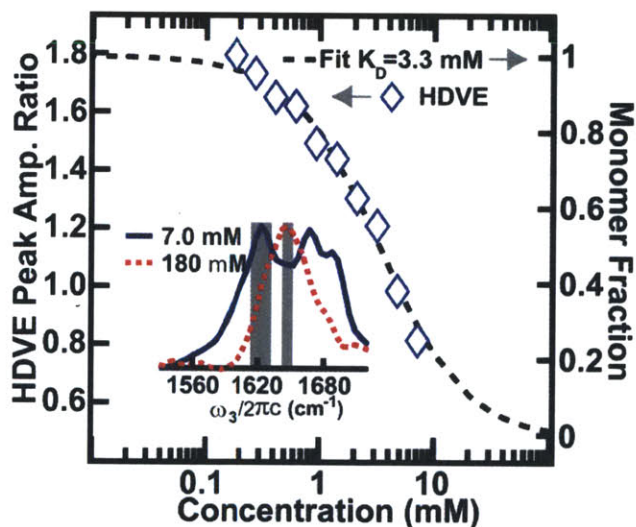


Figure 4.4: Left Axis: Insulin FTSI HDVE amplitude peak ratios are plotted versus concentration. Inlay: The lowest (180 μM , dotted red line) and highest (7.0 mM, solid blue line) concentration spectra are plotted. The grey bars indicate the integrated spectral regions used for the peak ratios. The 1642-1652 cm^{-1} integrated area is divided by the 1614-1633 cm^{-1} area. Right Axis: The ratios are fit to the Hill equation to extract a $K_D=3.3$ mM and an $n=1.1$. Insulin was dissolved in 10% EtOD and 0.27 M DCl. Each spectrum was collected on the less dispersive grating for 2 minutes at $\tau_{LO}=1.500$ ps with an internal LO passing through the sample ~ 30 ps after the other IR beams.

sheet at the dimer interface, allowing monomer and dimer to be distinguished from the relative intensities of α helical and β sheet regions of the amide I spectrum. Concentration-dependent HDVE amplitude spectra are shown for predominately dimeric, high (7.0 mM) and predominately monomeric, low (180 μ M) insulin concentrations. The ratio of HDVE amplitudes integrated across the α -helical region of the monomer (1642-1652 cm^{-1}) and the β -sheet region of the dimer (1614-1633 cm^{-1})⁶ are fit to the Hill equation: $\theta(c) = \frac{c^n}{K_D^n + c^n}$ where K_D is the

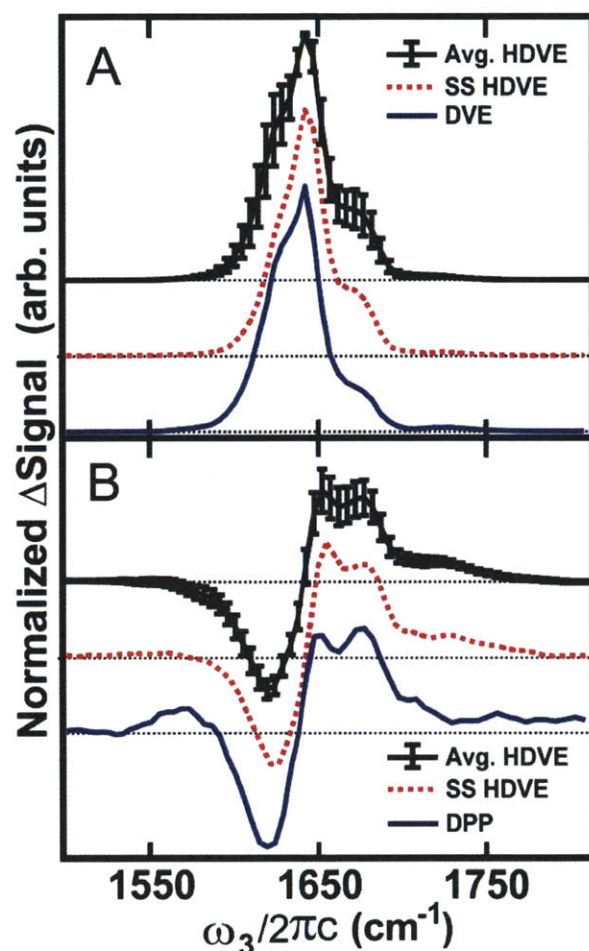


Figure 4.5: 15 mg/ml ubiquitin **A:** DVE (blue, solid line), single-shot HDVE-extracted amplitude squared (red, dotted line), 500-shot averaged HDVE-extracted amplitude squared (black, solid line), **B:** DPP (blue, solid line), single-shot HDVE-extracted DPP (red, dotted line), 500-shot averaged HDVE-extracted DPP (black, solid line) spectra are plotted. The standard deviation of the averaged HDVE-extracted spectra are shown. Spectra were collected at a 200 fs waiting time on the less dispersive grating at $\tau_{LO}=1.000$ ps with an internal LO. The DVE and DPP spectra were collected for 2.5 minutes.

dissociation constant, c is the insulin concentration, and n is the Hill coefficient.¹⁹ This fit leads to a value of $K_D = 3.3$ mM for $T = 22^\circ\text{C}$ and for a sample with 10% EtOD and 0.27 M DCl. For an ideal measurement of the dissociation constant, the chosen concentrations should span orders of magnitude across K_D . The amplification characteristics of HDVE allow for these necessary low concentration measurements, which are not accessible with current DVE intensities.

HDVE may also be used as a single-shot technique to extract the DVE and DPP spectra from a single measurement. HDVE is advantageous because it can be collected without scanning time delays and allows for the separation of phase and amplitude. Figure 4.5 compares the ubiquitin DVE and DPP spectra and their corresponding single-shot HDVE extracted counterparts. 500 shots are averaged and shown with standard deviations. Agreement between the single-shot HDVE and DVE and DPP spectra suggests that HDVE measurements may be used in low repetition or irreversible measurements where only unusually high absorption changes could be measured with conventional 2-beam pump-probe acquisition.

4.6 Conclusion

Heterodyne detection is a powerful technique that amplifies signals, allows for measurement of the amplitude and phase and generates a measurement that is linear to the electric field of the signal, which is in turn linear in concentration. The tradeoff is additional complexity in alignment, and an ambiguity in the absolute timing between LO and signal. We have demonstrated that HDVE is a one dimensional complement to 2D IR, which can be applied in single-shot mode. Through FTSI, PMSI, or a combination of the two, both the linear and third-order signal can be completely characterized as an amplitude and a phase, which allows for simultaneous extraction of the DVE and DPP or IR absorbance spectra. PMSI was also introduced as a complementary phase and amplitude determination technique that does not rely on pulse shapers or FT calculations. HDVE will be particularly valuable in experiments that have previously relied on DVE and DPP measurements, such as low concentration protein studies and low repetition rate temperature jump experiments, in which signal to noise ratios are small.

4.7 Acknowledgment

This work was supported by the National Science Foundation (CHE-0616575 and CHE-0911107).

4.8 References

- (1) Z. Ganim, H. S. Chung, A. W. Smith, L. P. DeFlores, K. C. Jones, and A. Tokmakoff, "Amide I Two-Dimensional Infrared Spectroscopy of Proteins," *Acc. Chem. Res.* **41**, 432-441 (2008).
- (2) H. S. Chung, M. Khalil, A. W. Smith, Z. Ganim, and A. Tokmakoff, "Conformational changes during the nanosecond to millisecond unfolding of ubiquitin," *Proc. Natl. Acad. Sci. U. S. A.* **102**, 612-617 (2005).
- (3) P. Hamm, M. Lim, W. F. DeGrado, and R. M. Hochstrasser, "Pump/probe self heterodyned 2D spectroscopy of vibrational transitions of a small globular peptide," *J. Chem. Phys.* **112**, 1907-1916 (2000).
- (4) A. T. Krummel and M. 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**, 13991-14000 (2006).
- (5) M. Khalil, N. Demirdoven, and A. Tokmakoff, "Coherent 2D IR Spectroscopy: Molecular Structure and Dynamics in Solution," *J. Phys. Chem. A* **107**, 5258-5279 (2003).
- (6) H. S. Chung, M. Khalil, and A. Tokmakoff, "Nonlinear infrared spectroscopy of protein conformational change during thermal unfolding," *J. Phys. Chem. B* **108**, 15332-15343 (2004).
- (7) W. P. de Boeij, M. S. Pshenichnikov, and D. A. Wiersma, "Heterodyne-detected stimulated photon echo: applications to optical dynamics in solution," *Chem. Phys.* **233**, 287 (1998).
- (8) M. C. Asplund, M. T. Zanni, and R. M. Hochstrasser, "Two-dimensional infrared spectroscopy of peptides by phase-controlled femtosecond vibrational photon echoes," *Proc. Natl. Acad. Sci. USA* **97**, 8219-8224 (2000).
- (9) A. Ferro Albrecht, J. D. Hybl, and D. M. Jonas, "Complete fs linear free induction decay, Fourier algorithm for dispersion relations and accuracy of the rotating wave approximation," *J. Chem. Phys.* **114**, 4649 (2001).
- (10) J. T. Kindt and C. A. Schmuttenmaer, "Far-Infrared Dielectric Properties of Polar Liquids Probed by Femtosecond Terahertz Pulse Spectroscopy," *J. Phys. Chem.* **100**, 10373 (1996).
- (11) M. Kuajwinska, "Spatial Phase Measurement Methods," in *Interferogram Analysis: Digital Fringe Pattern Measurement Techniques* (1993).
- (12) L. Lepetit, G. Chériaux, and M. Joffre, "Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy," *J. Opt. Soc. Am. B* **12**, 2467-2474 (1995).
- (13) C. Dorrer, N. Belabas, J.-P. Likforman, and M. Joffre, "Spectral Resolution and Sampling Issues in Fourier-transform Spectral Interferometry," *J. Opt. Soc. Am. B* **17**, 1795 (2000).
- (14) S. M. Gallagher, A. W. Albrecht, J. D. Hybl, B. L. Landin, B. Rajaram, and D. M. Jonas, "Heterodyne detection of the complete electric field of femtosecond four-wave mixing signals," *J. Opt. Soc. Am. B* **15**, 2338-2345 (1998).
- (15) S.-H. Shim, D. B. Strasfeld, Y. L. Ling, and M. T. Zanni, "Automated 2D IR spectroscopy using a mid-IR pulse shaper and application of this technology to the human islet amyloid polypeptide," *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14197-14202 (2007).

- (16) In balanced detection, after the signal and LO are combined on a beam splitter, both transmitted and reflected interferograms are collected and subtracted to remove homodyne components. A 180° phase shift is introduced between the reflected and the transmitted interferograms due to the presence of an internal reflection in one arm of the setup.
- (17) A. D. Bristow, D. Karaiskaj, X. Dai, and S. T. Cundiff, "All-optical retrieval of the global phase for two-dimensional Fourier-transform spectroscopy," *Opt. Express* **16**, 18017-18027 (2008).
- (18) E. H. G. Backus, S. Garrett-Roe, and P. Hamm, "Phasing problem of heterodyne-detected two-dimensional infrared spectroscopy," *Opt. Lett.* **33**, 2665-2667 (2008).
- (19) J. Barcroft and A. V. Hill, "The nature of oxyhaemoglobin, with a note on its molecular weight," *Journal of Physiology* **39**, 411-428 (1910).

Chapter 5

Transient Two Dimensional Spectroscopy with Linear Absorption Corrections Applied to Temperature Jump 2D IR

The work presented in this chapter has been published in the following paper:

- K. C. Jones, Z. Ganim, C. S. Peng, and A. Tokmakoff, "Transient Two Dimensional Spectroscopy with Linear Absorption Corrections Applied to Temperature Jump 2D IR," J. Opt. Soc. Am. B **29**, 118-129 (2012)

5.1 Abstract

Multidimensional spectroscopies provide increased spectral information but time resolution is often limited by the picosecond lifetimes of the transitions they probe. At the expense of additional complexity, *transient* multidimensional techniques extend the accessible timescales for studying non-equilibrium chemical and biophysical phenomena. Transient temperature-jump (T-jump) experiments are particularly versatile, since they can be applied to any temperature dependent change of state. We have developed a method to correct transient nonlinear techniques for distortions resulting from transient linear absorption of the probing pulses, distortions which can lead to false interpretations of the data. We apply these corrections in the collection of T-jump transient two-dimensional infrared spectra for the peptides diglycine and the β -hairpin peptide trpzip2. For diglycine, the T-jump induces changes in H-bonding, a

response which is inherent to all aqueous systems. The trpzip2 results probe the hairpin unfolding kinetics and reveal two time scales: <10 ns increased flexibility and 1.1 μ s β -hairpin disordering.

5.2 Introduction

Multidimensional spectroscopies such as two dimensional infrared (2D IR) and 2D electronic spectroscopy are becoming increasingly important for the study of structure^{2, 3}, dynamics^{4, 5}, and reactions⁶ in the condensed phase, from solutions⁷ to biological systems⁸⁻¹⁰. Multidimensional techniques are advantageous because they have fs time resolution, spectral information is spread over multiple frequency axes, peak intensities scale to high order with the dipole, and cross peaks reveal both coupling and dipole orientation¹¹. The use of multiple pulses also allows for time dependent experiments in which transitions are tagged and then probed after a variable waiting time¹². These waiting time experiments have been applied to study many problems including hydrogen bond switching in water⁴, photosynthetic energy transfer¹⁰, protein conformational dynamics^{13, 14}, and the chemical exchange of small molecule complexes in solution^{7, 15}. Although waiting time experiments are well suited to studying ultrafast timescales, their effective time window is limited by the vibrational or electronic lifetime of the chromophore - picoseconds in the case of condensed phase vibrations.

To overcome the lifetime-limited accessible time window of waiting time experiments, transient multidimensional experiments have been introduced in which spectra of a non-equilibrium system are collected following a triggering event¹⁶. For example, 2D IR has been used to track processes triggered by ultrafast or ns temperature-jump (T-jump) laser pulses¹⁷, such as for the study of photoswitchable peptide conformational changes^{16, 18}, ligand release^{19, 20}, charge injection²¹, and protein folding²²⁻²⁴. For these experiments, the accessible temporal window is not defined by spectroscopic relaxation properties of the oscillator, but rather by the time scale of the initiated process and the return to equilibrium. These transient 2D IR experiments are often done in conjunction with nonlinear experiments in one frequency dimension, such as Dispersed Pump Probe (DPP), Dispersed Vibrational Echo (DVE)^{1, 25}, or Heterodyned-detected Dispersed Vibrational Echo (HDVE), which are related to the real component, the power spectrum, and complex component, respectively, of the projection of the 2D IR spectrum onto the ω_3 axis²⁶. These one-dimensional experiments provide a reduced version of the 2D IR information, but they can be collected more quickly and easily to provide

better sampling of transient events.

The addition of a triggering pulse to a 3rd order nonlinear experiment increases the complexity of data collection and interpretation. Formally, transient multidimensional laser triggered experiments are 5th order nonlinear experiments, and they require a high signal-to-noise ratio to identify differences between the transient and equilibrium spectra, often necessitating dual frequency modulation strategies for double differential signal detection. Data collection is often slowed down to the repetition rate of the triggering event rather than the probing laser rate. Also, linear transient refractive index and absorption effects induced by the triggering pulse can lead to spurious peaks that overshadow the molecular changes of interest²⁷. These pulse propagation distortions are particularly problematic for solutes and solvents whose absorbance in the probing frequency range changes significantly in response to the triggering event. For example, in the T-jump triggered transient 2D IR (t-2D IR) studies described in this paper, the D₂O solvent linear absorbance is dependent on the temperature. Thus, the interrogating IR pulses and signal are shaped by the time-dependent temperature relaxation of the solvent. Although in this case the timescales are not expected to be significantly affected by the linear absorption, spurious peaks may alter the molecular interpretation. After overcoming the technical challenges, interpreting transient data is difficult due to the congested nature of the spectra, which have overlapping and interfering positive and negative peaks. Loss and gain features in the transient spectrum can show up as both positive and negative changes. Therefore, reducing spectral complexity by correcting for uninformative linear effects greatly aids the data analysis.

T-jump experiments have been used extensively in studies of protein and peptide folding kinetics, often by probing changes in the structurally sensitive amide I IR vibrational absorption spectrum. Relative to continuous-wave linear absorption T-jump techniques, the multidimensional T-jump amide I experiment is often challenged by greater noise, longer acquisition times, and technical complexity, but it offers enhanced spectral content, background free signals, and ultrafast time resolution. T-jump amide I IR experiments have almost exclusively measured linear absorption changes using techniques such as step-scan FTIR²⁸, single wavelength, continuous-wave IR lasers²⁹⁻³¹, or ultrafast IR pulse absorption³². The first two are limited to the time resolution of the IR detector and all three require additionally collecting the transient spectra of the solvent to isolate the solute response. The major advantage of the multidimensional T-jump amide I experiment is increased spectral content, allowing for

further disentanglement of the crowded amide I spectrum and simultaneous monitoring of crosspeaks which indicate coupling between vibrations. Although our T-jump t-2D IR experiment's time resolution is limited by the ns pulse length of the T-jump laser, the method is fundamentally limited only by the ps time-scale for vibrational relaxation in water. Also, since the nonlinear signal from the solvent is usually negligible, only data on one sample needs to be collected. All laser-induced T-jump experiments are limited by cavitation, thermal lensing, and beam homogeneity.

In a previous publication, we presented the design, construction, and testing of a transient 2D IR spectrometer for use in temperature-jump measurements²³. In this paper, we describe data acquisition methods that correct for non-equilibrium absorption and dispersion effects in T-jump t-2D IR spectroscopy to isolate the desired kinetics for chemical processes. We first apply these methods to the model dipeptide diglycine to measure the relaxation of a small molecule which should not exhibit a response in the ns-ms window of our T-jump. We then collect the t-2D IR spectra of the peptide trzip2 (TZ2) to interpret the disordering kinetics of an extensively studied fast folding β -hairpin. Although the techniques are used to study the amide I vibrations of peptides through nonlinear T-jump experiments, the topics discussed are generalizable for any temperature dependent equilibrium or transient multidimensional spectroscopy.

5.3 Experimental

5.3.1 Mid-Infrared Spectroscopy

All infrared probing methods reported are heterodyne-detected ultrafast third-order nonlinear spectroscopies performed in the boxcar geometry and detected using spectral interferometry with an external local oscillator (LO) pulse. Nonlinear infrared spectra were collected at 1 kHz using 90 fs pulses centered at $\sim 1650\text{ cm}^{-1}$ ($6\text{ }\mu\text{m}$) with 160 cm^{-1} FWHM bandwidth as described previously¹¹. The experimental layout of the sample and detection compartments is shown in Figure 5.1. Three pulses (E_1 , E_2 , and E_3), separated by sequential time intervals τ_1 and τ_2 were focused into the sample to a $\sim 75\text{ }\mu\text{m}$ diameter spot to generate a nonlinear signal field E_{sig} . The LO was focused through the same sample volume as the other infrared beams, but it was delayed $\sim 60\text{ ps}$ relative to E_3 . This delay is long enough not to interfere with the IR signal generation but short enough that all IR fields experience the same perturbations. The LO and signal were spatially overlapped by combining them on opposite sides

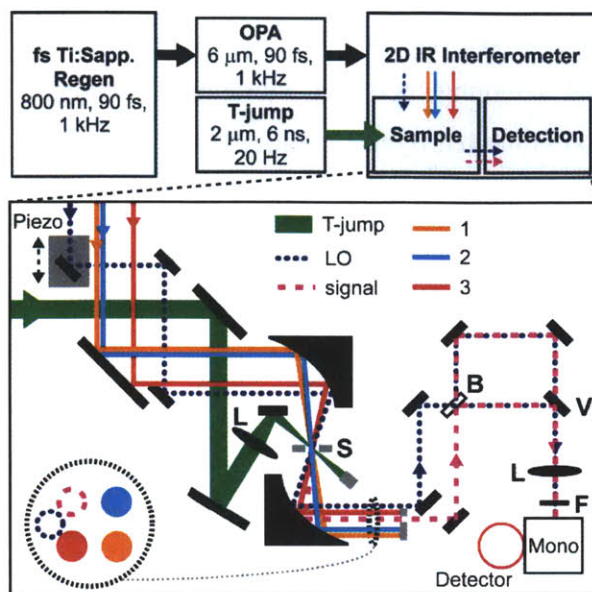


Figure 5.1: The output of a femtosecond Ti:sapphire regenerative amplifier is converted to the IR by difference frequency generation in an optical parametric amplifier, and split into four IR beams in the 2D IR interferometer. **Inlay:** Three IR beams generate IR 3rd order signal at the sample. The T-jump laser is overlapped spatially with the IR beams at the sample and delayed in time electronically. The LO beam is overlapped spatially with the IR beams at the sample, but delayed from the other IR pulses. The signal and LO beams are overlapped spatially and temporally at the detector and detected using balanced detection. The LO beam's delay can be stepped using the piezo-controlled stage for PMSI or FTSI measurements. **L:** lens, **S:** sample, **B:** beam-splitter, **V:** vertically displaced mirror, **F:** low pass $\lambda=3.4 \mu\text{m}$ filter.

of a beamsplitter. The signal field was detected using balanced detection of the two resulting heterodyned signals, which have a relative 180° phase shift. After diffracting the two resulting beams off of a 40 lines/mm grating in a 190 mm spectrograph (Triax 190, Jobin Yvon), interference between E_{sig} and E_{LO} was detected with $\sim 4.7 \text{ cm}^{-1}$ resolution on a 2 x 64 element mercury cadmium telluride array (IR-0144, Infrared Systems Development). Beam 2 or 3 was mechanically chopped at 500 Hz to remove scatter and the LO spectrum. LO timing and phasing were achieved by comparison to a two beam DPP spectrum. The resulting detection frequency axis is referred to as ω_3 . Both HDVE and 2D IR spectra were collected with all parallel polarization (ZZZZ) and at $\tau_2=150 \text{ fs}$ to minimize nonresonant signal from the CaF_2 windows.

5.3.2 Temperature-Jump

The T-jump laser system and its synchronization with the mid-IR system have been described previously²³. Briefly, the 6 ns, 1.98 μm , 9 mJ T-jump pulse was generated at 20 Hz

through the combination of a Q -switched Nd:YAG (yttrium aluminum garnet) laser and a β -barium borate based optical parametric oscillator (YG981c, OPOTEK) tuned to match the $\nu=2\leftarrow 0$ OD stretch of D_2O . The T-jump laser is focused to a $500\ \mu\text{m}$ waist at the same point in the sample as the mid-IR beams, creating a $10\text{-}12^\circ\text{C}$ jump in our sample in $<10\ \text{ns}$. Since the T-jump focus is approximately five times the size of the mid-IR waist, the T-jump in the region of mid-IR beam overlap is uniform. The electronic delay generator creates the time delay between the T-jump pulse and the first of the 50 mid-IR beam sets with an instrument response time of $6\ \text{ns}$ and a delay range of ns-ms. In addition to electrical isolation, we found it necessary to acoustically isolate the T-jump laser from the IR laser table to remove interferometrically-measurable noise created by flashlamp firing.

The difference between repetition rates resulted in a $50\ \text{ms}$ window with 50 collected mid-IR spectra between each T-jump shot. Figure 5.2 shows the time delays between the $1\ \text{kHz}$ IR and $20\ \text{Hz}$ T-jump pulses. We denote the time delay between the T-jump pulse and the j^{th} mid-IR pulse following the T-jump as

$$\tau^j = \tau + j \times (1\ \text{ms}), \text{ for } j=0,1,2,\dots,49. \quad (5.1)$$

τ is the primary time-delay between the T-jump pulse and the 2D IR probe event following it. The T-jump pulse was spatially and temporally overlapped with the mid-IR pulse by monitoring

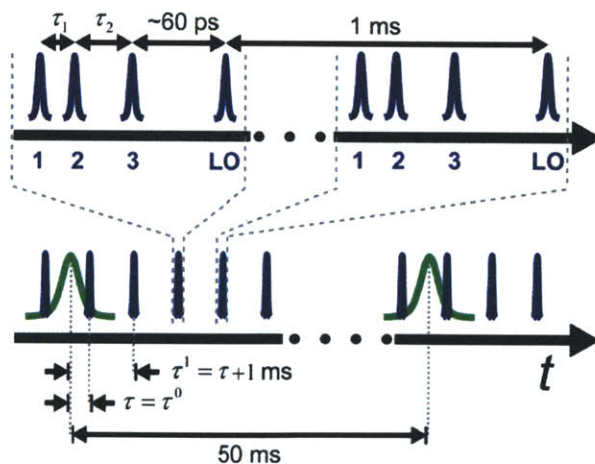


Figure 5.2: IR beams (blue) 1,2 and 2,3 are delayed by τ_1 and τ_2 . The LO reaches the sample $\sim 60\ \text{ps}$ after beam 3. The original IR pulse, from which beams 1,2,3, and LO are created, is generated at $1\ \text{kHz}$, which creates a $1\ \text{ms}$ lag between sets of IR pulses. The $20\ \text{Hz}$ T-jump pulse (green) is delayed electronically relative to the IR pulses. The first IR pulse set following the T-jump pulse arrives at a delay of $\tau = \tau^0$. The next 49 IR pulse sets arrive at $\tau^j = \tau + j \times (1\ \text{ms})$.

the transmittance of the LO through the transiently heated solvent. Zero time, $\tau^0=0$, was set at the delay when the transient transmittance signal of the LO was half of the maximum transient transmission, which was a ~5% increase over equilibrium. The T-jump re-equilibrates by thermal diffusion into the sample cell windows, reaching half its initial value at $t\sim 1.5$ ms, and the time-trace can be described by a stretched exponential of the form $\exp[-(t/\tau_s)^\beta]$ where $\tau_s = 2.05$ ms and $\beta=0.67$.

Transient data collection is complicated by chopping to remove the homodyned LO term of the collected signal. To collect the complementary 500 Hz chopped and unchopped mid-IR spectra at a given T-jump delay τ^j , the phase of the chopper was modulated between 0° and 180° . At a phase of $0^\circ(180^\circ)$, the odd(even) mid-IR pulses following the T-jump pulse are open, while the even(odd) pulses are blocked.

5.3.3 Sample Preparation

The tryptophan zipper β -hairpin peptide, TZ2 (NH₂-SWTWENGKWTWK-CONH₂) was synthesized as described previously³³. The TZ2 samples were dissolved in pH 2.5 DCl in D₂O solvent to 10 mg/mL concentrations (OD of ~ 0.05). Diglycine (L-glycyl-L-glycine) was purchased from Sigma and dissolved in D₂O (pH=6.5) to a concentration of 20 mg/mL (OD of ~ 0.9) unless noted otherwise. ~ 25 μ L peptide solution was sandwiched between two 1 mm thick CaF₂ windows with a 50 μ m pathlength defined by a teflon spacer. The TZ2 sample was then sealed in a brass cell held at $T_0=35^\circ\text{C}$ by a regulated water bath ($T_0=20^\circ\text{C}$ for diglycine). For these samples, any changes induced by the T-jump relax reversibly back to equilibrium within the 50 ms window, and flowing the sample was found to be unnecessary in this case.

5.4 Transient Data Collection and Processing

5.4.1 Theoretical Background

The purpose of our data acquisition process is to perform transient nonlinear spectroscopy in the absence of distortions or artifacts that may arise from propagation through the non-equilibrium environment prepared by the trigger pulse. As the mid-IR laser pulses pass through the sample, their group velocity is affected by the temperature-dependent (T-jump delay-dependent) real part of the refractive index, $n(\omega; \tau^j)$, and their magnitude is affected by temperature-dependent absorption, $\kappa(\omega; \tau^j)$. The complex refractive index experienced by the mid-IR pulses is given by:

$$\tilde{n}(\omega; \tau') = n(\omega; \tau') + i\kappa(\omega; \tau'). \quad (5.2)$$

We write this noting that this is not the equilibrium index of refraction, but an effective index that reflects an average over the dielectric response of the sample, $\epsilon(\omega)$, following the T-jump for frequencies $\omega > 2\pi/\tau^j$. As with other transient experiments, it will include changes in absorption and dispersion that arise from molecular dynamics on time scales faster than the T-jump delay, including acoustic and thermal responses.

The real, time-domain electric field of one of the incident pulses after passing through the sample, $E_\alpha(t; \tau')$, can be Fourier transformed (FT) into the frequency domain to give the complex field,

$$\tilde{E}_\alpha(\omega; \tau') = e_\alpha(\omega) \exp[i\phi_\alpha(\omega)] \exp[i\tilde{n}(\omega; \tau')\omega l / c], \quad \alpha = 1, 2, 3, LO \quad (5.3)$$

where e is the amplitude, ϕ is the phase, and the last term accounts for the temperature-dependent and T-jump delay-dependent linear refractive index change that accumulates as the field passes through the sample with path length l . c is the speed of light. Because beams 1, 2, 3, and the LO generate or interfere with the nonlinear signal, the linear \tilde{n} propagation effects manifest themselves in the resulting nonlinear spectrum.

To remove the relative effect of the $n(\omega; \tau')$, the LO is aligned through the sample along with the other pulses. As a result, the group velocity of each beam increases by the same amount, minimizing the phase difference between them. Although the LO and signal are intentionally delayed relative to each other at the sample, the ~ 60 ps difference is negligible compared to the ns T-jump pulse length. We therefore assume that the LO and signal are equally affected by the sample's linear refractive index as they pass through the sample.

Assessing how $\kappa(\omega; \tau')$ distorts the experimentally-measured nonlinear signal relative to an ideal (infinitely thin) sample signal requires a description of how absorption of each laser pulse affects the generated signal, which, in turn, is also affected by propagation through the sample³⁴⁻³⁶. The real time-domain signal electric field $E_{sig}(t_3; \tau_1, \tau_2 = 150 \text{ fs}, \tau')$ results from the polarization induced in the sample by the time-delayed beams 1, 2, and 3, which is generated gradually as the pulses propagate through the sample. To a first approximation, the *combined* absorption of beams 1 and 2 imprints a depletion term $K(\omega_1; \tau')$ along the ω_1 axis³⁶ of the frequency-domain two dimensional spectrum where

$$K(\omega_n; \tau^j) = \frac{1 - \exp[-2\kappa(\omega_n; \tau^j)\omega_n l / c]}{2\kappa(\omega_n; \tau^j)\omega_n l / c}. \quad (5.4)$$

The combined depletion effect of beams 3 and the signal is expressed by $\exp[-\kappa(\omega_3; \tau^j)\omega_3 l / c]$. This conventional absorption term can be explained in terms of a transient grating picture in which the signal is generated by diffraction of beam 3 off of a grating or in the picture in which the absorption accumulated by beam 3 is imprinted on the signal that is gradually generated in the sample. The LO can be treated independently as a pulse that accumulates an $\exp[-\kappa(\omega_3; \tau^j)\omega_3 l / c]$ absorption term, as is shown in eq. (5.3).

To a first approximation, absorbance of beams 1 and 2 distorts the resulting 2D spectrum along the ω_1 axis, whereas absorption of beam 3, the signal, and the LO affects ω_3 . Since the LO experiences the same index as the remaining beams, the effect of $n(\omega; \tau^j)$ vanishes upon detection. Collecting the above terms, which assume the sample is a weak absorber and ignore coherent energy transfer during τ_1 and τ_3 , we can approximate the effects of linear absorption as³⁶:

$$\tilde{S}^{1,3}(\omega_1, \omega_3; \tau_2, \tau^j) \approx \tilde{S}(\omega_1, \omega_3; \tau_2, \tau^j) \times K(\omega_1; \tau^j) \times \exp[-2\kappa(\omega_3; \tau^j)\omega_3 l / c]. \quad (5.5)$$

Here, $\tilde{S}^{1,3}$ is the directly detected 2D spectrum where the superscripts 1 and 3 indicate the existence of distortions along ω_1 and ω_3 . \tilde{S} without superscripts is the desired spectrum that would be measured for an optically thin sample unaffected by linear propagation effects, but still shaped by the finite bandwidth of our pulses, which acts to weaken apparent intensity on the wings of the spectrum. For the present work, τ_2 is always 150 fs, and it is not explicitly shown in the rest of the equations in the paper for simplicity. The second term on the right in eq. (5.5) represents the *combined* absorption of E_1/E_2 , while the third term accounts for the absorption of $E_{LO}+E_3/E_{sig}$, which results in the factor of 2. With negligible absorption, both correction terms approach 1.

With propagation effects accounted for, we can properly define the signal electric field emerging from the sample. \tilde{E}_{sig} depends parametrically on τ_2 , τ_1 , and τ^j and can be expressed as

$$\tilde{E}_{sig}^{1,3}(\omega_3; \tau_1, \tau^j) = e_{sig}^1(\omega_3; \tau_1, \tau^j) \exp[i\phi_{sig}(\omega_3; \tau_1, \tau^j)] \exp[i\tilde{n}(\omega_3; \tau^j)\omega_3 l / c]. \quad (5.6)$$

As above, the superscripts 1 and 3 indicate that the spectrum is distorted by propagation effects

along the ω_1 and ω_3 axes. The \tilde{E}_{LO} T-jump delay dependence is solely due to the transient linear absorbance of the sample, whereas the signal amplitude and phase are also expected to change due to molecular dynamics during τ_1 and τ_2 . Because propagation effects on the ω_1 axis cannot be cleanly expressed as a function of τ_1 , the ω_1 distortion is folded into $e_{sig}^1(\omega_3; \tau_1, \tau')$ while the ω_3 distortion is explicitly shown.

5.4.2 Data Collection

Our detector is a dual stripe array that collects data from 2x64 pixels at 1 kHz. The 1 kHz data is modulated by the electronically delayed 20 Hz T-jump laser and a 500 Hz chopper. The chopper phase is flipped periodically at ~ 0.1 Hz to fully characterize every combination of the 20 Hz and 500 Hz modulation. In practice, the detector measures intensity $I(\omega_3) = |\tilde{E}(\omega_3)|^2$ in the frequency domain after diffraction off of a grating. Characterization of \tilde{E} requires interferometric detection. In both 2D IR and HDVE spectroscopy, \tilde{E}_{sig} was characterized by combining the signal and LO at a delay τ_{LO} , and measuring the ω_3 difference spectrum of the LO with and without the signal present. In balanced detection, intensities are observed for the out-of-phase intensity from the upper (+) and lower (-) stripes. The spectrum of the LO (chopper blocked signal, °) is

$$I_{\pm}^{\circ}(\omega; \tau') = |\tilde{E}_{LO}(\omega; \tau')|^2 = [e_{LO}(\omega)]^2 \exp[-2\kappa(\omega; \tau')\omega l / c] \quad (5.7)$$

and that of the combined E_{LO} and E_{sig} (chopper open) is

$$\begin{aligned} I_{\pm}^{1,3}(\omega_3; \tau_1, \tau', \tau_{LO}) &= |\tilde{E}_{sig}^{1,3}(\omega_3; \tau_1, \tau') + \tilde{E}_{LO}(\omega_3; \tau') \exp[-i\omega_3 \tau_{LO}]|^2 \\ &= \exp[-2\kappa(\omega_3; \tau')\omega_3 l / c] \times \left\{ [e_{LO}(\omega_3)]^2 + [e_{sig}^1(\omega_3; \tau_1, \tau')]^2 \right. \\ &\quad \left. \pm 2e_{LO}(\omega_3)e_{sig}^1(\omega_3; \tau_1, \tau') \cos[\Delta\phi(\omega_3; \tau_1, \tau') + \tau_{LO}\omega_3] \right\} \end{aligned} \quad (5.8)$$

Here we have defined

$$\Delta\phi(\omega_3; \tau_1, \tau') = \phi_{sig}(\omega_3; \tau_1, \tau') - \phi_{LO}(\omega_3). \quad (5.9)$$

By sending the LO through the sample, the shared change in real refractive index induced by the T-jump delays the signal, LO, and beams 1, 2, and 3 by the same relative amount, and $n(\omega; \tau')$ drops out of eq. (5.8)²³. If the LO is not passed through the sample, the relative time delay between the LO and other beams changes with the temperature and the resulting spectra are not

phased properly, which results in the mixing of the real and imaginary parts.

If the interference cross-term in eq. (5.8) (which is proportional to the desired nonlinear signal field) were Fourier transformed along τ_1 , it would result in $\tilde{S}_{2D}^{1,3}$ given in eq. (5.5). Instead, the cross term in eq. (5.8) is separated and the ω_3 -axis linear absorption distortion is directly removed using the following data acquisition steps:

$$S^1(\omega_3; \tau_1, \tau^j, \tau_{LO}) = \left[\left\langle \frac{I_+^{1,3} - I_-^{1,3}}{I_+^{1,3} + I_-^{1,3}} \right\rangle - \left\langle \frac{I_+^c - I_-^c}{I_+^c + I_-^c} \right\rangle \right] \times [e_{LO}(\omega_3)]^2 \quad (5.10)$$

$$= 2e_{LO}(\omega_3)e_{sig}^1(\omega_3; \tau_1, \tau^j) \cos[\Delta\phi(\omega_3; \tau_1, \tau^j) + \tau_{LO}\omega_3] \quad (5.11)$$

$$\approx S(\omega_3; \tau_1, \tau^j, \tau_{LO}) \times K(\omega_1; \tau^j) \quad (5.12)$$

The out-of-phase detection arrays are subtracted to remove the homodyne terms and divided by the sum. The denominator divides out the linear absorption term due to LO, signal, and beam 3 and normalizes the heterodyne signal for each single shot to remove shot-to-shot fluctuations before averaging. These operations are performed at 1 kHz, and the data is then averaged, as indicated by triangular brackets. Finally, the subtracted, shot-normalized, and averaged chopped terms are subtracted. The resulting signal has been corrected for linear absorption along ω_3 . The 1 superscript indicates that there are still distortions along ω_1 . The $[e_{LO}(\omega_3)]^2$ term, the IR pulse spectrum in the absence of the sample, is calculated using a separate FTIR absorbance spectrum collected at the equilibrium temperature and the T-relaxed $I_{\pm}^c(\omega; \tau^{ref})$, where $ref=48$ or 49 (eq. (5.7)).

Theoretically, the 2nd term on the right of eq. (5.10) involving I^c is not necessary assuming balanced detection, but we found it essential for removing scatter and electrical noise as well as for correcting for residual intensity mismatch between the stripes. To remove any small electrical noise that might show up in the denominator of the 1st term on the right of eq. (5.10), a background was collected without any IR beams beforehand and subtracted directly from all the raw data. Data whose integrated LO intensity was less than a preset cutoff of ~50% of the average was rejected to remove errant shots.

Due to the difference in T-jump and mid-IR laser repetition rate, for each τ we actually collect 50 time delays, τ^j . Each of these 50 data sets are collected, saved, and processed independently, and the last two spectra, $j=48$ and 49 , are assumed to be equilibrium spectra

collected after the T-jump has relaxed back to the initial temperature.

5.4.3 Post Processing

To correct for the residual distortion in S^1 due to absorption of beams 1 and 2, the imaginary refractive index is calculated using the τ^j -dependent LO spectra. The differential absorption relative to the equilibrium LO was calculated and added to the reference absorption measured using an FTIR collected at the equilibrium temperature:

$$\kappa^j(\omega; \tau^j) = \Delta\kappa^{j,ref}(\omega; \tau^j) + \kappa_{FTIR}^{Eq}(\omega) \quad (5.13)$$

$$= \frac{-c}{2\omega l} \ln \left[\frac{I_+^o(\omega; \tau^j) + I_-^o(\omega; \tau^j)}{I_+^o(\omega; \tau^{ref}) + I_-^o(\omega; \tau^{ref})} \right] + \kappa_{FTIR}^{Eq}(\omega) \begin{cases} ref = 48, \text{ for } j = 0, 2, \dots, 48 \\ ref = 49, \text{ for } j = 1, 3, \dots, 49 \end{cases} \quad (5.14)$$

$\Delta\kappa^{j,ref}(\omega; \tau^j)$ represents the relative change in linear absorption between the τ^j transient spectra and the reference spectrum, $I^o(\omega; \tau^{48})$ or $I^o(\omega; \tau^{49})$, which is analogous to measuring the linear absorption using $I(\tau^j)/I(ref)$. Ideally, the sample has returned to equilibrium or been replaced with a fresh equilibrium sample within 48 ms. $\kappa^j(\omega; \tau^j)$ takes into consideration the absorption due to the solvent, solute, and the CaF₂ windows. For the t-2D IR, the ω_1 correction is applied after the FT into ω_1 :

$$\tilde{S}(\omega_1, \omega_3; \tau^j, \tau_{LO}) \approx \tilde{S}^1(\omega_1, \omega_3; \tau^j, \tau_{LO}) \times [K(\omega_1; \tau^j)]^{-1} \quad (5.15)$$

The correction in eq. (5.15) removes the τ -dependent absorption due to beams 1 and 2.

For t-HDVE, we cannot directly apply a correction for ω_1 because we only have 1D ω_3 -slices. Instead, we indirectly correct for ω_1 -distortions by multiplying the correction along ω_3 :

$$S(\omega_3; \tau_1, \tau^j, \tau_{LO}) \approx S^1(\omega_3; \tau_1, \tau^j, \tau_{LO}) \times [K(\omega_3; \tau^j)]^{-1}. \quad (5.16)$$

Since this is formally only correct for diagonal peaks ($\omega_1 = \omega_3$) without spectral diffusion, this assumption is tenuous and should be applied cautiously.

The corrections along ω_1 and ω_3 described above approximately correct for linear absorbance of the finite-bandwidth pulses by the sample. Measuring the ideal 2D spectrum that would have been generated with infinite bandwidth pulses would require further manipulations to remove the effect of the limited spectral bandwidth of our pulses, which accentuates peaks near the middle of the pumping and probing spectra. Using the same approximations described above, to correct for this artificial intensity accentuation we would divide both the ω_1 and ω_3

axes by $[e_{LO}(\omega)]^2$.

Once processed, averaged, and corrected, the data is presented as a difference signal, in which the last two spectra are subtracted out as equilibrium spectra:

$$\Delta S(\tau^j) = S(\tau^j) - S(\tau^{ref}), \begin{cases} ref = 48, & \text{for } j = 0, 2, \dots, 48 \\ ref = 49, & \text{for } j = 1, 3, \dots, 49 \end{cases} \quad (5.17)$$

For 2D IR, $S(\tau^j)$ is either the real or absolute value t-2D IR signal after FT along τ_1 : $S(\omega_1, \omega_3; \tau^j)$ and $|\tilde{S}(\omega_1, \omega_3; \tau^j)|$. For HDVE, $S(\tau^j)$ is either the real or absolute value HDVE signal at $\tau_1 = 0$ fs after exploiting the τ_{LO} dependence to calculate the complex signal: $S(\omega_3; \tau_2, \tau^j)$ and $|\tilde{S}(\omega_3; \tau_2, \tau^j)|$.

Although the methods described above were discussed specifically for transient T-jump 2D IR experiments, they are generalizable to temperature-dependent 2D IR studies or any heterodyne-detected transient third order IR or electronic spectroscopy, particularly in cases where there is a strong linear transient absorbance response. For example, in studying protein folding problems, the linear absorbance of the D₂O solvent strongly depends on the temperature. Separating the temperature-dependent response of the protein requires correcting for the linear absorption distortions.

5.4.4 t-2D IR

For t-2D IR collection, τ_1 delays were undersampled in 14 fs steps scanned from 0 to >2.4 ps and to >1.3 ps for rephasing and non-rephasing surfaces, respectively, using delay stages (Aerotech, ANT-50L) with 100 nm reproducibility²³. The τ_{LO} was set to 0. The data was collected in three nested 'for' loops: (τ ; τ_1 ; chopper phase). 7000 shots were collected at chopper=0° before flipping to chopper=180° and collecting another 7000 shots. The τ_1 delay was then stepped to the next delay, and another 7000+7000 shots were collected. This stepping of τ_1 continued until the entire rephasing and non-rephasing surfaces were mapped out, at which point τ was changed and the process was repeated. Data was processed on the fly (eq. (5.10)), the τ_1 axis was then fast Fourier transformed after zero padding rephasing and nonrephasing interferograms to 16 ps, and the undersampled data was reflected about the Nyquist frequency, $\omega_1 = 2\omega_{Nyq} - \omega_1^{FT}$, where ω_1^{FT} is the frequency that results from direct FT²³. The resulting spectrum was corrected along ω_1 (eq. (5.15)) to give $\tilde{S}(\omega_1, \omega_3; \tau^j)$. Because all 50 of the τ -set

spectra are collected with the same delay stage positions, the phasing correction is determined for the reference spectrum and then applied to the rest. Multiple spectra collected at the same τ are averaged in the frequency domain after correcting for the phasing error.

At each τ , a set of 50 2D IR spectra was collected in ~ 1.25 hours, or 1.5 minutes per τ' 2D IR spectrum. The τ spectra presented here for TZ2(diglycine) are the result of $\geq 8(1)$ averages. Spectra with $\tau' \geq 1$ ms but different τ were averaged together if the difference in time-delay was ≤ 1 μ s. For example, the presented 1 ms T-jump delay spectra are the average of τ' spectra ranging from $\tau' = 10$ ns +1 ms to $\tau' = 1$ μ s +1 ms.

5.4.5 t-HDVE

For t-HDVE collection, the τ_j delay was fixed to 0, and the τ_{LO} was stepped from 0 to 25 fs in 5 fs steps so that the complex spectrum could be calculated during data processing. The stepping was done using a piezo stage (THORLABS, NF5DP20S) for its 10 nm repeatability. The data was collected in three nested 'for' loops: (τ ; τ_{LO} ; chopper phase). 10000 shots were collected at chopper= 0° before flipping to chopper= 180° and collecting another 10000 shots. The τ_{LO} was then stepped by 5.0 fs, for which the carrier phase shift $\omega \times \Delta\tau_{LO} = \pi/2$, and another 10000+10000 shots were collected. Once all the τ_{LO} delays were collected, τ was changed and the process was repeated. Data was processed on the fly (eq. (5.10)) and the resulting spectra were corrected for ω_1 distortions (eq. (5.16)). The complex HDVE signal, $\tilde{S}(\omega_3; \tau', \tau_{LO})$ was then calculated using either FT spectral interferometry (SI) or phase modulation SI, in which the difference and sum of τ_{LO} -stepped spectra were used to calculate the real and imaginary spectra followed by a phase correction back to $\tau_{LO}=0$ ²⁶. The real HDVE signal $S(\omega_3; \tau', \tau_{LO})$ is equivalent to DPP, which is equal to the projection of the real 2D IR surface, $S(\omega_1, \omega_3; \tau')$, onto ω_3 . The absolute value of the HDVE signal, $|\tilde{S}(\omega_3; \tau', \tau_{LO})|$, is equivalent to the square root of the dispersed vibrational echo (DVE)²⁶.

Each τ set of 50 HDVE spectra was collected in ~ 2 minutes, or ~ 2.5 s per τ' HDVE spectrum. The τ spectra presented here are the result of ≥ 6 averages. As with the 2D IR data, spectra with $\tau' \geq 1$ ms but different τ were averaged together if the difference in time-delay was ≤ 1 μ s.

5.5 Results and Discussion

5.5.1 Diglycine

The dipeptide diglycine provides a baseline for the sorts of changes we should expect from a T-jump experiment on a peptide in which all dynamics are faster than the T-jump pulse, and no conformational changes are observed in the ns-ms T-jump window. We also study diglycine to compare the effects of correcting the nonlinear spectrum for linear absorption.

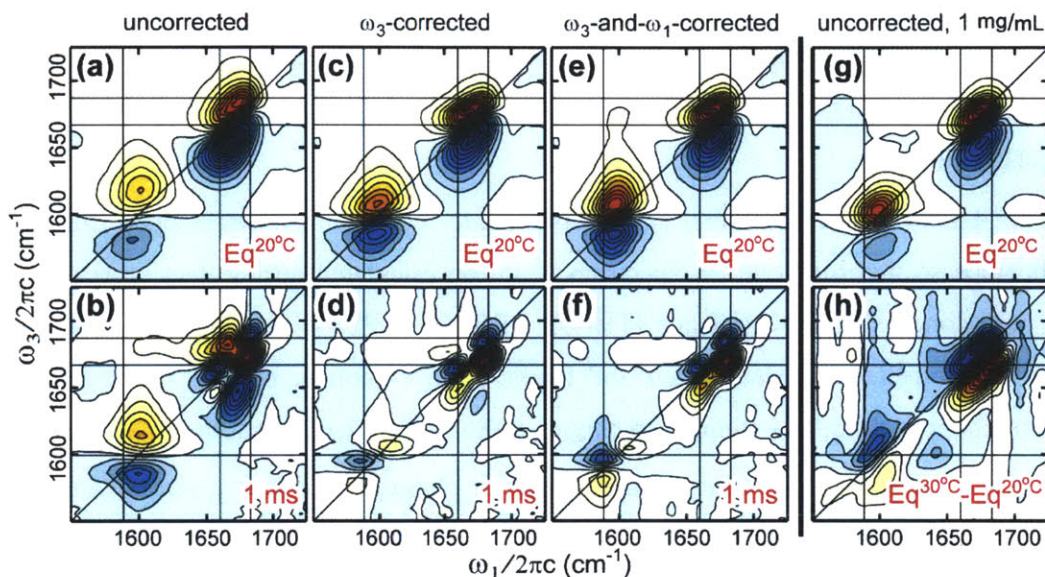


Figure 5.3: The absorptive (real) spectra of t-2D IR of diglycine are shown. Grid lines mark the three loss features shown in **f**. The first 3 columns (**a-f**) show data taken of 20 mg/mL diglycine ($OD = 0.9$) following a T-jump from 20 to 30°C. The 1st row (**a,c,e**) shows equilibrium spectra at 20°C. The 2nd row (**b,d,f**) shows the 1 ms *difference* spectra. Columns 1-3 are calculated from the same data but with different corrections for linear absorption of the IR beams. Column 1 (**a,b**) is uncorrected ($S^{1,3}$, eq. (5.5)). Column 2 (**c,d**) is ω_3 -corrected (S^1 , eq. (5.10)). Column 3 (**e,f**) is ω_1 -and- ω_3 -corrected (S , eq. (5.15)). Column 4 (**g,h**) shows the real equilibrium 2D IR spectra of diglycine at 1 mg/mL ($OD \sim 0.05$). **g** is the equilibrium 20°C spectrum. **h** shows the difference between the 30 and 20°C equilibrium spectra. The spectra are individually normalized and scaled such that 25 linearly-spaced contours span the entire magnitude range of $\sinh(S_{Norm.} \times 0.1)$.

At the pH of 6.5 used here, diglycine exists in its zwitterionic form and shows two discernible peaks in the equilibrium 2D IR spectrum (Figure 5.3a), a homogeneously broadened peak at 1595 cm^{-1} , which is assigned to the antisymmetric stretch of the C-terminal carboxylate, and an inhomogeneously broadened amide I' peak at 1674 cm^{-1} . Previous work on N-methylacetamide (NMA) has shown that the amide I' peak frequency strongly depends on the surrounding electrostatic environment defined by the multiple hydrogen-bond (H-bond)

configurations to the backbone, as should the carboxylate³⁷. Diglycine's amide I' peak is more inhomogeneous compared to NMA's peak, which suggests that there is a wider variety of H-bond configurations for diglycine as is expected given the proximity of the terminal H-bond-donating NH₃⁺. These peaks are expected to blue shift with temperature as the average number of hydrogen bonds decrease.

Diglycine: Corrections for Linear Absorption

Linear absorption affects the nonlinear equilibrium spectrum by reducing intensity along both ω_1 and ω_3 wherever there is absorption in the FTIR. In order to compare the effects of linear absorption on the equilibrium spectrum of diglycine, different corrections were applied to the same raw data, as is shown in the columns of Figure 5.3. The equilibrium spectra differ in the 1600 cm⁻¹ region because the sample absorption (peak optical density at $\omega = 1596$ cm⁻¹ is 0.9) depletes intensity along ω_3 , which results in an artificially large separation between the positive and negative peaks of the carboxylate stretch. This distortion also causes the ω_1 displacement between the positive and negative peak maxima observed in Figure 5.3a for the amide I' vibration. Corrections along ω_3 and ω_1 (Figure 5.3c and e) remove the distortions. The effect of ω_3 -correction is larger than ω_1 -correction, and both decrease the apparent anharmonicity.

Linear absorption in the transient spectra is more complicated than the equilibrium case because both the solvent and the solute respond to the T-jump. At our given initial temperature and T-jump size, the T-jump causes a ~5% increase in transmission due to the temperature dependency of the linear absorption, shown in Figure 5.4a. This ~5% transmission increase magnifies the underlying signal, which consists of the equilibrium spectrum and ~5% changes due to the T-jump. As a result, the difference spectra in Figure 5.3b is due to two interfering spectral features: ~5% molecular changes resulting from the T-jump and ~5% magnified equilibrium spectrum due to the increased transmission. Most of the mixed-in equilibrium spectrum is removed by correction for beam 3, the LO, and the signal (Figure 5.3d), where the τ -dependent ω_3 -correction factor, $\exp[2\kappa(\omega; \tau')\omega l / c]$, is shown in Figure 5.4b. Full correction along both ω_1 and ω_3 reveals the undistorted difference spectrum (Figure 5.3f), which differs from Figure 5.3d primarily in the reweighting of features in the 1600 cm⁻¹ region. Qualitatively, spurious spectral peaks like those seen in Figure 5.3b and d can lead to false molecular interpretations. Figure 5.4c shows the ω_1 -correction factor, $[K(\omega; \tau')]^{-1}$. Both the ω_3 and ω_1 -

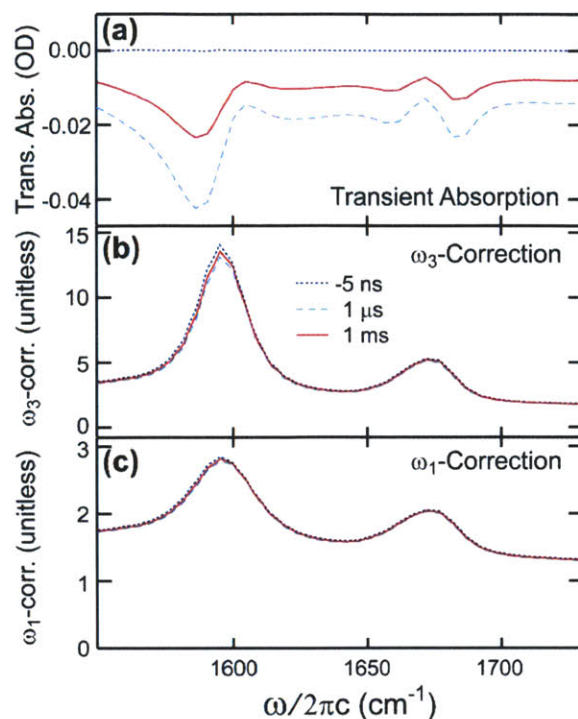


Figure 5.4: -5 ns , 1 μ s, and 1 ms difference spectra for diglycine in D_2O . **a** The transient difference absorption spectra of diglycine and solvent calculated from the transient LO spectra $(-\log[I^\circ(\omega; \tau')/I^\circ(\omega; \tau^{ref})])$. **b** The ω_3 -correction factor, $\exp[2\kappa(\omega; \tau')\omega l / c]$. **c** The ω_1 -correction factor, $[K(\omega; \tau')]^{-1}$.

correction factors have a significant contribution from the broad solvent HOD bend, and the ω_1 -correction factor is ~ 5 x smaller than the ω_3 -correction. Due to the gradual nature of depletion of beams 1 and 2, the ω_1 -correction factor is expected to be smaller than the ω_3 -correction, which results from depletion of beam 3, the signal, and the LO over the entire length of the sample. The sharp features in Figure 5.4a are the result of transient absorption due to the solute, while the broad baseline transient change is due to the solvent. Corrections for both the linear absorbance of the solute and solvent are necessary to reveal the underlying nonlinear signal.

Figure 5.3g and h show the 2D IR spectra from equilibrium experiments of diglycine at a much lower concentration of 1 mg/mL, where the effects of beam absorption by the solute should be negligible given the low ~ 0.05 OD of the diglycine. The equilibrium spectrum (Figure 5.3g) is more similar to the ω_1 -and- ω_3 -corrected spectrum than it is to the uncorrected spectrum, as is evident in the peak positions and lineshapes. The equilibrium carboxylate peak is weak because the beam spectrum was weaker on the red side on the day these data were collected. The

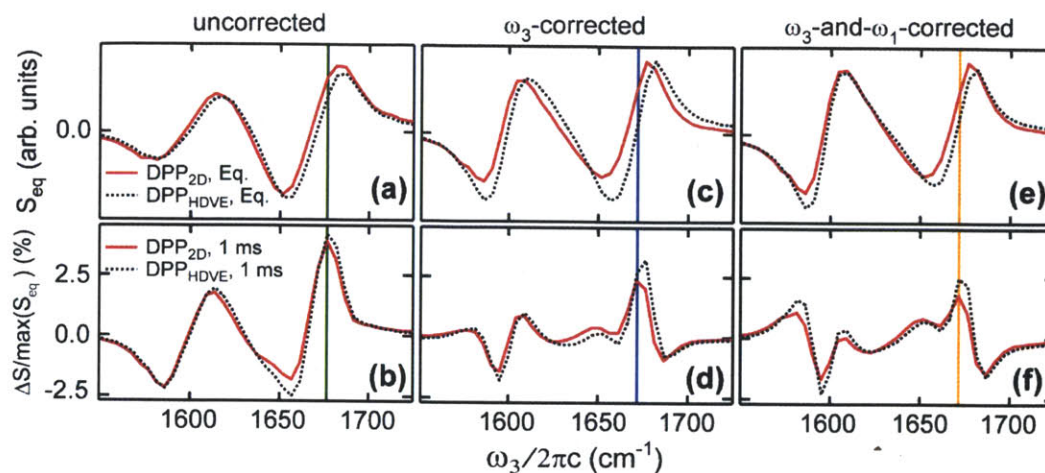


Figure 5.5: For 20 mg/mL diglycine, HDVE-derived DPP slices are compared to projections of the real 2D IR spectrum onto ω_3 . Spectra are shown at equilibrium (first row) and at 1 ms delay time (second row). The three columns represent uncorrected (**a,b**), ω_3 -corrected (**c,d**), and ω_3 -and- ω_1 -corrected (**e,f**) data analogous to the data shown in Figure 5.3. The vertical lines indicate the maximum of the amide I' peak difference spectrum, whose traces are shown in Figure 5.6.

30°C - 20°C equilibrium difference spectrum (Figure 5.3h) is most similar to the ω_1 -and- ω_3 -corrected transient difference spectrum, which suggests that the full correction returns the least distorted representation of the spectrum. Differences between the Figure 5.3h and f spectra may be due to noise, slight phasing errors, uncorrected solvent transmission changes in Figure 5.3h, and possible deviations from linear response in the T-jump case. Due to the experimentally fixed relative phasing and back-to-back collection of the transient spectra, we find that collecting T-dependent data transiently provides better signal-to-noise than collecting in equilibrium.

To understand the effects of linear absorption correction on the HDVE spectra, projections of the 2D surfaces onto ω_3 are compared to DPP spectra obtained from HDVE measurements in Figure 5.5. Theoretically, the projection of the absorptive (real) 2D spectrum onto ω_3 is equal to the DPP. There is good agreement between the 2D projections and DPP. This is particularly interesting in Figure 5.5e and f, where agreement indicates that the assumptions made for eq. (5.16) are reasonable for this case, where diagonal features dominate the 2D spectrum. Just as in the case for the 2D surfaces, corrections to account for linear absorption cause significant changes in the difference spectra, where Figure 5.5b only contains a few peaks compared to Figure 5.5d and f. Differences between Figure 5.5c(d) and e(f) are less pronounced, and again manifest themselves as reweighting of the carboxylate peaks.

Corrections for linear absorption are necessary for properly interpreting difference spectral features. As is shown in Figure 5.6, the timescales observed in the DPP of diglycine change negligibly with correction because both the timescales of diglycine's T-jump response and the timescales of the transient absorption follow the temperature. The presence of spurious peaks in the uncorrected data, however, changes the molecular interpretation. Previous mid-IR T-jump experiments have not explicitly corrected for the linear transient absorption. Instead, previous work has attempted to differentiate between these transmission changes and temperature-induced sample changes by identifying temperature dependent spectral features or timescale differences between the observed signal and the temperature's time profile^{22, 25, 38}. We aim to correct for as much of the transient absorption distortions as possible using reasonable corrections to avoid having to differentiate the desired signal from the transient absorption effects.

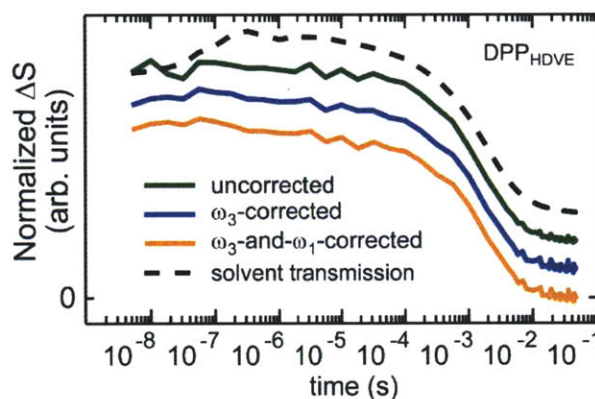


Figure 5.6: Slices taken from the DPP spectra shown in Figure 5.5 at 1677, 1672, and 1672 cm^{-1} , respectively. Traces are offset to allow comparison. The dashed line shows the solvent transmission.

Diglycine: Interpretation of Transient Results

Once the full correction is applied, the ‘truest’ transient and equilibrium spectra are given in Figure 5.3e and f. The carboxylate peak (lower left corner) of the ω_3 -and- ω_1 -corrected difference spectrum (Figure 5.3f) shows loss on the red side (marked with the vertical line) and a slight gain on the blue side. The amide I' peak difference spectrum (upper right corner) appears to be the result of two loss features (marked with vertical lines), but the observed doublet can be more consistently described by loss of the equilibrium peak and gain of a narrower center peak, which results in the low amplitude, broad off-diagonal gain intensity. The dominance of loss

features over gain is consistent with the transient absorption spectrum, and may be the result of non-Condon effects, like those observed in temperature dependent FTIR studies of NMA^{39, 40}, or shifting of oscillator strength to other coupled vibrational modes that lie outside of our spectral window.

The time-dependence of frequency slices from the HDVE-derived DPP are shown in Figure 5.6. Diglycine's spectral response tracks the solvent temperature profile, indicating that the observed changes happen on the sub-pulse width limited, <10 ns timescale, which causes the 'immediate' jump away from equilibrium. Due to this fast re-equilibration time, the diglycine sample is always equilibrated to the temperature of the solvent. The ~100 ns rise in the transient transmission spectrum shown in Figure 5.6 is thought to be the result of cavitation in the sample²³. This rise is largely removed from the nonlinear signal through linear absorption corrections.

Based on the fast response and the localized nature of the vibrations, we assign the difference peaks in Figure 5.3f to a decrease in the average number of H-bonds. The addition of T-jump energy causes an increase in the magnitude of the fluctuations both in the solvent and in the peptide. These fluctuations decrease the average number of H-bonds and cause a slight blue shift of the peak, which is observed as a loss on the red side and gain on the blue. However, a simple frequency shift cannot describe all changes in the amide I' peak, such as the dominance of loss without equivalent gain in this spectral region. Because the amide I' frequency, linewidth, and transition dipole strength reflect the population of hydrogen bonding configurations, the inhomogeneous nature of the equilibrium peak indicates that there are multiple H-bonding configurations. The temperature change causes population to shift between these configurations, and the resulting interference features result from the shifted frequency and decreased transition dipole as the average number of H-bonds is decreased at higher temperature. Previous work on NMA has indicated that T-dependent amide I' FTIR peak shifts are explained by electrostatic changes that can be linked to fewer H-bonds due to the increased kinetic energy⁴⁰. We expect similar changes for the amide I' peak of diglycine, although the terminal NH_3^+ group adds further sensitivity to the H-bonds at the N-terminus.

The transient results are compared to equilibrium temperature-dependent changes in Appendix 5.A.

Regardless of the exact assignments, diglycine allows us to set a baseline for the simplest

response that we can expect for a sample where we expect the primary response to T-jump to be picosecond solvation, including hydrogen bonding changes with the solvent. The t-2D IR also provides an explanation and a spectrum for the ‘burst phase’ signal that other transient experiments have observed⁴¹. For larger, more complicated peptides, we expect to see a similar pulse-width limited response. Deviations from this T-jump tracking signal indicate conformational changes.

5.5.2 Trpzip2

The t-HDVE spectra of TZ2 are shown in the top row of Figure 5.7. The t-HDVE data is displayed in two forms: as the real spectra(DPP) (Figure 5.7a) and the absolute value squared DVE spectra (Figure 5.7b). Due to the presence of both negative and positive peaks, negative and positive features in the DPP difference spectrum can be the result of loss or gain. The DPP difference spectra show three major spectral features: a positive(loss compared to equilibrium) peak at 1612 cm^{-1} , a negative(loss) peak at 1636 cm^{-1} , and a positive(gain) feature that appears to shift with time from 1655 to 1670 cm^{-1} . The DVE difference spectra, whose equilibrium spectrum is entirely positive, have two major features: a negative(loss) feature that shifts from 1621 to 1626 cm^{-1} and a positive(gain) peak that shifts from 1650 to 1665 cm^{-1} . All three DPP and both DVE peaks are immediately present at 10 ns , grow in on the 1.1 - $1.6\text{ }\mu\text{s}$ timescale, and then decay away following the temperature relaxation, as is shown by the normalized time traces in Figure 5.8. The μs timescale is consistent with previous T-jump experiments on TZ2^{31, 42} We interpret the DPP 1612 and 1636 cm^{-1} and the DVE 1626 cm^{-1} loss features as loss of the most

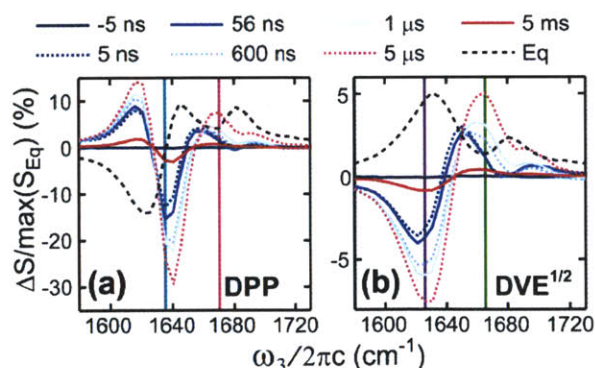


Figure 5.7: TZ2 transient spectra from t-HDVE measurements. **a,b** DPP and DVE spectra are shown at selected delays. The DVE spectra are actually presented as the square root of the DVE power spectrum (the square root is taken before subtraction gives the difference spectra) to remove the apparent factor of two rate of population change between the absolute-value-squared measurement and the DPP, which is linear in concentration¹.

intense equilibrium spectrum feature, the peak at 1636 cm^{-1} . Shifting of the higher frequency peaks in both *difference* data representations indicates the appearance of at least two peaks, which is interpreted as evidence of at least two species that appear concomitant with loss of the

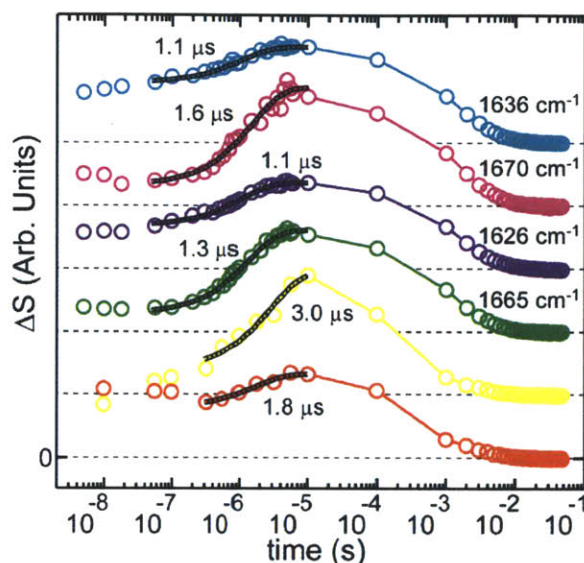


Figure 5.8: The TZ2 time dependent trace and exponential fit. The top two traces are from the DPP (Figure 5.7a), the middle two traces are from the square root of the DVE (Figure 5.7b), and the bottom two traces are from integrated boxes shown in Figure 5.9a.

equilibrium population; one species appears at $<10\text{ ns}$ followed by at least one other on the 1-1.6 μs exponential timescale.

The real and absolute value t-2D IR spectra are shown in Figure 5.9. The projection of the complex difference t-2D IR are equivalent to the complex 1D difference t-HDVE data, which means we can directly compare the DPP to the projection of the real t-2D IR onto the ω_3 axis. As with the t-HDVE data, there are changes in the t-2D IR at the earliest accessible times. The 56 and 100 ns difference t-2D IR are dominated by loss along the diagonal, in particular loss of the two peaks at 1636 and 1670 cm^{-1} , whose presence is indicative of anti-parallel β -strand structure⁴³. Aside from the loss, there are also two peaks indicating gain: the broad red/yellow peak above the diagonal and the blue ridge below the diagonal that extends along the $\omega_3=1636\text{ cm}^{-1}$ horizontal line. These peaks are particularly clear in the absolute value difference spectra (Figure 5.9f and g), where red(blue) is gain(loss). The loss and gain features are indicative of broadening into the anti-diagonal, or homogeneous, dimension, which is consistent with

increased flexibility and decreased numbers of H-bonds to the solvent, two effects that are expected with increased temperature. In the absence of our linear absorption corrections, contributions from direct solvent changes could not have been definitely ruled out of this assignment. As in the case with diglycine, we expect that these <10 ns changes in the spectrum are due to small amplitude conformational changes and response of the solvent interacting with the peptide, not large amplitude peptide motion. Although H-bonds across the backbone might be weakened, we do not expect that they are entirely lost.

Comparison of the 100 ns, 1 μ s, and 5.6 μ s spectra t-2D IR spectra reveals that peaks shift on the μ s timescale. In the real t-2D IR spectra, the positive(gain) peak in the upper diagonal shifts up and increases in size and intensity, which causes the positive 1665 cm^{-1} peak in the DPP spectra. The negative(gain) ridge along $\omega_3=1636 \text{ cm}^{-1}$ also changes; it shifts up, broadens, and increases in intensity. These changes are reflected in the absolute value spectra where gain is stronger than loss along the diagonal in the 1660 cm^{-1} region at 5.6 μ s. Based on the empirical correlations between amide I frequency and protein secondary structure⁴⁴, the appearance of intensity in the 1660 cm^{-1} region suggests an increase of disordered chains. Inspection of the integrated area shown in Figure 5.9a (Figure 5.8, bottom two traces) shows that peaks in the real t-2D IR also relax on the μ s timescale, although the t-2D IR data is much noisier than the t-HDVE. The μ s timescale is consistent with unfolding timescales, and we attribute the changes in this timeframe to loss of the native β -hairpin structure.

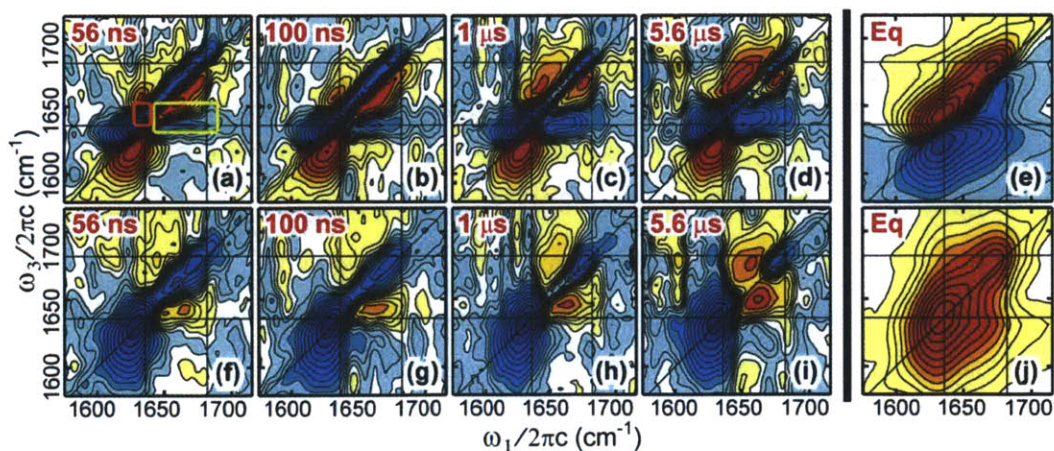


Figure 5.9: a-e The absorptive (real) correlation t-2D IR spectra at selected time delays for TZ2. f-j The t-2D IR square-root power spectra at selected time delays. The spectra are individually normalized and scaled such that 25 linearly-spaced contours span the entire magnitude range of $\sinh^{-1}(S_{Norm.} \times 27.29)$.

The t-HDVE and t-2D IR spectra and timescales can be explained by a three state kinetic model based on the presence of just two distinguishable timescales that are distinct from the thermal relaxation profile. Although there may be more states that are hidden within the noise, share the same spectral and temporal response as those described, or do not map well onto the H-bond sensitive amide I vibration, we can only distinguish relaxation rates of <10 ns and 1-1.6 μ s. As in the case with diglycine, we hypothesize that the <10 ns response is due to H-bond changes resulting from increased local fluctuations due to the increased temperature, which causes a loss of the equilibrium spectrum and the gain of a broader, two peaked spectrum. On the μ s timescale we see a shift in intensity to the 1660 cm^{-1} region, which indicates thermal disordering and unfolding of the backbone.

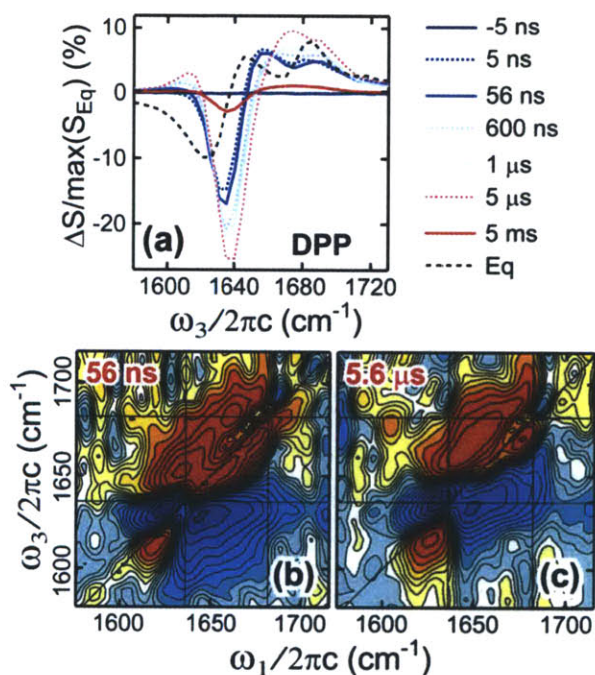


Figure 5.10: TZ2 uncorrected ($S^{1,3}$, eq. (5.5)) difference DPP (a) and difference 2D IR (b,c) spectra, which can be compared to the corrected data in Figure 5.7a and Figure 5.9 a and d. The 2D contours are plotted as described in Figure 5.9.

Due to the low TZ2 concentration and ~ 0.05 OD, the majority of the transient linear absorption distortion is due to the solvent. The temperature profile is relatively flat during the 100 ns – 10 μ s time frame, and the distorting linear absorption is not expected to significantly affect the observed relaxation timescales. The relaxation measured at 1636 and 1670 cm^{-1} from the uncorrected DPP can be fit to an exponential relaxation with time constants of 0.9 and 1.3 μ s,

which are similar to the 1.1 and 1.6 μs measured in the ω_3 -and- ω_1 -corrected DPP traces shown in Figure 5.8. The main differences between the corrected and uncorrected TZ2 data are in the spectra, not the timescales. The uncorrected TZ2 transient DPP and transient 2D IR spectra are shown in Figure 5.10. In the uncorrected DPP, there are more intense changes on the blue side and less on the red relative to the corrected DPP (Figure 5.7). The uncorrected transient 2D IR spectra show additional strong red and blue peaks above and below the diagonal, respectively, which obscure the off diagonal gain features as well as much of the diagonal loss features discussed above and shown in Figure 5.9. As discussed before, the linear absorption distortions primarily add equilibrium spectral features into the transient spectra. This is consistent with the uncorrected DPP and 2D IR transient spectra, which appear to contain the undistorted features from Figure 5.7 and Figure 5.9 masked by broad equilibrium peak features. Although the observed timescales are the same within fitting error, the uncorrected spectra contain different peaks and different relative amplitudes which may have affected the spectral interpretation. Even in this case with a low OD solute, we find that the linear absorption correction is necessary to properly interpret the data.

5.6 Conclusion

Transient multidimensional spectroscopy is an attractive experimental technique which combines the rich information content of multidimensional spectroscopies with the ability to access arbitrarily long timescales. The increase in complexity inherent to transient experiments, however, makes it hard to both cleanly measure the desired signal and to interpret the results once collected. Through linear absorption corrections, we have shown a method for simplifying transient nonlinear spectra by removing distortions that can potentially lead to false conclusions. These linear absorption corrections are general and can be applied to any transient 2D or DPP spectroscopy, including electronic spectroscopies, particularly in cases with a large transient absorption response. When applied to T-jump t-2D IR experiments on diglycine, these corrections reveal that amide I spectroscopy is sensitive to <10 ns changes in local H-bonds, a spectral signature that is relevant to all T-jump spectroscopies. When applied to the peptide TZ2, T-jump t-2D IR indicates at least three kinetic states: on the <10 ns timescale, H-bonds are strained and the β -hairpin begins to loosen and on the 1.1 μs timescale, the peptide disorders. The spectral analysis is enriched by the increased information content provided by t-2D IR spectroscopy. The combination of t-HDVE and t-2D IR spectroscopy provides both high time

resolution and detailed spectra that allow us to interpret congested spectra.

5.7 Acknowledgment

We thank Carlos Baiz and Andrew D. Horning for helpful comments. This work was supported by the National Science Foundation (Grants CHE-0616575 and CHE-0911107).

5.8 References

- (1) C. R. Baiz, R. McCanne, and K. J. Kubarych, "Transient Vibrational Echo versus Transient Absorption Spectroscopy: A Direct Experimental and Theoretical Comparison," *Appl. Spectrosc.* **64**, 1037-1044 (2010).
- (2) S. Woutersen and P. Hamm, "Structure determination of trialanine in water using polarization sensitive two-dimensional vibrational spectroscopy," *J. Phys. Chem. B* **104**, 11316-11320 (2000).
- (3) A. Remorino, I. V. Korendovych, Y. Wu, W. F. DeGrado, and R. M. Hochstrasser, "Residue-Specific Vibrational Echoes Yield 3D Structures of a Transmembrane Helix Dimer," *Science* **332**, 1206-1209 (2011).
- (4) C. J. Fecko, J. D. Eaves, J. J. Loparo, A. Tokmakoff, and P. L. Geissler, "Ultrafast Hydrogen-Bond Dynamics in the Infrared Spectroscopy of Water," *Science* **301**, 1698-1698 (2003).
- (5) H. Ishikawa, K. Kwak, J. K. Chung, S. Kim, and M. D. Fayer, "Direct observation of fast protein conformational switching," *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8619-8624 (2008).
- (6) J. F. Cahoon, K. R. Sawyer, J. P. Schlegel, and C. B. Harris, "Determining Transition-State Geometries in Liquids Using 2D-IR," *Science* **319**, 1820-1823 (2008).
- (7) J. Zheng, K. Kwak, J. Asbury, X. Chen, I. R. Piletic, and M. D. Fayer, "Ultrafast dynamics of solute-solvent complexation observed at thermal equilibrium in real time," *Science* **309**, 1338-1343 (2005).
- (8) Z. Ganim, H. S. Chung, A. W. Smith, L. P. DeFlores, K. C. Jones, and A. Tokmakoff, "Amide I Two-Dimensional Infrared Spectroscopy of Proteins," *Acc. Chem. Res.* **41**, 432-441 (2008).
- (9) A. T. Krummel, P. Mukherjee, and M. T. Zanni, "Inter and intrastrand vibrational coupling in DNA studied with heterodyned 2D-IR spectroscopy," *J. Phys. Chem. B* **107**, 9165-9169 (2003).
- (10) T. Brixner, J. Stenger, H. M. Vaswani, M. Cho, R. E. Blankenship, and G. R. Fleming, "Two-dimensional spectroscopy of electronic couplings in photosynthesis," *Nature* **434**, 625-628 (2005).
- (11) M. Khalil, N. Demirdoven, and A. Tokmakoff, "Coherent 2D IR Spectroscopy: Molecular Structure and Dynamics in Solution," *J. Phys. Chem. A* **107**, 5258-5279 (2003).
- (12) 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 (2009).
- (13) J. N. Bandaria, S. Dutta, M. W. Nydegger, W. Rock, A. Kohen, and C. M. Cheatum, "Characterizing the dynamics of functionally relevant complexes of formate dehydrogenase," *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17974-17979 (2010).
- (14) C. Fang, J. D. Bauman, K. Das, A. Remorino, E. Arnold, and R. M. Hochstrasser, "Two-dimensional infrared spectra reveal relaxation of the nonnucleoside inhibitor TMC278 complexed with HIV-1 reverse transcriptase," *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1472-1477 (2008).
- (15) J. M. Anna, M. R. Ross, and K. J. Kubarych, "Dissecting Enthalpic and Entropic Barriers to Ultrafast Equilibrium Isomerization of a Flexible Molecule Using 2DIR Chemical

- Exchange Spectroscopy," *J. Phys. Chem. A* **113**, 6544-6547 (2009).
- (16) J. Bredenbeck, J. Helbing, R. Behrendt, C. Renner, L. Moroder, J. Wachtveitl, and P. Hamm, "Transient 2D-IR spectroscopy: Snapshots of the nonequilibrium ensemble during the picosecond conformational transition of a small peptide.," *J. Phys. Chem. B* **107**, 8654-8660 (2003).
- (17) 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**, 990-991 (2004).
- (18) C. Kolano, J. Helbing, M. Kozinski, W. Sander, and P. Hamm, "Watching hydrogen-bond dynamics in a β -turn by transient two-dimensional infrared spectroscopy," *Nature* **444**, 469-472 (2006).
- (19) R. Kania, A. I. Stewart, I. P. Clark, G. M. Greetham, A. W. Parker, M. Towrie, and N. T. Hunt, "Investigating the vibrational dynamics of a 17e- metalcarbonyl intermediate using ultrafast two dimensional infrared spectroscopy," *Phys. Chem. Chem. Phys.* **12**, 1051-1063 (2010).
- (20) C. R. Baiz, M. J. Nee, R. McCanne, and K. J. Kubarych, "Ultrafast nonequilibrium Fourier-transform two-dimensional infrared spectroscopy," *Opt. Lett.* **33**, 2533-2535 (2008).
- (21) W. Xiong, J. E. Laaser, P. Paoprasert, R. A. Franking, R. J. Hamers, P. Gopalan, and M. T. Zanni, "Transient 2D IR Spectroscopy of Charge Injection in Dye-Sensitized Nanocrystalline Thin Films," *J. Am. Chem. Soc.* **131**, 18040-18041 (2009).
- (22) H. S. Chung, Z. Ganim, K. C. Jones, and A. Tokmakoff, "Transient 2D IR spectroscopy of ubiquitin unfolding dynamics," *Proc. Natl. Acad. Sci. U. S. A.*, 14237-14242 (2007).
- (23) H. S. Chung, M. Khalil, A. W. Smith, and A. Tokmakoff, "Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics," *Rev. Sci. Instrum.* **78**, 063101 (2007).
- (24) E. R. Andresen and P. Hamm, "Site-specific difference 2D-IR spectroscopy of bacteriorhodopsin," *J. Phys. Chem. B* **113**, 6520-6527 (2009).
- (25) H. S. Chung, M. Khalil, A. W. Smith, Z. Ganim, and A. Tokmakoff, "Conformational changes during the nanosecond to millisecond unfolding of ubiquitin," *Proc. Natl. Acad. Sci. U. S. A.* **102**, 612-617 (2005).
- (26) K. C. Jones, Z. Ganim, and A. Tokmakoff, "Heterodyne-detected dispersed vibrational echo spectroscopy," *J. Phys. Chem. A* **113**, 14060-14066 (2009).
- (27) Although the transient absorption signal is formally 3rd order, here we take a phenomenological approach in which the T-jump pulse is not considered in the nonlinear spectral response.
- (28) J. Wang and M. A. El-Sayed, "Temperature jump-induced secondary structural change of the membrane protein bacteriorhodopsin in the premelting temperature region: a nanosecond time-resolved fourier transform infrared study," *Biophys. J.* **76**, 2777-2783 (1999).
- (29) S. Williams, T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff, and R. B. Dyer, "Fast Events in Protein Folding: Helix Melting and Formation in a Small Peptide," *Biochemistry* **35**, 691-697 (1996).
- (30) 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).

- (31) K. Hauser, C. Krejtschi, R. Huang, L. Wu, and T. A. Keiderling, "Site-Specific Relaxation Kinetics of a Tryptophan Zipper Hairpin Peptide Using Temperature-Jump IR Spectroscopy and Isotopic Labeling," *J. Am. Chem. Soc.* **130**, 2984-2992 (2008).
- (32) H. Ma, J. Ervin, and M. Gruebele, "Single-sweep detection of relaxation kinetics by submicrosecond midinfrared spectroscopy," *Rev. Sci. Instrum.* **75**, 486-491 (2004).
- (33) A. W. Smith, J. Lessing, Z. Ganim, C. S. Peng, A. Tokmakoff, S. Roy, T. L. C. Jansen, and J. Knoester, "Melting of a β -Hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations," *J. Phys. Chem. B* **114**, 10913-10924 (2010).
- (34) J. D. Hybl, A. Albrecht Ferro, and D. M. Jonas, "Two-Dimensional Fourier Transform Electronic Spectroscopy," *J. Chem. Phys.* **115**, 6606-6622 (2001).
- (35) N. Belabas and D. M. Jonas, "Three-dimensional view of signal propagation in femtosecond four-wave mixing with application to the boxcars geometry," *J. Opt. Soc. Am. B* **22**, 655-674 (2005).
- (36) M. K. Yetzbacher, N. Belabas, K. A. Kitney, and D. M. Jonas, "Propagation, beam geometry, and detection distortions of peak shapes in two-dimensional Fourier transform spectra," *J. Chem. Phys.* **126**(2007).
- (37) S. Ham, J.-H. Kim, H. Lee, and M. Cho, "Correlation between electronic and molecular structure distortions and vibrational properties. II. Amide I modes of NMA-nD₂O complexes.," *J. Chem. Phys.* **118**, 3491-3498 (2003).
- (38) Adam W. Smith and A. Tokmakoff, "Probing Local Structural Events in β -Hairpin Unfolding with Transient Nonlinear Infrared Spectroscopy," *Angew. Chem., Int. Ed. Engl.* **46**, 7984-7987 (2007).
- (39) K. E. Amunson and J. Kubelka, "On the Temperature Dependence of Amide I Frequencies of Peptides in Solution," *J. Phys. Chem. B* **111**, 9993-9998 (2007).
- (40) J. Kaminsky, P. Bour, and J. Kubelka, "Simulations of the Temperature Dependence of the Amide I Vibration," *J. Phys. Chem. A* **115**, 30-34 (2011).
- (41) R. M. Ballew, J. Sabelko, and M. Gruebele, "Direct observation of fast protein folding: The initial collapse of apomyoglobin," *Proc. Natl. Acad. Sci. USA* **93**, 5759 (1996).
- (42) C. D. Snow, L. Qiu, D. Du, F. Gai, S. J. Hagen, and V. S. Pande, "Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy," *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4077-4082 (2004).
- (43) N. Demirdöven, C. M. Cheatum, H. S. Chung, M. Khalil, J. Knoester, and A. Tokmakoff, "Two-dimensional infrared spectroscopy of antiparallel β -sheet secondary structure," *J. Am. Chem. Soc.* **126**, 7981-7990 (2004).
- (44) A. Barth and C. Zscherp, "What vibrations tell us about proteins," *Q. Rev. Biophys* **35**, 369-430 (2002).

Appendix 5.A: Equilibrium Temperature Dependence

The changes observed in the t-2D IR spectra of diglycine agree with the changes observed in the temperature dependent equilibrium spectra. To separate these changes, SVD analysis was applied to the temperature-dependent equilibrium spectra of diglycine. The results of SVD are shown on the top row of Figure 5.A.1. SVD is discussed in detail in Chapter 6. SVD reveals that the ratio of the first three singular values is 84.7:8.89:2.75. Based on these ratios, the equilibrium spectra are fairly well described by just the first two spectra whose temperature dependence is shown at the top right of Figure 5.A.1. The U_1 is approximately the average spectrum over the temperature range, and its magnitude is fairly temperature independent. In this case, where we assume the 3rd and above components are noise, the U_2 spectrum represents the temperature dependent changes. The linearity of the V_2 temperature dependence and the two

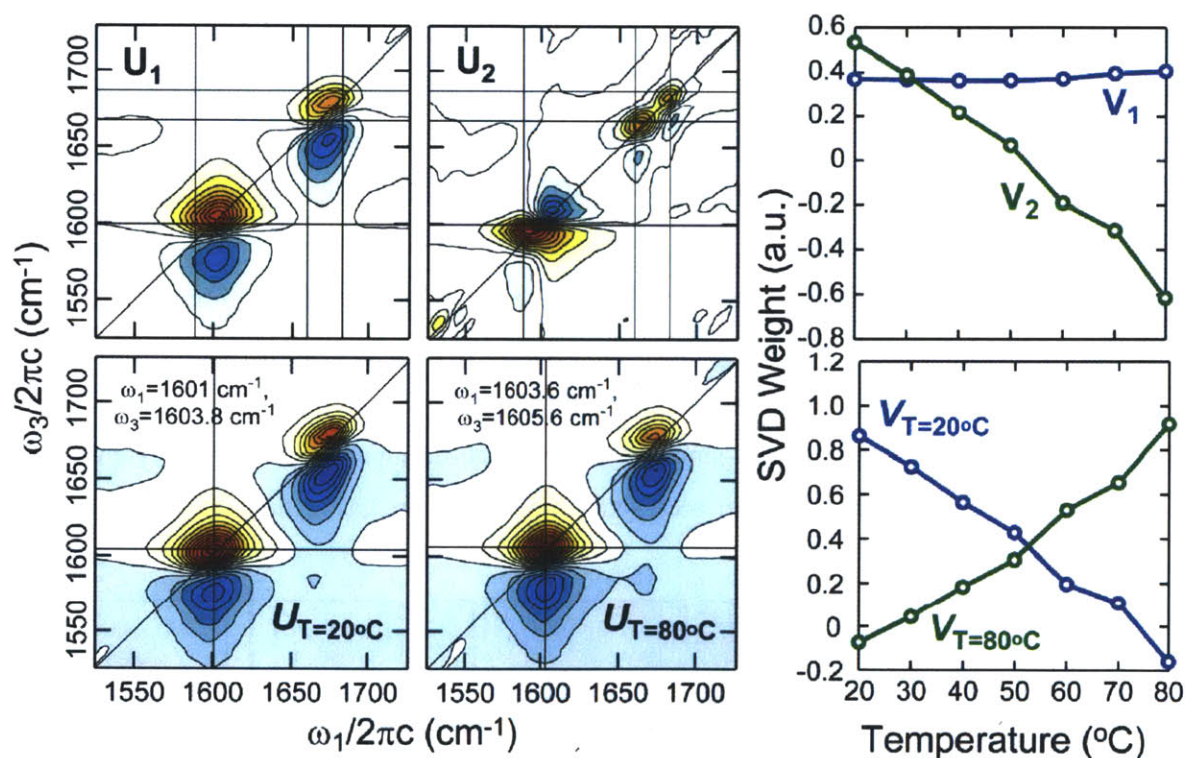


Figure 5.A.1: *Top row* – The first two SVD component spectra and temperature traces (*right*) that result from SVD analysis of the equilibrium, temperature dependent absorptive (real) 2D IR spectra of diglycine. Grid lines are consistent with Figure 5.3. Spectra are ω_1 -and- ω_3 -corrected and plotted on 25 evenly spaced intervals. *Bottom Row* – The predicted 20 and 80°C spectra that result from re-weighting the $U_{1,2}, V_{1,2}$ mathematical results. Grid lines are plotted at the frequencies indicated and defined by the peak of the carboxylate peak fundamental.

component dominance of the SVD analysis are consistent with the presence of two states – a low temperature and high temperature population with distinct spectra. Under this interpretation, raising the temperature leads to shifting of the relative populations of these two states, which causes the observed apparent blue-shifting of frequencies. This picture is consistent with the conclusions from the T-jump experiments, where two states were identified for diglycine – an initial state and a thermally excited state with a net drop in H-bonds to the solvent. Re-casting the SVD mathematical results such that the temperature dependence best approximates a linear dependence that shifts from 1 to 0 (or 0 to 1) over the 20 – 80°C temperature range results in the basis spectra shown on the bottom row of Figure 5.A.1. These spectra represent our predictions for the low (20°C) and high (80°C) temperature state spectra. As expected, the carboxylate spectrum at higher temperature is blueshifted relative to the low temperature predicted spectrum.

Comparison of U_2 to the 1 ms t-2D IR GG spectrum in Figure 5.3f shows the same spectral features, although the frequencies, relative peak amplitudes, and sign differ. A strict comparison of the equilibrium and transient changes would allow for assessing whether the non-equilibrium and equilibrium processes differ. Unfortunately, there are too many other reasons that can explain the observed differences: day-to-day spectral differences, a breakdown in our two-state approximation, and phasing errors. Based on the qualitative agreement between the transient and equilibrium difference spectra, we continue to assume that the equilibrium and non-equilibrium accessed states are the same. Any differences between the equilibrium and equivalent temperature non-equilibrium spectra are expected because the populations have not had enough time to adjust to the T-jump in the non-equilibrium case. For diglycine, the timescales for adjusting to the T-jump accessed temperatures are faster than the pulse-width,

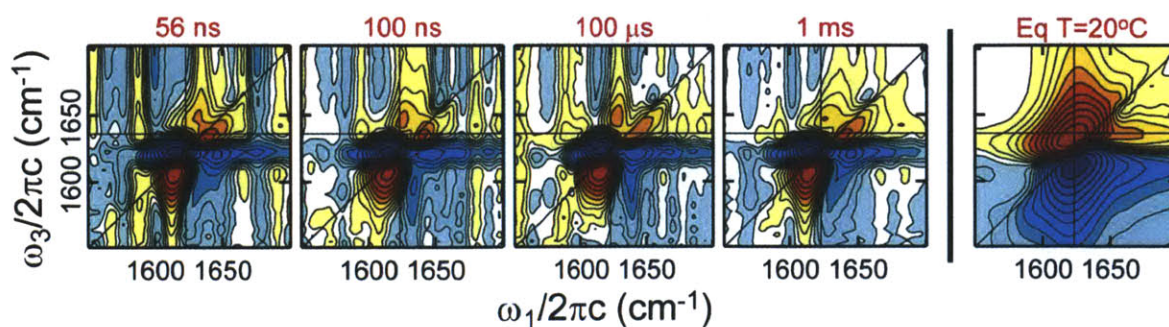


Figure 5.A.2: The ω_3 -corrected t-2D IR spectra of 1% NMA in D_2O . The 2D contours are plotted as described in Figure 5.9.

suggesting that the populations should always be equilibrated to the instantaneous temperature, and we expect (and observe here) that the equilibrium and t-2D IR spectra match.

T-jump experiments were also performed at $T_i=20^\circ\text{C}$ for 1% (by volume) N-methylacetamide (NMA) in D_2O . The t-2D IR spectra are shown in Figure 5.A.2. Unfortunately, before collecting this data, the sample was allowed to sit in the sample cell for over 24 hours at an elevated temperature, and there appears to be contamination that results in a blue-shifted shoulder in the spectrum (perhaps scatter or protonation). The t-2D IR results are consistent with a homogeneous peak blue-shifting.

As with diglycine, the transient spectral changes can be compared to the equilibrium changes identified by SVD analysis, which are shown in Figure 5.A.3. For the NMA temperature dependent equilibrium data, the first three singular values have a ratio of

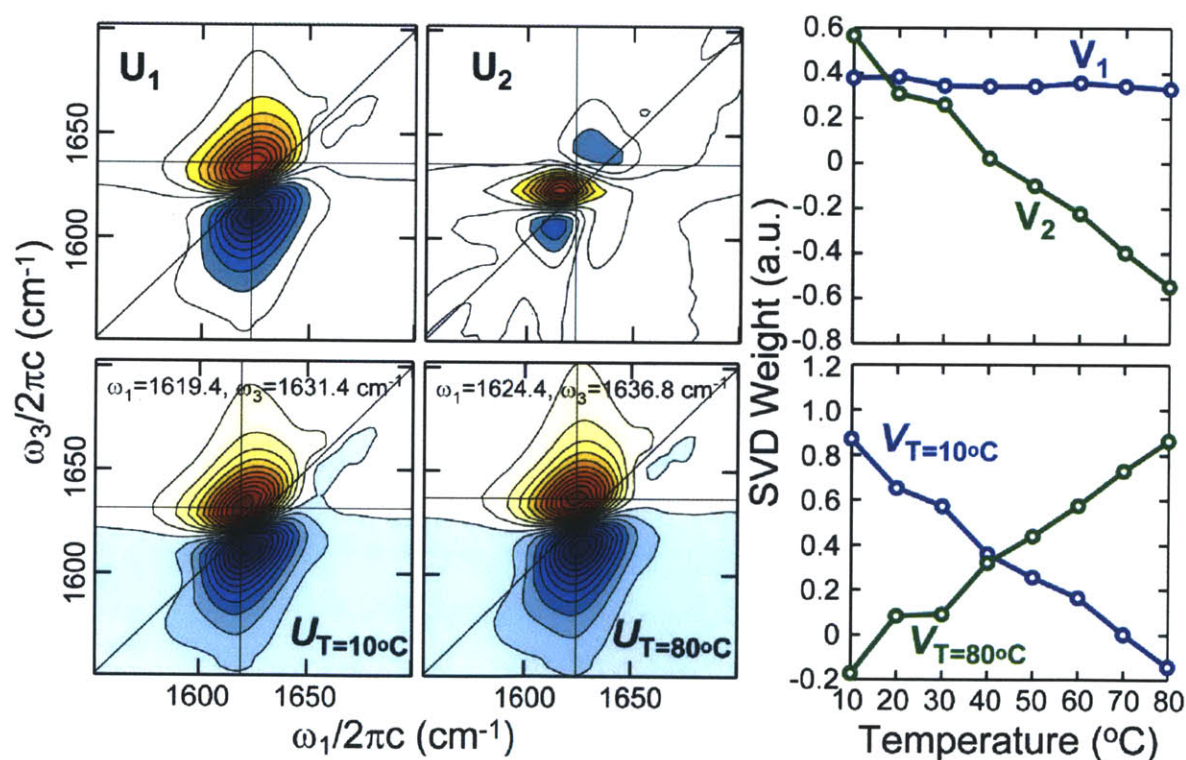


Figure 5.A.3: *Top row* – The first two SVD component spectra and temperature traces (*right*) that result from SVD analysis of the equilibrium, temperature dependent absorptive (real) 2D IR spectra of NMA. Spectra are ω_3 -corrected and plotted on 25 evenly spaced intervals. *Bottom Row* – The predicted 10 and 80°C spectra that result from re-weighting the $U_{1,2}, V_{1,2}$ mathematical results. Grid lines are plotted at the frequencies indicated and defined by the NMA fundamental peak.

81.4:10.6:2.39, which again suggests a two state system. As with diglycine, the temperature dependence associated with the SVD spectra indicate two states whose populations reciprocate with a linear temperature dependence. Agreement between the transient spectra and the 2nd component SVD spectrum suggest the equivalence between the equilibrium and non-equilibrium changes. We again predict a high and low temperature NMA spectrum by re-weighting the SVD results. As with diglycine, the predicted temperature dependences do not fit exactly to our postulated $V_{T=10C} = 1 - (T-10)/70$. If the dependences did match, then we could directly hypothesize that the $U_{T=10C}$ and $U_{T=80C}$ spectra are the low and high temperature NMA spectra.

Chapter 6

Singular Value Decomposition

6.1 Introduction

Singular Value Decomposition (SVD) is a linear decomposition method used in statistics to analyze rectangular matrices for rank determination, least-squares fitting, and decomposition into basis components. SVD is related to eigenvalue decomposition, and the two are equivalent for special cases. Also, SVD and principal component analysis are equivalent, although SVD calculation is more robust¹⁻³. For physical chemists who are trained in thinking about Fourier transforms and orthogonal basis sets, SVD can be thought of as a transformation of a data set into orthogonal basis sets. Similarly, many of the associated concepts such as projections and transformations are applicable, although orthogonality may not be retained

In spectroscopy, SVD is used to analyze spectra for their dependence on an *independent variable*, γ . For example, γ may be secondary structure content (circular dichroism)⁴, pH or voltage (UV-vis)⁵, or time following photolysis (transient absorption⁶, time-resolved X-ray crystallography⁷). In this thesis, SVD is used to analyze the variation in a series of spectra collected as a function of either temperature (T), or time-delay following the T-jump (τ). The analysis allows us to 1) separate signal from noise, 2) determine the number of distinguishable molecular species, and 3) interpret the data using physically meaningful models. SVD has been used in previous Tokmakoff group publications for doorway state calculation^{8, 9}, melting curve determination⁹⁻¹⁶, removing noise in transient data^{14, 17}, and secondary structure determination¹⁸.

The number of SVD-determined basis spectra necessary to describe the data is related to the number of underlying distinguishable spectra.

In this chapter, we begin by describing the application of SVD to general spectroscopy data. In order to differentiate different steps in the analysis, it is useful to define three basis sets, all of which can be used to represent the same original data. The *mathematical* basis is the numerical result of the SVD, and it serves as the starting point for analysis. The end goal of analysis is to calculate the *state* basis, which represents the spectra associated with chemically distinct species. Lying between these two is the *intermediate* basis, a physically meaningful basis calculated with a minimum number of assumptions. The distinctions between these bases and the assumptions made to transform sequentially from the *mathematical* into the *intermediate* and then *state* bases are described below and illustrated with examples. Finally, SVD is compared to an analogous projection decomposition technique, and some shortcomings of SVD are described and illustrated with simple examples. A table describing the many variables in the following discussion is given in Appendix 6.A.

6.2 Methods and Discussion

6.2.1 SVD application and the state and mathematical bases

In spectroscopies where amplitude is linear in population, we assume that the measured spectrum of the sample is a linear combination of the spectra of individual species scaled by their populations. For example, in 2D IR, each non-interacting molecule linearly contributes to the signal electric field. If we try to analyze the 2D IR spectrum based on sub-molecular structures, such as β -sheet content for example, the linear relationship does not apply (or only approximately applies) because the spectrum does not solely depend on the secondary structure. Due to coupling and delocalization, the spectrum of a β -sheet will depend on its size and the surrounding environment. Thus, although each molecule might have a β -sheet, we cannot define a universally shared β -sheet spectrum. In our discussion of SVD, we assume the linear relationship between the amplitude of the spectrum associated with the species and the population of the species. Each non-interacting molecule (or group of molecules for dimers, for example) can be treated independently with its own spectrum.

Measuring the spectra, S , as a function of the general variable γ allows for determination of the γ -dependence. Although this γ may be temperature, pH, time, voltage, pressure, or any

other variable, the underlying assumption is that each variable is affecting the population. Written explicitly for a sample containing h species, S will be the sum of the spectra of each individual species i , $\sigma_i(\omega)$, scaled by the γ -dependence of each species' population, $\zeta_i(\gamma)$. Since we use this method to analyze experimental data, we also include an additional term for noise, Ξ :

$$S(\omega, \gamma) = \sum_{i=1}^h \sigma_i(\omega) \zeta_i(\gamma) + \Xi(\omega, \gamma) = \sum_{i=1}^h \sigma_i \zeta_i^T + \Xi(\omega, \gamma) \quad (6.1)$$

where σ_i and ζ_i are column vectors. Eq. (6.1) can also be expressed in matrix form as

$$S = \Sigma Z^T + \Xi \quad (6.2)$$

where Σ and Z are the *state basis* matrices containing spectra $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_h\}$ and populations $Z = \{\zeta_1, \zeta_2, \dots, \zeta_h\}$. Although the term population is intentionally used for generalization, later, we also use the standard notation $[i] = \zeta_i$ for the specific (although not necessarily rare) case in which concentration is directly proportional to population. These σ and ζ vectors represent the underlying spectra and γ -dependence of the molecular species in the sample. Although in reality conceptualizing the number of species h depends on how the non-interacting, independent molecules are binned into separate states, we do not limit ourselves to any definitions. Rather, we view h and Σ and Z hypothetically - h can range from the number of oscillators in the sample down to 1. The goal of SVD (and spectroscopy) is to determine the number of distinguishable spectra and assign them to species within the system.

SVD is an objective mathematical decomposition technique that separates the γ -dependent spectral data into the *mathematical basis*, which consists of three matrices, one that carries spectra, one that carries global scaling factors, and one that carries the γ -dependence:

$$S = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (6.3)$$

The equation is depicted in Figure 6.1 along with the size of the matrices.

\mathbf{U} contains columns of spectra $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_y\}$ that are orthonormal or zero. \mathbf{V} contains columns of γ -dependence $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_z\}$ that are either orthonormal or zero. The number y is defined as the maximum number of independent pairs of $\mathbf{u}_i/\mathbf{v}_i$ necessary for reproducing the original data set S . y is limited by the dimensions of the γ -dependent spectral data matrix: either the length of ω (the number of pixels) or the number of γ -dependencies measured (the number of

$$\begin{array}{c}
 \left(\begin{array}{c} \leftarrow S \rightarrow \\ \uparrow \omega \downarrow \end{array} \right) \\
 x \times z
 \end{array}
 =
 \begin{array}{c}
 \left(\begin{array}{c} \mathbf{u}_1 \ \mathbf{u}_2 \quad \dots \quad \mathbf{u}_y \\ \downarrow \ \downarrow \quad \dots \quad \downarrow \\ \ominus \ \ominus \quad \dots \quad \ominus \\ \downarrow \ \downarrow \quad \dots \quad \downarrow \end{array} \right) \\
 x \times y
 \end{array}
 \begin{array}{c}
 \left(\begin{array}{c} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \dots \\ \mathbf{s}_y \end{array} \right) \\
 y \times z
 \end{array}
 \begin{array}{c}
 \left[\begin{array}{c} \left(\begin{array}{c} \mathbf{v}_1 \ \mathbf{v}_2 \quad \dots \quad \mathbf{v}_z \\ \downarrow \ \downarrow \quad \dots \quad \downarrow \\ \gamma \ \gamma \quad \dots \quad \gamma \\ \downarrow \ \downarrow \quad \dots \quad \downarrow \end{array} \right) \right]^T \\
 [z \times z]^T
 \end{array}
 \end{array}$$

Figure 6.1: Depiction of the mathematical basis SVD terms shown in eq. (6.3). The off-diagonal terms of the singular value matrix S are 0. The matrix dimensions are shown below each matrix. $y = \min(x,z)$.

collected spectra). Although the number of columns in the matrix V can be $\geq y$ (in the case where y is determined by the size of U), any V components at $> y$ are effectively redundant in reconstructing the original data set. S is a rectangular matrix whose only nonzero values are arranged in decreasing order along the diagonal $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_y\}$. Each s value is called a *singular value*, and it can be thought of as a global scaling factor. The number of nonzero singular values determines the number of non-trivial components, where each component i has its associated \mathbf{u}_i spectrum, singular value s_i , and γ -dependence, \mathbf{v}_i . Due to the off-diagonal zero values in the singular matrix S , the γ -dependent data series S can be recreated by summing the product of each component set:

$$S(\omega, \gamma) = \sum_{i=1}^y \mathbf{u}_i(\omega) \mathbf{s}_i \mathbf{v}_i(\gamma) \tag{6.4}$$

This equation is depicted in Figure 6.2. A column of S (the spectrum at a particular γ , time in this example) is recalculated by first multiplying each component spectrum \mathbf{u}_i by the associated singular value s_i and the associated γ -dependent coefficient $\mathbf{v}_i(\gamma)$. Each component product is then summed to give the spectrum at a particular γ . The process can be thought of equivalently in reverse: a row in S (the γ -dependence at a given frequency) can be calculated by first multiplying each γ -dependent trace \mathbf{v}_i by the associated singular value s_i and the associated ω -dependent coefficient $\mathbf{u}_i(\omega)$. Each component product is then summed to give the γ -dependent trace at a particular ω .

The nonzero columns of the spectral component matrix U and the γ -dependent matrix V are orthonormal, and their magnitudes are limited. The importance of each of the i^{th} components

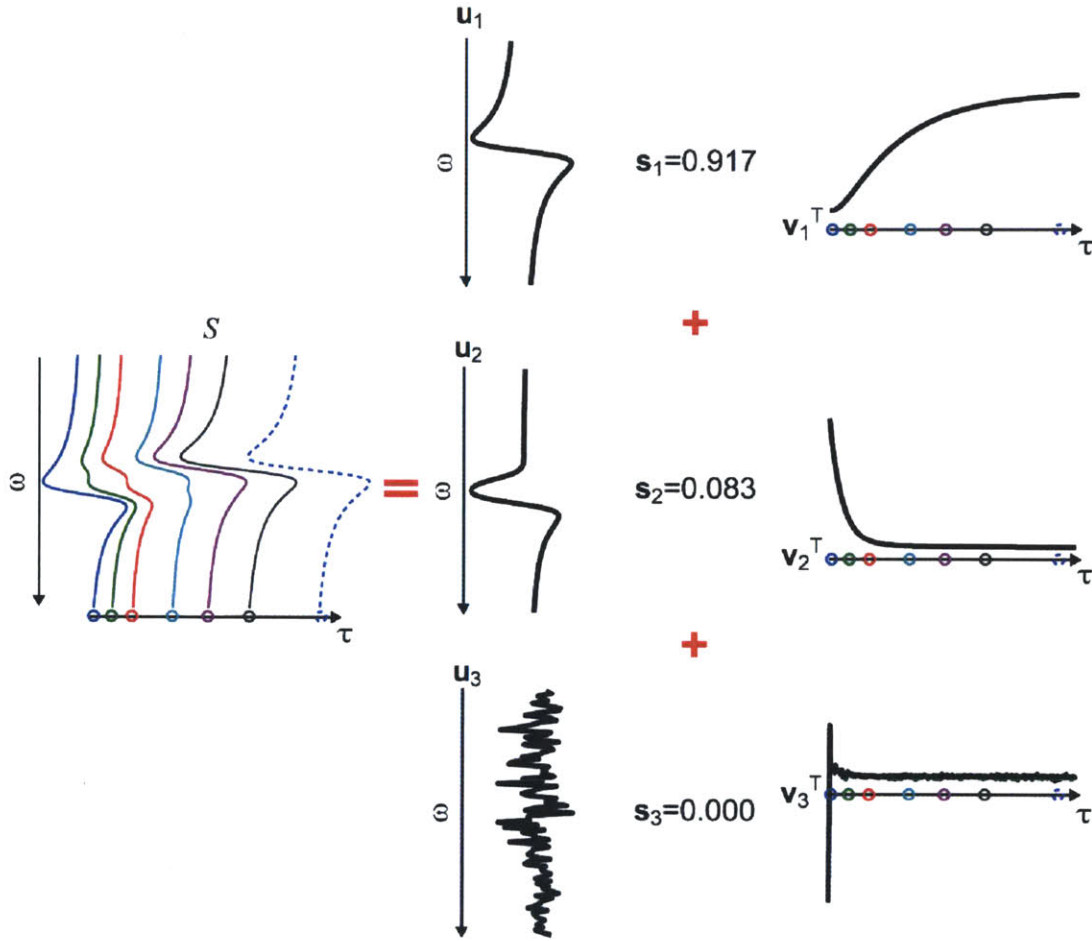


Figure 6.2: Eq. (6.3) is depicted using the SVD mathematical basis generated from the S from Figure 6.3 and shown in Figure 6.4. On the right, the circles along τ indicate the times for the *selected* spectra recalculated on the far left.

is therefore reflected in the magnitude of the singular values. Because the singular values are ordered by size, the first component spectrum \mathbf{u}_1 and γ -dependence \mathbf{v}_1 are the most important components. We can define this importance mathematically: if S is fit by a matrix of rank q , then $\mathbf{U}_{1,\dots,q}\mathbf{S}_{1,\dots,q}\mathbf{V}_{1,\dots,q}^T$ produces the best least-squares approximation of S , where $\mathbf{U}_{1,\dots,q}$ and $\mathbf{V}_{1,\dots,q}$ represent the 1st through q th columns of \mathbf{U} and \mathbf{V} , and $\mathbf{S}_{1,\dots,q}$ represents the 1st through q th columns and rows of \mathbf{S}^3 . For a two component state (where $s_i = 0$ for $i > 2$), the first component spectrum will be the average of S over γ . The second component spectrum will be the difference spectrum.

From a mathematical perspective, the SVD results are related to eigendecomposition. \mathbf{V} contains the eigenvectors of $S^T S$, \mathbf{U} contains the eigenvectors of SS^T , and the non-zero \mathbf{S}^2 diagonal terms are the non-zero eigenvalues of $S^T S$ and SS^T .

As described above, $\mathbf{U}, \mathbf{S}, \mathbf{V}$ is the mathematical basis that is objectively calculated through SVD. Comparing eqs. (6.1) and (6.3) shows that the mathematical basis $\mathbf{U}, \mathbf{S}, \mathbf{V}$ and the state basis Σ, Z both reproduce the initial data set S . Importantly, if we ignore the noise, they should each have the same rank. The unique characteristic of the mathematical basis is that the nonzero columns of \mathbf{U} and \mathbf{V} are orthonormal, while in the state basis, Σ and Z are generally not because they correspond to actual real species. The sets are also related because they each contain a matrix of spectra (Σ and \mathbf{U}) and a matrix of γ -dependence for each corresponding spectrum (Z and \mathbf{V}). As a result of this pairing, the mathematical basis set matrices \mathbf{U} and \mathbf{V} are a linear combination of their paired state basis analogues. If we collect a data set S we can perform SVD objectively to calculate the mathematical basis $\mathbf{U}, \mathbf{S}, \mathbf{V}$. Our goal is to take this mathematical result and transform it into the state basis, where we have made assumptions about some aspect of our system. We could assume that we know the spectrum of some of our species. Maybe we know the γ -dependence of some of our species. Combining and applying these assumptions allows us to calculate the appropriate transformation matrices that convert from the mathematical basis to the state basis, all under the umbrella restriction that we can still reconstruct the original data set S from our different basis matrices.

The other sections of this chapter analyze the information content contained within $\mathbf{U}, \mathbf{S}, \mathbf{V}$, the number of assumptions necessary to transform into the state basis Σ, Z , and the process of transformation.

6.2.2 Information content and the intermediate basis

Although both sets describe S , the state basis Σ, Z contain h components while the mathematical basis $\mathbf{U}, \mathbf{S}, \mathbf{V}$ (and, equivalently, S) is *at most* of rank h in the absence of noise. For example, in a noiseless two-state system where net $A \longrightarrow B$, there will be two sets of SVD components. If species of the system share indistinguishable spectra, the number of SVD components necessary to mathematically describe S decreases. For example, in the two component spectrum described above, if A and B have indistinguishable spectra, the spectra would not change with γ and there will be one SVD component with a trivial constant γ -dependence. Similarly, if the γ -dependence was indistinguishable, we would also end up with one SVD component set. We define the rank of S in the absence of noise as g which is $\leq h$. Given S , g is the maximum number of species that we can distinguish without interpreting the data for physicality (e.g. assigning two exponential kinetics to at least three states even if there

are only two SVD components).

Although it complicates our discussion, we need to define an integer, f , which represents the number of spectra that we can separate from the noise. Ordering the three integers by size gives: h (number of species) $\geq g$ (number of species with distinguishable spectra) $\geq f$ (number of distinguishable spectra that can be separated from noise).

For a well characterized spectroscopic system, the spectra or γ -dependence of the individual species are known or hypothesized, which allows f state basis spectra (Σ) and f associated γ -dependent traces (Z) to be determined based on a fit to the mathematical $\mathbf{U}, \mathbf{S}, \mathbf{V}$ results. In more complicated systems, we must often balance the uncertainty introduced by assumptions with the increased information they provide.

We find it useful to define a third basis set, the *intermediate basis*, with rank f that represents a physically meaningful set of spectra, U , and γ -dependence, V , which bridges the gap between the other two basis sets. Although the intermediate basis is physically meaningful, it does not necessarily correspond to individual specific states – it is not the state basis. With this intermediate basis, we enter an uncomfortable realm where some of the restricting information is assumed (or known) and some is not. For example, if we start with the mathematical basis $\mathbf{U}, \mathbf{S}, \mathbf{V}$ and we know the *difference* spectrum between two of the molecular species, but not their actual spectra σ , we can still transform into some intermediate basis U, V , while we could not transform fully into the state basis Σ, Z . In the intermediate basis, this one specified known difference spectrum, can still constrain the other components such that additional information about the system can be inferred. Extracting this additional information from the mathematical basis $\mathbf{U}, \mathbf{S}, \mathbf{V}$ would be more difficult.

Working with the mathematical basis $\mathbf{U}, \mathbf{S}, \mathbf{V}$ and converting to the intermediate basis U, V ensures that both sets of matrices describe the original dataset S . Thus, converting to the intermediate basis U, V can be thought of as beginning to model the data. We assume certain features of our data, and necessary constraints on the remaining components are enforced automatically.

Thus, the nonzero columns of the mathematical $\mathbf{U}, \mathbf{S}, \mathbf{V}$ represent the calculated orthonormal basis set, U, V represents a basis set with *intermediate* physical meaningfulness calculated using some assumptions about the system, and the state basis Σ, Z represents the full molecular information, which only results from enough assumptions to specify the full system.

Up to our limiting f number, each set can be calculated from the previous with the addition of assumptions regarding the nature of the molecular species. The intermediate set U, V is left intentionally vague, and it is discussed in more detail later for specific cases. Stated differently, the mathematical $\mathbf{U}, \mathbf{S}, \mathbf{V}$ is objective whereas U, V is subjective given that it depends on assumptions about the system. Although there is a true, objective underlying state basis Σ, Z , we usually have to make many assumptions to hypothesize it.

Writing out each basis set explicitly to show its number of components gives:

$$S(\omega, \gamma) = \sum_{i=1}^h \sigma_i(\omega) \zeta_i(\gamma) + \Xi = \sum_{i=1}^f u_i(\omega) v_i(\gamma) + \Xi' = \sum_{i=1}^f \mathbf{u}_i(\omega) \mathbf{s}_i \mathbf{v}_i(\gamma) + \Xi' \quad (6.5)$$

with $\Xi' = \Xi + \sum_{i=f+1}^g \mathbf{u}_i(\omega) \mathbf{s}_i \mathbf{v}_i(\gamma)$ representing the sum of the noise and any components that are inseparable from the noise. Although the number of variables is confusing we find it useful to explicitly separate the hypothetical from the real-world, measurable results. It should be noted that the Ξ' noise places a limit on the accuracy of the calculated spectra and γ -dependence. For example, due to the orthonormality imposed by SVD, in a mixture the true spectrum of species i , σ_i , can conceivably be split into the linear combination of two \mathbf{u} spectra ($\sigma_i = c_1 \mathbf{u}_j + c_2 \mathbf{u}_k$), one of which might be thrown away with the noise. Then, in the recalculation, the predicted “ σ_i ” = $c_1 \mathbf{u}_j$ might differ from the true spectrum. If there is clearly anti-correlated noise in multiple components (upward spikes in one component that correspond with downward going spikes in another), then it indicates that the original noise (spikes) has been partitioned into more than one component. In the ideal case, the noise would have ended up in a single, rejectable component. In this case, the two (or more) components with the anti-correlated noise should be kept, as mixing them will result in the desired components, one which does not have the noise, and one that does.

Example

As an example of SVD, and to relate the mathematical and intermediate bases, we start by generating a noiseless S using the middle part of eq. (6.5). We start in the intermediate basis with two basis horizontally-offset Lorentzian spectra $U_{1,2} = \{u_1, u_2\}$ with associated $V_{1,2} = \{v_1, v_2\}$ exponential time-traces shown in Figure 6.3. One spectrum decays with a fast timescale, while the other grows in more slowly. These are in the intermediate basis because they do not actually correspond to all the molecular species – we only have two components, but the time

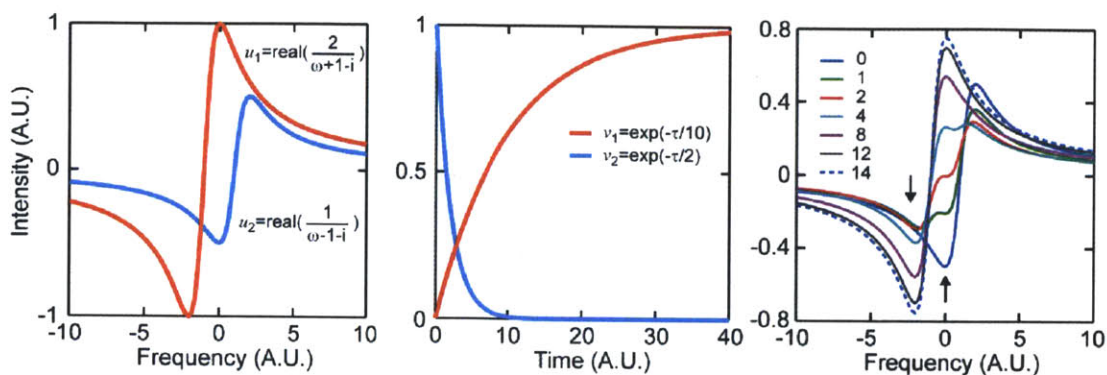


Figure 6.3: *Left* – Two model basis Lorentzian $U_{1,2}$ spectra generated with the shown equations. *Center* – Two model basis exponential $V_{1,2}$ traces generated with the shown equations. *Right* – The model S spectra at selected delay times generated by $S = U_{1,2} V_{1,2}^T$.

traces suggest at least three molecular species. Although the model might be simpler and easier to understand if we generated the data using the hypothetical, state basis (Σ, Z) we find it demonstrative to explicitly analyze the more realistic intermediate basis case.

For the example, which continues in the following sections, we start with the given intermediate basis matrices and we calculate the noiseless data set S . From this S , we forget about our original generating spectra, calculate the mathematical basis $\mathbf{U}, \mathbf{S}, \mathbf{V}$, and then try to transform the results back into an intermediate basis by feeding in assumptions. The process is described by $U, V \rightarrow S; S \rightarrow \mathbf{U}, \mathbf{S}, \mathbf{V} \xrightarrow{\text{Ass. 1}} "U", "V" \xrightarrow{\text{Ass. 2}} "U", "V" \propto U, V$, where the semi-colon represents us forgetting about the initial generating basis set and the “ U, V ” represent our intermediate basis sets predicted after putting in one and two assumptions, described below. As is shown in the above schematic, we find that only two assumptions are necessary to transform our two component ($f=2$) mathematical basis $\mathbf{U}, \mathbf{S}, \mathbf{V}$ into a generated intermediate basis set “ U, V ” that is proportional to our desired, original basis set U, V . Given that this is a contrived example, no real discussion is made on where the assumptions come from because the assumptions depend on the specific system.

The mathematical basis \mathbf{U} , \mathbf{S} , and \mathbf{V} are calculated for the noiseless S shown in Figure 6.3. The results of the SVD are shown in Figure 6.4. To begin, we will briefly compare the mathematical basis with the generating (forgotten) intermediate basis, $U_{1,2}, V_{1,2}$. \mathbf{u}_1 is dominated by the generating u_1 (orange in Figure 6.3), which contributes more to the S integrated over time because the u_1 spectrum has a larger amplitude and is associated with a longer exponential time dependence, v_1 , than u_2 . Given that there are two input basis spectra without noise, we expect $g =$

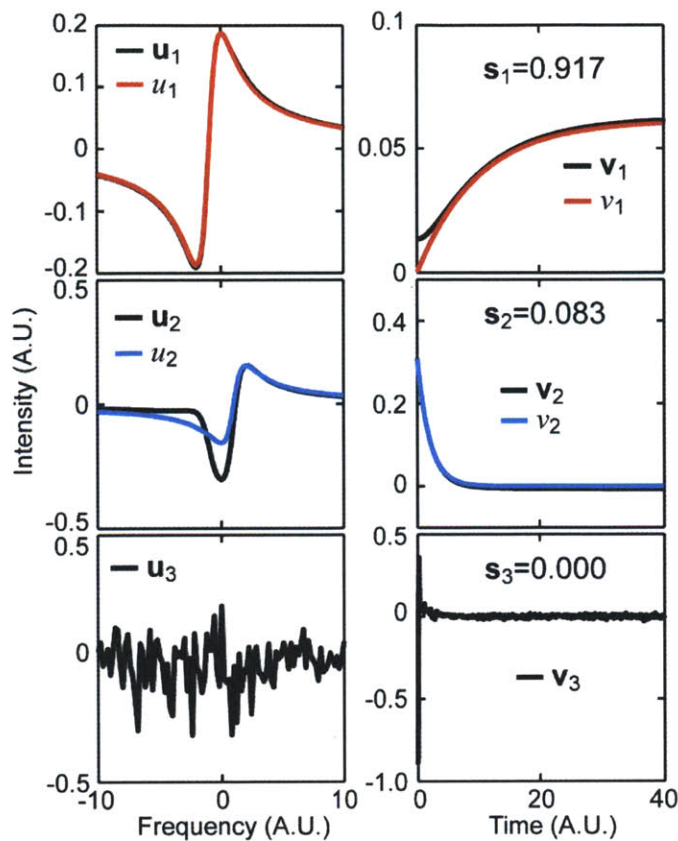


Figure 6.4: The first three mathematical \mathbf{U} and \mathbf{V} SVD results (Column 1 and 2, respectively) from the S shown in Figure 6.3. For comparison, the first two mathematical basis components are compared to the two generating intermediate basis spectra and time-traces from Figure 6.3.

$f=2$ and indeed two SVD basis spectra with non-zero singular values are observed. However, if there were just two species, then the loss of the first would be directly tied to the gain of the second, resulting in mirrored time behavior. Rather, SVD analysis shows two distinct timescales, suggesting that there are more than two physical underlying species ($h > g$) despite the fact that S is generated from two basis spectra. The $h=3$, $g=2$ case can be expected for a kinetic scheme $A \rightleftharpoons B \rightleftharpoons C$ with only two distinguishable spectra. This is illustrative of how the number of underlying species, h , is related to the number of observed components, $f \leq g \leq h$. We reserve the state basis Σ, Z notation for later, when we begin to hypothesize the spectra or γ -dependence of species.

For a spectroscopist, the ideal SVD separates S into the physically meaningful input spectra and time-traces, $\mathbf{U}_{1,2} = U_{1,2}$ and $\mathbf{V}_{1,2} = V_{1,2}$. In this example, the SVD time-dependent $v_{1,2}$ match the input time-traces well except for v_1 at early times. The u_1 spectrum matches well while

\mathbf{u}_2 does not. These mismatches reflect the mathematical nature of the SVD. This example is a special case in which the dominance of the u_1 spectrum results in a \mathbf{u}_1 that almost exactly matches the input spectrum. However, when $\mathbf{u}_1 \propto u_1$, it is guaranteed that $\mathbf{v}_2 \propto v_2$, as shown in Appendix 6.B and described in the following sections.

The similarity between the calculated mathematical $\mathbf{U}_{1,2}, \mathbf{V}_{1,2}$ and the generating intermediate $U_{1,2}, V_{1,2}$ is not guaranteed, and it depends on the relative magnitude of the different ‘true’ components, the differences in timescale, and the spectral overlap of the ‘true’ spectra. Generally, decreased spectral overlap (closer to orthogonal) will result in better separation of the underlying spectra. If the two spectra overlap entirely, then they will not be distinguishable except perhaps in their exhibition of multiple γ behaviors. That is, there may only be one component, \mathbf{u}_1 and \mathbf{v}_1 , but multiple exponentials or melting points may indicate the presence of more than two species.

Although the SVD components do not generally match the true underlying components, the untouched SVD decomposition is useful. The time constants measured from the observed γ -dependence are physically meaningful, and first component spectrum \mathbf{u}_1 and γ -dependence \mathbf{v}_1 will capture the most important features of the data. Perhaps more importantly, the number of significant singular values, f , identifies the number of distinguishable spectra, or the minimum kinetic species. In the case of real data, components may be lost in the noise, and the SVD analysis indicates the minimum number of states. Once the spectra are rearranged in descending mathematical importance, as is determined by \mathbf{s} , a subset of components can be rejected based on their insignificance or noisiness. In this way, noise can be removed.

Assessing the number of significant components, f , is challenging, and it lies in the gray area between quantitative and qualitative determination. Keeping too many components will introduce unnecessary noise into the analysis while keeping too few will throw away meaningful information. To avoid throwing the baby out with the bath water, the components can be assessed based on relative \mathbf{s} singular values, continuity or noisiness of the \mathbf{U} spectra and \mathbf{V} γ -dependence, consistency between spectra or γ -behavior of S and a reconstruction using the significant components (eq. (6.4) with y replaced by f), and noisiness of the reconstructed S without the significant components (eq. (6.4) with $i = f + 1, \dots, y$). Others describe a quantification of noise and an autocorrelation method which can be used to assess component importance^{2, 3, 5}.

6.2.3 Calculating intermediate basis U, V from mathematical basis U, S, V

Here we describe the transformation from the mathematical to the intermediate basis. This represents an intermediate step towards transformation to the state basis. In some situations, if we are not comfortable specifying enough assumptions to hypothesize a state basis, the analysis will end after calculating the intermediate basis.

The transformation process is subjective – we choose some features of our intermediate basis and then calculate the transformation to give us our chosen features and to fill in the remaining holes such that the intermediate basis still describes our original data set S .

As described above, the mathematical basis U, S, V results are objective, and, in the example, represent linear combinations of the generating, intermediate U, V components. To transform the mathematical basis U, S, V into physically meaningful results involves two steps: reducing the rank of U, S, V down to f and specifying the shape (implying proportionality) of some of the physical intermediate basis u_i and/or v_i components. These conditions are described specifically for the two component state in the next section. Once a transformation matrix, W , has been determined based on the input assumption constraints, the intermediate basis U, V can be calculated through

$$\begin{aligned} S &\approx U_{1,\dots,f} S_{1,\dots,f} (\mathbf{W}\mathbf{W}^{-1}) \mathbf{V}_{1,\dots,f}^T = (\mathbf{U}_{1,\dots,f} S_{1,\dots,f} \mathbf{W}) (\mathbf{W}^{-1} \mathbf{V}_{1,\dots,f}^T) \\ &= (\mathbf{U}_{1,\dots,f} S_{1,\dots,f} \mathbf{W}) (\mathbf{V}_{1,\dots,f} (\mathbf{W}^{-1})^T)^T = UV^T \end{aligned} \quad (6.6)$$

where $\mathbf{W}\mathbf{W}^{-1} = \mathbf{I}$. More discussion of W is given in Appendix 6.C. Note that the transformation folds the singular value matrix S into the intermediate basis spectrum matrix U . It can also be multiplied into the intermediate basis γ -dependence matrix V .

While the mathematical $U_{1,\dots,f}$ spectra and $V_{1,\dots,f}$ γ -dependencies are orthonormal, the intermediate U spectra and V traces are generally not. If we know the intermediate basis U spectra to which we would like to transform, then the coefficients of the W matrix can be determined by projecting the mathematical basis $(US)^{-1}$ onto each u spectrum:

$$\begin{aligned} U_{1,\dots,f} S_{1,\dots,f} \mathbf{W} &= U \\ (\mathbf{S}_{1,\dots,f}^{-1} \mathbf{U}_{1,\dots,f}^{-1}) U_{1,\dots,f} S_{1,\dots,f} \mathbf{W} &= (\mathbf{S}_{1,\dots,f}^{-1} \mathbf{U}_{1,\dots,f}^{-1}) U \\ \mathbf{W} &= (\mathbf{S}_{1,\dots,f}^{-1} \mathbf{U}_{1,\dots,f}^T) U \end{aligned} \quad (6.7)$$

Here, the W is determined by the intermediate basis U , and the V would be transformed such that

S is still given by UV^T . An equivalent routine can be written out if a desired V is known:

$$\begin{aligned}
 \mathbf{W}^{-1}\mathbf{V}_{1,\dots,f}^T &= V^T \\
 \mathbf{W}^{-1}\mathbf{V}_{1,\dots,f}^T(\mathbf{V}_{1,\dots,f}^T)^{-1} &= V^T(\mathbf{V}_{1,\dots,f}^T)^{-1} \\
 \mathbf{W}^{-1} &= V^T\mathbf{V}_{1,\dots,f}
 \end{aligned}
 \tag{6.8}$$

In the constrained U case of eq. (6.7), each column of \mathbf{W} contains the coefficients that scale each mathematical basis \mathbf{u}_i spectrum appropriately such that the linear combination sum best gives u . These coefficients, and therefore \mathbf{W} , can also be determined with a fitting routine. Because the \mathbf{W} is used to create U and V , if a mix of u and v are specified, then \mathbf{W} is usually split up into multiple transformation matrices, $\mathbf{W}=\mathbf{W}_a\mathbf{W}_b\dots$, one \mathbf{W}_i for each specified u or v . In this way, the proper \mathbf{W}_i can be determined consecutively such that the overall \mathbf{W} produces the desired transformation. Because the first \mathbf{W}_i breaks the orthogonality of the intermediately transformed “ U_i ” = $\mathbf{U}_{1,\dots,f}\mathbf{S}_{1,\dots,f}\mathbf{W}_i$, then the projection determination breaks down, and a fitting routine should be used thereafter.

In the consecutive process of determining $\mathbf{W}_a, \mathbf{W}_b, \dots$ care must be taken in constructing $\mathbf{W}_{>a}$ such that the intermediate basis component created by the preceding \mathbf{W}_{i-1} is not affected by the current transforming matrix \mathbf{W}_i . For example, if the first column of \mathbf{W}_a is chosen to give a particular u_1 , $\mathbf{U}_{1,\dots,f}\mathbf{W}_a=\{u_1, \mathbf{u}_2, \mathbf{u}_3, \dots \mathbf{u}_f\}$ then \mathbf{W}_b should be constructed such that the first column is the same as the first column of the identity matrix such that the multiplication still gives u_1 as the first column of the intermediate spectral basis matrix: $\mathbf{U}_{1,\dots,f}\mathbf{W}_a\mathbf{W}_b=\{u_1, \dots\}$. If each \mathbf{W}_i only differs from the identity matrix in just one column, \mathbf{W}_i will only affect the component that it is creating in the intermediate basis matrix. In this consecutive process – fitting (to determine \mathbf{W}_a), applying one constraint (multiplying by \mathbf{W}_a), fitting the result to appropriately apply a second constraint (determining \mathbf{W}_b), multiplying, ... – if each \mathbf{W}_i only affects one component (it is equivalent to the identity matrix on all but one column), then no components will be mistakenly zeroed and the fits should be unique.

Example

As an example, the re-weighting is applied to the results of the SVD analysis shown in Figure 6.4. In the transformation procedure, we input assumptions to calculate the underlying components. The true intermediate basis components used to generate S are denoted as u_1, u_2, v_1, v_2 while the eq. (6.6) calculated components are denoted with quotation marks as

" u_1 ", " u_2 ", " v_1 ", " v_2 ." The transformation proceeds through two sequential steps with $\mathbf{W}\mathbf{W}^{-1} = \mathbf{W}_a\mathbf{W}_b\mathbf{W}_b^{-1}\mathbf{W}_a^{-1}$:

- 1) \mathbf{W}_a is found so that " u_1 " is proportional to the latest time spectrum of S , which also means that v_2 decays to 0 at later times. (" u_1 " $\propto S(\tau_{\text{end}})$)
- 2) \mathbf{W}_b is found so that " u_2 " calculated from the results of 1) is proportional to the earliest time spectrum of S , which also means that " v_1 " is 0 at early times. (" u_2 " $\propto S(\tau_0)$)

The results from each sequential step are shown in Figure 6.5 and Figure 6.6, and the specific \mathbf{W} 's are described in Appendix 6.C in eq. (6.C.1) and (6.C.2), respectively for 1) and 2). The input criteria reflect the real world application, where separation of timescales may allow for attributing spectra at early and late γ to different species. As described in Appendix 6.B, for the special two component case, correctly specifying that the predicted " u_1 " $\propto u_1$ ensures that " v_2 " $\propto v_2$. Thus, although only one criterion is input into 1), we directly learn about two components.

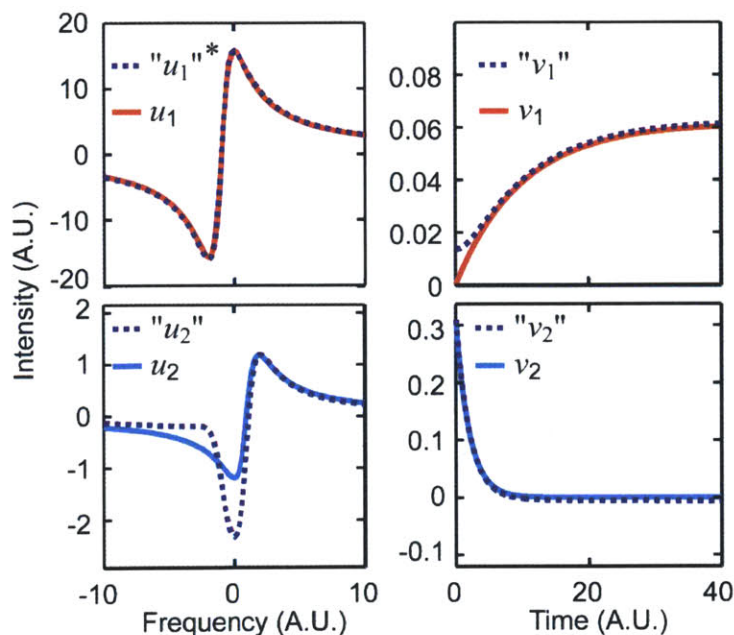


Figure 6.5: Illustration of the 1st step of transformation shown in eq (6.6). The predicted intermediate basis " $U_{1,2}$ " and " $V_{1,2}$ " SVD results (*Column 1* and *2*, respectively) from the S shown in Figure 6.3 after constraining " u_1 " $\propto S(\tau_{\text{end}})$ are shown. For comparison, the first two predicted intermediate basis components are compared to the two generating intermediate basis spectra and time-traces from Figure 6.3. The * indicates the one feature fit to the selected criterion described in the text. For this case, the mathematical basis \mathbf{U}, \mathbf{V} from Figure 6.4 closely match the predicted intermediate basis " U ", " V " results shown

Although this is a simple example, Figure 6.5 and Figure 6.6 show that the re-weighting results better approximate the wanted underlying spectra and γ -dependence. The success of the recovery depends on the input criterion and the appropriateness of the \mathbf{W} . We stress that the procedure works well (“ U ” and “ V ” match U and V) here because our assumptions were accurate and our criterion provided enough constraints to match the remaining components. There is no guarantee that this reweighting procedure will work as well in all cases.

Similarly, specifying incomplete information may also lead to valuable results. For example, if we only had one constraining assumption, such as specifying “ u_1 ” $\propto u_1$, we can calculate “ v_2 ” $\propto v_2$, but the remaining “ u_2 ” and “ v_1 ” will be linear combinations of $u_{1,2}$ and $v_{1,2}$ as shown in Figure 6.5. These components still allow for useful interpretation, however, because differences in behavior between “ u_2 ” and “ v_1 ” and u_1 and v_2 can be attributed to u_2 and v_1 .

Although the examples given above and below in the main text apply the transformation to time dependent data, another example of this transformation procedure for secondary structure

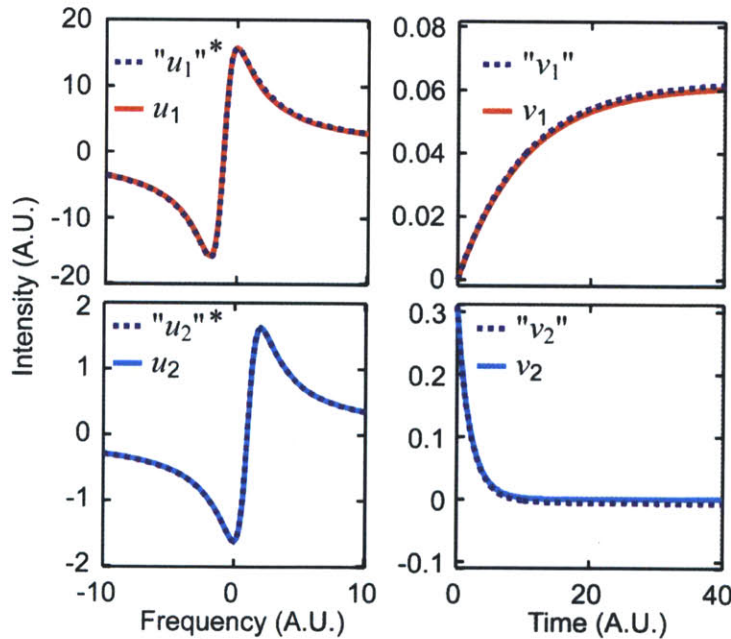


Figure 6.6: Illustration of the 2nd step of transformation shown in eq (6.6). The predicted intermediate basis “ $U_{1,2}$ ” and “ $V_{1,2}$ ” SVD results (*Column 1* and *2*, respectively) from the S shown in Figure 6.3 after constraining both “ u_1 ” $\propto S(\tau_{\text{end}})$ and “ u_2 ” $\propto S(\tau_0)$ are shown. For comparison, the first two predicted intermediate basis components are compared to the two generating intermediate basis spectra spectra and time-traces from Figure 6.3. The * indicates the two features fit to the selected criterion described in the text.

determination is given in Appendix 6.D, where the specific application is discussed and another model is analyzed.

6.2.4 Information content for a two component system

As with the above example from Figure 6.6, if we know the two underlying spectra in a two component system, then we can find the γ -dependence associated with each spectrum. This can be done through SVD and conversion to the intermediate basis, or it can be done with modeling, in which we assume that the spectrum at every γ point is a linear combination of the two model spectra and we find the coefficients through fitting each spectrum. In terms of the intermediate basis, this situation can be described as follows: if we specify u_1 and u_2 , we can calculate v_1 and v_2 . Similarly, if we know the two γ -dependencies, we can find their associated unknown spectra again through SVD with conversion or through fitting – specifying v_1 and v_2 allows us to calculate u_1 and u_2 . These sorts of intuitive thought experiments lead us to question how the different components are linked, or – expressed differently – how many unknown components we can calculate given 0, 1, or 2 known components. These questions are addressed below, where we show that $u_1 \leftrightarrow v_2$ and $u_2 \leftrightarrow v_1$ are linked pairs of components.

Although a two component data set ($f=2$) is described by 6 mathematical basis quantities $\mathbf{U}_{1,2}$, $\mathbf{S}_{1,2}$, and $\mathbf{V}_{1,2}$, there are only two independent variables *if* we are concerned with only the shape (and not the magnitude) of the spectra and γ -dependence. A scalar magnitude constant can be moved from the intermediate basis spectrum u to the γ -dependence v through multiplication and corresponding division, which makes the magnitude of u or v unimportant. Only the shape of the intermediate basis $U_{1,2}$ and $V_{1,2}$ and the absolute magnitude of $u_1 v_1^T$ and $u_2 v_2^T$ are necessary for properly describing S .

For a two component system, $u_1 \leftrightarrow v_2$ and $u_2 \leftrightarrow v_1$ are linked pairs; knowing one of a set allows for the determination of the other (Appendix 6.B). Thus, if we hypothesize 0, 1, or 2 of the unpaired intermediate basis u or v components, then varying amounts of information can be extracted, as is described in Figure 6.7.

Knowing the mathematical basis $\mathbf{U}_{1,2}$, $\mathbf{S}_{1,2}$, and $\mathbf{V}_{1,2}$ and the shape of two of the unpaired components provides enough information for the calculation of the other two. This is analogous to the example shown in Figure 6.6 and described in the bottom case of Figure 6.7.

In the less informed case (middle case of Figure 6.7), in which only the intermediate basis $u_1(v_1)$ is known or estimated, $v_2(u_2)$ can be determined from the SVD results. In this special

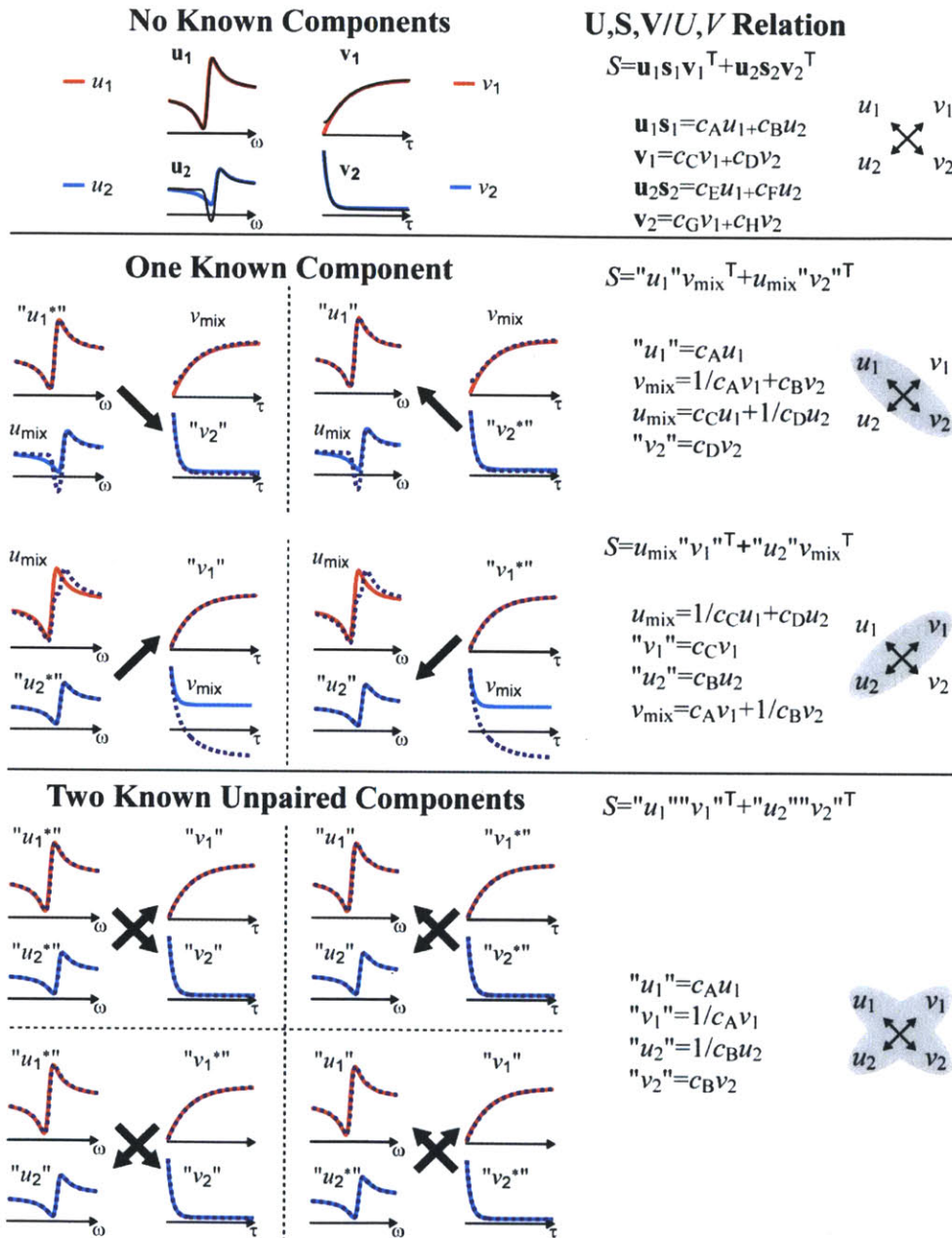


Figure 6.7: Using the spectra from Figure 6.3, the SVD results are related to the true underlying intermediate basis components based on the number of known spectral components. Cases in which 0, 1, and 2 known components (marked with *) are discussed. In the 'one known component case,' specifying any of the components places constraints on the paired component. Similarly, in the bottom case, specifying any two unpaired components specifies (proportionally) all four. c_i are constants, and their number can be reduced (for example, in the 'one known component' case, $c_A c_B + c_C c_D = 0$). For the second 'one known component' case, the mixed components are mixed further to highlight deviations. The grey ovals at right indicate which pair of components is constrained. Components are normalized for comparison to the generating $U_{1,2}, V_{1,2}$.

case where one known component allows for calculation of its linked complement, the two remaining components are linear combinations of $U_{1,2}$ and $V_{1,2}$. Although the unspecified, mixed spectral ($c_C u_1 + 1/c_D u_2$) and γ -dependence ($1/c_A v_1 + c_B v_2$) components do not necessarily correspond to pure u or v components, differences relative to the known u/v pair are unique. That is, if u_1 , and by association, v_2 are known, then unique features observed in u_{mix} and v_{mix} must be the result of u_2 and v_1 . Thus, knowing or estimating the spectrum or γ -dependence of one component can allow for separating the characteristics of all the species, as is shown in the example displayed in Figure 6.5.

6.2.5 Two component re-weighting applied to t-DPP and t-2D IR

In this section, the discussion turns to the specific case of using SVD to analyze the T-jump τ -dependent t-DPP and t-2D IR spectra described in Chapters 7-9. The process is diagrammed in Figure 6.8.

Each of the Chapter 7-8 data sets reduces well into three components, which suggests that the experiments are sensitive to three sets of populations. In order to highlight the changes, the equilibrium spectrum is subtracted from the data set to give the difference data, ΔS . This is equivalent to defining an intermediate basis u_{eq} spectrum equal to the equilibrium spectrum with

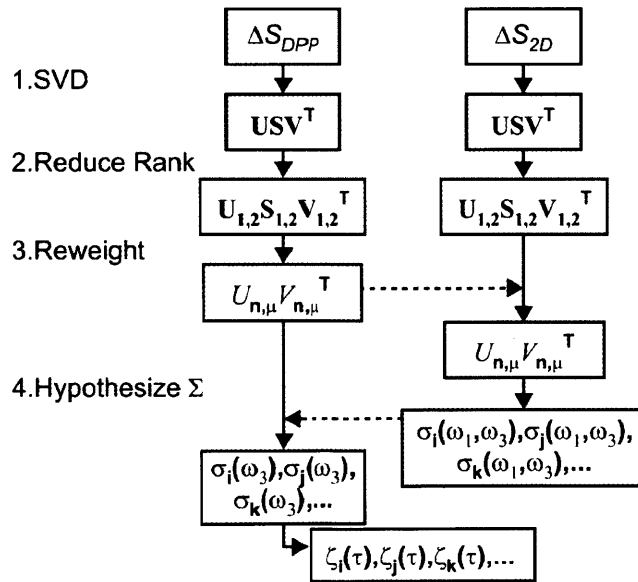


Figure 6.8: Diagram of the two-component SVD re-weighting applied to complementary DPP and 2D IR data sets. Input assumptions allow the mathematical basis U, S, V to be transformed to the intermediate basis U, V , which can then be transformed to the state basis Σ, Z .

an associated constant γ -dependence trace v_{eq} of 1. Although this makes sense for this T-jump data case, for equilibrium studies, for example, it might make more sense to subtract the raw (possibly non-constant) mathematical first component if one assumes that it will contain noise due to laser intensity drift, such that subtraction removes this noise. After the constant equilibrium spectrum subtraction, the number of significant SVD components is reduced by one, down to two. Applying SVD to the ΔS results in mathematical basis spectra $U_{1,2}$ with mathematical τ -dependence $V_{1,2}$. Due to the subtraction of the equilibrium, the remaining $U_{1,2}$ represent difference spectra between the three distinguishable species. As described before, these distinguishable species represent a lower limit on the total number of states in the system.

For the T-jump t-DPP and t-2D IR data, we take a different conceptual approach than is described in the previous example, in which an intermediate basis is assumed without discussion of where it comes from. Here, we propose an intermediate basis that appropriately maps onto the data and makes sense within our physical interpretation of the system's response to a T-jump. In theory, we can choose any intermediate basis such that its product reproduces S . We carefully choose an appropriate model such that the intermediate basis spectra and τ -dependence correspond to predicted physical quantities. Assessment of the legitimacy of our proposed framework intermediate basis begins on the intuitive level, and it is described in the next section. We then analyze the proposed intermediate basis and the assumptions that allow us to calculate it using kinetic models. Although we present the process as first proposing the intermediate basis followed by describing the assumptions that allow us to calculate it, the problem can be flipped - if we start with the appropriate assumptions, the intermediate basis and physical framework necessarily fall into place. That is, if we constrain two of the components to known spectra or γ -dependence traces, by extension we can assign the other two components to physical phenomena.

Proposed intermediate basis

In our T-jump experiments, we observe a separation of timescales, and we use it to help define our intermediate basis. On the pulse-width limited timescale, there is a <5 ns response that we attribute to primarily H-bond breaking between the solvent and the protein as described in Chapter 5 for diglycine and TZ2. Conformational change within the protein samples, such as unfolding or disassociation, is expected to occur on longer, >1 μ s timescales. Thus, the spectrum at early time is expected to be the sole result of loss of the equilibrium species and gain of the H-

bond-broken species' spectrum. The spectrum at longer timescales after conformational change will be the linear combination of all three spectra.

Based on the existence of three-components in S and the ns and μ s timescale separation, we propose three intermediate basis u and v components (equilibrium, ns, and μ s contributions) to describe the t-DPP and t-2D IR data:

$$S = \sum_{i=n,\mu,eq} u_i v_i^T + \Xi' \quad (6.9)$$

with the u_{eq} spectrum equal to the equilibrium spectrum, $v_{eq} = 1$ at all time-delays, τ , and the Ξ' (noise+rejected components) is the same as described in eq. (6.5). The relationship between the intermediate basis U, V and state basis Σ, Z is described below. ΔS is the difference spectrum from equilibrium, which is given by just summing the n and μ components.

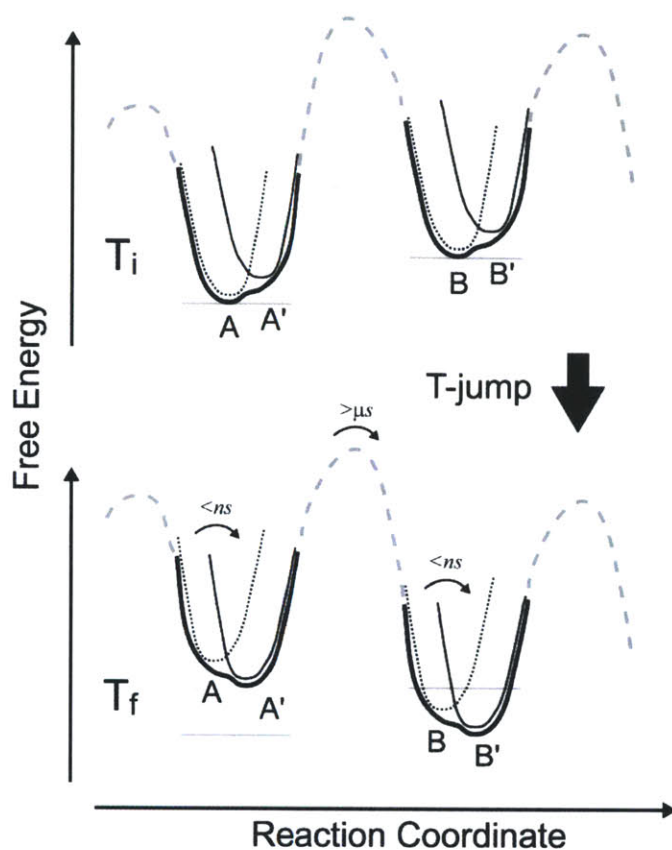
The intermediate basis n and μ components can be assigned to combinations of species in the system. Based on the empirically observed timescales, we separate our system into macrostates and microstates depending on the barriers that separate the species. The macrostates, $\{\tilde{I}, \tilde{J}, \dots\}$ are defined by energy separation of $> k_B T$. The macrostates are composed of the microstates $\tilde{I} = \{I, I', \dots\}$, which are separated by small $< k_B T$ barriers. We assume and observe just two microstates per macrostate (one whose loss feeds the gain of the other), although our discussion can be generalized to more. Due to the small barriers, re-equilibration between the microstates occurs on a < 5 ns observed timescale. On the other hand, conversion between different macrostate populations gives rise to a > 1 μ s response. A cartoon of free energy surfaces illustrating the separation of these states is depicted in Figure 6.9.

In this way, the system timescales and spectra can be separated into the proposed intermediate basis. The intra-well conversion has spectrum u_n and τ -dependence v_n , and inter-well conversion has u_μ and v_μ . At the bottom of Figure 6.9, we list the proposed intermediate basis set and how it relates to the physical model described above.

Assumptions

Although we have proposed the ideal intermediate basis set that we hypothesize should exist (i.e. $U_{n,\mu}, V_{n,\mu}^T = \Delta S$), we have not described the specific assumptions that allow us to constrain the transformation such that we are confident that the calculated intermediate basis matches the proposed.

The assumptions for converting from the mathematical basis to the T-jump specific



$$\begin{aligned}
 \text{"n"} &= c_A^n(A-A') + c_B^n(B-B') + \dots \\
 u_n &\approx c_A^n(\sigma_A - \sigma_{A'}) + c_B^n(\sigma_B - \sigma_{B'}) + \dots \\
 v_n &\approx \propto c_A^{\text{in}}([A'(\tau)] - [A(\tau)]) + c_B^{\text{in}}([B'(\tau)] - [B(\tau)]) + \dots \\
 \text{"}\mu\text{"} &= c_A^\mu(A+A') + c_B^\mu(B+B') + \dots \\
 u_\mu &\approx c_A^\mu(\sigma_A + \sigma_{A'}) + c_B^\mu(\sigma_B + \sigma_{B'}) + \dots \\
 v_\mu &\approx \propto c_A^{\text{tr}}([A(\tau)] + [A'(\tau)]) + c_B^{\text{tr}}([B(\tau)] + [B'(\tau)]) + \dots
 \end{aligned}$$

Figure 6.9: Exaggerated illustration of the free-energy intra- (I, I') and inter- ($I+I', J+J'$) well states at initial and final temperatures, T_i and T_f . Arrows indicate the net flow of population following the T-jump. The states can be grouped into difference states, n and μ , based on the separation of timescales. The spectra of these clumped states will be u_n and u_μ while the τ -dependence is given by v_n and v_μ . Thus, our proposed clumping of states maps directly onto the proposed intermediate basis. $[I'(\tau)] = [I(\tau)] - [I(\text{eq})]$.

intermediate $U_{n,\mu}, V_{n,\mu}$ are shown in Figure 6.10 and discussed below.

To separate the spectral components contributing to the difference spectra, we make the first assumption that the early time difference spectrum, $\Delta S(\tau \leq 5 \text{ ns})$, is the sum of the difference spectra between the I and I' pairs. Based on the complementarity (Appendix 6.B), setting $u_n \propto \Delta S(\omega, \tau \leq 5 \text{ ns})$ ensures that v_μ is proportional to the inter-well kinetic time behavior. We expect

Mathematical Basis

$$\begin{pmatrix} \Delta S \\ \leftarrow \tau \rightarrow \\ \uparrow \ominus \downarrow \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \\ \downarrow \ominus \downarrow \\ \downarrow \ominus \downarrow \end{pmatrix} \begin{pmatrix} \mathbf{s}_1 & 0 \\ 0 & \mathbf{s}_2 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 \xrightarrow{\tau} \\ \mathbf{v}_2 \xrightarrow{\tau} \end{pmatrix}$$

Intermediate Basis

$$\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \\ \downarrow \ominus \downarrow \\ \downarrow \ominus \downarrow \end{pmatrix} \begin{pmatrix} \mathbf{s}_1 & 0 \\ 0 & \mathbf{s}_2 \end{pmatrix} \begin{pmatrix} \mathbf{w} \end{pmatrix} = \begin{pmatrix} u_n^* \propto \\ \Delta S(\omega, \tau \leq 5 \text{ ns}) \\ \downarrow \ominus \downarrow \end{pmatrix} \begin{pmatrix} u_\mu \\ \downarrow \ominus \downarrow \end{pmatrix}$$

$$\begin{pmatrix} \mathbf{w}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 \xrightarrow{\tau} \\ \mathbf{v}_2 \xrightarrow{\tau} \end{pmatrix} = \begin{pmatrix} v_n^* \propto \\ \mathbf{T}(\tau) - \mathbf{T}(\text{eq}) \\ v_\mu \xrightarrow{\tau} \end{pmatrix}$$

Figure 6.10: The transformation process and assumptions for generating an appropriate intermediate basis $U_{n,\mu}, V_{n,\mu}$. *Top* – The two component mathematical basis is shown. *Middle* – The $\mathbf{U}_{1,2}\mathbf{S}_{1,2}$ matrices are transformed into the intermediate basis under the assumption (marked with asterisk and grey box) that $u_n \propto \Delta S(\omega, \tau \leq 5 \text{ ns})$. *Bottom* – The $\mathbf{V}_{1,2}^T$ matrices are transformed into the intermediate basis under the assumption (marked with asterisk and grey box) that $v_n \propto \mathbf{T}(\tau) - \mathbf{T}(\text{eq})$. Although not explicitly shown, constraining u_n and v_n also affects the calculated u_μ and v_μ .

that v_μ is zero at early times, indicating that it is solely the result of the $>\mu\text{s}$ behavior. Deviation of v_μ from zero at early times indicates the need to re-assess the assumptions described here.

The second assumption is based on the relationship between $[I]$ and $[F]$. The equilibrium constant that describes the ratio between $[I]$ and $[F]$ depends on the temperature, and we assume that interconversion is so fast that $[I]$ and $[F]$ are always at equilibrium within the well at times $> 5 \text{ ns}$. We first assume that $[I]/[F] \propto C_1 + C_2\{\mathbf{T}(\tau) - \mathbf{T}(\text{eq})\}$. We further assume that $\{[I(\tau)] - [I(\text{eq})]\} - \{[F(\tau)] - [F(\text{eq})]\}$ will be proportional to the change in temperature. To 0th order, these statements are correct – both $[I]/[F]$ and $\{[I(\tau)] - [I(\text{eq})]\} - \{[F(\tau)] - [F(\text{eq})]\}$ will begin at the T_{final} values and monotonically return to the T_{initial} values. The simplest assumption is that this

monotonic return will be proportional to the change in temperature. These assumptions are discussed further in Appendix 6.E.

Furthermore, because of our definition of intermediate basis spectrum u_n as proportional to the difference spectrum between I and I' , then we expect the associated intermediate basis τ -dependence v_n to also be proportional to $T(\tau) - T(\text{eq})$. Linking these pieces together results in our second assumption, that $v_n \propto T(\tau) - T(\text{eq})$. Thus, the expected time dependence (v_n) is matched to what we expect for the difference spectra assigned in our first assumption (u_n).

These two assumptions provide enough constraints to approximate the two remaining unknown components, u_μ and v_μ . We expect the assumptions to lead to the proposed intermediate basis $U_{n,\mu}$ and $V_{n,\mu}$ generalized at the bottom of Figure 6.9. Based on the assumptions, v_μ is expected to show the time-behavior of the conformational changes and u_μ should be the difference spectrum associated with the conformational change. Interestingly, if the v_μ kinetics extend beyond the temperature relaxation, then at long times after the temperature equilibrates, the observed difference spectrum will be approximately solely the result of u_μ . This case is discussed in Chapter 7 for ubiquitin.

Example

As an example, two simple kinetic schemes were modeled for first-order, reversible, sequential reactions between three and four states (where A' is analogous to I' , A to I). These scenarios represent the simplest models that would give rise to \sim three component SVD results (two after subtraction of the equilibrium spectrum). The four-component system actually has four SVD components, but one of them is relatively small. The method of simulation is described in Appendix 6.F.

In Figure 6.11 the entire process is laid out for the three and four component system. We begin at the top of the figure in the state basis (I). We then describe the ideal intermediate basis that represents what we expect each intermediate component to correspond to after the assumptions are applied (II). Finally, the assumptions are applied, and the calculated results are verified to be in good agreement with the predictions from the ideal case (III).

The results of the simulation and SVD analysis are shown in Figure 6.12 and Figure 6.13. The three state case is chosen for its simplicity, while the four state case is chosen because it is more realistic.

These kinetic models allow us to test our assumptions by calculating the mathematical

Three-state Case

$$[i^*(\tau)] = [i(\tau)] - [i(\text{eq})]$$

I. State Basis

$$S = \sigma_A[A(\tau)]^T + \sigma_{A'}[A'(\tau)]^T + \sigma_B[B(\tau)]^T$$

$$[A(\tau)] + [A'(\tau)] + [B(\tau)] = 1$$

$$\Delta S(\omega, \gamma) = S(\omega, \gamma) - \sigma_A(\omega)[A(\text{eq})] + \sigma_{A'}(\omega)[A'(\text{eq})] + \sigma_B(\omega)[B(\text{eq})]$$

$$\Delta S = \sigma_A[A^*(\tau)]^T + \sigma_{A'}[A'^*(\tau)]^T + \sigma_B[B^*(\tau)]^T$$

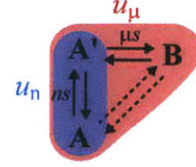
$$[A^*(\tau)] + [A'^*(\tau)] + [B^*(\tau)] = 0$$

II. Proposed Intermediate Basis

$$\Delta S = u_n v_n^T + u_\mu v_\mu^T$$

$$u_n = \sigma_A - \sigma_{A'} \quad v_n = \{[A^*(\tau)] - [A'^*(\tau)]\} / 2$$

$$u_\mu = 2\sigma_B - (\sigma_A + \sigma_{A'}) \quad v_\mu = [B^*(\tau)] / 2$$



III. Calculated

$$u_n^* = \Delta S(t < 10 \text{ ns}) \quad v_n^* \propto T(\tau) - T(\text{eq})$$

$$\approx \sigma_A - \sigma_{A'} \quad \approx \{[A^*(\tau)] - [A'^*(\tau)]\}$$

$$u_\mu \approx R_1 \sigma_C - (\sigma_A + \sigma_{A'}) \quad v_\mu = [B^*(\tau)] / R_1$$

Four-state Case

I. State Basis

$$S = \sigma_A[A(\tau)]^T + \sigma_{A'}[A'(\tau)]^T + \sigma_B[B(\tau)]^T + \sigma_{B'}[B'(\tau)]^T$$

$$[A(\tau)] + [A'(\tau)] + [B(\tau)] + [B'(\tau)] = 1$$

$$\Delta S = \sigma_A[A^*(\tau)]^T + \sigma_{A'}[A'^*(\tau)]^T + \sigma_B[B^*(\tau)]^T + \sigma_{B'}[B'^*(\tau)]^T$$

$$[A^*(\tau)] + [A'^*(\tau)] + [B^*(\tau)] + [B'^*(\tau)] = 0$$

II. Proposed Intermediate Basis

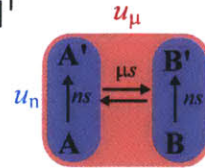
$$\Delta S = u_n v_n^T + u_\mu v_\mu^T + u_3 v_3^T$$

$$\Delta S = u_n v_n^T + u_\mu v_\mu^T + u_x v_x^T \approx u_n v_n^T + u_\mu v_\mu^T$$

$$u_n = (\sigma_{A'} - \sigma_A) + (\sigma_{B'} - \sigma_B) \quad v_n = 1/4 \{ [A'^*(\tau)] - [A^*(\tau)] \} + \{ [B'^*(\tau)] - [B^*(\tau)] \}$$

$$u_\mu = (\sigma_B + \sigma_{B'}) - (\sigma_A + \sigma_{A'}) \quad v_\mu = 1/4 \{ [B^*(\tau)] + [B'^*(\tau)] \} - \{ [A^*(\tau)] + [A'^*(\tau)] \}$$

$$u_x v_x^T = -1/2 \{ \{ \sigma_{A'} + \sigma_B \} \{ [A'^*(\tau)] + [B^*(\tau)] \}^T + \{ \sigma_A + \sigma_{B'} \} \{ [A^*(\tau)] + [B'^*(\tau)] \}^T \}$$



III. Calculated

$$u_n^* = \Delta S(t < 10 \text{ ns}) \quad v_n^* \propto T(\tau) - T(\text{eq})$$

$$\approx (\sigma_{A'} - \sigma_A) + (\sigma_{B'} - \sigma_B) \quad \approx 1/4 \{ \{ [A'^*(\tau)] - [A^*(\tau)] \} + \{ [B'^*(\tau)] - [B^*(\tau)] \} \}$$

$$u_\mu \approx (\sigma_B + \sigma_{B'}) - (\sigma_A + \sigma_{A'}) \quad v_\mu \approx 1/4 \{ \{ [B^*(\tau)] + [B'^*(\tau)] \} - \{ [A^*(\tau)] + [A'^*(\tau)] \} \}$$

Figure 6.11: The three-state and four-state cases are described in terms of state basis and the intermediate basis difference spectra and time-dependencies $U_{n,\mu}$ and $V_{n,\mu}$. *'s indicate which spectra and time dependencies are specified in the assumptions. $[i(\tau)]$ is the concentration of species i . σ_i is the spectrum of species i .

basis U, S, V , transforming to the intermediate basis $U_{n,\mu}, V_{n,\mu}$, and comparing the intermediate basis results to the linear combination of spectra and τ -dependent concentration traces that we assume for real data sets. For example, we assume that the early time difference spectrum

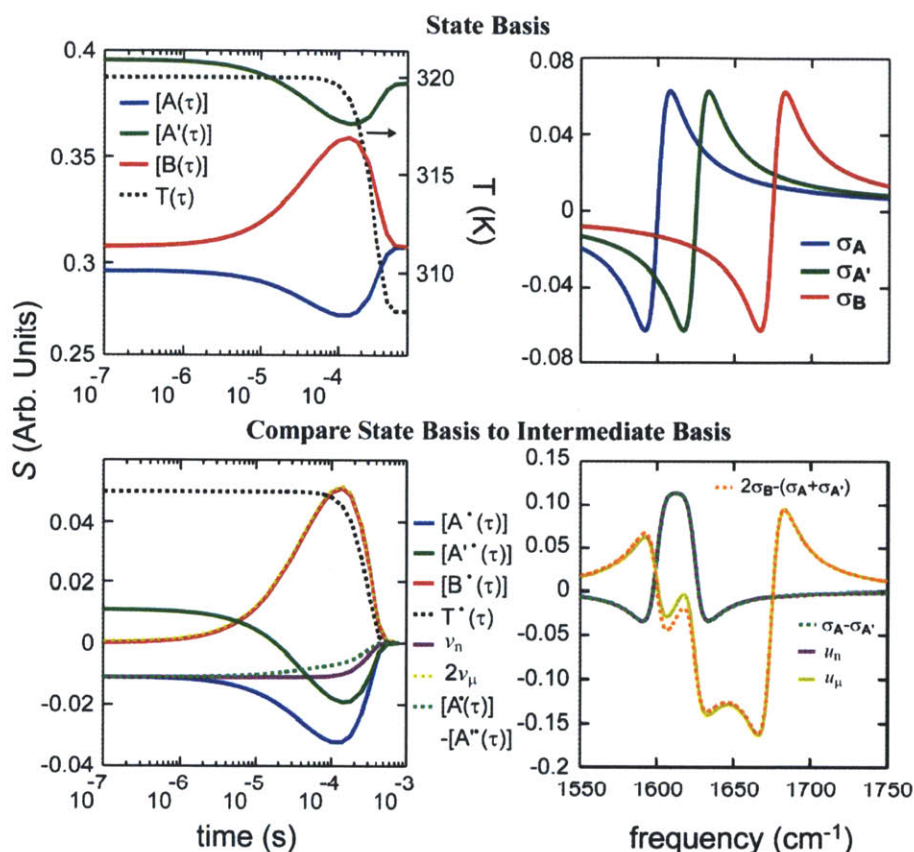


Figure 6.12: The three-state T-jump kinetic model: *Top*: The state basis concentration dependence (left) and spectra (right) for A, A', and B. *Bottom*: The state basis difference concentration dependence and expected difference spectra are compared to the intermediate basis SVD $V_{n,\mu}$ and $U_{n,\mu}$.

(which we use to define our intermediate basis spectrum u_n) should be proportional to the spectrum of loss and gain of the intra-well states.

Although the rates and thermodynamic parameters are highly contrived and the temperature profile drops unphysically fast, the models illustrate the spectral changes and SVD results we expect for three and four state systems. In the three-state case, A and A' are intra-well states with ns conversion rates. A and A' are separated from B by a larger barrier that results in 10's of μ s conversion. As expected, at early times population moves from A to A'. At longer times, population moves from A+A' to B. Although in the simulation population is only allowed to move sequentially, we expect the same results for the fuller scheme in which movement is allowed directly between A and B; the conversion between A and A' is so fast relative to movement to B, that A+A' can be viewed as a single state at longer times. In the four-state case, A,A' and B,B' are pairs of intra-well states while A+A' and B+B' convert on the 10's of μ s

timescale.

As shown in Figure 6.12 and Figure 6.13, the SVD re-weighting intermediate basis results match the expected difference spectra and timescales (for the three state case: $u_n \propto \sigma_A - \sigma_{A'}$; $v_n \propto T(\tau) - T(\text{eq})$; $u_\mu \propto 2\sigma_B - (\sigma_A + \sigma_{A'})$; $v_\mu \propto [B'(\tau)]$) indicating that the assumptions are reasonable for these model scenarios. However, the slight differences between expected and calculated results suggest that the assumptions are not entirely correct. For the four-state case, there are actually four SVD components (three for ΔS) which can explain some differences because we only use three of them (two for ΔS). In both cases, however, we do not expect the second assumption ($v_n \propto T(\tau) - T(\text{eq})$) to be exact. The $[A''(\tau)] - [A'(\tau)]$ is not expected to be exactly proportional to $T(\tau) - T(\text{eq})$. Rather, as $A+A'$ feeds B (and B'), the $[A''(\tau)] - [A'(\tau)]$ will change with some behavior not proportional to the temperature – instead the changes will be

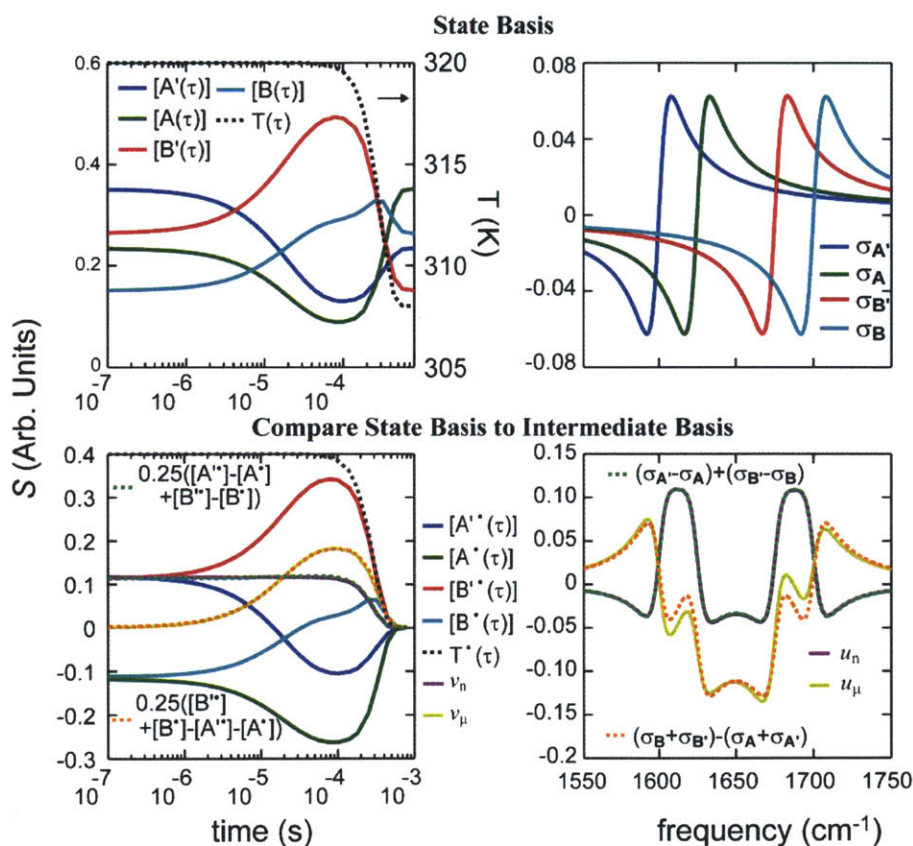


Figure 6.13: The four-state T-jump kinetic model: *Top:* The state basis concentration dependence (left) and spectra (right) for A, A', B, and B'. *Bottom:* The state basis difference concentration dependence and expected difference spectra are compared to the intermediate basis SVD $V_{n,\mu}$ and $U_{n,\mu}$.

proportional to the $>1 \mu\text{s}$ timescale behavior. The relative magnitude of the temperature-proportional behavior is expected to be large enough to dominate the response.

Another way to check the reasonability of the assumptions is to examine the resulting intermediate basis τ -dependence v_μ , which is expected to start at 0. In these cases it does, which suggests that the assumptions are consistent with the anticipated results.

The overarching goal of this SVD analysis is to extract as much information as possible from the data with minimal assumptions, which is why we restrict ourselves here to the intermediate basis rather than trying to calculate the state basis. In the model cases, we have specified six thermodynamic variables (forward and reverse $\Delta G^\ddagger, \Delta S^\ddagger, \Delta H^\ddagger$) and a spectrum for each kinetic state. Using the un-manipulated SVD results we can identify the minimum number of distinguishable states and the observed rates that govern the conversion between them at T_f . The tedious re-weighting transformations described above allow us the added benefit of identifying the spectral changes that occur on the fast and the slow timescales. These changes should be apparent in the ΔS , but the SVD analysis allows for a more careful separation of the features.

The benefit of t-DPP and t-2D IR results is that they provide complementary information. Because the t-DPP is related to the projection of the t-2D IR onto ω_3 , the SVD of each data set can be used to directly inform the processing of its complement. The t-DPP provide better τ -resolution and signal-to-noise while the t-2D IR provides increased spectral content. As shown in Figure 6.8, the SVD analysis and reduction of size down to two components is applied independently to the t-DPP and t-2D IR data. Because the t-DPP is generally less noisy, the re-weighting criterion is applied to the t-DPP to determine the intermediate basis $U_{n,\mu}$ and $V_{n,\mu}$. The t-2D IR SVD components are re-mixed to find the intermediate basis $U_{n,\mu}$ and $V_{n,\mu}$ based on matching the t-DPP $U_{n,\mu}$ and $V_{n,\mu}$. The dashed horizontal line in the 3rd step of Figure 6.8 indicates that information from the t-DPP analysis is used to inform the t-2D IR analysis. Thus, using **steps 1-3** from Figure 6.8, the better time spacing of the t-DPP τ -dependence $V_{n,\mu}$ can be directly linked to the increased spectral information of the t-2D IR $U_{n,\mu}$ spectra.

6.2.6 Hypothesizing the state basis Σ_{ijk} and Z_{ijk} from t-DPP and t-2D IR

At this point in the procedure described in Figure 6.8, we have calculated the intermediate $U_{n,\mu}$ and $V_{n,\mu}$ for the t-DPP and t-2D IR spectra.

The goal of **step 4** in Figure 6.8 is to hypothesize the spectra of individual populations, which allows for the calculation of the state basis τ -dependent concentration. The intermediate $U_{n,\mu}$ spectra represent difference spectra – loss of population of some species and gain of population of others species. Based on their spectral features, if we assume that we know how both intermediate $U_{n,\mu}$ relate to the state basis Σ , we can calculate the hypothesized spectra of the underlying species. We use this analysis on TZ2 in Chapter 8, and more specific details are given there.

Briefly, we describe the simplest A,A',B three-state case (analogous to the three-state system described in Appendix 6.F, Figure 6.11, and Figure 6.12) under the assumption that we have one state, A, at equilibrium. In the previous treatment this assumption was not made. Although we discuss this case as an example, the treatment is not explicitly performed on the model data from Figure 6.13.

We begin by attempting to hypothesize the spectra of the individual species A, A', B by converting from the intermediate basis U, V components to the state basis Σ, Z . Because the t-2D IR spectra provide more spectral content than the t-DPP, we focus on using the $S^{2D}(\omega_1, \omega_3)$ for the following steps.

This simplifying assumption that the equilibrium state is entirely A allows for determination of

$$\sigma'_{A'}(\omega) = u_{eq}(\omega) \quad (6.10)$$

where the prime indicates that σ is not normalized properly. Although we have focused on the difference data and not discussed the equilibrium spectrum much (section 6.2.5, eq. (6.9)), the u_{eq} has been defined as the equilibrium spectrum. Because we guess that the equilibrium concentration is purely state A, from the equilibrium spectrum, we know the shape of σ_A , but not its absolute magnitude. Following results (II) in Figure 6.11, the other two spectra can be approximated as $u_n \propto \sigma_{A'} - \sigma_A$ and $u_\mu \propto 2\sigma_B - (\sigma_{A'} + \sigma_A)$. Therefore, to get $\sigma_{A'}$, we sum u_n and σ_A with a scaling factor c_n :

$$\begin{aligned} \sigma'_{A'}(\omega) &= u_n(\omega) + c_n \times \sigma'_{A'}(\omega) \\ &= u_n(\omega) + c_n \times u_{eq}(\omega) \end{aligned} \quad (6.11)$$

Written differently, we hypothesize the spectrum of B by first scaling the spectrum of A with the arbitrarily, subjectively determined c_n such that its intensity matches the spectrum of A', then

scale their sum with c_μ , and add it to u_μ to give a hypothesized (but not properly scaled) B spectrum:

$$\begin{aligned}\sigma'_B(\omega) &= u_\mu(\omega) + c_\mu \times [c_n \times \sigma'_{A'}(\omega) + \sigma'_{A'}(\omega)] \\ &= u_\mu(\omega) + c_\mu \times [c_n \times u_{eq}(\omega) + u_n(\omega) + c_n \times u_{eq}(\omega)] \\ &= c_\mu \times u_n(\omega) + u_\mu(\omega) + 2c_\mu \times c_n \times u_{eq}(\omega)\end{aligned}\tag{6.12}$$

In both eqs. (6.11) and (6.12), the spectra are not normalized appropriately (and negative signs may be transferred into the scaling factors). The only proportionality factors that matter are c_n , which equates the magnitude of the A and A' spectra, and c_μ , which equates the magnitude of σ_B (the factor of two is absorbed into the scaling factor) with the sum of the A and A' spectra. At this point, as in the intermediate basis, we are only concerned with the shape of the spectra. The transformation equations keep track of all scaling constants, as is described later in eq. (6.17).

We are unaware of a definitive way of objectively quantifying c_i . Rather, we determine c_i by subjectively assessing the continuity and reasonableness of the hypothesized results. Although the arbitrariness of this procedure places high error bars on the results, general features can be observed. One can immediately reject extremely high or low values of c_i and reject values that give discontinuous or overly sharp features.

The manipulation described in eq. (6.10), (6.11), and (6.12) can be performed in a single matrix transformation by multiplying $U_{n,\mu,eq}$ by T :

$$T = \begin{pmatrix} 0 & 1 & c_\mu \\ 0 & 0 & 1 \\ 1 & c_n & 2c_\mu c_n \end{pmatrix}\tag{6.13}$$

where T 's columns are taken from eq. (6.10), (6.11), and (6.12) such that:

$$U_{n,\mu,eq} T = \Sigma'_{A,A',B}\tag{6.14}$$

is a re-writing of the set of linear equations. At this point, the predicted state basis spectra, σ'_i , are not normalized appropriately. Ideally, each σ_i will be in an intensity-per-molar unit such that the concentration dependence can be calculated in molar terms. We settle for a crude normalization of each spectrum by defining ω_{max} , the frequency at which the equilibrium spectrum, $|\sigma'^{DPP}_A|$, is maximized. We then assume that the dipole is conserved at this ω_{max} , which allows for normalization by multiplying $\Sigma'_{A,B,C}$ by L :

$$L = \begin{pmatrix} 1/\sigma'_{A'}(\omega_{\max}) & 0 & 0 \\ 0 & 1/\sigma'_{A'}(\omega_{\max}) & 0 \\ 0 & 0 & 1/\sigma'_{B'}(\omega_{\max}) \end{pmatrix}. \quad (6.15)$$

The units depend on how S is normalized. In this thesis, ΔS is usually normalized to the $\max(|S(\text{eq})|)$. The spectrum determination and normalization from eqs. (6.13), (6.14), and (6.15) are combined in:

$$\Sigma_{A,A',B} = U_{n,\mu,\text{eq}} T L. \quad (6.16)$$

At this point, we have determined the hypothesized 2D IR spectra. In order to calculate the concentration dependence, Z , we return to the DPP data, which has spectra collected at more τ values. We can calculate the σ_i DPP counterparts (horizontal dashed line after **step 4** in Figure 6.8) by using the same eq. (6.16) but fitting the resulting σ_i^{DPP} to the projections of the $\sigma_i^{2\text{D}}$.

Similar to the process described in eq. (6.6), the DPP state basis concentration dependence, $Z_{A,A',B}$, can be calculated through:

$$Z_{A,A',B} = V_{n,\mu,\text{eq}} (T^{-1})^T (L^{-1})^T. \quad (6.17)$$

At the end of this procedure, we end up with hypothesized spectra and population traces which reproduce the original data set S to within the noise. The number of assumptions littering our path like dejected graduate students (and postdocs), however, suggests enormous errors. Along the way we assume 1) there are only 3 meaningful components, 2) the criterion for re-weighting to get the intermediate basis U, V , 3) A, A', B are the only states, 4) the equilibrium population is a single state, 5) we can subjectively determine c_i , and 6) structures in each population have the same dipole at ω_{\max} . It seems as if a further constraint could be placed on this procedure if we assume conservation of dipole strength/spectral area (or volume for 2D IR), but this is difficult considering the positive/negative nature of the spectra and it is not necessarily true, as is indicated in the dominance of loss features in all of our spectra.

At the end of the procedure, as long as the hypothesized state basis spectra and population traces make sense within our proposed model (as they do to some degree above), then they represent one possible situation that can explain the data. As with any generic proposed model, proving that one's hypothesized model uniquely satisfies the evidence highly depends on the specifics of each situation. Thus, the described SVD analysis and transformation presented

in this chapter is one technique that aids in generating a model. SVD helps remove noise, helps to determine the number of species, and ensures that the product of the generated state basis components faithfully re-creates the original data set.

Although the above specific procedure assumed three states, we can also modify it for more states by defining different, expanded T and L matrices, but this involves further assumptions to specify more states, as would be necessary in a four-state case as described at the bottom of Figure 6.11.

A procedure such as the one described is equally valid for equilibrium temperature-dependent studies. Perhaps temperature-dependent changes observed with the finer temperature steps of DPP can be interpreted using the complementary spectra from 2D IR. Spectra for the different states could be hypothesized, perhaps with lower error if the melting curve baselines are sampled.

6.2.7 Comparison of re-weighting to projections

SVD decomposes the data set S into orthogonal, mathematical basis sets $\mathbf{U}_{\leq f}$ and $\mathbf{V}_{\leq f}$ where appropriately left- or right-multiplication of eq. (6.3) shows that $\mathbf{U}\mathbf{S}$ or \mathbf{V} can be determined through projections of each other onto S : $\mathbf{U}_{\leq f}\mathbf{S}_{\leq f} \approx \mathbf{S}\mathbf{V}_{\leq f}$ and $\mathbf{V}_{\leq f}^T \approx \mathbf{S}_{\leq f}^{-1}\mathbf{U}_{\leq f}^T\mathbf{S}$. Here we discuss the case in which we want to take a simpler approach of decomposing the data through analogous projections using our own, self-defined basis sets without using singular value decomposition.

This sort of analysis can be thought of as a more intuitive way of correlating spectra with γ -dependence. If we have the same γ -dependent data set S , and we want to separate the data set into specific spectra with associated γ -dependence, we would project the known spectra onto each spectrum in S . The overlap coefficient for each projection is related to the γ -dependent weighting of each known spectrum. The conceptually reverse process is also valid, in which a known γ -dependent trace is projected onto each row of S to find the frequency-dependent overlap coefficient. The equations that summarize the procedure described above are shown below.

This described technique for analyzing spectral data sets involves projections or cross-correlations. If a spectrum or γ -behavior, \mathbf{p}_ω or \mathbf{p}_γ is known (or hypothesized), its complementary γ -behavior dependence or spectrum can be measured with assumptions by calculating its overlap with the data:

$$\mathbf{p}_\gamma^i = \frac{S^T \mathbf{p}_\omega^i}{\mathbf{p}_\omega^{i T} \mathbf{p}_\omega^i} \quad (6.18)$$

or

$$\mathbf{p}_\omega^i = \frac{S \mathbf{p}_\gamma^i}{\mathbf{p}_\gamma^{i T} \mathbf{p}_\gamma^i} \quad (6.19)$$

where S is the same (ω, γ) matrix described previously, \mathbf{p}_ω and \mathbf{p}_γ are column vectors, and the denominators are for normalization. \mathbf{p}_ω and \mathbf{p}_γ are analogous to the intermediate u and v , but we want to discuss this procedure without the SVD constraints. Thus, we use different variables.

In this procedure, a hypothesized spectrum or γ -dependence is specified. For analyzing for multiple components, the overlap of a predicted, specified component would be calculated and subtracted from S . This prediction, overlap calculation, and subtraction can be repeated until the remainder is zeroed or drops below the noise level. S can be reconstructed from these projections through:

$$S = \sum_{i=1}^o \mathbf{p}_\omega^i \mathbf{p}_\gamma^{i T} = \mathbf{P}_\omega \mathbf{P}_\gamma^T \quad (6.20)$$

where $\mathbf{P}_j = \{\mathbf{p}_j^1, \mathbf{p}_j^2, \dots, \mathbf{p}_j^o\}$. Unlike SVD, whose number of significant components will match the rank of S (f), poor selection of projection basis spectra will result in an arbitrarily large number of repetitious 'components' ($o > f$). If there are only two distinguishable components and only two well-chosen sets of projection components, then \mathbf{p}_ω and \mathbf{p}_γ are linear combinations of the intermediate basis $u_{1,2}$ and $v_{1,2}$. If, as with the two component SVD case $\mathbf{p}_\omega^1(\mathbf{p}_\gamma^1)$ is chosen to be proportional to $u_1(v_1)$ then $\mathbf{p}_\gamma^2(\mathbf{p}_\omega^2)$ will be proportional to $v_2(u_2)$ (see Appendix 6.B).

An example of projection analysis applied to the S in Figure 6.3 is shown in Figure 6.14, in which \mathbf{p}_ω^1 was first specified as $S(\omega, 40)$ and \mathbf{p}_ω^2 was then specified as the remainder at $\tau=0$ after subtracting out the first component, $S(\omega, 0) - \mathbf{p}_\omega^1(\omega) \mathbf{p}_\gamma^1(0)$. The results are similar to the SVD mathematical basis components $\mathbf{U}_{1,2}$ and $\mathbf{V}_{1,2}$ of Figure 6.4 because in the original SVD analysis $\mathbf{u}_1 \propto u_1$. Based on our specified projection basis spectra, \mathbf{p}_ω^1 is proportional to u_1 and \mathbf{p}_γ^2 is necessarily proportional to v_2 .

If the proper basis spectra are chosen, projection analysis can reveal similar results as SVD or the re-weighting analysis. SVD seems to be generally more useful because it is better at

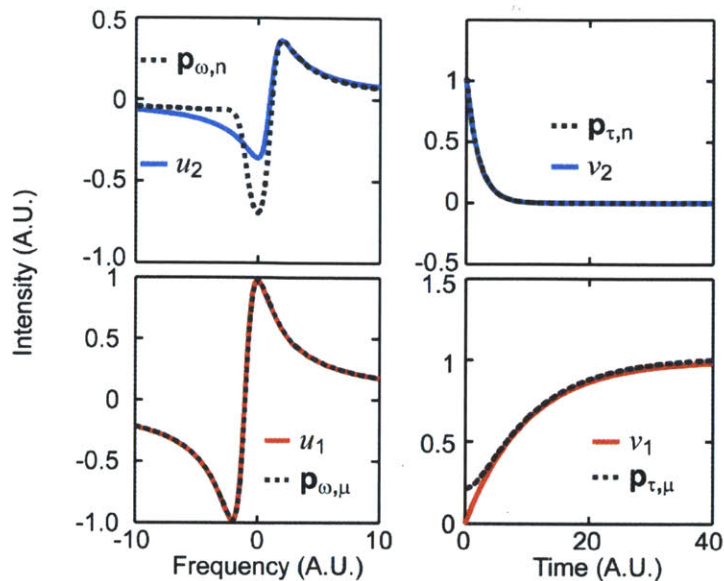


Figure 6.14: The $\mathbf{p}_{n,\mu}$ projection results from the S shown in Figure 6.3. For comparison, the first two projection components are compared to the two generating intermediate basis spectra and time-traces from Figure 6.3.

separating out noise. If a noisy projection basis component is specified, it can lead to noisier results. If however, the data is relatively noisy and there are multiple consistent data sets, as is the case for insulin in Chapter 9, then SVD analysis of each individual data set may result in noise being mixed into different components in different data sets. Projection decomposition

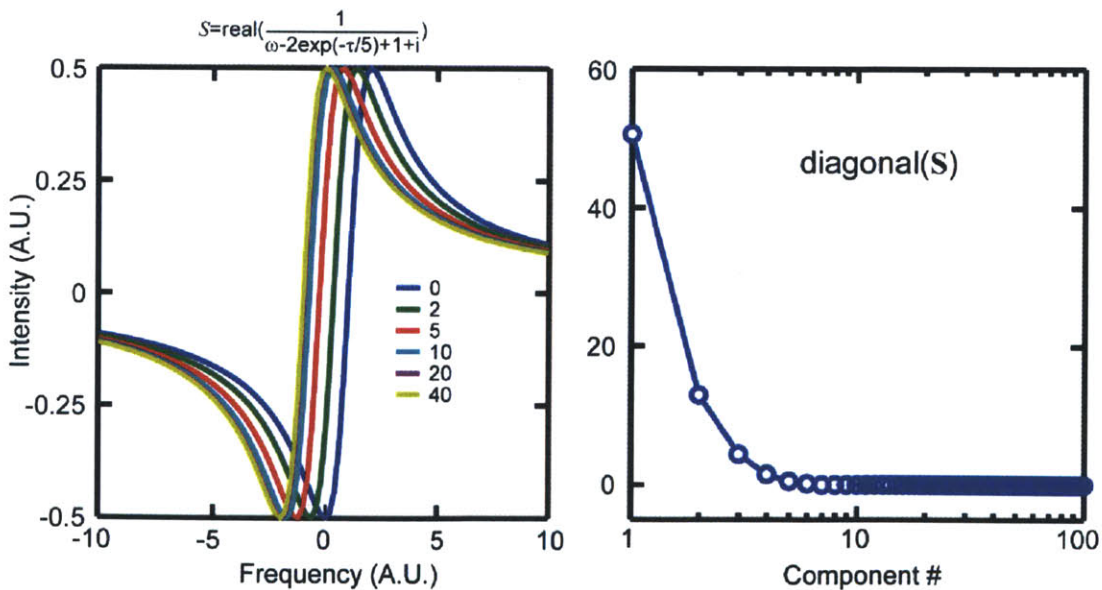


Figure 6.15: *Left* – Select spectra from the S peak-shift data. *Right* – The mathematical basis singular values resulting from the SVD analysis.

may be better suited in this case because it allows for consistent selection of the specified component spectra. In the SVD procedure, components above f are rejected and lost. In the projection decomposition, none of the data has been thrown away, which is advantageous if one is worried that important data is being thrown out with the noise. Similarly, projection analysis might be advantageous if the fitting for the SVD re-weighting procedure is not converging consistently across multiple data sets.

Projection decomposition has more flexibility than SVD, but its reliance on subjectively specified components can introduce higher risks in interpretation.

6.2.8 Shortcomings of SVD

SVD reduces data sets to the fewest number of possible components. The previous examples have not addressed the cases in which this reduction may lead to false interpretations,

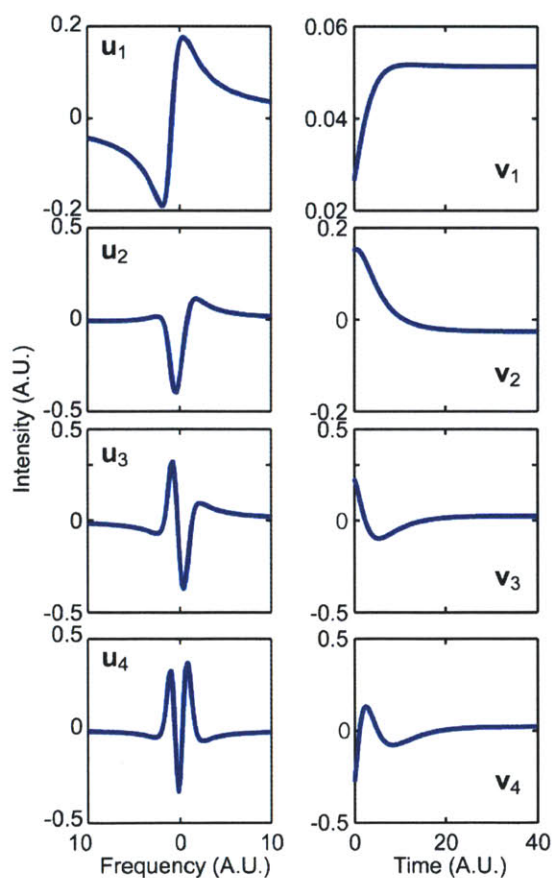


Figure 6.16: Mathematical basis U_{1-4} and V_{1-4} SVD results from the S peakshift data from Figure 6.15.

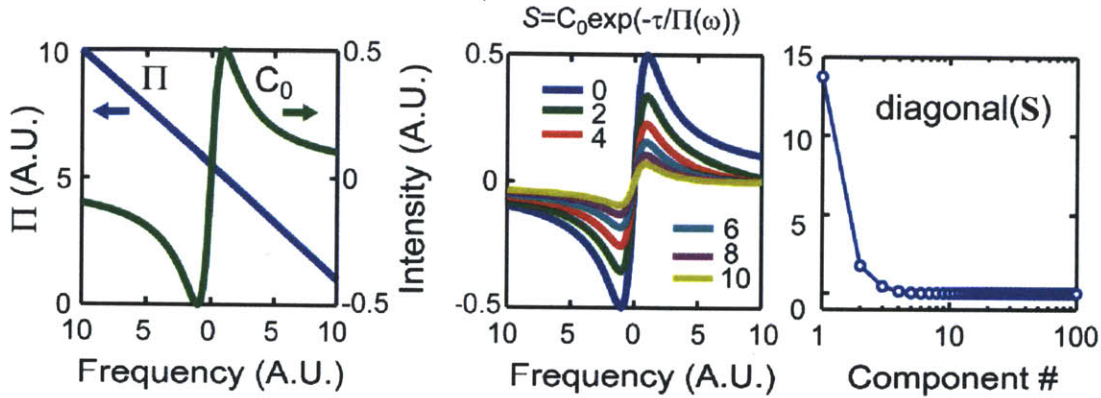


Figure 6.17: *Left* – The generating components for the center data. *Center* – The $S \Pi(\omega)$ data. The *Right* – The mathematical basis s singular values resulting from the SVD analysis.

especially when noise will set a limit on the number of significant components. As two examples, SVD was applied to 1) a data set that results from an exponential peak-shift and 2) data with a frequency-dependent exponential decay rate.

The peak-shift S is shown in Figure 6.15 - a Lorentzian peak shifts with an exponential dependence. Each spectrum at a given τ can be thought of as an intermediate basis u spectrum component with a delta function τ -dependence v . An exact description of this data would require the same number of components as the minimum dimension of S . SVD analysis reveals multiple components in Figure 6.16, although the first 2(3) s -values account for 90.5%(96.8%) of the sum of all the s -values, which are shown in Figure 6.15. The dominance of just a few SVD components, especially in the case of added noise, would indicate the presence of a few states, not the special peak-shifting phenomenon that is the actual cause of this data. Although this case is largely considered unphysical, generally, this example indicates that SVD may fail when components have similar spectra.

The S shown in Figure 6.17 result from a Lorentzian peak whose amplitude decays with a frequency-dependent exponential rate, $\Pi(\omega)$. Because each frequency decays at a different rate, an exact description of this data would require the same number of components as the minimum dimension of S . SVD analysis reveals multiple components in Figure 6.18, but the first 2(3) s -values account for 96.3%(99.0%) of the sum all the s -values, which are shown in Figure 6.17. As with the previous peak-shift case, the dominance of a few SVD components would lead to the interpretation that there are only a few states. Although the initial data seems unphysical and especially pernicious, it raises concerns over SVD's ability to separate states with similar time-

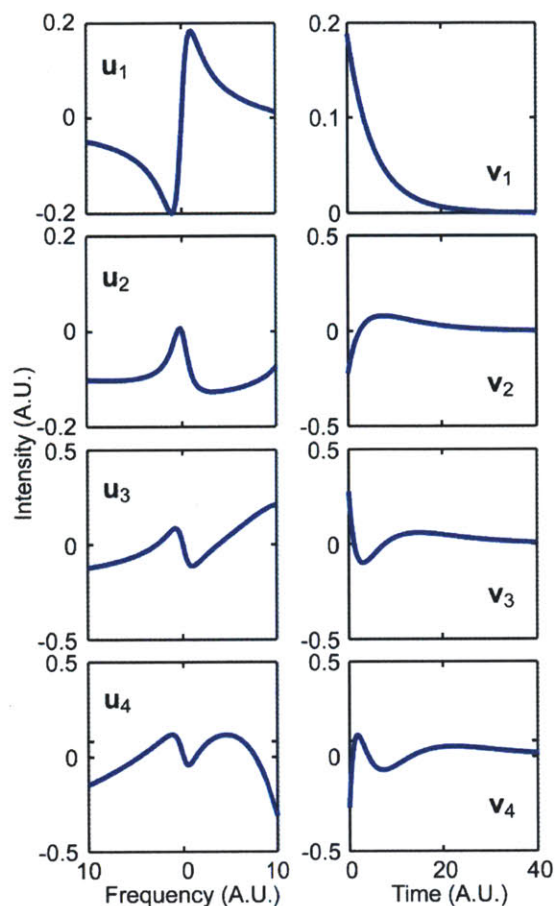


Figure 6.18: Mathematical basis \mathbf{U}_{1-4} and \mathbf{V}_{1-4} SVD results from the $S \Pi(\omega)$ data from Figure 6.17.

behavior.

Both of the generated S data sets above are noiseless. As a result, there are an equivalent number of nonzero components (although the singular values asymptotically approach zero) as there are spectra in S . The SVD mathematical basis spectra and τ -dependent traces, \mathbf{U} and \mathbf{V} , are orthonormal. As is observed in other orthonormal basis sets (Fourier transforms, wavefunctions), the number of nodes in the mathematical basis \mathbf{U} and \mathbf{V} increases with each component.

6.3 Conclusion

SVD is a mathematical tool for linearly decomposing spectral data. SVD is useful for removing noise and assessing the minimum number of significant, distinguishable basis spectra. Although they are mathematical results, the SVD basis components can be transformed to reveal physically interpretable results. This re-weighting is discussed for the special case of two meaningful components, particularly as applied to DPP and 2D IR results.

For the T-jump experiment, we make two assumptions: 1) $\Delta S(\omega, \tau < 5 \text{ ns})$ is equal to the difference spectrum between intra-well components, and 2) the time dependence of this difference spectrum is proportional to $T(\tau) - T(\text{eq})$. These assumptions appear to be reasonable based on our intuition and kinetic modeling. As shown with the peak-shift and frequency-dependent rate decay data, SVD can lead to oversimplified explanations, although perhaps considering unlikely situations such as these is an overcomplication to begin with. Analyzing SVD results requires care and proper assessment of the input assumptions.

6.4 Acknowledgment

I would like to thank Chunte Sam Peng and Carlos Baiz for their help in reading over this chapter (multiple times) and providing much needed feedback and suggestions.

6.5 References

- (1) G. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische Mathematik* **14**, 403-420 (1970).
- (2) R. I. Shrager, "Chemical transitions measured by spectra and resolved using singular value decomposition," *Chemom. Intell. Lab. Syst.* **1**, 59-70 (1986).
- (3) E. R. Henry, J. Hofrichter, and M. L. J. Ludwig Brand, "[8] Singular value decomposition: Application to analysis of experimental data," in *Methods in Enzymology* (Academic Press, 1992), pp. 129-192.
- (4) J. P. Hennessey and W. C. Johnson, "Information content in the circular dichroism of proteins," *Biochemistry* **20**, 1085-1094 (1981).
- (5) R. I. Shrager and R. W. Hendler, "Titration of individual components in a mixture with resolution of difference spectra, pKs, and redox transitions," *Anal. Chem.* **54**, 1147-1152 (1982).
- (6) J. Hofrichter, J. H. Sommer, E. R. Henry, and W. A. Eaton, "Nanosecond absorption spectroscopy of hemoglobin: elementary processes in kinetic cooperativity," *Proc. Natl. Acad. Sci. U. S. A.* **80**, 2235-2239 (1983).
- (7) S. Rajagopal, M. Schmidt, S. Anderson, H. Ihee, and K. Moffat, "Analysis of experimental time-resolved crystallographic data by singular value decomposition," *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 860-871 (2004).
- (8) H. S. Chung and A. Tokmakoff, "Visualization and characterization of the infrared active amide I vibrations of proteins," *J. Phys. Chem. B* **110**, 2888-2898 (2006).
- (9) A. W. Smith, J. Lessing, Z. Ganim, C. S. Peng, A. Tokmakoff, S. Roy, T. L. C. Jansen, and J. Knoester, "Melting of a β -Hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations," *J. Phys. Chem. B* **114**, 10913-10924 (2010).
- (10) H. S. Chung, M. Khalil, and A. Tokmakoff, "Nonlinear infrared spectroscopy of protein conformational change during thermal unfolding," *J. Phys. Chem. B* **108**, 15332-15343 (2004).
- (11) H. S. Chung, M. Khalil, A. W. Smith, Z. Ganim, and A. Tokmakoff, "Conformational changes during the nanosecond to millisecond unfolding of ubiquitin," *Proc. Natl. Acad. Sci. U. S. A.* **102**, 612-617 (2005).
- (12) A. W. Smith, H. S. Chung, Z. Ganim, and A. Tokmakoff, "Residual Native Structure in a Thermally Denatured β -Hairpin," *J. Phys. Chem. B* **109**, 17025-17027 (2005).
- (13) Adam W. Smith and A. Tokmakoff, "Probing Local Structural Events in β -Hairpin Unfolding with Transient Nonlinear Infrared Spectroscopy," *Angew. Chem., Int. Ed. Engl.* **46**, 7984-7987 (2007).
- (14) H. S. Chung, A. Shandiz, T. R. Sosnick, and A. Tokmakoff, "Probing the Folding Transition State of Ubiquitin Mutants by Temperature-Jump-Induced Downhill Unfolding," *Biochemistry* **47**, 13870-13877 (2008).
- (15) K. C. Jones, Z. Ganim, and A. Tokmakoff, "Heterodyne-detected dispersed vibrational echo spectroscopy," *J. Phys. Chem. A* **113**, 14060-14066 (2009).
- (16) Z. Ganim, K. C. Jones, and A. Tokmakoff, "Insulin dimer dissociation and unfolding revealed by amide I two-dimensional infrared spectroscopy," *Phys. Chem. Chem. Phys.* **12**, 3579-3588 (2010).
- (17) H. S. Chung and A. Tokmakoff, "Temperature-dependent downhill unfolding of ubiquitin. I. Nanosecond-to-millisecond resolved nonlinear infrared spectroscopy,"

- Proteins **72**, 474-487 (2008).
- (18) C. R. Baiz, C. S. Peng, M. E. Reppert, K. C. Jones, and A. Tokmakoff, "Coherent two-dimensional infrared spectroscopy: Quantitative analysis of protein secondary structure in solution," *Analyst* **137**, 1793-1799 (2012).
 - (19) R. I. Shrager, "Optical Spectra from Chemical Titration: An Analysis by SVD," *SIAM Journal on Algebraic and Discrete Methods* **5**, 351-358 (1984).

Appendix 6.A: List of Variables

S	matrix	Spectral data matrix (ω, γ) with columns consisting of spectra collected as a function of ω at a particular γ . Each row contains the γ -dependence of a particular frequency.
y	scalar	Maximum number of independent basis spectral components necessary for reproducing the data set, S . y is determined by the minimum dimension of S . Without noise, the number of nonzero y components will be less than or equal to g , the number of distinguishable species.
h	scalar	The number of independent species in the system, each of which has an independent, although not necessarily distinct, spectrum. The hypothetical maximum of h is defined by the number of non-interacting species. Usually, a non-interacting species is defined at the molecular level. If sub-domains within the molecule have independent spectra, then h could be as large as the number of sub-domains in the probed sample volume.
g	scalar	The number of non-zero components that result from SVD analysis applied to a noiseless data set S . ($g \leq h$). If two species have indistinguishable spectra or γ -dependence, then $g < h$. Even in this $g < h$ case, h species may still be identified by interpreting the data (i.e. a bi-exponential may indicate more than one species).
f	scalar	Number of distinguishable components that can be separated from the noise. $f \leq g$ (equality holds for noiseless data set). f sets a lower limit on the number of species in the system
Σ, Z	matrix	<i>State basis</i> set matrices representing each state of the system. Σ contains spectra $\{\sigma_1(\omega), \sigma_2(\omega), \dots, \sigma_h(\omega)\}$ and Z contains γ -dependence $\{\zeta_1(\gamma), \zeta_2(\gamma), \dots, \zeta_h(\gamma)\}$. For example, in a system with three states ($h = 3$), A,B,C, each of the three states would have an associated spectrum $\sigma(\omega)$ and γ -dependence $\zeta(\gamma)$. In hypothetical cases where a true σ is known, the calculated, predicted species' spectrum is denoted with quotation marks as " σ ." In real data sets, the predicted spectrum is denoted without quotation marks. σ' and ζ' indicates un-normalized components.

U,S,V	matrix	<i>Mathematical basis</i> set matrices that result from the SVD analysis of S . \mathbf{U} contains spectra $\{\mathbf{u}_1(\omega), \mathbf{u}_2(\omega), \dots, \mathbf{u}_y(\omega)\}$ and \mathbf{V} contains γ -dependence $\{\mathbf{v}_1(\gamma), \mathbf{v}_2(\gamma), \dots, \mathbf{v}_y(\gamma)\}$. \mathbf{S} is a diagonal rectangular matrix with singular values $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_y$.
U,V	matrix	<i>Intermediate basis</i> set matrices that result from a basis set transformation from the first f mathematical basis $\mathbf{U}, \mathbf{S}, \mathbf{V}$ components. In the general case, U contains spectra $\{u_1(\omega), u_2(\omega), \dots, u_f(\omega)\}$ and V contains γ -dependence $\{v_1(\gamma), v_2(\gamma), \dots, v_f(\gamma)\}$. For the special two component t-DPP and t-2D IR case described here for a ΔS data set, $U = \{u_n(\omega), u_\mu(\omega)\}$ and $V = \{v_n(\tau), v_\mu(\tau)\}$. U, V result from an intermediate number of assumptions about the system. For example, in the A,B,C state case described in Figure 6.11, u_n is approximately proportional to $\sigma_A - \sigma_A$ while u_μ is approximately proportional to $2\sigma_B - (\sigma_A + \sigma_A)$. Thus, U, V are generally still linear combinations of state basis Σ, Z , but the relationship linking the two is hypothesized. In hypothetical cases where a true u or v is known, the calculated, predicted components are denoted with quotation marks as " u " or " v ."
W	matrix	Transformation matrix that transforms \mathbf{US} to U and \mathbf{V} to V .
T	matrix	Transformation matrix that transforms U to Σ' and V to Z' .
L	matrix	Transformation matrix that transforms Σ' to Σ and Z' to Z .
o	scalar	Number of user-defined basis components that make up \mathbf{P}_ω and \mathbf{P}_γ . If the components are chosen poorly, then $o > f$. If the components are chosen wisely, then $o = f$.
$\mathbf{P}_\omega, \mathbf{P}_\gamma$	matrix	User-defined projection basis set matrices that describe S . \mathbf{P}_ω contains spectra $\{\mathbf{p}_\omega^1, \mathbf{p}_\omega^2, \dots, \mathbf{p}_\omega^o\}$ and \mathbf{P}_γ contains γ -dependence $\{\mathbf{p}_\gamma^1, \mathbf{p}_\gamma^2, \dots, \mathbf{p}_\gamma^o\}$. The basis set is defined entirely by us rather than through the SVD analysis and rejection of components. Because the columns of \mathbf{P}_ω and \mathbf{P}_γ are explicitly chosen to match predicted components, \mathbf{P}_ω and \mathbf{P}_γ are analogous to U and V respectively, although U and V are calculated from the mathematically generated $\mathbf{U}, \mathbf{S}, \mathbf{V}$.

Appendix 6.B: Special case of $\mathbf{u}_j \propto u_j$

Theorem

If the data decomposes into two components

$$S = \sum_{i=j,k} \mathbf{u}_i \mathbf{s}_i \mathbf{v}_i^T = \sum_{i=l,m} u_i v_i^T \quad (6.B.1)$$

and the spectrum of one of the SVD spectra (for our discussion $i=[1, 2]$ or $[n, \mu]$) is proportional to the spectrum of one of the 'true' underlying component spectra

$$\mathbf{u}_j \propto u_j \quad (6.B.2)$$

then \mathbf{v}_k is proportional to v_m :

$$\mathbf{v}_k \propto v_m \quad (6.B.3)$$

Proof

To show this, we first scale \mathbf{u}_i by its \mathbf{s}_i value:

$$\mathbf{u}'_i = \mathbf{u}_i \mathbf{s}_i \quad (6.B.4)$$

which when inserted into eq. (6.B.1) gives

$$S = \mathbf{u}'_j \mathbf{v}_j^T + \mathbf{u}'_k \mathbf{v}_k^T = u_l v_l^T + u_m v_m^T \quad (6.B.5)$$

Using the specified proportionality from eq. (6.B.2) and expressing the other \mathbf{u} and \mathbf{v}_k components as linear combination of u and v components scaled by factors A-G gives

$$\mathbf{u}'_j = E u_l; \quad \mathbf{u}'_k = F u_l + G u_m; \quad \mathbf{v}_j = A v_l + B v_m; \quad \mathbf{v}_k = C v_l + D v_m \quad (6.B.6)$$

Combining eqs. (6.B.5) and (6.B.6) gives

$$\begin{aligned} S &= E u_l (A v_l^T + B v_m^T) + (F u_l + G u_m) (C v_l^T + D v_m^T) \\ &= (EA + FC) u_l v_l^T + (EB + FD) u_l v_m^T + (GC) u_m v_l^T + (GD) u_m v_m^T \end{aligned} \quad (6.B.7)$$

Inspection allows us to place constraints on the constants based on the distinguishability of the l and m components:

$$EA + FC = 1; \quad EB + FD = 0; \quad GC = 0; \quad GD = 1. \quad (6.B.8)$$

The important result is that $C=0$, which shows that \mathbf{v}_k is proportional to v_m

$$\mathbf{u}'_j = E u_j; \quad \mathbf{u}'_k = F u_l + \frac{1}{D} u_m; \quad \mathbf{v}'_j = \frac{1}{E} v_l + B v_m; \quad \mathbf{v}'_k = D v_m \quad (6.B.9)$$

Intuitively, this result makes sense. Each u_i spectral component must be paired with a linear combination of its γ behavior. If \mathbf{u}_j consists purely of u_l , then the only place to pair u_m and v_m is in \mathbf{u}_k and \mathbf{v}_k , and \mathbf{v}_k must be proportional to v_m because there is nowhere else to 'correct' for mixing.

Appendix 6.C: Transformation matrices, \mathbf{W}

The choice of \mathbf{W} depends on the nature of the input criteria and the SVD components. \mathbf{W} 's used in the thesis isolate the floating scalar variable that is used to fit the SVD results, ζ , to the single specified criterion so that it affects only the specified component and its complement ($\mathbf{u}_1, \mathbf{v}_2$ and $\mathbf{u}_2, \mathbf{v}_1$ are paired):

$$\mathbf{W} = \begin{pmatrix} 1 & 0 \\ \zeta & 1 \end{pmatrix} \text{ with } \mathbf{W}^{-1} = \begin{pmatrix} 1 & 0 \\ -\zeta & 1 \end{pmatrix} \quad (6.C.1)$$

specifies that $u_1 = \mathbf{u}_1 \mathbf{s}_1 + \zeta \mathbf{u}_2 \mathbf{s}_2$, $u_2 = \mathbf{u}_2 \mathbf{s}_2$ (no change in shape), $v_1 = \mathbf{v}_1 \mathbf{s}_1$ (no change in shape), $v_2 = -\zeta \mathbf{v}_1 \mathbf{s}_1 + \mathbf{v}_2 \mathbf{s}_2$ while

$$\mathbf{W} = \begin{pmatrix} 1 & \zeta \\ 0 & 1 \end{pmatrix} \text{ with } \mathbf{W}^{-1} = \begin{pmatrix} 1 & -\zeta \\ 0 & 1 \end{pmatrix} \quad (6.C.2)$$

specifies that $u_1 = \mathbf{u}_1 \mathbf{s}_1$ (no change in shape), $u_2 = \zeta \mathbf{u}_1 \mathbf{s}_1 + \mathbf{u}_2 \mathbf{s}_2$, $v_1 = \mathbf{v}_1 \mathbf{s}_1 - \zeta \mathbf{v}_2 \mathbf{s}_2$, $v_2 = \mathbf{v}_2 \mathbf{s}_2$ (no change in shape). Therefore, the \mathbf{W} 's eq. (6.C.1) and (6.C.2) mix the two $\mathbf{U}_{1,2}$ to give one specified u_i . Inspection of eq. (6.C.1) and (6.C.2) shows the complementarity of $\mathbf{u}_1, \mathbf{v}_2$ and $\mathbf{u}_2, \mathbf{v}_1$.

Ultimately, the procedure depends on the specified criteria for re-weighting. If a particular re-weighting routine results in unphysical results, a different routine should be used.

If two components are specified, then a mixture of $\mathbf{W} = \mathbf{W}_a \mathbf{W}_b \dots$ can be used. In this consecutive case, \mathbf{W}_a should be found such that $\mathbf{U} \mathbf{W}_a$ or $\mathbf{W}_a^{-1} \mathbf{V}^T$ gives a desired intermediate component u_{des}^1 or v_{des}^1 ($\mathbf{U} \mathbf{W}_a = \{\dots, u_{des}^1, \dots\}$ or $\mathbf{W}_a^{-1} \mathbf{V}^T = \{\dots, v_{des}^1, \dots\}^T$). Then \mathbf{W}_b should be found such that $\mathbf{U} \mathbf{W}_a \mathbf{W}_b$ or $\mathbf{W}_b^{-1} \mathbf{W}_a^{-1} \mathbf{V}^T$ gives the second desired intermediate component without affecting the first ($\mathbf{U} \mathbf{W}_a \mathbf{W}_b = \{\dots, u_{des}^1, \dots, u_{des}^2, \dots\}$ or $\mathbf{W}_b^{-1} \mathbf{W}_a^{-1} \mathbf{V}^T = \{\dots, v_{des}^1, \dots, v_{des}^2, \dots\}^T$). Thus, if the i th column of \mathbf{W}_a gives the desired intermediate component, then the i th column of \mathbf{W}_b will be the same as the i th column from the identity matrix (zeros off diagonal, one on the diagonal).

A higher dimension \mathbf{W} might look something like this:

$$\mathbf{W} = \begin{pmatrix} 1 & 0 & 0 \\ \zeta_1 & 1 & 0 \\ \zeta_2 & 0 & 1 \end{pmatrix} \text{ with } \mathbf{W}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -\zeta_1 & 1 & 0 \\ -\zeta_2 & 0 & 1 \end{pmatrix} \quad (6.C.3)$$

where only the first column creates (transforms into) a constrained, desired intermediate component, while the other columns are equivalent to the identity matrix.

Appendix 6.D: Re-Weighting applied to structure determination

An example of the re-weighting routine can be applied to the structural determination of proteins, as is done in Ref. 18. . If we have a library of protein spectra, S , then applying SVD analysis will lead to separation into the mathematical basis U spectra, S scaling factors, and the V coefficients which indicate how much of each spectral component is contained in each protein spectrum:

$$USV^T = S \quad (6.D.1)$$

where $U=\{\mathbf{u}_1(\omega), \mathbf{u}_2(\omega), \dots\}$ is made up of columns of spectra, S is zero except for its diagonal singular value elements, and $V=\{\mathbf{v}_1(\pi), \mathbf{v}_2(\pi), \dots, \mathbf{v}_p(\pi)\}$ is made up of columns of coefficients where $\pi = 1$ is affiliated with the first protein and $\pi = p$ is affiliated with the last of the p proteins in the library. Using this notation, the spectra will have elements given by $S(\omega, \pi)$. The number of important components indistinguishable from noise, f , (as determined by the relative S values, for example) corresponds roughly to the number of separable structural elements. For example, if we assume that the β -sheet, α -helix, and random coil have unique spectra, then we would observe three spectral components whose spectra are a mix of the three unique 'pure' spectra associated with each structural component. If structural components overlap greatly, then we will observe a reduced number of components. If, as we expect due to coupling, the spectrum of a β -sheet in one protein is different than the spectrum of a β -sheet in another, then we expect to see an increased number of components.

If we observe three important components and we assume that the spectra can be decomposed into β -sheet, α -helix, and random coil spectra, then it follows that we should be able to decompose S into the state basis:

$$\Sigma Z^T = S \quad (6.D.2)$$

where, as in the main text, $\Sigma=\{\sigma^\beta(\omega), \sigma^\alpha(\omega), \sigma^r(\omega)\}$ with columns containing the 'pure' spectra of the secondary structures and $Z=\{\zeta^\beta(\pi), \zeta^\alpha(\pi), \zeta^r(\pi)\}$ with columns containing the fraction β -sheet, α -helix, and random coil contained in each protein π . Depending on the units of the σ and S spectra, the ζ fraction may have different units. We assume that the S are taken at the same concentrations, such that each ζ fraction can be written as a percentage such that

$\zeta^\beta(\pi) + \zeta^\alpha(\pi) + \zeta^r(\pi) = 1$. If the spectrum of protein $\pi = 1$ is at twice the concentration of protein $\pi = 2$, then $\zeta^\beta(1) + \zeta^\alpha(1) + \zeta^r(1) = 2[\zeta^\beta(2) + \zeta^\alpha(2) + \zeta^r(2)]$.

If we reject the components above $f = 3$, and we know the Z of our library of proteins, then we can calculate the \mathbf{W}^{-1} such that:

$$\mathbf{U}_{1,2,3} \mathbf{S}_{1,2,3} \mathbf{W} \mathbf{W}^{-1} \mathbf{V}_{1,2,3}^T \approx \mathbf{S} \quad (6.D.3)$$

with

$$\begin{aligned} \mathbf{W}^{-1} \mathbf{V}_{1,2,3}^T &\approx \mathbf{Z}^T \\ \mathbf{W}^{-1} &\equiv \mathbf{Z}^T \mathbf{V}_{1,2,3} \end{aligned} \quad (6.D.4)$$

It should be noted that the $\mathbf{W} \mathbf{W}^{-1}$ could be inserted between \mathbf{U} and \mathbf{S} , but this makes calculation of \mathbf{W}^{-1} harder (because \mathbf{S} must be inverted). We use \approx because only in the perfect case (noiseless, just three components) will the left and right sides of eq. (6.D.3) be equivalent. Once, \mathbf{W}^{-1} has been calculated, then Σ can be calculated through

$$\Sigma \approx \mathbf{U}_{1,2,3} \mathbf{S}_{1,2,3} \mathbf{W} \quad (6.D.5)$$

If the spectrum of a protein with unknown structure, $s^{(\text{unk})}$, is collected, the structure of the protein $Z^{(\text{unk})} = \{\zeta^\beta(\text{unk}), \zeta^\alpha(\text{unk}), \zeta^r(\text{unk})\}$ can be estimated by solving:

$$\Sigma (Z^{(\text{unk})})^T \approx s^{(\text{unk})} \quad (6.D.6)$$

As an example, the spectra of nine model proteins (8 known, 1 unknown) were first

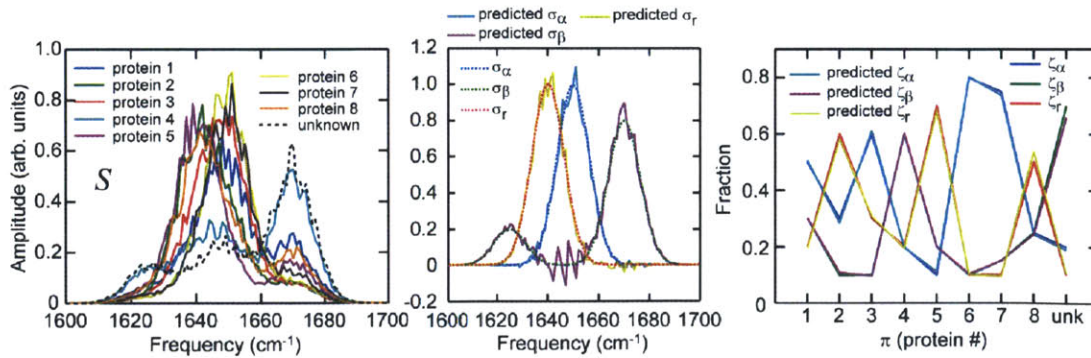


Figure 6.D.1: SVD is applied to eight generated spectra (left). The pure spectra associated with three secondary structural elements are predicted (center) based on the spectra of the proteins with known structures. The structural content of the eight proteins and an unknown spectrum are calculated from the SVD results (right).

generated by using eq. (6.D.2) as shown in Figure 6.D.1. Random noise ranging from 0.85 - 1.15 was multiplied into the spectra. The model random coil and α -helix spectra are gaussian peaks with linewidths $\sim 15 \text{ cm}^{-1}$ centered at 1640 and 1650 cm^{-1} , respectively. The β -sheet spectrum consists of two unequal amplitude gaussian peaks with $\sim 15 \text{ cm}^{-1}$ linewidths centered at 1625 and 1670 cm^{-1} . Despite the large noise and overlapping of the 'pure' spectra, the structure of the unknown protein is estimated with good accuracy (within $\sim 5\%$). In Figure 6.D.1, the slight mismatch between the predicted and known amplitudes of proteins 1-8 results from the \approx in eq. (6.E.4) - the spectra do not perfectly decompose into three parts because of the noise. Removing the noise results in perfect agreement between the real and predicted structure.

The real world application of this sort of analysis is done in Ref. 18. The real world application is complicated by the fact that the data S does not decompose into just three components. In this case, more than three components can be kept throughout the analysis – the number of secondary structure elements predicted might be fewer than the number of retained components. If we want to determine the b secondary structure components $Z^{(\text{unk})} = \{\zeta^1(\text{unk}), \zeta^2(\text{unk}), \dots, \zeta^b(\text{unk})\}$ while keeping f SVD mathematical basis components ($f > b$), then we can reduce the rank of the mathematical basis to f . In the example above $f = 3 = b$. In the $f > b$ case, substituting eq. (6.D.4) into eqs. (6.D.5) and (6.D.6) gives

$$\begin{aligned}
 \mathbf{W}^{-1} &= \mathbf{Z}_{\leq b}^T \mathbf{V}_{\leq f} \\
 \mathbf{U}_{\leq f} \mathbf{S}_{\leq f} [\mathbf{W}^{-1}]^{-1} &= \Sigma \\
 \Sigma^{-1} &= \mathbf{W}^{-1} \mathbf{S}_{\leq f}^{-1} \mathbf{U}_{\leq f}^T \\
 \mathbf{Z}_{\leq b}^{(\text{unk})T} &= \Sigma^{-1} s^{(\text{unk})} = \mathbf{Z}_{\leq b}^T \mathbf{V}_{\leq f} \mathbf{S}_{\leq f}^{-1} \mathbf{U}_{\leq f}^T s^{(\text{unk})}
 \end{aligned} \tag{6.D.7}$$

where \mathbf{W} does not have to be calculated to determine $Z^{(\text{unk})}$ – the last line can be used to bypass intermediate steps. \mathbf{W} is only necessary for determining Σ .

Appendix 6.E: Temperature dependence of [I(τ)] and [I'(τ)]

In this appendix, we take a closer look at the assumption that $[A^*(\tau)] - [A'(\tau)] \propto \Delta T(\tau)$. That is, the change of population within a free energy well, where A and A' are separated by barriers of $< kT$, is proportional to the change in temperature. This assumption is useful because we have already assumed that the intermediate basis nanosecond component spectrum is proportional to the difference of the spectra of A and A': $u_n \propto \sigma_A - \sigma_{A'}$. Based on this, the associated concentration dependence should also be proportional to the change in concentration of the species: $v_n \propto [A^*(\tau)] - [A'(\tau)]$. Comparing this assumption with the assumption above allows us to make the leap to $v_n \propto \Delta T(\tau)$.

This is a particularly powerful result for analyzing the T-jump t-DPP and t-2D IR data because it allows us to relate the change in temperature, which we know, to the concentration dependence of our assumed kinetic species. In order to assess the quality of this approximation, we begin by showing that $[A'(\tau)]/[A(\tau)] \propto C_1 + C_2 \Delta T(\tau)$.

6.E.1 Taylor expansions and system definition

Two Taylor expansions are used here:

$$\frac{1}{1+x} = \sum_{n=0}^{\infty} (-1)^n x^n, \text{ for } x < 1 \quad (6.E.1)$$

$$\approx 1 - x, \text{ for } x \ll 1$$

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad (6.E.2)$$

$$\approx 1 + x, \text{ for } x \ll 1$$

Following a T-jump, for the kinetic equilibration between two states A and A',



if k_1 and $k_{-1} > (5 \text{ ns})^{-1}$, then A and A' will always be in equilibrium with each other at long times, and

$$K(\tau) = \frac{[A'(\tau)]}{[A(\tau)]} = \exp\left(\frac{-\Delta G_{A'A}(\tau)}{RT(\tau)}\right) \quad (6.E.4)$$

will describe the ratio of A and A' populations at all times $\tau > 5 \text{ ns}$.

6.E.2 $[A'(\tau)]/[A(\tau)] \propto C_1 + C_2 \Delta T(\tau)$

If we assume that the Gibb's free energy is linear in temperature:

$$\Delta G_{AA'}(\tau) \approx \Delta H_{AA'} - T(\tau)\Delta S_{AA'} , \quad (6.E.5)$$

and that the relative T-jump is small compared to the equilibrium temperature, $\Delta T(\tau) \ll 1$, where:

$$\Delta T(\tau) = \frac{T(\tau) - T(\infty)}{T(\infty)} = \frac{T(\tau)}{T(\infty)} - 1 \quad (6.E.6)$$

then $K(\tau)$ can be approximated as:

$$K(\tau) \approx \exp\left(\frac{-\Delta H_{AA'}}{RT(\infty)} \frac{1}{1 + \Delta T(\tau)} + \frac{\Delta S_{AA'}}{R}\right) \quad (6.E.7)$$

Using eq (6.E.1), $K(\tau)$ can be further approximated to:

$$\begin{aligned} K(\tau) &\approx \exp\left(\frac{-\Delta H_{AA'}}{RT(\infty)} \{1 - \Delta T(\tau)\} + \frac{\Delta S_{AA'}}{R}\right) \\ &\approx K(\infty) \exp\left(\frac{\Delta H_{AA'} \Delta T(\tau)}{RT(\infty)}\right) \end{aligned} \quad (6.E.8)$$

Because $\max(\Delta T(\tau))$ is $\sim 3\%$ for a 10 K T-jump, as long as $\Delta H_{AA'}$ is $< \sim 30$ RT, then the argument of the exponential will be < 1 . If we further assume that $\Delta H_{AA'}$ is $\ll \sim 30$ RT, the result can be approximated using eq. (6.E.2):

$$K(\tau) \approx K(\infty) + K(\infty) \frac{\Delta H_{AA'} \Delta T(\tau)}{RT(\infty)} . \quad (6.E.9)$$

We define $K(\tau)$ as the sum of a constant term and τ -dependent term:

$$K(\tau) = K(\infty) + \Delta K(\tau) \quad (6.E.10)$$

with

$$\Delta K(\tau) \approx K(\infty) \frac{\Delta H_{AA'} \Delta T(\tau)}{RT(\infty)} . \quad (6.E.11)$$

Inverting eq (6.E.4) results in:

$$K^{-1}(\tau) = \frac{[A(\tau)]}{[A'(\tau)]} \approx K^{-1}(\infty) - K^{-1}(\infty) \frac{\Delta H_{AA'} \Delta T(\tau)}{RT(\infty)} . \quad (6.E.12)$$

Thus, based on the above assumptions, $[A'(\tau)]/[A(\tau)] \propto C_1 + C_2 \Delta T(\tau)$ (eq. (6.E.9)) if $\Delta H_{AA'}$ is $\ll \sim 30 RT$. As an illustration, the three-state kinetic model described in Appendix 6.F is used to test this conclusion. $K(\tau)$ is compared to eq. (6.E.9) in Figure 6.E.1. The small value of $\Delta H_{AA'} = 1.72 RT$ results in good agreement.

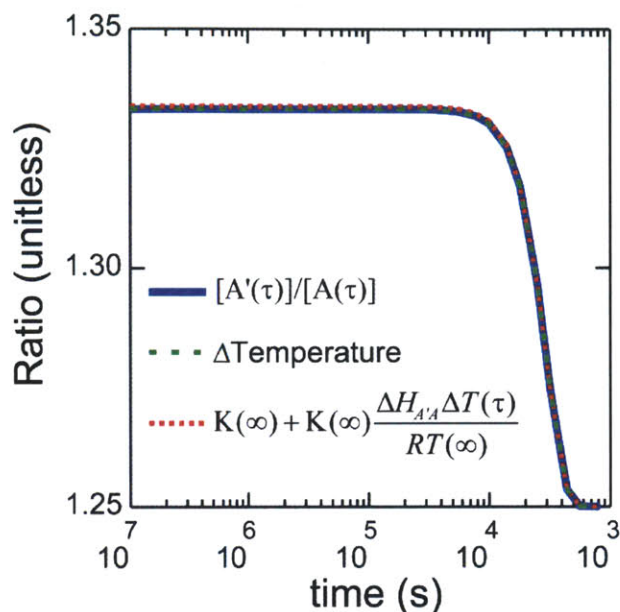


Figure 6.E.1: Eq. (6.E.9) is compared to the results from the three-state case kinetic model described in Appendix 6.F. The change in temperature is scaled and offset to best match the concentration ratio.

Even in the case where $\sim RT < \Delta H_{AA'} < \sim 30 RT$, the Taylor expansion of the exponential in eq. (6.E.8) leads to a series of decreasing terms that depend on $(\Delta T(\tau))^2$, $(\Delta T(\tau))^3$, ... Because $\Delta T(\tau)$ is either constant or monotonically decreasing, each of these terms will have the general shape of $\Delta T(\tau)$, with the differences in shape appearing in the non-constant, $\tau > 100 \mu s$ region. Thus, the sum of these similarly shaped terms, $K(\tau)$, will still approximately be proportional to $\Delta T(\tau)$. In the four-state example shown in the text, the $\Delta H_{AA'}$ is $\sim 22RT$, and the Taylor expansion approximates the real trace well only after including terms up to $\sim 2^{\text{nd}}$ order. Despite the high $\Delta H_{AA'}$, the $[A'(\tau)]/[A(\tau)]$ is still approximately proportional to $C_1 + C_2 \Delta T(\tau)$, and our approximations still give good results, as shown in Figure 6.13.

6.E.3 $[A'(\tau)] - [A'(\tau)] \propto \Delta T(\tau)$

Due to their fast interconversion rates, we group A and A' together, and define the sum of their concentrations as:

$$[AA'(\tau)] = [A(\tau)] + [A'(\tau)] . \quad (6.E.13)$$

Re-arrangement of the first part of eq. (6.E.4) to isolate $[A]$ or $[A']$ leads to:

$$[A(\tau)] = \frac{[AA'(\tau)]}{1+K(\tau)} \text{ and} \quad (6.E.14)$$

$$[A'(\tau)] = \frac{[AA'(\tau)]}{1+K^{-1}(\tau)} . \quad (6.E.15)$$

Defining

$$[I^*(\tau)] = [I(\tau)] - [I(\infty)] \quad (6.E.16)$$

leads to:

$$\begin{aligned} [A^*(\tau)] - [A^*(\infty)] &= \{[A(\tau)] - [A(\infty)]\} - \{[A'(\tau)] - [A'(\infty)]\} \\ &= [AA'(\tau)] \left\{ \frac{1}{1+K(\tau)} - \frac{1}{1+K^{-1}(\tau)} \right\} - \\ &\quad [AA'(\infty)] \left\{ \frac{1}{1+K(\infty)} - \frac{1}{1+K^{-1}(\infty)} \right\} \end{aligned} \quad (6.E.17)$$

appropriate multiplication by K/K leads to

$$[A^*(\tau)] - [A^*(\infty)] = [AA'(\tau)] \left\{ \frac{1-K(\tau)}{1+K(\tau)} \right\} - [AA'(\infty)] \left\{ \frac{1-K(\infty)}{1+K(\infty)} \right\} \quad (6.E.18)$$

On the right side of this equation, the second term is the equilibrium value at the initial (infinite time) temperature, whereas the first term carries the time dependence. This time dependence is the product of $[AA'(\tau)]$, the concentration within the $A+A'$ free energy well, and the $K(\tau)$ term, the ratio of the two populations.

If we assume that $\Delta K(\tau)/\{1+K(\infty)\} \ll 1$, then the first term on the right side of eq. (6.E.18) can be re-written using eq. (6.E.1) as:

$$\begin{aligned} [AA'(\tau)] \left\{ \frac{1-K(\tau)}{1+K(\tau)} \right\} &= [AA'(\tau)] \left\{ \frac{1-K(\infty) - \Delta K(\tau)}{1+K(\infty)} \frac{1}{1 + \frac{\Delta K(\tau)}{1+K(\infty)}} \right\} . \\ &\approx [AA'(\tau)] \left\{ \frac{1-K(\infty) - \Delta K(\tau)}{1+K(\infty)} \left(1 - \frac{\Delta K(\tau)}{1+K(\infty)} \right) \right\} \end{aligned} \quad (6.E.19)$$

To further separate, we define $[AA'(\tau)]$ as the sum of a constant and a τ -dependent term:

$$[AA'(\tau)] = [AA'(\infty)] + [\Delta AA'(\tau)] . \quad (6.E.20)$$

Substituting the results from eq. (6.E.19) and eq. (6.E.20) into eq. (6.E.18) and removing the $\{\Delta K/(1+K)\}^2$ terms (and higher order) gives:

$$\begin{aligned}
 [A^*(\tau)] - [A'^*(\tau)] \approx & [\Delta AA'(\tau)] \left\{ \frac{1-K(\infty)}{1+K(\infty)} \right\} \\
 & - [\Delta AA'(\tau)] \left\{ \frac{\Delta K(\tau)}{1+K(\infty)} \right\} \left\{ 1 + \frac{1-K(\infty)}{1+K(\infty)} \right\} \\
 & - [AA'(\infty)] \left\{ \frac{\Delta K(\tau)}{1+K(\infty)} \right\} \left\{ 1 + \frac{1-K(\infty)}{1+K(\infty)} \right\}
 \end{aligned} \tag{6.E.21}$$

This equation separates the results into three different terms with different τ -dependencies: one that is linear in $[\Delta AA'(\tau)]$, one that is the product of $[\Delta AA'(\tau)]$ and $\Delta K(\tau)$, and one that is linear in $\Delta K(\tau)$. We note that eqs. (6.E.9), (6.E.11), (6.E.19), and (6.E.21) assume $\Delta H_{AA'} \Delta T(\tau) / RT \ll 1$. Consistent with this previous assumption, inspection of eq. (6.E.11) shows that $\Delta K(\tau)$ should be proportional to $\Delta T(\tau)$.

If we assume that $[AA'(\infty)] \gg [\Delta AA'(\tau)]$ or that $K(\infty) \approx 1$, then eq. (6.E.21) will be dominated by the last term, and $[A^*(\tau)] - [A'^*(\tau)] \propto \Delta T(\tau)$. Stated more completely, for a 10°C T-jump, this assumption relies on $\Delta H_{AA'}$ is $\ll \sim 30$ RT and the change in A+A' concentration is small relative to its equilibrium value, $[\Delta AA'(\tau)] / [AA'(\infty)] \ll 1$.

As an illustration, the three-state kinetic model described in Appendix 6.F is used to test this conclusion. The left side, right side sum, and individual terms from the right side of eq. (6.E.21) are plotted in Figure 6.E.2. For this case, $[AA'(\infty)] > [\Delta AA'(\tau)]$, but $[AA'(\infty)]$ doesn't dominate. As a result, each of the three terms contributes appreciably to the total. As expected, the second term is the smallest because it is the product of two small values, $[\Delta AA'(\tau)]$ and $\Delta K(\tau)$. The first term, which carries $[\Delta AA'(\tau)]$ dependence, is proportional to $[B^*(\tau)]$ because for this three state case $[A^*(\tau)] + [A'^*(\tau)] + [B^*(\tau)] = 0$. The third term best approximates the $[A^*(\tau)] - [A'^*(\tau)]$ result, particularly at early and later times. In the analysis of the SVD results, we assume that $[A^*(\tau)] - [A'^*(\tau)] \propto \Delta T(\tau)$. Although in this case, this approximation is tenuous, the resulting separation of components shown in Figure 6.12 is remarkably successful. Thus, the approximation is justified by the results.

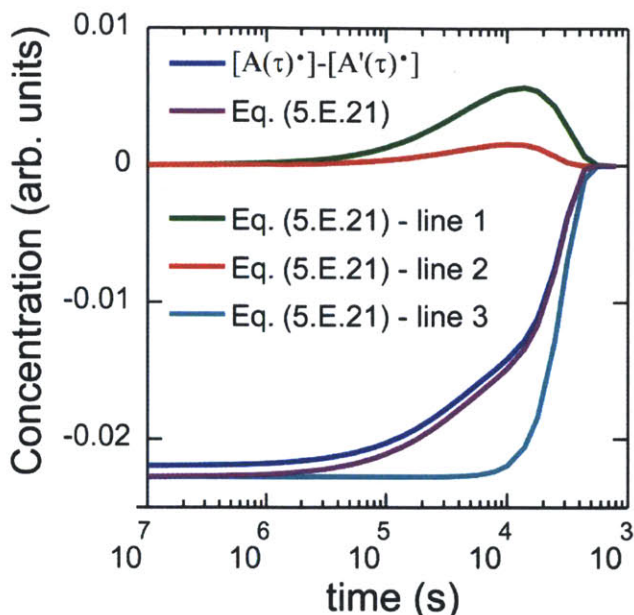


Figure 6.E.2: Eq. (6.E.21) is compared to the results from the three-state case kinetic model described in Appendix 6.F. Aside from the $[A^*(\tau)]-[A'(\tau)]$ plot, each trace is taken from the right side of the equation.

The above discussion has focused on analyzing the time dependence of two intra-well states. If our assumption holds, that the timescale of the difference between these two generic intra-well states is proportional to the change in temperature, then we can generalize to many states. For example, in the four state case described as an example in the main text, if $[A^*(\tau)]-[A'(\tau)] \propto \Delta T(\tau)$ and $[B^*(\tau)]-[B'(\tau)] \propto \Delta T(\tau)$, then $[A^*(\tau)]-[A'(\tau)] + [B^*(\tau)]-[B'(\tau)] \propto \Delta T(\tau)$. This can be extended indefinitely for more sets of intra-well states.

In fact, for the four-state case, our assumption that $[A^*(\tau)]-[A'(\tau)] + [B^*(\tau)]-[B'(\tau)] \propto \Delta T(\tau)$ is even better than in the three-state case. As described before, if the top right term of eq. (6.E.21) vanishes, the proportionality to temperature change becomes a better and better assumption (the top right term causes most of the departure from this proportionality). For the four state case, $[\Delta AA'(\tau)] = -[\Delta BB'(\tau)]$. Writing out the top right term of eq. (6.E.21) for $[A^*(\tau)]-[A'(\tau)] + [B^*(\tau)]-[B'(\tau)]$ leads to:

$$\begin{aligned}
 & [\Delta AA'(\tau)] \left\{ \frac{1 - K_{AA'}(\infty)}{1 + K_{AA'}(\infty)} \right\} + [\Delta BB'(\tau)] \left\{ \frac{1 - K_{BB'}(\infty)}{1 + K_{BB'}(\infty)} \right\} \\
 & = [\Delta AA'(\tau)] \left\{ \frac{1 - K_{AA'}(\infty)}{1 + K_{AA'}(\infty)} - \frac{1 - K_{BB'}(\infty)}{1 + K_{BB'}(\infty)} \right\}
 \end{aligned} \tag{6.E.22}$$

which, due to the subtraction of the BB' equilibrium constant term, will be smaller than the analogous term in the three-state case. We can even speculate further that if the intra-well states differ by H-bonding to the solvent induced by the temperature change, than the $K_{AA'}(\infty) \approx K_{BB'}(\infty)$ because the A to A' and B to B' transition are due to a similar mechanism. In this $K_{AA'}(\infty) \approx K_{BB'}(\infty)$ situation, the term shown in (6.E.22) will approach zero, and eq. (6.E.21) will be dominated by the temperature change proportional third line term.

Appendix 6.F: Kinetic Models

The three- and four-state kinetic models are calculated numerically for the two reactions shown below:



k_{ij} 's at initial and final temperature are specified. $\Delta G^\ddagger, \Delta S^\ddagger, \Delta H^\ddagger$ are then calculated to match the specified rates by rearranging the Eyring equation:

$$k_{ij} = \frac{k_B T}{h} \exp \left[\frac{-\Delta G_{ij}^\ddagger}{RT} \right] \quad (6.F.3)$$

with

$$\Delta G_{ij}^\ddagger = \Delta H_{ij}^\ddagger - T \Delta S_{ij}^\ddagger \quad (6.F.4)$$

Once $\Delta G^\ddagger, \Delta H^\ddagger, \Delta S^\ddagger$ are calculated, the rate constants as a function of temperature are calculated by plugging T back into eq. (6.F.3) and (6.F.4). The function for T (in K) is:

$$T(t) = 12 \exp \left[- \left(\frac{\tau}{300 \mu s} \right)^3 \right] + 308 \quad (6.F.5)$$

$T(\tau)$ is plotted in Figure 6.12 and Figure 6.13.

To numerically calculate the population of species $[I]$, we use first-order reaction kinetics propagated in discrete steps of $\tau_{i+1} - \tau_i = 10$ ns. For example, for $[A(\tau)]$:

$$[A(\tau_{i+1})] = \{-k_{AA'}(T)[A(\tau_i)] + k_{A'A}(T)[A'(\tau_i)]\}(\tau_{i+1} - \tau_i) + [A(\tau_i)] \quad (6.F.6)$$

The simulations are run twice: once with equally distributed initial concentrations and again using the equilibrated concentrations from the end of the first simulation as the initial concentrations for the second simulation. Initial (and τ -dependent) concentrations sum to 1.

The specified rates are:

<u>Model</u>		<u>$1/k_{AA'}$ (ns)</u>	<u>$1/k_{A'A}$ (ns)</u>	<u>$1/k_{A'B}$ (μs)</u>	<u>$1/k_{B'A}$ (μs)</u>		
3-state	T _i	40	50	75	60		
	T _f	30	40	67.5	67.5		
		<u>$1/k_{A'A}$ (ns)</u>	<u>$1/k_{AA'}$ (ns)</u>	<u>$1/k_{AB'}$ (μs)</u>	<u>$1/k_{B'A}$ (μs)</u>	<u>$1/k_{B'B}$ (ns)</u>	<u>$1/k_{BB'}$ (ns)</u>
4-state	T _i	40	60	70	30	35	61
	T _f	60	40	10	60	61	35

which, using RT at T = 308 K leads to:

<u>Model</u>		<u>$i,j=A,A'$</u>	<u>$i,j=A',A$</u>	<u>$i,j=A',B$</u>	<u>$i,j=B,A'$</u>		
3-state	ΔH_{ij}^\ddagger (J mol ⁻¹)	6.65RT	4.93RT	1.79RT	-4.16RT		
	ΔS_{ij}^\ddagger (mJ mol ⁻¹ K ⁻¹)	-18.8RT	-25.2RT	-59.1RT	-77.7RT		
		<u>$i,j=A',A$</u>	<u>$i,j=A,A'$</u>	<u>$i,j=A,B'$</u>	<u>$i,j=B',A$</u>	<u>$i,j=B',B$</u>	<u>$i,j=B,B'$</u>
4-state	ΔH_{ij}^\ddagger (J mol ⁻¹)	-11.8RT	9.79RT	50.9RT	-19.5RT	-15.8RT	13.8RT
	ΔS_{ij}^\ddagger (mJ mol ⁻¹ K ⁻¹)	-78.9RT	-9.96RT	100RT	-125RT	-91.4RT	-2.98RT

Although these numbers reproduce the temperature dependent rates and ratios of populations that we want, they are not physically reasonable or meaningful. For example, some enthalpies are negative and some are much higher than we would have expected.

Chapter 7

Equilibrium and Transient Unfolding of Ubiquitin

7.1 Introduction

The preceding Chapter 5 describes improvements in the collection and processing of transient two-dimensional (t-2D IR) spectra that leads to cleaner results undistorted by linear absorption effects. Chapter 6 presents Singular Value Decomposition (SVD) as an analysis technique that aids in the separation of meaningful spectra associated with different kinetic states.

The purpose of this chapter is twofold: to benchmark more recent techniques to those used in the past and to compare the conclusions from each technique. In particular, the downhill unfolding conclusions from previous transient 2D IR experiments are compared to conclusions from similar, recent experiments. The more recent experiments are not only processed differently, but the results are also interpreted by focusing on full spectral changes rather than individual frequency ranges, a shift facilitated by the SVD analysis.

In this chapter, the described linear absorption corrections and SVD analysis are applied to the collection of equilibrium 2D IR, t-transmission, transient heterodyned dispersed

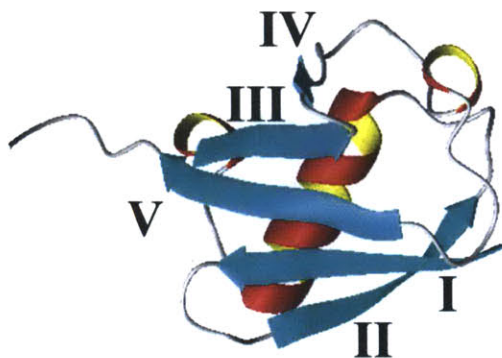


Figure 7.1: The structure of ubiquitin from PDB file 1UBQ¹. The β -strands are labeled.

vibrational echo (t-HDVE), and t-2D IR spectra of ubiquitin, which has been extensively studied previously by the Tokmakoff group using equilibrium and transient dispersed vibrational echo (t-DVE) and t-2D IR spectroscopy²⁻⁷. The results presented here are limited to a single temperature, a single concentration, and a single data collection. A full analysis of ubiquitin's folding pathway and thermodynamics would require temperature dependence and increased averaging. Rather, this work is presented to compare the results of previous and current experimental techniques and to illustrate the potential of the SVD analysis. Specifically, conclusions based on previous t-DVE and t-2D IR results can be assessed based on analyzing similar data sets processed using newer methods that better isolate the molecular responses. First, previous and recent data sets are assessed for consistency using the same data processing techniques. Then, the conclusions from the previous analysis are compared to conclusions from the improved analysis of recent data.

As described here and in Chapter 5, the linear absorption introduces spurious peaks and changes that complicate analysis. Once the linear absorption is corrected, the t-DPP and t-2D IR ubiquitin results can be interpreted using multi-exponentials and four kinetic states (folded, broken solvent/protein H-bonds, partially unfolded, and unfolded), a picture that does not require invoking the downhill unfolding phenomenon. This newer interpretation results after both correcting for the linear absorption and assigning full spectra to different states rather than focusing on individual frequencies.

7.1.1 Background

Ubiquitin is a 76 residue, α/β protein that exists in almost all eukaryotic cells, where it

tags other proteins to mark them for destruction. The structure of ubiquitin is known from NMR¹ and it is shown in Figure 7.1 - five β -strands form a β -sheet that surrounds an α -helix in a β -grasp motif. Its small size, commercial availability, and stability have made ubiquitin the focus of molecular dynamics simulations⁸ and protein folding experiments, where its observed ms relaxation rates have made it a target for mixing⁹⁻¹⁵, H/D exchange¹⁶, and T-jump experiments^{17, 18}. The N-terminal β -hairpin formed from β -strands I and II is expected to be a site for folding nucleation based on its high stability¹⁹.

Due to the large amount of background published experimental and theoretical work, its availability, its reversible folding properties, its resistance to aggregation, and its sub-50 ms kinetics, ubiquitin was studied extensively in the Tokmakoff group, where it is viewed as a model small protein. Experiments and theory were aimed at understanding ubiquitin's folding dynamics with the hope that the results can be generalized towards understanding small protein folding.

Ubiquitin was the focus of the first published T-jump t-DVE experiments in the Tokmakoff group, where the third order nonlinear homodyne signal was measured as a function of time-delay between the probing IR beams and the T-jump pump. From the T-jump t-DVE results, probe-dependent, fast, and stretched exponential kinetics were interpreted as signatures of downhill unfolding³. In standard Arrhenius kinetics, the exponential rate of interconversion between two states is determined by the height of the separating barrier. In the proposed downhill unfolding scenario, the majority of population unfolds through Arrhenius kinetics. A small sub-population initially at the top of the barrier, however, unfolds in a ballistic fashion that maps out the pathway for unfolding as its structures sequentially unravel. The downhill folding scenario is shown in Figure 7.2. In the t-DVE experiments, stretched exponentials were observed and may be explained by downhill folding, where the absence of clear barriers may lead to the convolution of multiple timescales²⁰. The $<100 \mu\text{s}$ timescales are also close to the fastest timescales expected for proteins of ubiquitin's size²¹, which suggests barrierless motion. Lastly, different, decreasing timescales were observed for loss of the α -helix/coil region and the two predominant β -sheet modes, v_{\parallel} and v_{\perp} , whose dipoles are aligned parallel and perpendicular to the strand. This probe-dependency indicates the existence of intermediate states, and it was interpreted with molecular dynamics simulations and statistical models to indicate sequential unfolding of β -strands III, IV, V, followed by unraveling of strands I and II. The results from

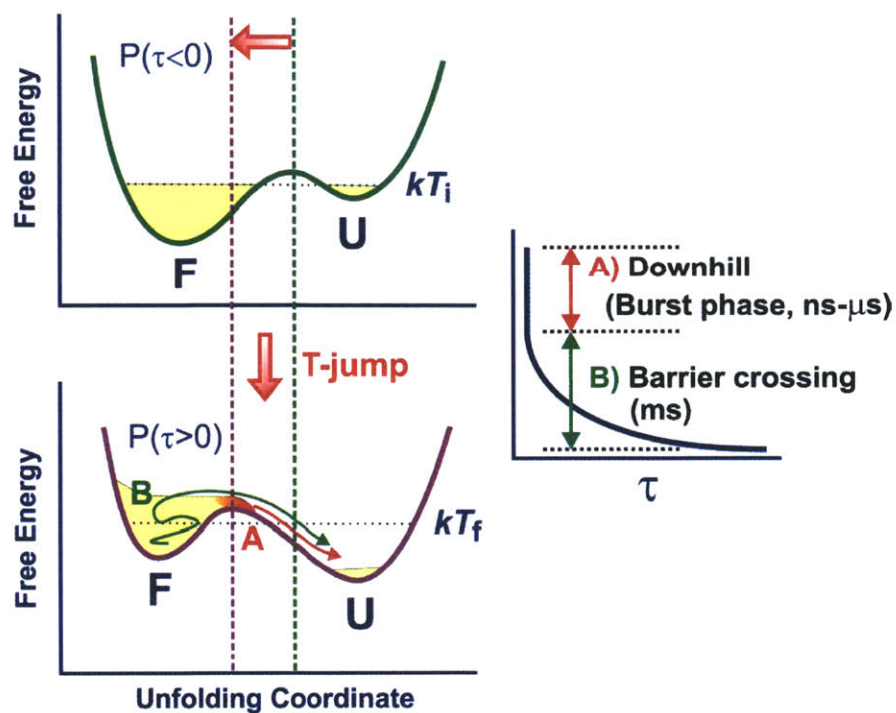


Figure 7.2: Cartoon of the downhill unfolding scenario. After the T-jump from T_i to T_f , the free energy surface changes. A small sub-set of structure (A) moves ballistically from the folded (F) to the unfolded (U) well on the ns- μ s timescale. The majority of the net movement from F to U occurs through ms barrier crossing in the form of B. The figure is reproduced from Ref. 2.

the t-DVE experiments were re-affirmed with t-2D IR spectra, and further t-DVE measurements were taken at variable T_i , T_f , and ΔT and with different ubiquitin mutants⁴. All of the experiments were interpreted within the framework of the downhill unfolding scenario with strands III-V unraveling first.

7.2 Experimental

Briefly, as described previously, DVE is a homodyne measurement, while 2D IR and HDVE are heterodyned experiments in which the 3rd order signal is interferometrically detected by overlapping it with a reference pulse. DVE measurements scale with the concentration squared while HDVE experiments scale linearly. DVE and HDVE measurements are theoretically single-shot measurements in which the signal at one $\tau_1=0/\tau_2=0$ or 150 fs delay is collected after dispersing it off a grating to reveal ω_3 . In 2D IR measurements, the signal is collected along ω_3 at many τ_1 . In the transient experiments, the DVE, HDVE, or 2D IR spectra are collected at time-delays following a 5-6 ns 10-12°C T-jump.

The previously published t-DVE data was generated with three time-coincident beams in a boxcar geometry electronically delayed relative to the T-jump laser. As a DVE measurement, there was no local oscillator reference, and the intensity was collected in homodyne mode. The t-2D IR data was collected similarly as was described in Chapter 4, although there was no shot-normalization and no correction for linear absorption. The previous t-DVE data was collected at $T_i=58^\circ\text{C}$ with a 33 mg/mL ubiquitin sample dissolved in a DCl in D_2O solvent at $\text{pH}^*=1$. The t-2D IR data was collected at zero waiting time and $T_i=63^\circ\text{C}$ with a 30 mg/mL ubiquitin sample dissolved in a 0.35% DCl in D_2O (w/w) solvent at $\sim\text{pH}^*=1$.

The recent equilibrium 2D IR data was collected on a 20 mg/mL ubiquitin sample dissolved in 0.35% DCl in D_2O (w/w) solvent. Rephasing and non-rephasing spectra were stepped to 2.5 and 1.5 ps, respectively, in 4 fs steps.

The recent t-HDVE and t-2D IR spectra were collected at a waiting time $\tau_2=150$ fs, and a $T_i=63^\circ\text{C}$ with a 15 mg/mL ubiquitin sample dissolved in a 0.35% DCl in D_2O (w/w) solvent. Rephasing and non-rephasing spectra were stepped to 2.4 and 1.4 ps, respectively, in 14 fs steps. The t-HDVE data results from 8 averages with 10000 shots per chopped and un-chopped spectrum. The t-2D IR spectra are the result of 1 average with 7000 shots per per chopped and un-chopped spectral slice. The t-transmission spectra are collected incidentally and simultaneously with the t-HDVE data.

After using FTSI of $\tau_{LO}=0:5:25$ fs spectra to calculate the real and imaginary spectra, the recent t-HDVE data is presented in two forms, as t-DVE (amplitude squared) and ω_3 -corrected t-DPP. The recent t-2D IR is also presented in two forms, as uncorrected and ω_3 -corrected. For ease of calculation, the ω_3 -corrected form differs slightly from that presented in Chapter 5. Namely, the data presented here comes from eq. 5.10 without multiplication of the $[e_{LO}(\omega_3)]^2$ term. ω_1 -corrections were not applied and are expected to result in small changes to the spectra.

All the ubiquitin samples were purchased from Sigma (U6253), H/D exchanged, and lyophilized before data collection. All polarizations were ZZZZ.

7.3 Results and Discussion

7.3.1 Equilibrium 2D IR spectra

Recent equilibrium 2D IR spectra of ubiquitin at 63 and 75°C are shown in Figure 7.3. Examination of these equilibrium data help to interpret the changes from the transient

experiments. The spectra show 1580 and 1710 cm^{-1} features assigned to arginine and deuterated carboxylic acid side-chains, respectively. The 75°C spectrum shows gain of a peak in the 1610 cm^{-1} region, which is attributed to aggregation. As the temperature is raised, the 1635 cm^{-1} ν_{\perp} diagonal peak is lost, intensity grows in the 1660 cm^{-1} random coil region, and the $\nu_{\perp}/\nu_{\parallel}$ cross peak ridge along $\omega_3=1680 \text{ cm}^{-1}$ is lost, all of which indicate melting of β -sheet structure and gain of random coil. These changes are reflected in the difference spectrum, in which loss is most apparent as a blue/red (above/below the diagonal) doublet in the 1635 cm^{-1} ν_{\perp} region. Gain appears in the off diagonal region as a red peak above the diagonal and a blue peak below the diagonal at $\omega_1=1650 \text{ cm}^{-1}$. There are also subtle overlapping β -sheet loss features that appear along the diagonal in the 1680 cm^{-1} ν_{\parallel} and $\nu_{\perp}/\nu_{\parallel}$ cross-peak $\omega_3=1635 \text{ cm}^{-1}$ regions.

The difference spectrum is similar to the 2nd component SVD spectrum also shown in Figure 7.3, although the features have the opposite sign due to the monotonically decreasing nature of the normalized SVD 2nd component ‘melting curves’ shown in Figure 7.4. The \mathbf{u}_2 spectrum represents the major changes in the spectra over the 10-70°C temperature range from

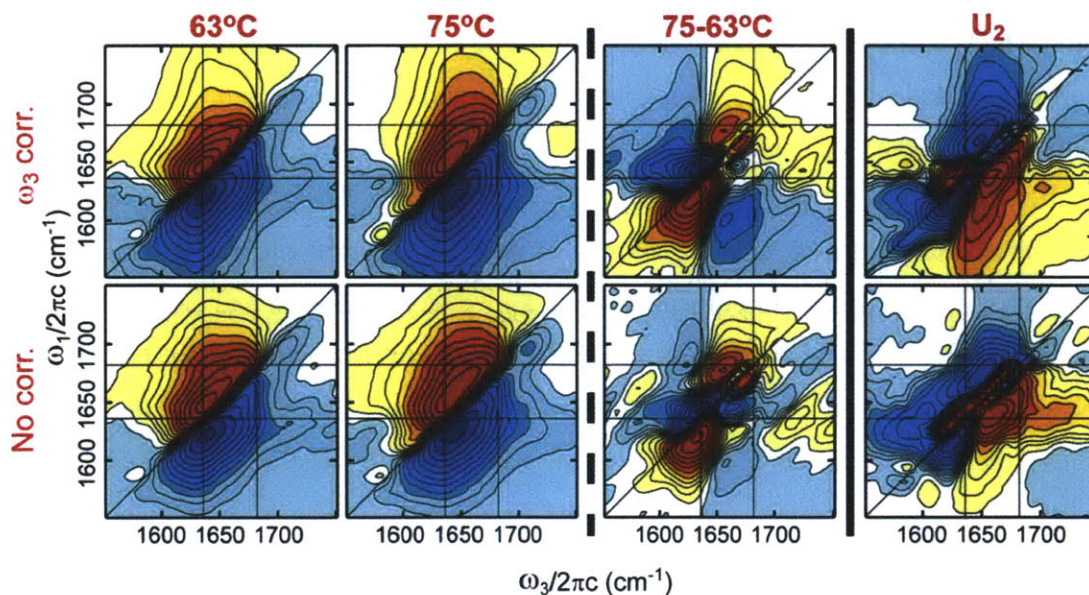


Figure 7.3: Column 1+2 – The equilibrium 2D IR ZZZZ spectra at 63 and 75°C. Column 3 – The difference spectrum 75-63°C. Column 4 – The 2nd SVD component spectrum that results from analysis of 10 2D IR spectra ranging from 20-70°C. Row 1 – After ω_3 correction. Row 2 – Without linear absorption correction. The spectra are individually normalized and scaled such that 25 linearly-spaced contours span the entire magnitude range of $\sinh^{-1}(S_{Norm.} \times 27.29)$.

which the data was analyzed. Higher temperatures were not used in the SVD analysis to avoid the aggregation observed in the 1610 cm^{-1} peak. Because it analyzes the changes in more spectra over a wider temperature range, the 2nd SVD component spectrum may be a better representation of the spectral features of melting, but it is a mathematical result which will also miss some of the non-two-state changes. Ultimately, the similarities between the difference spectrum and the \mathbf{u}_2 spectrum indicate that they are both reasonably accurate representations of the changes associated with melting. The SVD melting curve indicates a melting temperature $\sim 65^\circ\text{C}$ as has been observed previously for this pH³.

As with transient data, the temperature dependent linear absorption of the solvent and solute will affect equilibrium spectra. As is observed in Figure 7.3, the uncorrected 2D IR spectra differ from the ω_3 -corrected spectra. The uncorrected 63°C spectrum has a more prominent v_\perp/v_\parallel cross peak ridge along $\omega_3=1680\text{ cm}^{-1}$ and the major diagonal peak at 75°C is at a higher frequency than the peak in the corresponding ω_3 -corrected spectrum. There are also subtle changes in the difference and \mathbf{u}_2 spectra, although all features seem to be consistent albeit with different amplitudes and linewidths. The differences may be attributed to artificial weighting imposed by the limited Gaussian spectral bandwidth of the IR pulses. ω_3 -correction helps to remove some of this weighting, particularly due to linear absorption of ubiquitin, whose

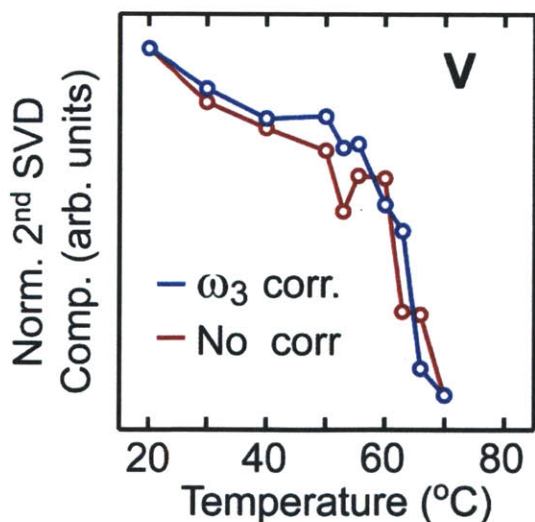


Figure 7.4 Normalized 2nd SVD component temperature dependent ‘melting curve’ that results from analysis of the 2D IR spectra. The complementary spectra are shown in Figure 7.3.

amide I peak is centered around 1640 cm^{-1} . Regardless of linear absorption, both the uncorrected and ω_3 -corrected spectra are shaped by the spectrum of the pumping IR beams, which will exaggerate the nonlinear response at the center of the IR spectrum, which for these experiments is around 1660 cm^{-1} .

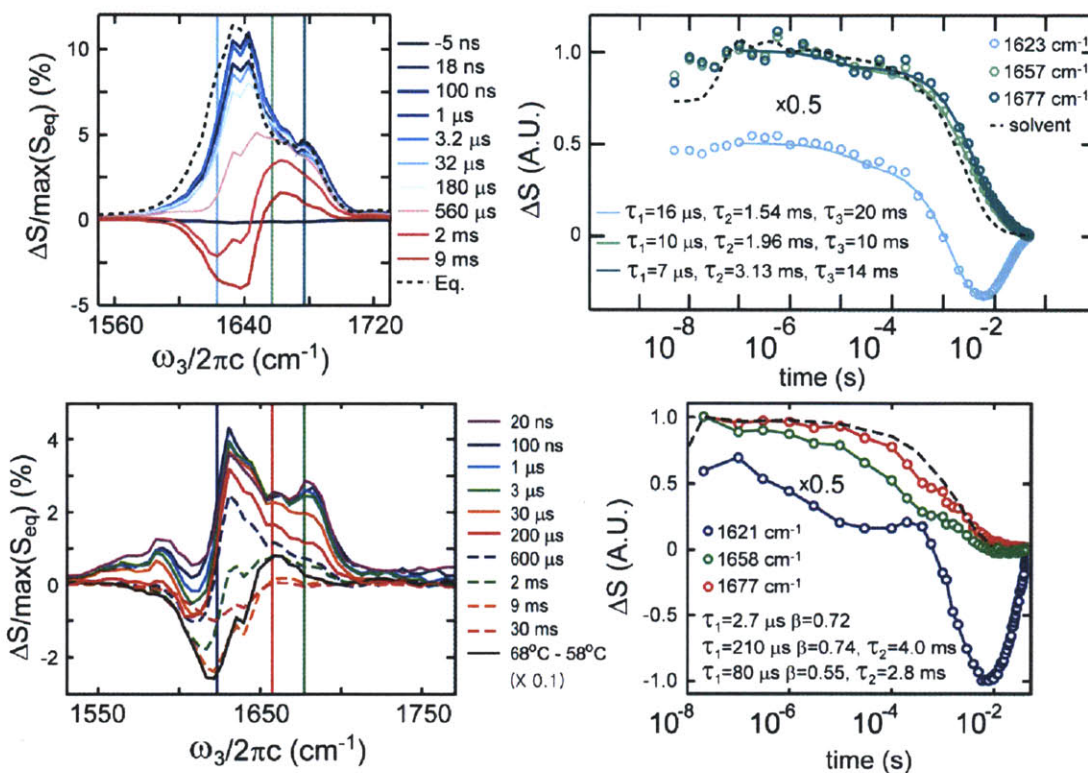


Figure 7.5: *Column 1*– Selected delays from the difference t-DVE spectra. *Column 2* – Normalized traces from the indicated frequencies. *Row 1* – Data collected at $T_i=63^\circ\text{C}$, 15 mg/mL using t-HDVE without linear absorption correction. The traces were fit to tri-exponentials without allowing stretched exponentials. *Row 2* – Data collected at $T_i=58^\circ\text{C}$, 33 mg/mL using t-DVE and modified from Ref. 3.

Based on the 2D IR equilibrium spectra, after the T-jump, we expect to see loss of the ν_{\perp} and ν_{\parallel} peaks and gain of an underlying random coil peak.

7.3.2 Transient DVE spectra

The previous $T_i=58^\circ\text{C}$, 33 mg/mL ubiquitin t-DVE results are shown on the bottom row of Figure 7.5. The top row contains the t-DVE (actually calculated from t-HDVE) results from the recent $T_i=63^\circ\text{C}$, 15 mg/mL ubiquitin experiments. The differences in concentration, initial temperature, and the IR pulse spectra make direct comparison difficult, but general features are

conserved. Based on conclusions from Chapter 5, results from uncorrected t-DVE are the least reliable form of 2D IR spectroscopy not only because of distortions due to the linear absorption, but also because it is an amplitude squared measurement that is not linear in the concentration. The nonlinear scaling can magnify changes in the spectrum in non-intuitive ways.

The major difference between the spectra is in the 1610 cm^{-1} region, where the previous t-DVE spectra show a dip that is not clearly apparent in the more recent data. Given its frequency and its absence in the lower concentration, recent data, this dip might be explained as a peak due to aggregation, which might melt away following the T-jump to cause the negative peak. Relative to other spectral features, the magnitude of this feature, particularly in the spectra taken at higher T_i ⁴ (data not shown here) is larger than would be expected for a small aggregation peak. Instead, a more likely cause of the spectral difference is the IR pulse spectrum. It appears to have been shifted redder in the previous experiments relative to the more recent experiments. This is apparent in the equilibrium spectra, in which the ratio of ν_{\perp} to ν_{\parallel} peak intensities is 4:1 in the previous data and 2:1 in the more recent data. The bluer pumping spectrum in the recent experiment shifts intensity to higher frequencies, which explains the low intensity on the red side of the recent spectra and may explain the absence of the 1610 cm^{-1} feature. Although the previous data appears to show smaller changes (5% vs >10%) The magnitude of the normalized signal is highly dependent on the maximum of the equilibrium spectrum, which again depends on the IR pulse spectrum.

The general features of the frequency time traces are conserved between the previous and recent data. Namely, all the frequency traces deviate from the temperature profile. The 1621 cm^{-1} traces both start positive, drop negative, and return back to equilibrium after 5-10 ms. The 1621 cm^{-1} non-monotonic behavior was previously attributed to competing effects: the increased temperature causes an increase in the solvent transmission at early times, melting of the protein decreases the intensity below zero as the temperature drops to the initial value, and refolding causes the return to equilibrium. The 0.1 - 100 μs time range from the previous 1621 cm^{-1} trace was fit to a stretched exponential with $\tau=2.7\text{ }\mu\text{s}$ and $\beta=0.72^3$. The full recent 1623 cm^{-1} time trace is fit to a tri-exponential. The poor fit at early times suggests stretched exponential behavior, which is evident in the sloping decrease between 100 ns and 100 μs . Fitting the recent data over 0.1 - 100 μs with a single stretched exponential gives $\tau=2.7\text{ }\mu\text{s}$ and $\beta=0.78$. The most unusual feature of the previous 1621 cm^{-1} trace is the apparent increase at $\sim 200\text{ }\mu\text{s}$. Although

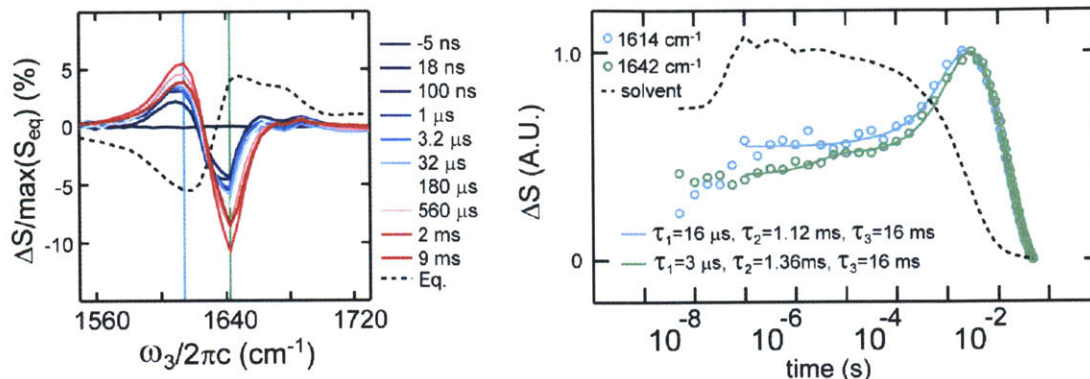


Figure 7.6: ω_3 -corrected t-DPP spectra calculated from t-HDVE.

this feature is not observed in the recent data at $T_i=63^\circ\text{C}$, the recent data stretched drop in the sub-ms time region is consistent with separate previous data sets taken at $T_i=58$ and 63°C ⁵, which also display no 200 μs increase.

The recent random coil 1657 cm^{-1} and $\nu_{\parallel} 1677\text{ cm}^{-1}$ traces deviate from the temperature profile and exhibit a stretched drop in magnitude during the 0.1 – 100 μs time range, similar to the previous traces. Although the previous data traces drop away faster than the temperature while the recent traces are slower, these differences may be consistent with an increase in timescales for these frequencies with increasing T_i , as was observed in previous work⁴.

Although the previous and recent data sets cannot be directly compared, the recent spectra contain stretched exponential behavior in the 0.1 – 100 μs time range. This behavior can be interpreted as downhill unfolding. To test if this unusual behavior might be the result of experimental artifacts such as nonlinearity in concentration or linear absorption distortions, the data is presented in different, less distorted representations in the following sections.

7.3.3 Transient DPP spectra

The recent ω_3 -corrected t-DPP spectra taken at 15 mg/mL, $T_i=63^\circ\text{C}$ are shown in Figure 7.6. These t-DPP spectra are calculated from the same data set as the t-DVE data shown on the top row of Figure 7.5. Due to differences in data acquisition, previous t-DPP spectra complementing the previous t-DVE are not available for comparison.

The equilibrium DPP spectrum at $T_i=63^\circ\text{C}$ shows three main features that can be attributed to overlapping pairs of negative/positive doublets associated with the two ν_{\perp} and ν_{\parallel} features. The ν_{\perp} doublet is stronger and appears at lower frequency than the ν_{\parallel} doublet, whose main feature is the positive peak at 1680 cm^{-1} . The low frequency negative peak of the

equilibrium HDVE-reconstructed DPP shown in Figure 7.6 is broader than the 2-beam DPP spectrum collected prior to T-jump data collection, which suggests that there is some aggregation or that the equilibrium spectrum shown is not entirely re-equilibrated following the 50 ms delay after the T-jump. The measured long-time behavior time constant of 17 ms (discussed later) suggests that 5% of the T-jump-affected population has not equilibrated in the 50 ms timescale. We expect that the system quickly reaches a steady state equilibrium, in which there is slightly more population in the unfolded well than is expected at $T_i=63^\circ\text{C}^4$. Although the populations will be affected by this non-ideal steady-state, time scales $\ll 50$ ms should be unchanged.

The difference t-DPP spectra shown in Figure 7.6 show strong v_\perp and weak v_\parallel loss features on the pulse-width limited <5 ns timescale, as is shown in the 18 ns difference spectrum. This spectrum shows loss features on the red sides of each of the equilibrium peak doublets, and a possible gain of intensity on the blue side of the v_\parallel peak at 1690 cm^{-1} . At $\sim 560\ \mu\text{s} - 9\text{ ms}$, the 1640 cm^{-1} negative feature blue-shifts, possibly due to the increase in population of random coil, melted structures whose overtone is in this frequency region. On the >0.5 ms timescale, there are few changes in the $>1650\text{ cm}^{-1}$ region, suggesting canceling out due to interference between loss of the v_\parallel peaks and gain of the random coil peaks. The dominance of loss features over gain is consistent with our work with TZ2, and it suggests that at higher temperature there are changes in dipole strength and vibrational mode character that result in decreased amplitudes.

The time traces for the 1614 and 1642 cm^{-1} t-DPP time traces are shown on the right in Figure 7.6. As described in the discussion above, the traces show a pulse-width limited change followed by a ~ 1.2 ms relaxation away from equilibrium and a final ~ 16 ms re-equilibration. The 1642 cm^{-1} trace shows some rise (actually decay before normalization) in the $0.1 - 100\ \mu\text{s}$ time period, but this rise seems to mirror the drop in the temperature profile, which is monitored by the solvent transmission. The noise in the solvent transmission during this timescale seems to be mirrored in the DPP 1642 cm^{-1} trace. The $<100\ \mu\text{s}$ shared behavior in the solvent transmission and DPP suggests that the cause of this behavior may possibly be convolution of the stretched exponential temperature profile with the protein response, as discussed in Chapter 3.

We can describe the DPP in terms of two responses. On the $0-100\ \mu\text{s}$ timescale, the spectral changes can be attributed partially, if not entirely, to temperature-tracking changes in solvent H-bonds to the protein and possible loosening of internal protein H-bonds, as was

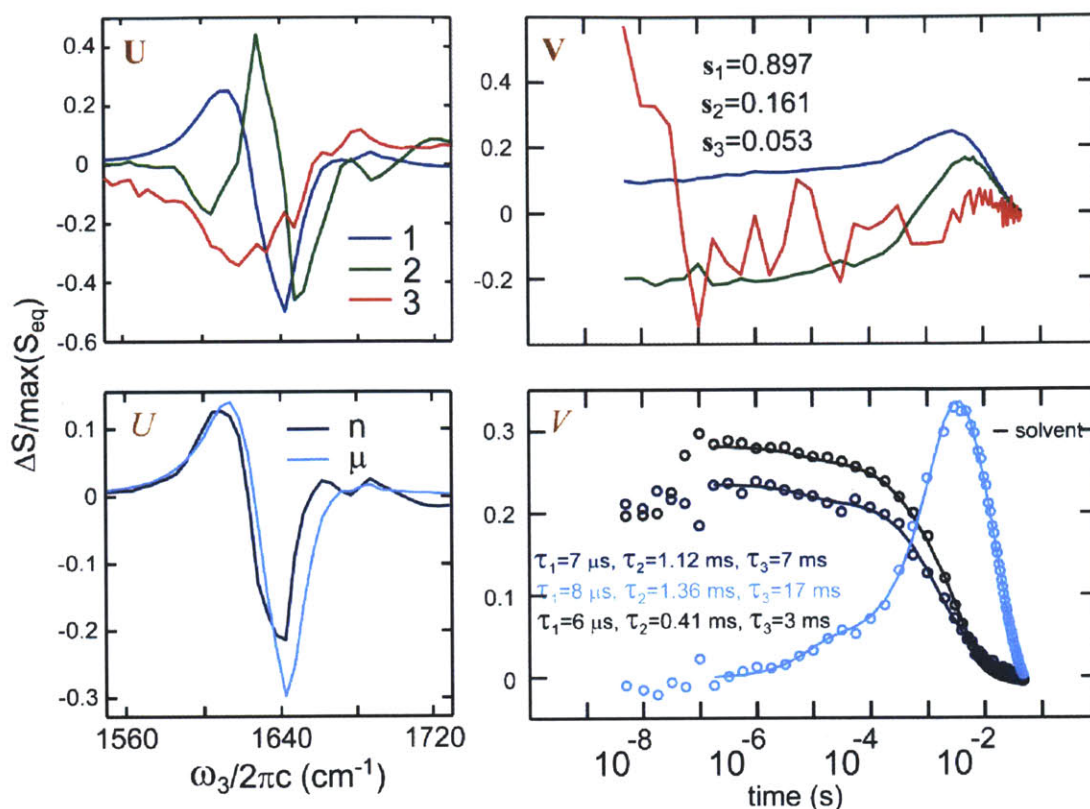


Figure 7.7: Row 1 – $U_{1,2,3}$ and $V_{1,2,3}$ (left and right, respectively) SVD components from the t-DPP shown in Figure 7.6. Row 2 – $U_{n,\mu}$ and $V_{n,\mu}$ (left and right, respectively) SVD components that result from re-mixing the 1st and 2nd components from row 1. In the bottom right, the transient transmission trace at 1725 cm^{-1} is also shown with its tri-exponential fit.

described in Chapter 5 for diglycine and TZ2. Interestingly, based on the convolution study described in Chapter 3, the ~ 1.2 ms timescale and > 10 ms timescale can be assigned to the same process – an exponential response with a > 10 ms time constant at the initial temperature. That is, the ~ 1.2 ms timescale is entirely due to convolution of a > 10 ms response with the temperature profile. Thus, on the > 10 ms timescale we are observing the net return (refolding) of a disordered population to the equilibrium state. Separating the protein melting response, which is assumed to track the temperature-profile (Chapter 5), from the solvent H-bond response is difficult when analyzing individual frequencies. Separation is clearer when taking the response of every frequency into consideration, as is done using the entire spectrum in SVD analysis.

7.3.4 SVD analysis of t-DPP spectra

To highlight the described changes in the t-DPP spectra and to evaluate the minimum

number of important kinetic species, we applied the SVD analysis described in Chapter 6 to the t-DPP spectra shown in Figure 7.6. The results of SVD analysis are shown in Figure 7.7. The fast drop in s value magnitude and the noisiness of \mathbf{v}_3 suggest that ubiquitin's t-DPP spectra can be decomposed into two basis spectra with two associated time traces. The first two components describe the spectra at times >100 ns to within 0.5% where the maximum changes are $\sim 12\%$. Re-creating the data with the first two components leads to almost identical tri-exponential fits as were measured in the original data: $\tau_1=1 \mu\text{s}$, $\tau_2=1.26 \text{ ms}$, $\tau_3=16 \text{ ms}$ at 1614 cm^{-1} and $\tau_1=5 \mu\text{s}$, $\tau_2=1.34 \text{ ms}$, $\tau_3=16 \text{ ms}$ at 1642 cm^{-1} . This agreement between the un-decomposed and decomposed/reconstructed data indicates that the first two components capture the complex, multi-exponential time behavior.

In order to re-cast the mathematical results of the SVD analysis shown in the top row of Figure 7.7 into physically meaningful spectra and population traces, we placed two constraints on the resulting components: 1) the ns SVD spectrum, u_n , matched the 32 ns (not shown) difference t-DPP spectrum and 2) the ns time trace, v_n , matched the temperature profile (as determined from the transient transmission). The first criterion was enforced using eq. (6.6) with the \mathbf{W} given in eq. (6.C.1). The second criterion was then applied to the results using the \mathbf{W} given in eq. (6.C.2). The first criterion assumes that the sole response at early times is due to the H-bond breaking/loosening response, whose spectrum should not change with temperature. The second criterion assumes that the H-bond breaking/loosening response will follow the temperature profile, which is not strictly true considering that all the states are linked. As the unfolded state, represented by difference spectrum u_μ , re-equilibrates, population will be fed back into the H-bond broken state, and its population trace will reflect the observed refolding rate. We expect this long timescale imprinting onto v_n , to be small compared to the shaping due to the temperature profile.

The re-casting results in the u_n spectrum with associated time trace v_n , which captures the ns behavior, and the u_μ spectrum with associated time trace v_μ , which captures the significant remaining behavior over the other (longer) time periods. As a consequence of the restrictions that result in this re-casting, v_n carries all of the weight at early times of $<1 \mu\text{s}$.

The u_n and u_μ spectra share the same characteristics as the 18 ns and 560 μs spectra discussed above, respectively. Better spectral assignment of the u_n and u_μ spectra is made later in the discussion of the t-2D IR complementary spectra. Briefly, the u_n spectrum has five

features from low to high frequency: a positive, negative, positive, dip, and positive feature that can respectively be assigned to loss on the red-side of the overtone and fundamental of ν_{\perp} doublet peak, loss on the red-side of the overtone and fundamental of the ν_{\parallel} doublet peak, and gain on the blue-side of the ν_{\parallel} fundamental. Based on these spectral changes, we assign the u_n spectrum to the combination of loss of equilibrium and gain of a state with loosened solvent-protein and possibly protein-protein H-bonds. This loss/gain shows up as a blue-shifting of the original spectrum.

The u_{μ} spectrum has two main features, a positive and a negative peak. The negative peak appears to result from loss of the ν_{\perp} fundamental and gain of a random coil overtone. The only features in the $>1650 \text{ cm}^{-1}$ u_{μ} spectral region are too small for analysis. As is shown later with the 2D IR spectra, the lack of features in this frequency region is due to the cancellation of multiple peaks, which indicates the importance of interpreting t-DPP spectra with the increased spectral content provided by 2D IR.

Now that the spectra have been assigned, we can begin to interpret their associated time traces. The u_{μ} spectrum is due to unfolding of the protein. As expected from our analysis of the t-DPP spectra, the ν_{μ} trace shows paired 1.36 / 17 ms exponential time constants, which can be attributed to a temperature-profile distorted 17 ms long timescale observed rate (see Figure 3.19), which is assigned to a re-equilibration of disordered structures (net refolding back to the initial temperature population distribution). Because it occurs after the temperature has re-equilibrated, 17 ms is an accurate timescale for the relaxation of ubiquitin at the T_i , 63°C, and it matches the timescale previously observed using t-DVE⁴.

The long 17 ms observed relaxation is also unique relative to TZ2 and insulin results because it extends beyond the temperature relaxation profile. Thus, any spectral changes at long times are expected to be predominately due to conformational changes rather than H-bond solvent breaking.

Of further particular interest is that ν_{μ} indicates that there is an unfolding process on the μs timescale with a fit exponential of 8 μs . Although the difference in timescale between the 8 μs and 17 ms time constants suggests that there are two separate unfolding processes separated by an intermediate, we cannot spectrally resolve the two. Either the fast and slow unfolding involves loss and gain of similar structures with the same spectra, or the signal-to-noise ratio is too low to separate the low amplitude 8 μs species' spectrum. The former possibility seems

reasonable: both the 8 μ s and 17 ms unfolding involves melting of separate β -strand structures and gain of random coil conformations, which is why the two unfolding processes share the same spectrum.

The ν_n trace can be related to the population of the state with fewer solvent-protein H-bonds. The trace mirrors the behavior of the temperature profile, as defined by the criterion used to calculate it. The ν_n trace is populated on the <5 ns pulse-width limited timescale, and it is depleted on the 1.12 and 7 ms timescales.

The fact that the t-DPP data can be well described with two spectra indicates that there is no frequency-dependence to the response. Rather, any differences in timescales at different frequencies are due to different ratios of the two time traces.

7.3.5 Transient 2D IR spectra

t-2D IR spectroscopy provides increased spectral information relative to t-DPP, which is equivalent to the projection of the t-2D IR spectra onto the ω_3 axis. In order to make a connection with previous $T_i=63^\circ\text{C}$, 30 mg/mL t-2D IR data, which was not corrected for the linear absorption, the recent t-2D IR data was collected at the same T_i , although at a lower 15 mg/mL concentration to minimize the possibility for aggregation. The previous uncorrected t-2D IR is shown on the left side of Figure 7.8, while the recent uncorrected spectra are shown on the right. Although the early time recent spectra are much noisier than the previous spectra, the same features are observed. At early times, the difference t-2D IR spectra show *i*) a red-side loss doublet in the diagonal ν_{\perp} region, *ii*) a pair of red/blue (above/below diagonal) off-diagonal wings, and *iii*) narrow, diagonally elongated loss features along the diagonal, although the above-diagonal negative feature overlaps the positive wing. At longer times, feature *ii*) drops in intensity, a loss doublet appears above *i*), and a doublet appears in the lower right anti-diagonal region. As with the t-DVE spectra, the tuning of the IR pulse spectrum affects the relative magnitude of the spectral features, which is why the *i*) features are more prominent in the previous spectra relative to the recent spectra. The *ii*) feature is due to the linear absorption distortion. At higher temperature the solvent has an increased transmission, and the equilibrium spectrum intensity increases by $\sim 5\%$. In the difference data, this 5% equilibrium spectrum is mixed in with the true 5% molecular spectral changes.

The previous t-2D IR data were interpreted within the framework of the downhill folding hypothesis. Stretched exponential time dependence during the 0.1-100 μ s timescale observed in

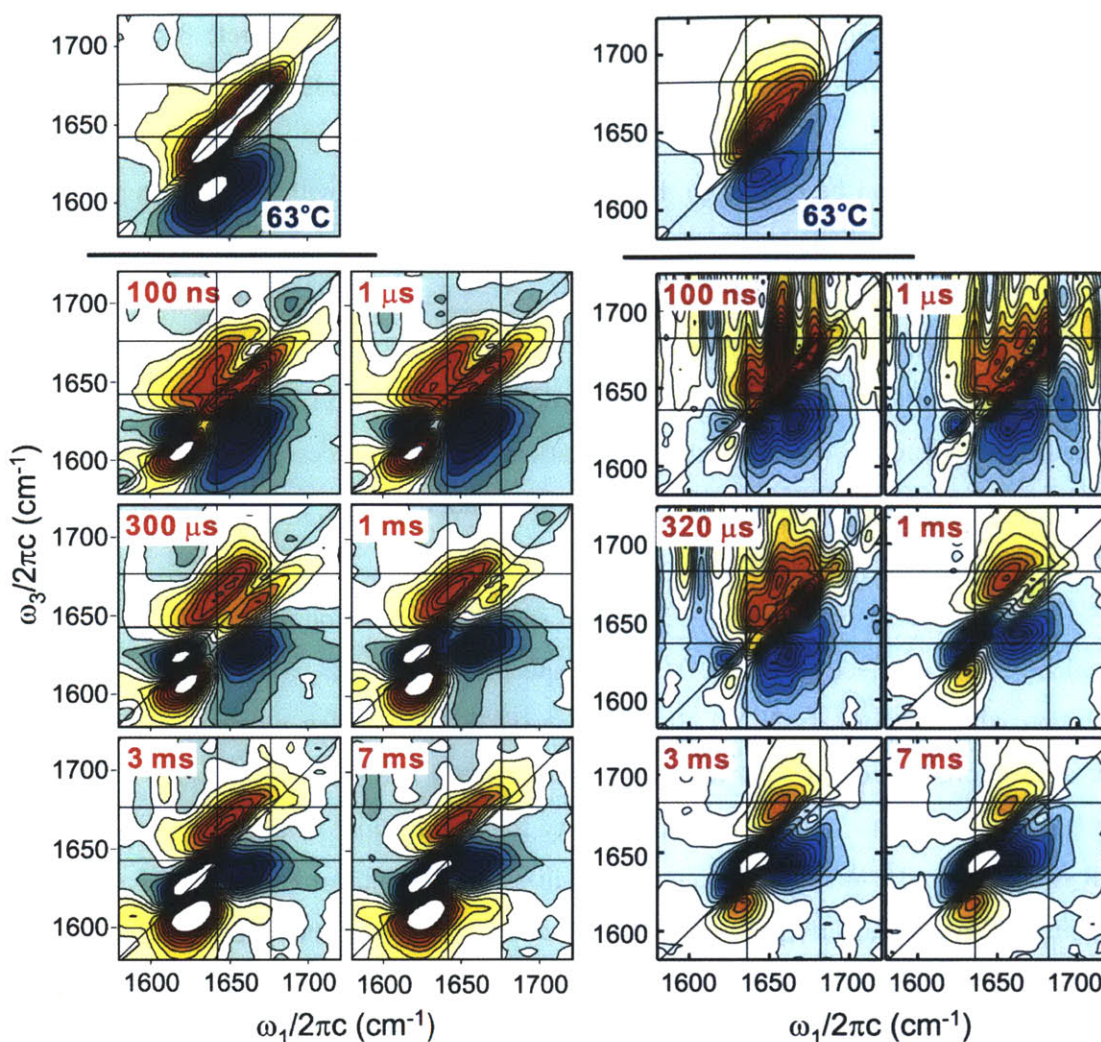


Figure 7.8: *Left* – The 33 mg/mL t-2D IR without linear absorption correction modified from Ref. 2. The spectra are plotted with 21 evenly spaced contours between $\pm 1.5\%$ after normalization to the maximum of the equilibrium. *Right* – The 15 mg/mL t-2D IR without linear absorption correction. The difference spectra are plotted with 25 evenly spaced contours between $\pm 4\%$ after normalization to the maximum of the equilibrium. Both sets of data were taken at $T_i=63^\circ\text{C}$ with ZZZZ polarization. Notice that frequency axes are slightly different on the left and right.

the 1st SVD component, traces from different points in the spectrum, homogeneous linewidths, and peak positions were all interpreted as signatures of downhill folding². The most interesting of these is the observed peak shift, which was interpreted as resulting from the decreasing size of the β -sheet, which is expected to be accompanied by a frequency shift. The consistency between the previous and recent t-2D IR spectra suggests the possibility of a more direct comparison between the two data sets than was possible for the t-DVE/t-DPP comparison. The conclusions

from results based on the previous uncorrected t-2D IR can be compared to conclusions from the more recent ω_3 -corrected t-2D IR data set

To remove the distortions due to the linear absorption effects, the difference t-2D IR data is also presented after ω_3 -correction in Figure 7.9. As with the t-DPP, the ω_1 -correction was not applied for ease of processing, but it is expected to cause minor changes. The ω_3 -corrected spectra in Figure 7.9 represent the true molecular changes that occur after the T-jump. The early time 560 ns and 100 μ s spectra are dominated by on-diagonal loss on the red-side of the ν_{\perp} and ν_{\parallel} peaks, which is consistent with blue-shifting attributed to average loss of solvent-protein H-bonds. The later ms spectra show loss of the β -strand ν_{\perp} diagonal doublet and gain of an on-diagonal peak in the 1660 cm^{-1} region, which is attributed to gain of random coil structures. As expected, along with loss of the ν_{\perp} diagonal doublet the spectra also show loss in the $\nu_{\perp}/\nu_{\parallel}$ cross-peak regions. In the upper left corner, this loss manifests itself in a weak blue feature, which is suppressed due to overlap with gain of the random coil fundamental. In the lower right corner, the loss shows up as strong doublet peaks, although the upper blue loss peak overlaps with gain of the random coil overtone (also blue).

The spectral changes observed in the equilibrium difference and \mathbf{u}_2 spectra in Figure 7.3 appear to be a combination of the early time (~ 560 ns) and late time (> 1 ms) spectra. This is expected, as the equilibrium spectra will be a mixture of the four populations whose difference spectra are observed in the transient experiment: solvent H-bond broken and unbroken, and the folded and unfolded populations. Given that the T-jump profile relaxes more quickly than the unfolded protein refolds, the T-jump experiments separate the two responses based on their time dependence. That is, the early and late spectra can be approximately assigned directly to the

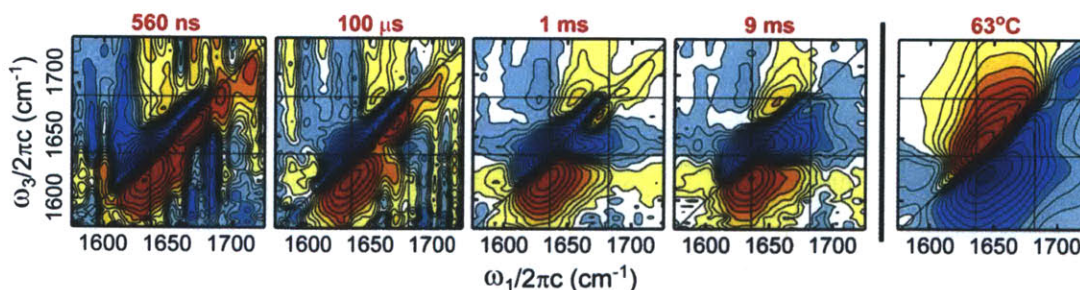


Figure 7.9: The difference t-2D IR spectra after ω_3 -correction. The 2D IR spectra are plotted with the same contour spacings described in Figure 7.3.

difference between H-bonded/H-bond broken and folded/unfolded populations, respectively. This is done more carefully with SVD analysis. For proteins like TZZ and insulin, where re-equilibration is faster or on the same timescale as the T-jump cooling, this separation is still possible, but there is not as clear a separation, and SVD is more essential to try to accurately differentiate the spectra.

Based on the hypothesis that the early spectrum is due to H-bond breaking, collecting T-jump spectra of ubiquitin well below the melting temperature, where there is no conformational change, should yield results entirely due to the < 5 ns H-bond breaking response. Thus, we would expect to see the same difference spectrum at every time delay scaled proportionally to the change in temperature. Although we have not explicitly done this for ubiquitin recently, similar and consistent results have been observed for N-methylacetamide and diglycine (Chapter 5,²²), two molecules that do not exhibit conformational change. Previous data was taken of ubiquitin at 25°C, however, $\sim 40^\circ\text{C}$ below the melting point². As expected the $T_i=25^\circ\text{C}$, $\tau = 100$ ns and 1 ms difference t-2D IR spectra are similar to each other and to the $T_i=63^\circ\text{C}$, $\tau = 100$ ns spectrum, which is consistent with our analysis and expectations.

7.3.6 SVD analysis of t-2D IR spectra

As with the t-DPP spectra, SVD analysis of the t-2D IR spectra reveals the predominance of two component spectra. As described in Chapter 6, the first two t-2D IR SVD component spectra were re-cast into physically meaningful spectra by best matching $V_{n,\mu}^{2D}$ to $V_{n,\mu}^{DPP}$ and $\int U_{n,\mu}^{2D} d\omega_1$ to $U_{n,\mu}^{PP}$. These four quantities as well as the resulting $U_{n,\mu}^{2D}$ are shown in Figure 7.10. The $U_{n,\mu}^{2D}$ spectra highlight the spectral changes described above at early and later times, respectively. u_n^{2D} is the combined spectrum of loss of the equilibrium conformations and gain of structures with fewer solvent-protein and possibly protein-protein H-bonds. u_μ^{2D} is the combined spectrum of loss of the equilibrium conformations and gain of melted structures.

The $V_{n,\mu}^{2D}$ traces match the $V_{n,\mu}^{DPP}$ traces. Of particular importance is the matching of the 0.1-100 μs behavior in V_μ^{2D} , which reaffirms the t-DPP conclusion that there is melting on the ~ 10 μs and >10 ms (1.2 / 17 ms) timescales.

The fact that the t-2D IR spectra can be decomposed into two basis spectra suggests that

the shifting of peaks interpreted as signatures of downhill unfolding may instead be due to the loss and gain of two overlapping features with different center frequencies.

At 63°C, ubiquitin is expected to have significant folded and unfolded conformations,

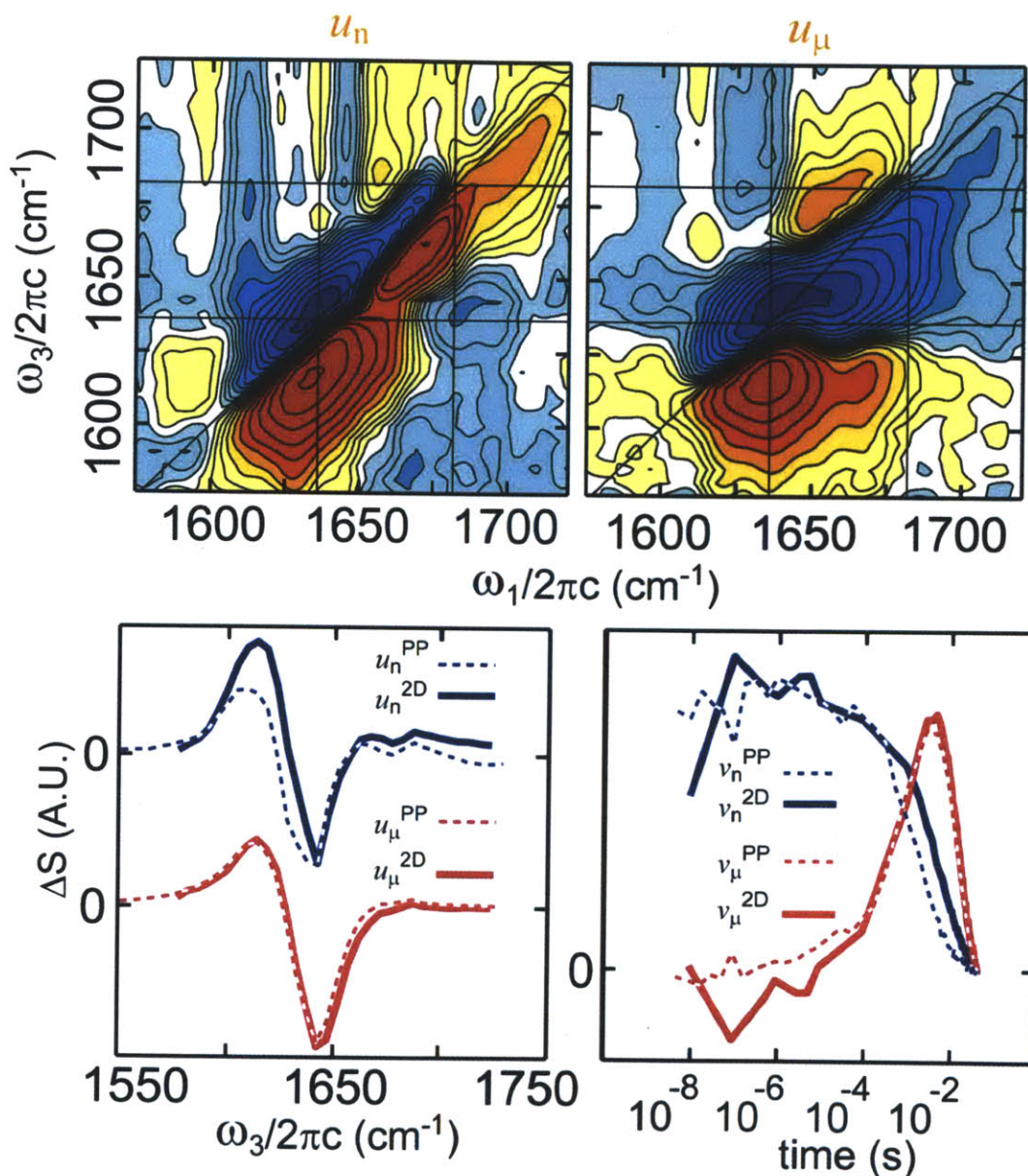


Figure 7.10: Row 1 – The $U_{n,\mu}$ 2D IR spectra that result from re-weighting the 1st and 2nd SVD components. The $U_{n,\mu}$ 2D IR spectra are chosen based on matching their ω_3 -projections to the t-DPP $U_{n,\mu}$ (bottom left) and matching $V_{n,\mu}$ from 2D IR and t-DPP (bottom right). The t-DPP SVD results are also shown in Figure 7.7. The 2D IR spectra are plotted with the same contour spacings described in Figure 7.3.

which makes hypothesizing the folded and unfolded spectrum by adding the weighted

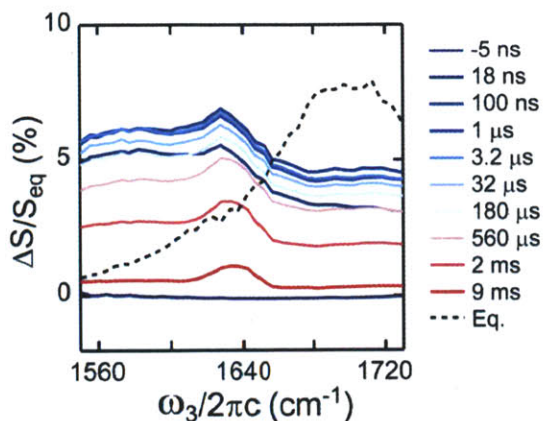


Figure 7.11: The transient transmission spectra of ubiquitin and the DCI/D₂O solvent.

equilibrium spectrum to $U_{n,\mu}^{2D}$ (described in Chapter 6) difficult, and we do not attempt it in this case.

7.3.7 SVD analysis of t-transmission spectra

As a byproduct of the t-HDVE data collection, the LO spectrum is collected and saved as a function of time delay following the T-jump. This LO spectrum is actually the t-transmission spectrum, which we consider as a linear spectrum although it is formally 3rd order. The difference t-transmission spectra are shown at selected delays in Figure 7.11. In standard T-jump linear spectroscopy, the t-transmission spectra of the solvent without the solute is separately collected and subtracted from the solute+solvent spectra to isolate the solute response. Instead of re-running our experiment on just the solvent, we used SVD analysis to approximately subtract the solvent response. Although the solute molecular changes are merely a drop in the bucket overflowing with the solvent transmission response, the SVD does a reasonable job of separating the different spectra.

We expect two different time traces corresponding to three responses: *A*) the solvent response should exactly follow the temperature profile, *B*) the ns protein-solvent H-bond breaking should approximately follow the T-jump profile, and *C*) the melting response should start at 0 and then grow in with small and large amplitude on the 10 μs and 1.2 ms timescale (17 ms underlying, 1.2 ms because of the temperature profile). Because *A* and *B* both follow the T-jump profile, we expect the sum of their spectral signatures to appear in one component.

We apply SVD to separate out different features. Figure 7.12 shows the results of the SVD. The analysis indicates that the t-transmission spectra are dominated by the solvent

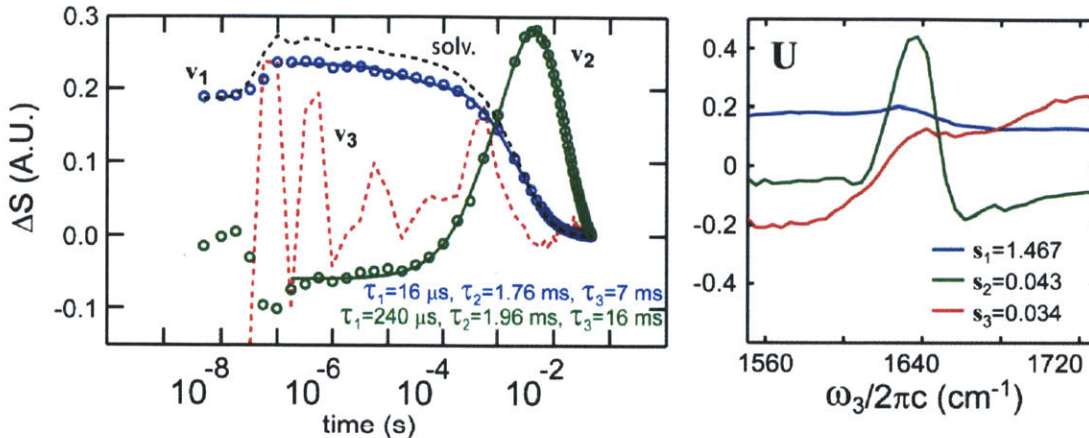


Figure 7.12: *Left* – SVD $\mathbf{V}_{1,2}$ components from transient transmission measurements. The dashed-line solvent response is taken from the t-transmission spectrum at 1725 cm^{-1} , where we expect the majority of the response is due to the solvent. *Right* – The corresponding SVD $\mathbf{U}_{1,2,3}$ components.

response, as is indicated by the 34:1 ratio of the 1st and 2nd s value, the largely broad, featureless nature of the \mathbf{u}_1 spectrum, and the temperature-tracking character of the \mathbf{v}_1 time trace. Because of the shared timescales between A and B , the first component carries some of the pulse-width limited $<10\text{ ns}$ protein-solvent H-bond breaking response, which is evident in the sharp positive feature at $\sim 1630\text{ cm}^{-1}$ in the \mathbf{u}_1 spectrum. This sharp positive feature corresponds to an increase in transmission (loss) of something similar to the equilibrium spectrum. Because the solvent response dies away by $\sim 10\text{ ms}$, any spectral changes that persist beyond can be attributed to the protein conformational response. The second component, whose \mathbf{v}_2 persists with high amplitude out beyond the temperature-profile, captures the protein response at long times, which is predominantly melting. The \mathbf{u}_2 spectrum shows loss at 1640 cm^{-1} and broad gain above 1650 cm^{-1} except for a dip at $\sim 1675\text{ cm}^{-1}$, three features attributed to loss of ν_{\perp} , gain of random coil, and loss of ν_{\parallel} . These spectral changes are consistent with the FTIR protein changes observed previously¹⁸. The negative character of the \mathbf{v}_2 trace at early times is due to mixing of the H-bond (B from above) response into both \mathbf{u}_1 and \mathbf{u}_2 .

To better separate the melting from the solvent and H-bond response, the first two components were re-cast. The approximation that the first two components are the only significant components is a worse assumption than in the t-DPP and t-2D IR cases because of the closeness of the s_2 and s_3 values shown in Figure 7.12. The criterion for the re-casting is that 1) ν_{μ} is zero at 180 ns , or, equivalently, that the u_n matches the 180 ns difference t-transmission

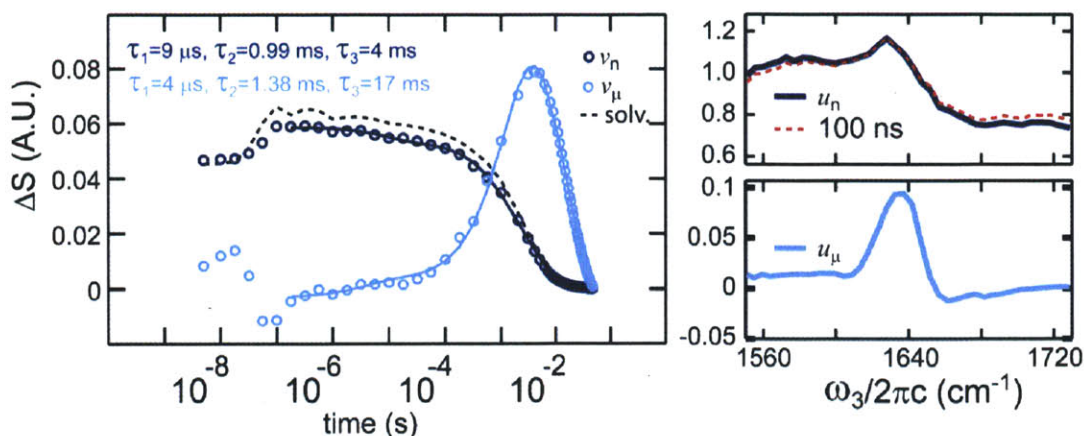


Figure 7.13: *Left* – SVD $V_{n,\mu}$ components from transient transmission measurements. *Right* – The corresponding SVD $U_{n,\mu}$ components. u_μ is compared to the arbitrarily scaled 100 ns transient transmission difference spectrum.

spectrum AND that 2) the B H-bond break time response will follow the solvent absorption profile. The results are shown in Figure 7.13. The 1) criterion was chosen based on our expectations for the melting response, which, unlike the A and B responses, should begin at zero on the ns timescale and grow in on the μ s and ms periods. The uniqueness of this early time melting behavior ensures that, within our approximations, v_μ will trace out the response of the melting, but its u_μ spectrum (without more constraints) will be a mixture of the full melting response and the A , B , and C spectra. The 2) criterion allows for a better approximation of the u_μ spectrum by shifting out the A and B character. Hence, the u_μ has less of a baseline than u_2 , which is more physically reasonable.

Features of the melting can be observed in the u_μ spectrum shown in Figure 7.13; increased transmission at $\sim 1635\text{ cm}^{-1}$ and decreased transmission at 1660 cm^{-1} indicates loss of the v_\perp peak and gain of the random coil, respectively. The v_μ trace also shows melting on the $\sim 4\text{ }\mu\text{s}$ and $1.38 / 17\text{ ms}$ timescale, although the μs phase seems to have a lower amplitude relative to the ms than in the nonlinear spectroscopy case.

Although the dominating presence of the solvent response casts a shadow of doubt over the t-transmission analysis, the general observations are consistent with the nonlinear results. The timescales and spectral features of unfolding can be identified and quantified. This example serves as both an illustration of the t-transmission potential and an application of SVD analysis. Perhaps measuring just the solvent t-transmission spectrum would provide a necessary 2nd constraint to aid the SVD analysis in separating the three expected components.

7.3.8 Comparison of results

As stressed earlier, the recent data is the result of a single data run under one set of concentration, pH, and temperature conditions. A better characterization of time constants and the thermodynamic variables requires increased averaging and a greater sampling of variable phase-space.

Previous and recent t-DVE data were collected under different conditions, which allows differences to be explained through different T_i , different concentrations, or different pulse spectra. Ideally, the previous and recent t-DVE data would match exactly, and artifactual spectral features due to linear absorption could be traced directly from the previous to the recent to the corrected t-DPP spectra. Instead, we focus on the semi-quantitative agreement between the t-DVE sets and independently assess the recent data. We note that the recent and previous⁴ 63°C results match better than the data sets compared here (previous 58°C, recent 63°C).

Both recent and previous t-DVE show stretched <100 μ s kinetics and frequency-dependent responses that were previously attributed to downhill unfolding processes. Comparison of the recent t-DVE to t-DPP reveals that the t-DVE spectral changes, particularly in the 1660 cm^{-1} region, are strongly influenced by the linear absorption and the concentration-squared nature of DVE. The difference t-DPP data can be better fit to multi-exponential kinetics than the t-DVE, and the t-DPP frequency-dependency can be explained by two spectra whose concomitant loss and gain result in mixed timescales and an apparent peak shift.

Despite the difference in concentration, agreement between the previous and recent t-2D IR allows for a direct comparison. Comparison between the linear absorption uncorrected and ω_3 -corrected t-2D spectra shows that a broad temperature-profile-tracking spectrum similar to the equilibrium is added into the t-2D spectra due to linear absorption. ω_3 -corrections remove this spectral response to more clearly reveal the response of the protein. As with the t-DVE case, the t-2D IR previously assigned peak-shifts, frequency-dependency, and stretched exponentials can be explained through the growth and loss of two components, as is revealed by the SVD.

Neither the recent t-DPP nor t-2D IR spectra necessarily indicate downhill unfolding. Rather, interpreting the t-DPP, t-2D IR, and t-transmission data independently of the previous work reveals two kinetic phases: 1) the < 5 ns breaking of protein/solvent and intra-protein/protein H-bonds and 2) the \sim 10 μ s and 17 ms net observed unfolding of the β -strands. As described above, the 17 ms response results in a rise and decay of signal due to the distortion of

the temperature-profile. Thus, the 1-2 ms and 17 ms result from the same 17 ms response. The 2D IR difference spectra for these two phases are separated using SVD and are shown in Figure 7.10. Each timescale can be interpreted as movement between kinetic states. The spectrum and existence of the ~ 10 μ s timescale indicates the presence of an intermediate with partially unfolded β -strands, probably III, IV, and/or V based on previous fragmentation and simulation studies.

The 17 ms observation is in the range expected based on previous observation of a 5-10 ms disordering rate measured with previous stopped flow experiments under different conditions and lower temperatures^{16, 23}. Previous T-jump studies have observed 2 ms relaxation, although we speculate this is due to the convolution with the temperature profile¹⁸.

7.4 Conclusion

The ubiquitin data presented here illustrates the importance of properly processing the data to account for distortions due to linear absorption. Not correcting for the linear absorption results in significant spurious spectral features that can lead to misinterpretations of the data. For example, linear absorption in the t-DVE leads to exaggerated spectral changes in which the solute response and solvent response are convolved. These exaggerated spectral changes along with the concentration-squared dependence appear to magnify the frequency dependence and distort the kinetics. Assessing similar ω_3 -corrected t-DPP and t-2D IR data reveals simpler changes that can be interpreted with multi-exponential relaxation and four state kinetic states (folded, broken solvent/protein H-bonds, partially unfolded, and unfolded). Given our single t-HDVE and t-2D IR data sets, the spectral features can be interpreted without invoking the downhill unfolding mechanism.

The application of linear absorption correction and SVD analysis represents both a technological improvement and also a different way of interpreting the data. Previous analysis has assigned different frequency ranges to specific structural elements. Differences in the timescales observed in these ranges were interpreted as evidence of different timescales for disruption of the corresponding structures. Although this approach is ideally correct and reasonable, in reality, the overlapping nature of secondary-structure spectra and the non-ideal shape of the T-jump relaxation profile make it difficult to separate out the different responses, especially when linear absorption is thrown into the mix. For example, the sum of two individual exponential relaxations might be mis-assigned to a stretched exponential. Through

SVD analysis, different spectral components can be compared and assigned to different populations, which allows for a better separation of the data based on full spectra rather than individual frequency ranges. Although, as discussed in Chapter 6, the SVD has shortcomings, its accuracy can be assessed by comparing how faithfully a few spectral components can re-create the frequency-dependent responses observed by looking at individual pixels. In this way, SVD analysis seems to be less likely at over-interpreting individual frequencies, which can be affected by noise and represent a mixture of overlapping responses.

The discussions here provide an outline and example of our analysis of the proteins studied in the following chapters. In Chapter 6 a general model was described in which the data is separated into the sum of two difference spectra: the < 5 ns H-bond breaking and $> \mu$ s conformational change response. The ubiquitin results consistently fit into this model, which is particularly interesting because ubiquitin's kinetics extend beyond the temperature profile, suggesting that we can assign the spectra of both of these responses to early and late data. This extra information is not available in the TZ2 or insulin cases, and it can be used to test the model: agreement between the u_{μ} spectrum and the 9 ms t-2D IR spectrum affirms the model. Based on its size and complexity, ubiquitin appears to be an ideal candidate for t-2DIR experiments. Although the experiments described here provide interesting kinetics that can be linked with different spectra and mapped onto different folding mechanisms, there is much more information to be gained from altering different variables (such as T_i , T_f , ΔT) or inserting isotope labels to begin to differentiate more states through localized probes. Despite the potential information to be gained, a more ideal candidate than ubiquitin is a protein that reveals more than three spectral components (more than two for the difference spectra), which would allow for the assignment of more complicated folding mechanisms. Based on the signal to noise ratios observed here for ubiquitin, distinguishing more than three components will require an appreciable separation of timescales or spectra associated with different intermediates. Larger proteins provide more potential intermediates, but assigning different spectra to different sub-structures is difficult and the timescales generally become longer with increasing size. Based on the signal to noise, the spectral overlap between secondary structural elements, the current difficulties in isotope labeling, and the 50 ms upper limit, t-2D IR continues to be a technique limited to a small subset of biological systems. The potentially available unique information, however, is certainly worthy of the required effort and the continuing struggle to improve and further generalize the

technique.

7.5 References

- (1) S. Vijay-Kumar, C. E. Bugg, and W. J. Cook, "Structure of ubiquitin refined at 1.8 Å resolution," *J. Mol. Biol.* **194**, 531-544 (1987).
- (2) H. S. Chung, Z. Ganim, K. C. Jones, and A. Tokmakoff, "Transient 2D IR spectroscopy of ubiquitin unfolding dynamics," *Proc. Natl. Acad. Sci. U. S. A.*, 14237-14242 (2007).
- (3) H. S. Chung, M. Khalil, A. W. Smith, Z. Ganim, and A. Tokmakoff, "Conformational changes during the nanosecond to millisecond unfolding of ubiquitin," *Proc. Natl. Acad. Sci. U. S. A.* **102**, 612-617 (2005).
- (4) H. S. Chung and A. Tokmakoff, "Temperature-dependent downhill unfolding of ubiquitin. I. Nanosecond-to-millisecond resolved nonlinear infrared spectroscopy," *Proteins* **72**, 474-487 (2008).
- (5) H. S. Chung, "Thermal unfolding dynamics of proteins probed by nonlinear infrared spectroscopy," (Massachusetts Institute of Technology, Cambridge, MA, 2007).
- (6) H. S. Chung, A. Shandiz, T. R. Sosnick, and A. Tokmakoff, "Probing the Folding Transition State of Ubiquitin Mutants by Temperature-Jump-Induced Downhill Unfolding," *Biochemistry* **47**, 13870-13877 (2008).
- (7) H. S. Chung and A. Tokmakoff, "Temperature-dependent downhill unfolding of ubiquitin. II. Modeling the free energy surface," *Proteins* **72**, 488-497 (2008).
- (8) D. O. V. Alonso and V. Daggett, "Molecular dynamics simulations of hydrophobic collapse of ubiquitin," *Protein Sci.* **7**, 860-874 (1998).
- (9) E. Larios, J. S. Li, K. Schulten, H. Kihara, and M. Gruebele, "Multiple probes reveal a native-like intermediate during low-temperature refolding of ubiquitin," *J. Mol. Biol.* **340**, 115-125 (2004).
- (10) B. A. Krantz, R. S. Dothager, and T. R. Sosnick, "Discerning the structure and energy of multiple transition states in protein folding using psi-analysis," *J. Mol. Biol.* **337**, 463-475 (2004).
- (11) J. Jacob, B. Krantz, R. S. Dothager, P. Thiyagarajan, and T. R. Sosnick, "Early Collapse is not an Obligate Step in Protein Folding," *J. Mol. Biol.* **338**, 369-382 (2004).
- (12) T. R. Sosnick, R. S. Dothager, and B. A. Krantz, "Differences in the folding transition state of ubiquitin indicated by ϕ and ψ analyses," *Proc. Natl. Acad. Sci. USA* **101**, 17377-17382 (2004).
- (13) Z. Qin, J. Ervin, E. Larios, M. Gruebele, and H. Kihara, "Formation of a compact structured ensemble without fluorescence signature early during ubiquitin folding," *J. Phys. Chem. B* **106**, 13040-13046 (2002).
- (14) B. A. Krantz and T. R. Sosnick, "Distinguishing between two-state and three-state models for ubiquitin folding," *Biochemistry* **39**, 11696-11701 (2000).
- (15) H. M. Went, C. G. Benitez-Cardoza, and S. E. Jackson, "Is an intermediate state populated on the folding pathway of ubiquitin?," *FEBS Lett.* **567**, 333-338 (2004).
- (16) M. S. Briggs and H. Roder, "Early hydrogen-bonding events in the folding reaction of ubiquitin," *Proc. Natl. Acad. Sci. USA* **89**, 2017-2021 (1992).
- (17) J. Sabelko, J. Ervin, and M. Gruebele, "Observation of strange kinetics in protein folding," *Proc. Natl. Acad. Sci. USA* **96**, 6031-6036 (1999).
- (18) C. S. Colley, I. P. Clark, S. R. Griffiths-Jones, M. W. George, and M. S. Searle, "Steady state and time-resolved IR spectroscopy of the native and unfolded states of bovine ubiquitin: protein stability and temperature-jump kinetic measurements of protein folding

- at low pH," *Chem. Commun.* (2000).
- (19) M. M. Harding, D. H. Williams, and D. N. Woolfson, "Characterization of a partially denatured state of a protein by two-dimensional NMR: reduction of the hydrophobic interactions in ubiquitin," *Biochemistry* **30**, 3120-3128 (1991).
 - (20) W. Y. Yang and M. Gruebele, "Folding at the speed limit," *Nature* **423**, 193-197 (2003).
 - (21) J. Kubelka, J. Hofrichter, and W. A. Eaton, "The protein folding 'speed limit'," *Curr. Opin. Struct. Biol.* **14**, 76-88 (2004).
 - (22) K. C. Jones, Z. Ganim, C. S. Peng, and A. Tokmakoff, "Transient two-dimensional spectroscopy with linear absorption corrections applied to temperature-jump two-dimensional infrared," *J. Opt. Soc. Am. B* **29**, 118-129 (2012).
 - (23) S. Khorasanizadeh, I. D. Peters, T. R. Butt, and H. Roder, "Folding and stability of a tryptophan-containing mutant of ubiquitin," *Biochemistry* **32**, 7054-7063 (1993).

Chapter 8

Transient Studies of Isotope-Labeled Trpzip2

8.1 Introduction

The Occam's Razor argument – the simplest explanation is the right one until further information requires reassessment – may not be the correct reasoning when it comes to studying protein folding. For example, proteins are often labeled as two state folders if past experimental data is consistent with a model in which the vast configurational space is coarsely separated into two states: the folded and unfolded state. The advancement of experimental and simulation techniques has indicated that two state folders are generally the exception rather than the rule, which suggests that assuming proteins have multiple important states *a priori* may be a better default than the current two-state Occam's Razor default^{2, 3}. Here, we focus on studying the folding pathway of a small model peptide, trpzip2, which has previously been labeled a two state folder based on melting curves⁴. The observation of multiple states in this small peptide suggests that larger proteins also exist in more than two states.

Theoretical, computational, and experimental protein folding studies cover decades in time, span orders of magnitude in protein size, and vary in complexity from atomistic to course-grained detail. The experimental study of peptides is an attractive facet of protein folding studies because the folding mechanism of peptides is expected to mirror that of proteins, and their small size reduces the complexity of the problem and allows for the tractability of extensive molecular dynamics (MD) simulations that can be compared directly to experimental results^{6, 7}. The smaller size (<100 residues) generally corresponds with faster folding, which allows for better sampling

in MD simulations but requires submillisecond experimental time resolution. Experiments and simulations are aimed at identifying the number of significant states and understanding the mechanism of folding.

Although increasing computational power and technical advancements have allowed for the direct comparison of MD simulations to fast-folding experiments, there remains a disconnect between the physical pictures that emerge². Experiments tend to be interpreted using a few general states that are explained within the framework of paradigmatic mechanisms. Simulations with atomistic detail identify conformations at every time step, and analysis techniques, such as the emerging Markovian State Model (MSM) approach, sort the trajectory into many representative structures connected by kinetic rates⁸⁻¹⁰. The results from simulation indicate multiple parallel folding routes involving more states than observed in experiment^{10, 11}. The disconnect between simulation and experiment is largely due to the inherent differences between the sensitivity of atomistic simulations and experimental probes. There is a distinct lack of experimental validation for theoretical predictions. NMR, which has the requisite atomic resolution, lacks the sub-millisecond time resolution. Detailed experimental results require a technique with fast time resolution and sensitivity to both the general secondary structure and local structure at key locations along the peptide backbone.

Bridging the gap between theory and experiment requires developing experimental techniques that are sensitive to many states identified by simulation. In this work, we use isotope-edited temperature-jump (T-jump) amide I two dimensional infrared (2D IR) spectroscopy to identify multiple states in the unfolding kinetics of the engineered tryptophan-zipper peptide trpzip2 (TZ2). Amide I primarily consists of the C=O stretch along the peptide backbone. Coupling between oscillators results in delocalized vibrations that are sensitive to the secondary and tertiary structure of the peptide^{12, 13}. Isotope labeling shifts the frequency and localizes the vibrational mode, which becomes sensitive to the electrostatic environment surrounding the labeled residue. 2D IR correlates pumping and probing frequencies by spreading the amide I spectral information over two axes. Coupling results in cross peaks between different modes, and homogeneous and inhomogeneous broadening are separated along the anti-diagonal and diagonal direction, respectively. Although the 2D IR time resolution is limited by the ps vibrational lifetime, introduction of a T-jump laser allows for monitoring of timescales from 6 ns, the pulse-length of the T-jump laser, out to 50 ms, determined by the 20 Hz T-jump repetition

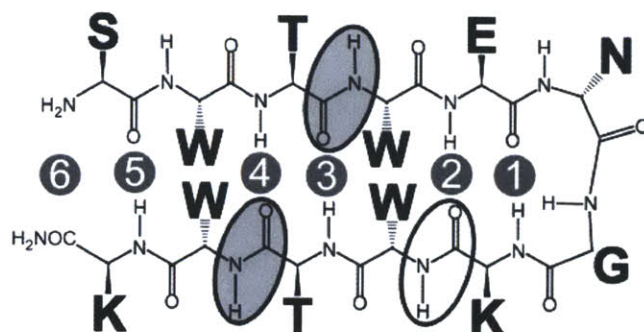


Figure 8.1: The structure of TZ2 showing native registry of cross-strand H-bonds. The white oval depicts the isotope labeled oscillator unit for the K8 species. The grey ovals depict the oscillators for the TT species. Numbers label the possible cross-strand H-bonds from turn outward.

rate. By electronically delaying the 2D IR pulses relative to the T-jump pulse, we collect ps snapshots to monitor the peptide's response at variable delays following the temperature change^{14, 15}. Transient 2D IR experiments have been performed in the past¹⁶⁻²¹, but our group is unique in collecting t-2D IR spectra following a T-jump^{14, 15, 22}. Vibrationally isolating individual residues through isotope labeling provides increased sensitivity to a wider variety of kinetic states and the increased information content of 2D IR allows for a more accurate assignment of these states.

TZ2, shown in Figure 8.1, is a twelve residue peptide that forms a type I' β -turn in its native state due to stabilization by cross-strand H-bonds and edge-to-face packing of two pairs of cross-strand tryptophan indole rings⁴. TZ2's small size and fast folding rates have made it an attractive system to study for both experimentalists and simulation^{1, 4, 10, 12, 15, 23-48}. The growing wealth of knowledge in both fields has fostered comparison and enriched interpretation of results, and TZ2 has become a test subject for new simulation and experimental techniques^{10, 15, 36, 41, 43, 44, 48, 49}. The ultimate goal of experiment and simulation is to identify the folding mechanism. Recent work has focused on discriminating between two general folding pathways: 1) hydrophobic collapse^{37, 47}, in which H-bond contacts are formed after hydrophobic contacts create a favorable collapsed state and 2) a zipping mechanism^{25, 36, 39} where native H-bond formation is the folding nucleation event which is then stabilized by sidechain packing. Although there is little consensus on their structures, many recent simulations identified multiple stable states - native, disordered, and partially folded. The partially folded states are usually either helical^{36, 42, 46} or they contain native H-bonds in the turn region^{39-41, 43, 48}. When measured, the transition state is usually identified as containing the inner turn H-bonds^{37, 41, 48}.

Early experimental work measured a sigmoidal melting curve⁴ and T-jump experiments

revealed single exponential relaxation²⁵, both of which were interpreted as signs of two-state folding. Recent experimental work has interpreted probe-dependent melting points measured with equilibrium CD, fluorescence, isotope-label FTIR, and isotope-label 2D IR as signatures of heterogeneous folding and multi-state behavior^{1, 23, 32}. Heterogeneous folding was also proposed based on frequency-dependent differences in isotope-labeled 2D IR experiments and rates in T-jump fluorescence and T-jump linear IR measurements^{1, 24, 31, 33}. Simulations have identified multiple important states including misregistered³⁵ and partially folded intermediates in addition to the native and disordered states, but neither simulation nor experiments have come to a consensus on the folding mechanism or the specific assignment of states. Both cross-strand H-bond formation and tryptophan packing are recognized as being important, but the order in which they form is unclear. In this work, multiple TZ2 isotopologues are investigated using ultrafast transient T-jump 2D IR spectroscopy. The isotope-label variants provide probes of different turn conformations and native H-bond contacts at different regions along the peptide. The time dependence of different 2D IR spectral signatures allows us to assign and interpret kinetic states. We identify native, misfolded, and disordered states and propose a free-energy surface depicting their connectivity.

8.2 Methods

8.2.1 TZ2 Samples

Unlabeled and K8 ¹³C C' labeled TZ2 variants were synthesized with C-terminal amides using an Fmoc-based solid state synthesis as described previously^{1, 12}. Each peptide was purified through HPLC and its purity was confirmed with mass spectrometry. The C-terminal amide T3T10 ¹³C C' labeled TZ2 was purchased from Anaspec Inc. (San Jose, CA). Residual TFA was removed through lyophilization. All peptides were H/D exchanged through heating to 60°C in D₂O and successive lyophilization. Each sample was dissolved in pH* 2.5 DCl in D₂O solution to a concentration of 10 mg/mL to avoid aggregation, which was observed after extended periods of time. Experiments were performed at an initial temperature of 35°C after sandwiching the sample between two CaF₂ windows separated by a ring Teflon spacer to give a 50 μm pathlength.

8.2.2 Optical setup

Experiments were performed using transient T-jump 2D IR, which has been described

previously^{14, 15, 22}. Briefly, three 1 kHz, 90 fs, 6 μm mid-IR pulses focused into the same sample volume in a boxcar geometry were varied to generate a time-delay dependent 3rd order nonlinear signal that was interferometrically measured against the 4th reference pulse using balanced detection. The dispersed pump-probe (DPP) spectra were calculated from the heterodyne-detected dispersed vibrational echo (HDVE) spectrum⁵⁰. HDVE and 2D IR measurements were performed with ZZZZ polarization at a waiting time of $\tau_2=150$ fs. HDVE and 2D IR spectra were collected as a function of electronically-generated time delay, τ , following a T-jump. The T-jump was generated using the 9-10 mJ, 1.98 μm , 20 Hz, 6 ns output from a Nd:YAG pumped optical parametric oscillator, which was used to heat the D₂O by $\sim 10^\circ\text{C}$ by exciting the $\nu=2\leftarrow 0$ OD stretch.

Transient HDVE (t-HDVE) measurements were collected using the Fourier-transform spectral interferometry (FTSI) and phase-modulation spectral interferometry (PMSI) methods with τ_{LO} stepping from 0 to 25 fs in 5 fs steps⁵⁰. t-2D IR surfaces were collected in 14 fs undersampled steps from 0 to ≥ 2.4 and to ≥ 1.3 ps for rephasing and non-rephasing surfaces, respectively. t-2D IR spectra are the result of ≥ 8 averages. The t-2D IR spectra used for this paper were the result of ~ 350 hours of data collection. Although each individually averaged t-2D IR surface was collected in ~ 1.5 minutes, which requires high stability and signal-to-noise to measure the double difference signals at $< 0.1\%$, surfaces could only be collected in sets of 50 as defined by the T-jump repetition rate. t-HDVE spectra are the result of ≥ 6 averages.

All equilibrium 2D IR spectra are shown as absorptive correlation spectra. All spectra for t-HDVE(t-DPP) and t-2DIR T-jump experiments are displayed as difference spectra between the transient spectrum and the equilibrium spectrum preceeding the T-jump. Both t-HDVE and t-2D IR spectra were corrected for linear absorption distortions along ω_1 and ω_3 ^{15, 51}.

As with previous t-2D IR experiments, we rely on t-DPP spectroscopy calculated from t-HDVE spectra to provide the detailed timescales for spectral changes we observe in t-2D IR. t-DPP spectra can be collected $\sim x250$ faster than t-2D IR spectra, which allows for increased averaging and finer sampling of time scales following the T-jump. Spectral changes in the one-dimensional t-DPP can be traced directly back to changes in the t-2D IR because the t-DPP spectra are equivalent to the projection of the t-2D IR spectra onto the vertical, ω_3 axis⁵⁰.

8.2.3 SVD Analysis

The t-DPP and t-2D IR difference spectra were analyzed using Singular Value

Decomposition (SVD), a linear decomposition that has been applied in spectroscopy⁵² to separate arrays of difference data, ΔS , into multiple basis components that have orthonormal spectra, U , a weighting factor, S , and an associated orthonormal time dependence, V :

$$\Delta S = USV^T \quad (8.1)$$

The SVD process is entirely mathematical and the identified distinguishable spectra are linear combinations of the spectra of individual species. The usefulness of the procedure is to identify a lower limit on the number of distinguishable spectra, or kinetic species, and to separate signal from noise. Once the spectra are rearranged in descending mathematical importance, as is determined by the weighting factor, S , a subset of components can be rejected based on their insignificance or noisiness. More discussion of SVD is given in Appendix 8.B.

8.3 Results and Interpretation of Experimental Spectra

Previously, our group has conducted equilibrium studies on unlabeled and isotope-labeled TZ2^{1, 12, 27}. At temperatures 10°C above the reported melting point, retention of homogeneously broadened peaks indicating turn structure, mid-strand native H-bonds, and general β -strand structure were interpreted as indications of a partially disordered structure with frayed N- and C-termini^{1, 27}. Aside from a partially disordered structure, a bulged turn structure was also observed, suggesting heterogeneous folding through multiple pathways¹. Although with increasing temperature they are both lost due to net conversion to a disordered or the partially-disordered state, the native type I' β -turn configuration population appeared to increase relative to the bulged turn population.

8.3.1 Equilibrium 2D IR Spectra

In order to understand how the general spectral features report on different structural elements, the 2D IR spectra of five Markov states were simulated previously¹ and reproduced in Figure 8.2. These states were selected to highlight general features. Structural considerations and spectral simulations suggest that other states will produce similar spectra.

The *folded* state corresponds to the native, type I' β -turn observed by NMR in which the

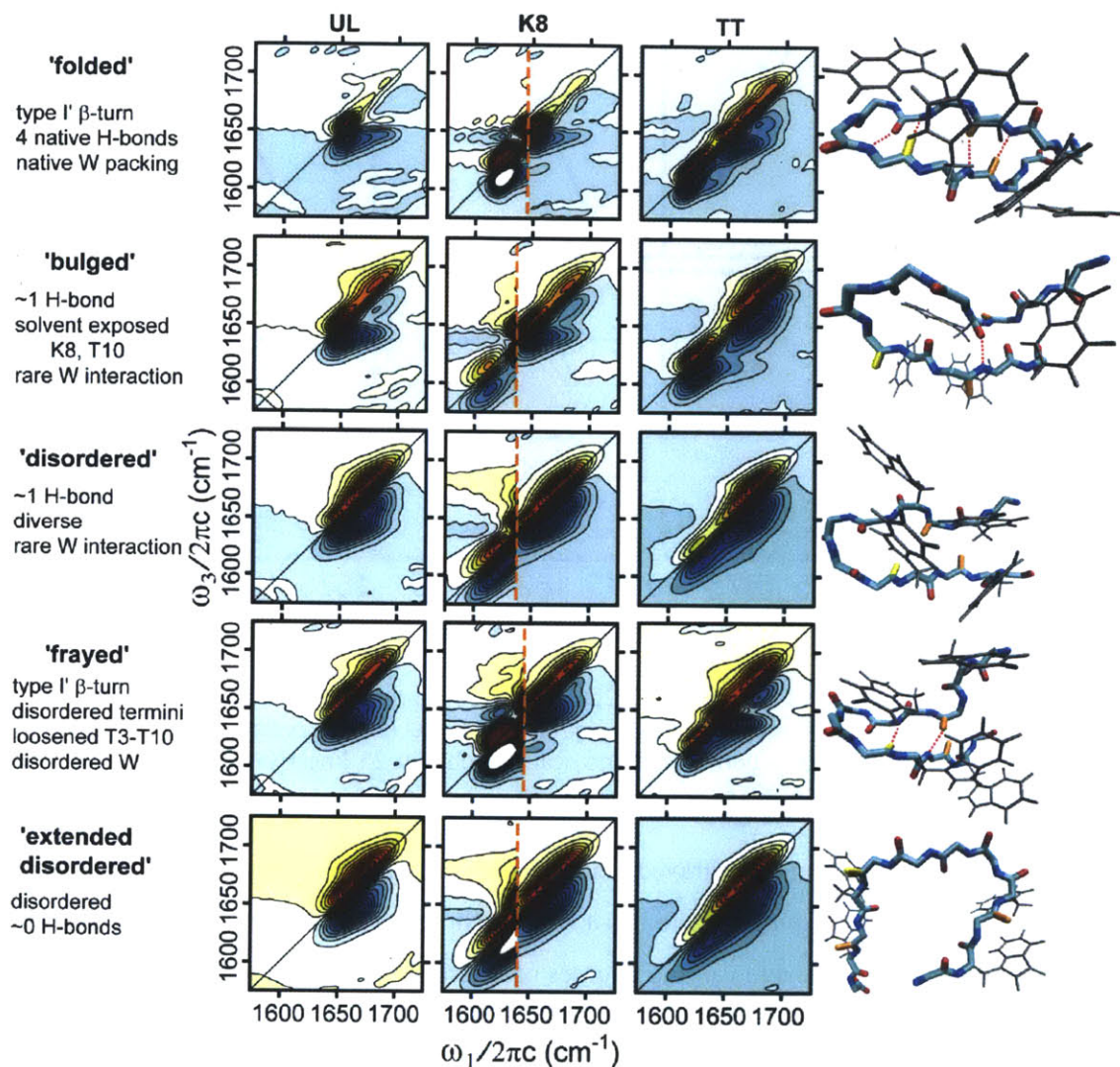


Figure 8.2: The spectra simulated from a few select separate markov states¹. Briefly, the *folded* state is an ordered state similar to the NMR structure^{4,5} with an average of 4 H-bonds (including the K8, T3, T10 C=O cross-strand H-bonds) and native tryptophan side-chain packing. The *bulged* state has solvent-exposed K8 and T10 peptide carbonyls, one average cross-strand H-bond, and the tryptophan side chains rarely interact. The *disordered* state is fairly diverse with a generally misaligned C-terminus registry, a bulged turn, one average cross-strand H-bond, and unorganized tryptophan side-chains. The *frayed* state has the native type I' turn region, two average cross-strand H-bonds, and disordered termini region with proper orientation of T3 and T10, but disordered tryptophan side chains. The 'extended disorder' ensemble is substantially disordered with rare cross-strand H-bonds. On the right, characteristic structures for each state are shown. The K8 carbonyl is yellow, while the T3 and T10 carbonyls are orange.

K8, T3, and T10 oscillators are H-bonded across the strand⁴. The *frayed* state is similar to the

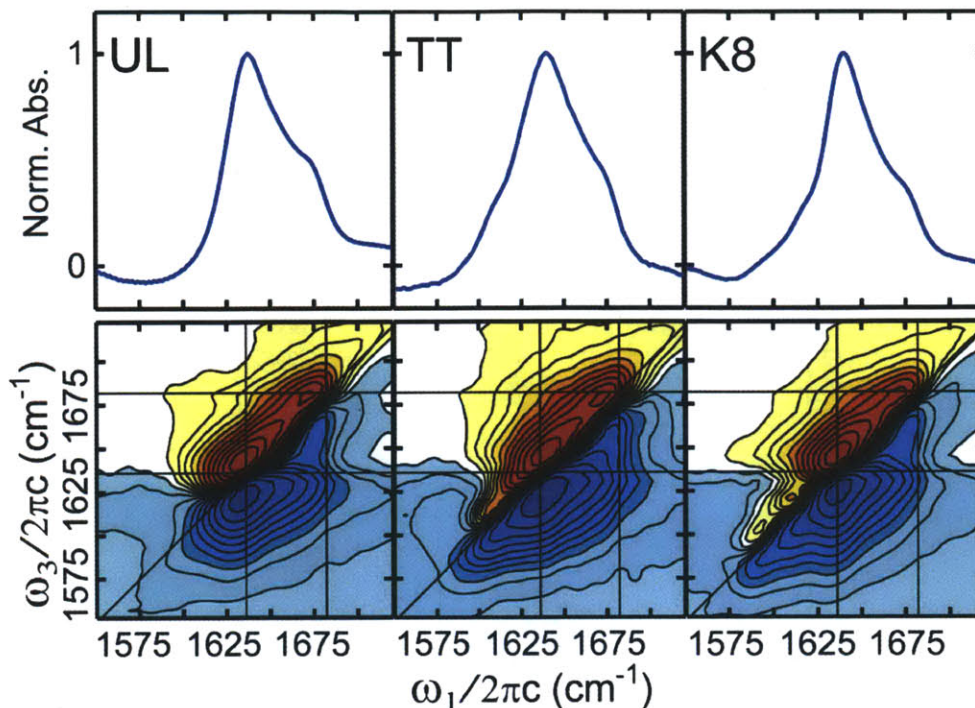


Figure 8.3: Equilibrium FTIR (top) and 2D IR (bottom) surfaces at 35°C for, from left to right, UL, TT, and K8. Spectra were acquired in the parallel (ZZZZ) polarization. The 2D IR spectra are individually normalized and scaled such that 25 linearly-spaced contours span the entire magnitude range of $\sinh^{-1}(S_{Norm.} \times 27.29)$.

folded state except the termini are frayed and the T3 and T10 contact are in the native orientation, but loosened. The *bulged* state has a K8 solvent exposed oscillator. The *disordered* and *extended disordered* states have varying degrees of disorder. We label these simulated states in italics throughout the text to compare general spectral features.

The equilibrium FTIR and 2D IR spectra are shown for the UL, TT, and K8 species in Figure 8.3. In 2D IR spectroscopy, each resonance has two peaks, a positive $1 \leftarrow 0$ peak (fundamental, yellow/red) and a negative, anharmonically shifted $2 \leftarrow 1$ (overtone, blue) peak at lower ω_3 frequency. The 2D IR spectrum can be thought of as a correlation map in which energy initially pumped into a system at ω_1 is emitted at a frequency indicated by the probe axis, ω_3 . Cross peaks that appear at $\omega_1 \neq \omega_3$ indicate coupling between different modes, fast changes in molecular environments, or fast energy transfer.

The UL equilibrium spectrum consists of an underlying broad amide I' peak, with two distinct AP β -sheet features along the diagonal: the 1635 cm^{-1} peak and the lower intensity 1670

cm⁻¹ peak. These are due to the v_{\perp} and v_{\parallel} modes, two delocalized vibrational modes whose dipoles are aligned perpendicular and parallel to the β -strands, respectively¹². These two doublet peaks have crosspeaks indicating coupling between the modes, which manifests itself in the upper left as a positive ridge along $\omega_3=1682$ cm⁻¹, and in the lower right as a negative ridge along $\omega_3 = 1625$ cm⁻¹. The combination of crosspeak ridges and doublet diagonal peaks results in a “Z” shape indicative of an anti-parallel β -sheet, and it is present in the spectra of all three isotopologues. The UL equilibrium spectrum shares many features with the *folded* simulated states. The peaks due to the isotope-labeled residues appear in the 1575-1620 cm⁻¹ region, which is highlighted by the left panels in Figure 8.6. For TT, there is a single diagonally elongated shoulder that extends to lower frequency, while for K8 there are two peaks.

8.3.2 Equilibrium structural assignment

We have previously characterized the TT and K8 peaks, and used structure-based modeling to assign them to peptide configurations^{1, 53}. An individual isotope label introduced into the T10 residue (not shown here) results in the appearance of an isotopically-shifted 1627 cm⁻¹ peak¹. Introduction of two isotope labels at TT results in the observed peak at 1610 cm⁻¹. The extra shifting due to the dual isotope label indicates coupling between the two isotope-labeled oscillators. The 1610 cm⁻¹ peak is assigned to the combined asymmetric stretch due to both oscillators. The coupling, which results in such an intense, red-shifted peak, suggests proximity between the two TT residues, indicating the presence of hydrogen bonding contacts within the central region of the peptide. This TT peak is reproduced in the *folded* simulated spectrum. Disordering of the midstrand TT H-bonds will reduce the coupling and cause the TT shoulder to blue-shift in frequency and decrease in amplitude, as is shown in the blue-shifting in the *bulged* and *frayed* states and the drop in intensity in the *disordered* and *extended disordered* simulated spectra¹.

The two K8 peaks have been attributed to two general turn configurations whose frequencies differ based on the number of accepted H-bonds made to the K8 isotope labeled C=O oscillator¹. The lower frequency (K8-1) peak at $\omega_1=1592$ cm⁻¹ is attributed to a bulged turn configuration in which the K8 carbonyl is pointed out to the solvent, where it can accept additional H-bonds from solvent molecules, as is indicated in the *bulged* simulated spectrum. The extensive diagonal elongation indicates that it is inhomogeneously broadened by a distribution of hydrogen bonding environments. The higher frequency (K8-2) peak at $\omega_1=1610$

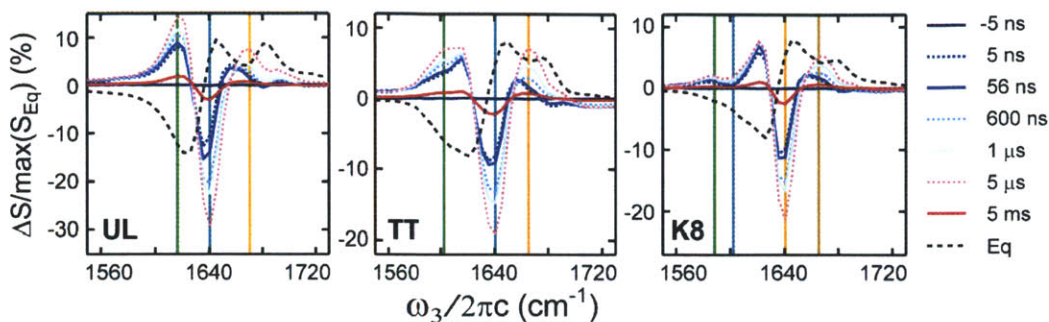


Figure 8.4: ZZZZ $T_i=35^\circ\text{C}$ t-DPP spectra at selected T-jump delay times. Vertical lines indicate the frequency of traces shown in the Figure 8.5.

cm^{-1} is assigned to the native type I' turn, in which TZ2 forms a H-bond across the turn to the N-H of E5, as is consistent with the *folded* simulated spectrum. The equilibrium 2D IR spectrum for the K8 isotope label also shows a crosspeak ridge along $\omega_3=1645 \text{ cm}^{-1}$ indicating strong coupling between the delocalized ν_\perp vibration of the peptide and the isotope labeled oscillator. If the K8 is free to sample a wide variety of environments, we expect a weak, inhomogeneously broadened peak, as is observed in the *disordered* and *extended disordered* simulated spectra.

The assignments illustrate how each spectral feature provides a unique structural marker of the peptide configuration: UL contains information on the global fold of β -sheet structure, TT reports on the mid strand region contacts, and K8 monitors both the native turn and bulged turn configurations. In response to the T-jump, decreasing intensity indicates loss of population corresponding to these reporters. Blue-shifting of peaks indicates loss of the initial population and gain of a population with fewer or weaker H-bonds. Broadening indicates increased conformational flexibility. Statistically significant differences in observed timescales observed for relaxation of different reporters indicates the presence of different kinetic states that are interconverting at different rates.

8.3.3 Transient dispersed pump-probe (t-DPP) Spectra

Although we focus on the t-DPP data, consistent conclusions are reached based on the transient-absorption spectra, which are presented in Appendix 8.C.

A time series of t-DPP difference spectra, $\Delta S = S(\tau) - S_{\text{eq}}$, from time delays between $\tau=-5$ ns and 5 ms are shown for the three isotopologues in Figure 8.4 along with the corresponding scaled equilibrium DPP spectra, S_{eq} . All equilibrium spectra show a positive doublet with peaks near 1645 cm^{-1} and 1680 cm^{-1} , which arise from the ν_\perp and ν_\parallel modes, and a broad negative

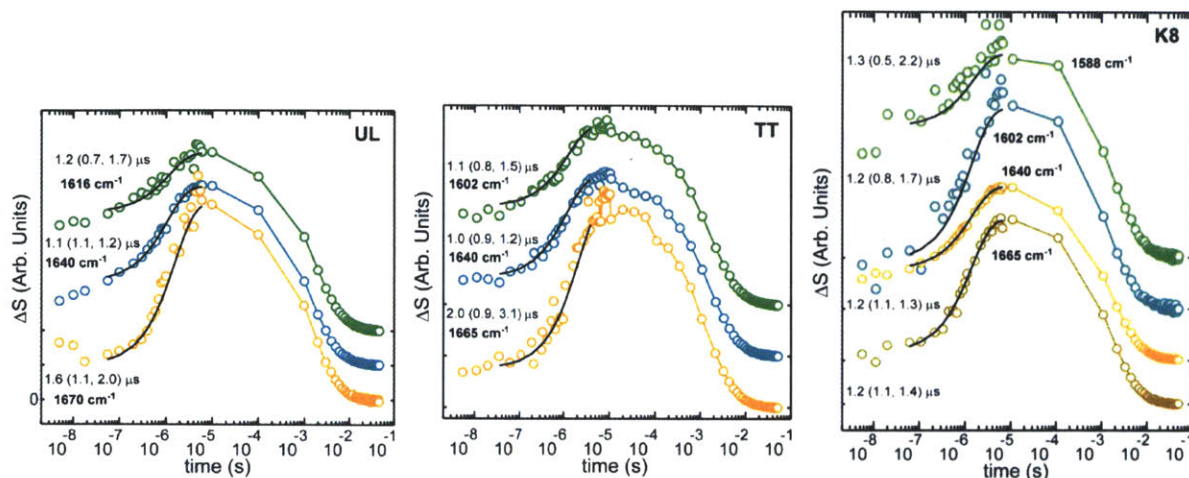


Figure 8.5: Time traces from data shown in Figure 8.4. *Left* – The τ -dependence of the UL t-DPP data is shown for 1616, 1640, and 1670 cm^{-1} . *Center* –The τ -dependence of the TT t-DPP data is shown for 1602, 1640, and 1670 cm^{-1} . *Right* –The τ -dependence of the K8 t-DPP data is shown for 1588, 1602, 1640, and 1670 cm^{-1} . Mono-exponential fits are shown in black with associated time constant and 95% confidence intervals.

feature for frequencies $<1635 \text{ cm}^{-1}$.

Briefly, for the main amide I' band, the 5 ns spectra show significant changes indicating a blue shift of the two AP β -sheet peaks, ν_{\perp} and ν_{\parallel} . On the μs timescale, there is a blue shift of intensity for peaks at 1635-1640 and 1655-1670 cm^{-1} , which suggests that there are multiple underlying peaks whose intensities are changing. Due to the presence of negative and positive overlapping peaks, assessing the percentage change at different points is difficult. The difference t-DPP are presented as $\Delta S / \max(\text{abs}(S_{eq}))$, and the changes of the largest difference peak are on the order of 20-30%.

These spectra and the slices shown in Figure 8.5 exhibit three general timescales: a $<5 \text{ ns}$ T-jump pulse-width limited response, a 1-2 μs response from the sample after the initial T-jump, and a ms return back to equilibrium. On the ms timescale, the temperature drops as the heat diffuses out of the sample. The ms time-dependence is complicated by the fact that it is a convolution of the peptide response and the temperature re-equilibration. Therefore we focus on analyzing the $<5 \text{ ns}$ and μs time-scales, which result from the peptide's response to the initial T-jump.

The isotope labeled TT and K8 peaks ($1590\text{-}1615 \text{ cm}^{-1}$) are apparent in the t-DPP spectra, but their changes are small and they overlap with the main 1621 cm^{-1} positive peak that results

from loss of the negative equilibrium spectral feature. The positions of the most apparent isotope label peaks (TT overtone, K8-1 overtone, and the K8-2 overtone/K8-1 fundamental) are marked with vertical lines in their respective spectra in Figure 8.4, and their traces are shown in the Figure 8.5. For all three of these isotope labeled peaks, there is immediate loss (positive) in the 5 ns spectrum, followed by increased loss on the μs timescale, although the amplitudes of these ns and μs responses are different.

UL		TT		K8	
<i>Metric</i>	τ (μs), (95% C.I.)	<i>Metric</i>	τ (μs), (95% C.I.)	<i>Metric</i>	τ (μs), (95% C.I.)
t-DPP		t-DPP		t-DPP	
1616 cm^{-1}	1.2 (0.7, 1.7)	1602 cm^{-1}	1.1 (0.8, 1.5)	1598 cm^{-1}	1.3 (0.5, 2.2)
1640 cm^{-1}	1.1 (1.1, 1.2)	1640 cm^{-1}	1.0 (0.9, 1.2)	1602 cm^{-1}	1.2 (0.8, 1.7)
1670 cm^{-1}	1.6 (1.1, 2.0)	1665 cm^{-1}	2.0 (0.9, 3.1)	1640 cm^{-1}	1.2 (1.1, 1.3)
P_n	-	P_n	-	1665 cm^{-1}	1.2 (1.1, 1.4)
P_μ	1.1 (1.1, 1.2)	P_μ	1.3 (1.2, 1.4)	P_n	-
t- 2D IR				P_μ	1.2 (1.2, 1.3)
orange	1.8 (0.7, 2.9)				
yellow	3.0 (0.2, 5.8)				

Table 8.1: Table of measured rates from indicated transient spectra. For TT, the P_n data was too noisy for an exponential fit. UL t-2D IR fits come from Ref. 15. C.I. stands for confidence interval.

Slices from the t-DPP spectra at the indicated frequencies are fit to single exponentials in Figure 8.5, and the time constants are listed in Table 8.1. All of the measured exponential rates are indistinguishable from 1.2 μs within their 95% confidence intervals. To identify the observed changes in our ensemble during the ns and μs timescales we interpret the spectra at early and late times. Inspection of different peaks in the isotope-labeled spectra gives insight into the ns and μs response of different parts of the peptide. t-2D IR spectra provides more spectral information than t-DPP, and we focus on analyzing the t-2D IR spectra to understand what states are visited on the separate <5 ns and 1-2 μs timescales.

8.3.4 Singular value decomposition (SVD) into spectral components

Rather than analyzing each frequency point individually, we look for shared timescales between many points. These shared timetraces are assigned a frequency-dependent amplitude, which, when linked together, creates the spectrum of species that share the identified time behavior. This decomposition into spectra is done through SVD, which was applied to the t-DPP and t-2D IR spectra (see Appendix 8.B).

The results from SVD show that the spectral data can be described by two spectra and two times-behaviors:

$$\Delta S \approx S_n P_n^T + S_\mu P_\mu^T \quad (8.2)$$

S_n matches the spectral changes at 10 ns while S_μ captures the spectral changes that grow in on the 1-2 μ s timescale. The P_n and P_μ are vectors that describe the time dependence of the two spectra. The underlying assumption is that we can segregate our observed species into states separated by large, $>kT$, free energy barriers and states separated by small, $<kT$, barriers (see Figure 6.9). The large-barrier separated states $\{A, B, C, \dots\}$ can be associated with wells in the free-energy surface. The smaller-barrier separated states $\{a, a', \dots, b, b', \dots\}$ result from further segregation within the larger wells such that $A = \{a, a', \dots\}$, $B = \{b, b', \dots\}$. Thus, any kinetics on the ns timescale can be assigned to intra-well conversion ($a \leftrightarrow a'$, $b \leftrightarrow b'$) with difference spectrum S_n , while any kinetics on μ s or longer timescales are due to inter-well reactions ($a + a' \leftrightarrow b + b'$) with difference spectrum S_μ . Based on these distinctions, any spectral changes observed at \sim ns timescales can be assigned to population redistribution within individual wells whereas any spectral changes observed at $>\mu$ s timescales can be assigned to interconversion between different wells. Thus, S_μ captures the changes associated with the larger barriers (see Appendix 8.B for details).

Similarly, after separating the ns and μ s spectra for the t-DPP data, we use the t-DPP kinetic traces to decompose the original t-2D IR spectra into new spectra associated with the ns and μ s timescales. The first two components of the t-2D IR SVD analysis were remixed such that their ω_3 projections and time traces matched the t-DPP reweighted SVD components.

Through this SVD analysis, we can encapsulate the ns and μ s changes we observe in the t-DPP and t-2D IR through two spectra. This helps to identify the observed changes. Analysis of the spectra without this decomposition results in an equivalent qualitative interpretation of the

data. By removing noise, the SVD aids the quantitative analysis and provides a clearer separation of spectra. The t-2D IR spectra before decomposition are shown in Appendix 8.A.

8.3.5 t-2D IR – ns response

The ns and μ s difference t-2D IR spectral components are shown in Figure 8.6 along with their matching t-DPP timetraces. As was specified in the separation of ns and μ s spectra, the UL, TT, and K8 P_μ traces start close to 0 and grow in on the ~ 1.2 μ s exponential timescale. The P_n traces start above 0 and stay constant on the μ s timescale. Both P_n and P_μ traces decay back to equilibrium concomitant with the temperature re-equilibration. There is also slight growth of the P_n traces at <100 ns, although this should not be overinterpreted, given that the 100ns timescale may be corrupted by cavitation of the sample which cause artifactual intensity changes¹⁴. T-jump

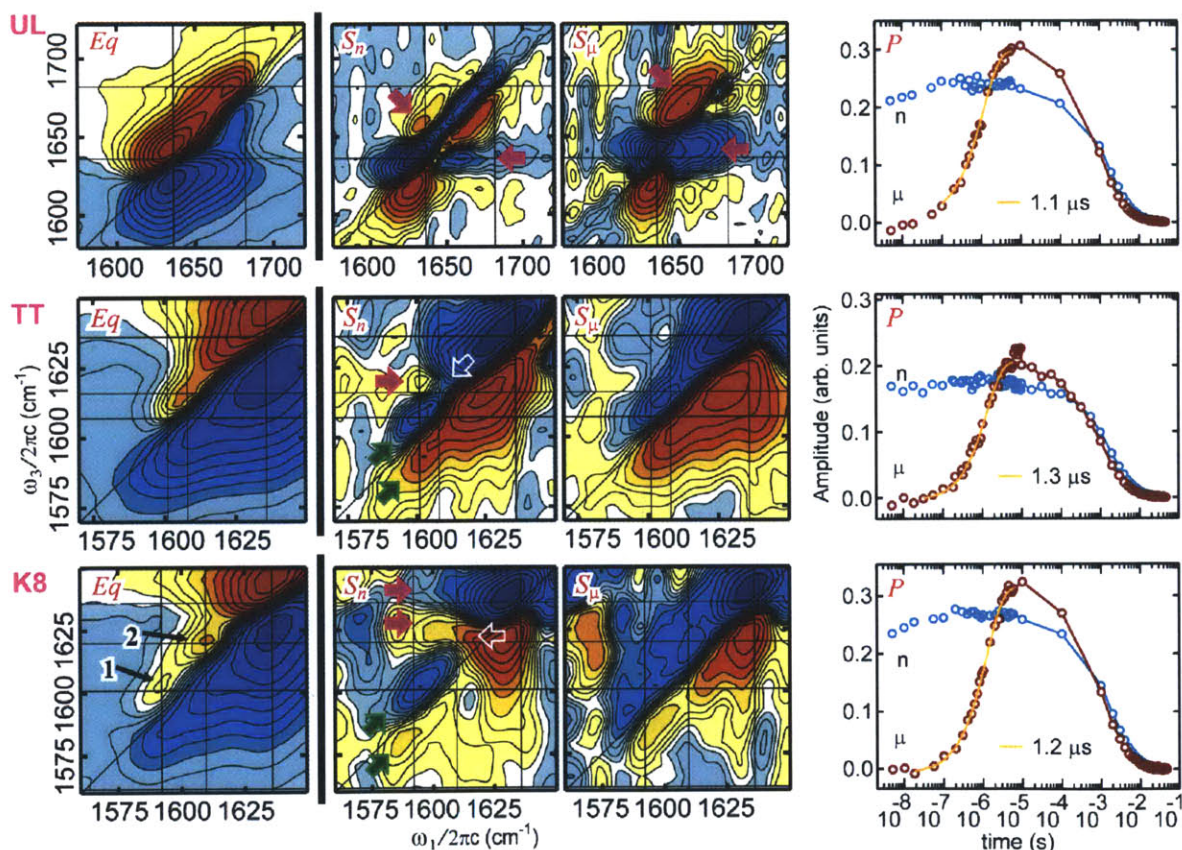


Figure 8.6: Column 1: The equilibrium 2D IR spectra. Column 2+3: The t-2D IR ns and μ s basis spectra resulting from reweighting SVD results. Column 4: The t-DPP SVD time traces corresponding to the spectra shown in columns 2 and 3. The 2D IR spectra are displayed as described in Figure 8.3. The 2D IR spectra are plotted over different frequency ranges.

fluorescence experiments have observed similar behavior which was interpreted as arising from

loosening of the tryptophan core^{24, 34}. Although there may be molecular changes on this timescale, we do not observe distinct amide I spectral changes in this time range, and therefore do not assign it to backbone conformational changes.

Inspection of the S_n spectra allows for the identification of the spectral changes that occur on the T-jump pulsewidth limited (< 5ns) timescale. We expect to see loss of the equilibrium population and gain of a new conformational population. The ns timescale places constraints on the magnitude of the conformational change. Because the timescale is related to the free energy differences, the ns timescale suggests an increase in a population structurally similar to the TZ2 equilibrium population. The fastest peptide unfolding (loss of secondary structure) occurs on the >50 ns timescale^{54, 55}. Based on the <5 ns response, we expect less drastic *net* changes, such as fluctuations in local structure with ~1 nm dimensions, changes in side chain conformations, diffusion limited interconversion, peptide reorientation, and solvent/protein H-bond disruption. Previous t-2D IR experiments have shown that the <5ns timescale may be partially, if not entirely, due to this last mechanism – changes in the solvent H-bonding to the peptide. Increased kinetic energy causes the average number of H-bonds between the D₂O and peptide to drop, which causes a blue shift and homogeneous broadening in the observed frequencies. This phenomenon is the microscopic manifestation of the decrease in water density with temperature, which causes an abrupt volume change in the solvent, the protein, and at the interface between the two. The increased kinetic energy may also cause increased fluctuations within the backbone, which would result in similar blue shifting and broadening. Thus, the S_n spectral response is due to localized changes rather than the global conformational changes that we think of for two-state folding/unfolding processes. We expect the ns changes are partially due to an average loss of solvent/peptide or intrapeptide H-bonds, and, for each spectrum, we discuss this possibility along with some others below.

Unlabeled (backbone)

The UL S_n spectrum has two sets of paired features. For t-2D IR difference spectra, the sign (positive or negative) of gain and loss peaks depends on the sign of the equilibrium peak¹⁶ (Figure 3.15). Generally, the equilibrium spectrum is positive (red) above and negative (blue) below the diagonal, which means oppositely signed peaks in the difference spectra indicate loss. Along the diagonal, there are loss features represented by doublets (blue above and red below the diagonal) that span the entire diagonal of the spectrum with narrow anti-diagonal widths. This

pair of features is most intense at $\omega_1=1635$ and 1670 cm^{-1} , which corresponds to loss on the red-side of the equilibrium anti-parallel β -strand v_{\perp} and v_{\parallel} modes. The other pair of spectral features (marked by purple arrows in Figure 8.6) occurs in the off-diagonal regions where there is a red/yellow(gain) feature above the diagonal and a blue(gain) ridge below the diagonal extending along $\omega_3=1636\text{ cm}^{-1}$. These features indicate gain of off-diagonal intensity. We attribute both the lower frequency diagonal loss features and homogeneous broadening to increased conformational flexibility and fluctuations that upset solvent to peptide H-bonds due to the increased kinetic energy at higher temperature. Because H-bonds between D_2O and the peptide are affected, we also suspect that there will be similar loosening of intrapeptide H-bonds, as is observed in the *frayed* simulated spectrum.

TT Label (mid-strand region)

For a loss of H-bonding surround the TT and K8 oscillators, we expect to see blue-shifting of each oscillator as a result of the strengthening of the C=O bond and broadening of the peak due to increased fluctuations. As in the UL case, the TT and K8 S_n spectra exhibit similar H-bond loosening behavior. In the TT S_n spectrum, green arrows indicate two loss features along the diagonal (blue above the diagonal, red below) that indicate loss of the low frequency equilibrium TT peak. Comparing the frequency of this feature with the peak in the equilibrium distribution indicates that only the very strongest of TT couplings are disturbed – or equivalently, that the equilibrium TT peak is lost at all frequencies, but there is gain of a population with weaker TT coupling, which produces a blue-shifted peak. A positive(gain) feature is marked with a purple arrow. There is also a dip in the blue, loss ridge (white arrow) which results from gain of an underlying positive peak that depletes the negative peak. These two features marked with arrows suggest the presence of the same underlying peak, which appears to be growth of a blue-shifted, broadened TT peak. Overall the observed spectral changes are consistent with the changes observed in equilibrium spectra as the temperature is increased - we expect the equilibrium changes to be a combination of the S_n and S_{μ} changes. One can explain all features through a loosening of the central contacts of folded β -turn conformers or through loosening of proximate H-bonds between the peptide and the solvent. Although fraying of the ends of the hairpin may lead to loosening of the central β -turn contacts, which would produce features predicted by the *frayed* simulated spectrum, we have no isolated, direct probe of this region. We do not think that the blue shifting is due to complete loss of the TT contacts, which we expect to

occur on longer timescales and result in a significant drop of intensity at all $< 1620 \text{ cm}^{-1}$ frequencies, as we observe in S_{μ} . The presence of the purple/white-arrow marked underlying peak indicates retention of the TT contacts.

K8 Label (β -turn region)

The K8 S_n spectrum shown in Figure 8.6 has five main features. In the K8-1 region, the diagonally-elongated two-peak doublet marked with green arrows indicates loss over the entire K8-1 peak and loss on the red side of K8-2. Another two-peak doublet (marked with purple arrows) indicates loss of the crosspeak ridge between the K8-2 peak and the ν_{\perp} band observed in the equilibrium K8 spectrum. Cross peaks to the main amide peaks are only expected for the β -turn conformers K8-2¹, and we attribute the majority of this off diagonal intensity to this phenomenon. Any residual intensity in the ν_{\perp} /K8-1 crosspeak region is weak, and can be attributed to the horizontal, low level homogeneous broadening attributed to the diagonal peak and observed in the simulated spectra¹. The fifth feature (marked with a white arrow) appears to be growth of a positive peak blue-shifted relative to the K8-2 peak position, which we interpret as the weakening of H-bonds within the type I' β -turn structures. The loss of K8-1 is paired with no obvious gain. Although we cannot definitively say that the loss of K8-1 does not lead to gain of K8-2, it is highly unlikely that bulged K8-1 structures would convert to β -turn K8-2 structures on the <5 ns timescale. Rather, we expect that upon losing H-bonds to the solvent, the solvent exposed K8-1 carbonyl would shift up in frequency, and it may be hidden under the K8-2 peak. The larger blue-shift expected for K8-1 versus K8-2 is expected because it is more sensitive to H-bond changes given its solvent exposed nature.

On the pulse-width limited timescale following the T-jump, the anti-parallel β -sheet (UL), mid-strand region (TT), native turn configuration (K8-2), and solvent exposed K8 carbonyl (K8-1) all exhibit loss on the low frequency side of peaks, which is consistent with loss of stronger H-bonds, that causes frequency blue-shifting. Both the TT peak and K8-2 peak appear to broaden and blue-shift consistent with loss of solvent H-bonds and possibly loosening of cross-strand H-bonds. The presence of the blue-shifted TT and K8-2 peaks indicates that the mid-strand and turn-region H-bonds are not lost on the ns timescale. For complete loss of these contacts, which is not anticipated on this timescale, we would not expect the appearance of the blue-shifted TT and K8-2 peaks, as the *disordered* spectrum shows a significant drop in intensity and large distribution of frequencies.

8.3.6 t-2D IR – μ s response

Now that the spectrum of the ns-populated species has been identified as S_n , rather than analyze the un-decomposed data, we interpret S_μ , which represents the remaining spectral features in the difference t-2D IR spectra after subtracting the S_n spectrum. Therefore, S_μ highlights the spectral changes that occur exclusively on the μ s timescale, and it can be thought of as a combination of loss of the equilibrium spectrum and gain of the spectrum of the μ s-populated configurations.

Unlabeled (backbone)

The UL S_μ spectrum displays loss and gain features. Along the diagonal, one observes loss of peaks at 1640 cm^{-1} and 1680 cm^{-1} corresponding to the AP β -sheet v_\perp and v_\parallel modes, and the loss of a β -sheet ridge below the diagonal. The loss of these features indicates loss of β -sheet structure based on their close correspondence to calculated spectra for the folded Type I' β -turn structure^{1, 46}. In the 1660 cm^{-1} region marked with purple arrows, there is a concomitant increase of a broad doublet peak, which we assign to disordered structures based on both empirical and simulation results¹. In the previous simulation results, the UL loss and gain features are consistent with a *folded, bulged, or/and frayed* \rightarrow *disordered, or/and extended disordered* transition. The observed spectral changes are also consistent with those predicted in simulation, although the timescales between experiment and simulation are different⁴⁶. Partial melting of a β -sheet structure is a possibility.

TT (mid-strand) and K8 (β -turn region) labels

The TT and K8 isotope-region S_μ spectra display loss along the diagonal without equivalent gain features. Loss of the TT and K8-2 peaks (blue above the diagonal, red below the diagonal) indicates that the mid-strand and turn region H-bonds are broken upon disordering. This assessment is consistent with the simulated spectra, where loss of the TT contacts and loss of the β -turn structure result in a significant loss of intensity in the isotope labeled TT and K8-2 regions (e.g. the *disordered* and *extended disordered* simulated spectra). If there were only partial unfolding and the TT or K8-2 H-bonds were retained while surrounding structures were lost, we would expect less uniform loss features - preferential loss on the red frequency side as is observed in the S_n case, where blue-shifted TT or K8-2 peaks appear due to the increased surrounding fluctuations loosening of the remaining H-bonds, as is observed by comparing the *folded* and *frayed* TT simulated spectra¹. Because we do not observe this blue-shifting on the

microsecond timescale, we expect that the folded structures are disordering rather than partially unfolding. Upon disordering and loss of contacts, the absence of gain features is expected for the TT case, in which loss of cross-strand coupling will result in a large $\sim 15\text{ cm}^{-1}$ blue-shift, which will hide the uncoupled T3 and T10 peaks under the main amide I' peaks. Solvent exposure of the K8 carbonyl is expected to result in large inhomogeneous broadening, which results in weak intensity spread over a large frequency range. Thus, the disordered state K8 peak may be elongated, red-shifted, and hidden underneath diagonal loss features or, more speculatively, it may be blue-shifted due to its hydrophobic burial inside a collapsed disordered state, which has been predicted in simulation.

The K8-1 peak is also lost in the K8 S_{μ} spectrum. As discussed above, based on its frequency, the K8-1 peak is due to the K8 carbonyl in a geometry in which it can accept multiple H-bonds from the solvent, at most two. The configurations responsible for the K8-1 are misfolded such that the K8 carbonyl is pointed out into the solvent. The K8 isotope labeled disordered state is also expected to have intensity in this low frequency region because some of its K8 oscillators will be exposed to solvent, where they can accept multiple H-bonds from D_2O . The intensity is expected to be less for the disordered structures, however, because of the range of solvent exposure in the disordered structure – the K8-1 bulged turn structures force the K8 oscillator into the solvent, whereas it is not necessary in the disordered configurations. This difference in intensity is observed in the simulated *bulged* and *disordered* spectra. The absence of increase of intensity in this region on either the ns or μs timescale indicates that either 1) the disordered TZ2 structure is a collapsed state in which the K8 carbonyl is not fully solvent exposed, which would blue-shift its frequency relative to the K8-1 configuration, or 2) the gain of intensity due to populating a disordered state does not outweigh the loss of the K8-1 peak due to loss of the solvent-exposed misfolded population. Previous equilibrium spectra collected at high temperatures, where a significant fraction of the population is disordered, do not exhibit an increase in intensity in this solvent exposed region, which may suggest that the disordered state is collapsed. The μs timescale associated with loss of these misfolded K8-1 configurations indicates a $> kT$ barrier separating the misfolded configurations from the disordered state.

Previous T-jump experiments have attributed the μs timescale response to disordering of the peptide^{24-26, 31}. Our results are consistent with this interpretation, and our increased spectral content can help to interpret the final state of the disordered peptide.

8.4 Discussion

8.4.1 Assignment of Kinetic States

Based on the spectral signatures and relaxation time-scales in our experiments, our prior analysis of equilibrium TZ2 2D IR spectra, prior characterization of this system, and structural constraints to conformational exchange, we propose a free energy surface for the folding of TZ2 in Figure 8.7. This surface is characterized by three thermodynamically distinct states: folded (F), misfolded (M), and disordered (D). We also assign two states corresponding to a change in the average structures *within* F and M between the initial (F₀,M₀) and final (F,M) temperature distributions spanned by the T-jump. For the purpose of characterizing structures, the hydrogen bonds in the following discussion (not in Figure 8.7) are numbered using the residue labeling convention shown in Figure 8.1.

Folded (F and F₀): Contain the type I' β -turn, hydrogen bonds 1-4 and perhaps 5, and edge-to-face Trp sidechain packing presented in the original NMR structure and in the *folded* state configuration from simulations. The F₀ spectrum contains ν_{\perp} and ν_{\parallel} peaks, the K8-2 peak, and the TT peak. On thermal excitation, the F state may include a *frayed* state with H-bonds 1-3 and loosened packing of Trp 2-11. Relative to the F₀ spectrum, the spectrum of F shows more homogeneously broadened ν_{\perp} , ν_{\parallel} , K8-2, and TT peaks, as well as blue-shifted K8-2 and TT frequencies.

Misfolded (M and M₀): These exhibit solvent-exposed K8 states with bulged turns, misregistered states, some non-native H-bonds near the termini and poor packing of tryptophans. These spectra feature the K8-1 peak, which is expected to blue shift on thermal excitation.

Disordered (D): Represents unstructured states with ≤ 1 cross-strand H-bond. The structures vary from compact to extended, and K8 is on average not entirely solvent-exposed. The D spectrum contains a single disordered main amide I peak at ~ 1660 cm⁻¹, and the TT and K8 peaks are expected to be broad and weak due to lost contacts. We expect a weak, inhomogeneous peak in the low frequency region due to a buried and partially solvent-exposed K8 carbonyl.

These assignments represent the minimum number of states that can explain the equilibrium and transient data. We assume that experiments conducted with sensitivity to orthogonal coordinates

may either reveal hidden states or may allow for discrimination of multiple subpopulations within our general states.

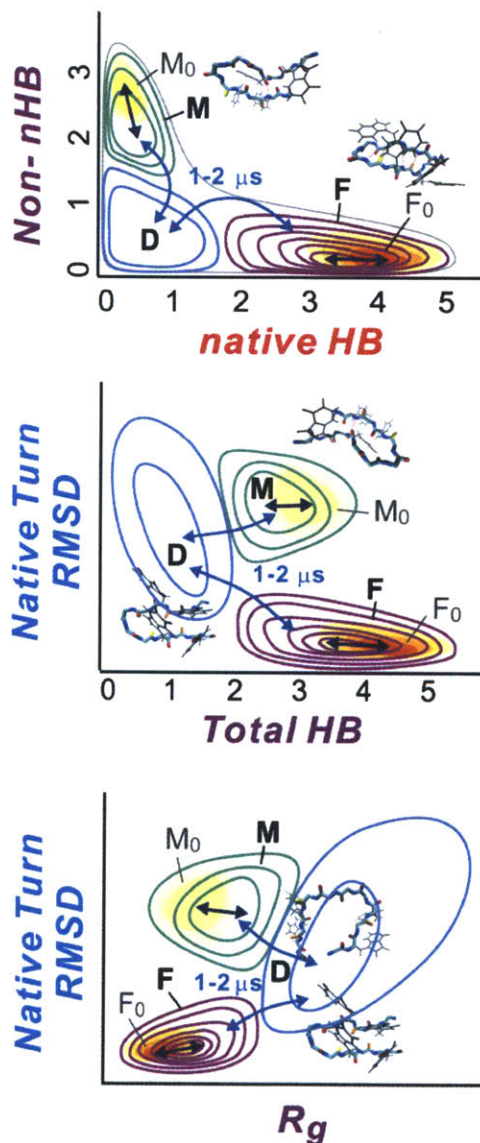


Figure 8.7: *Proposed Free Energy Surface* – Experimentally informed proposal for the TZ2 conformational free energy surface. Arrows indicate movement of population on either the $< 5\text{ ns}$ (dark blue) or $1\text{-}2\ \mu\text{s}$ timescale. Different coordinates are used for each plot. Non-native H-bonds (Non-nHB) and native H-bonds (native HB) refer to the total number of cross-strand H-bonds (not the labels from Figure 8.1). Native turn root mean squared displacement (Native Turn RMSD) describes the structure of the ENGK relative to the native type I' β -turn. Total H-bonds (Total HB) refers to the sum of native and non-native cross-strand H-bonds. Radius of gyration (R_g) describes the compactness of the full structure.

Based on the UL spectrum, the equilibrium ensemble contains structures with native type I' β -turn conformations. The TT and K8-2 peaks indicate hydrogen bonds 2-4 are intact in these equilibrium β -turn conformations, and the resulting geometric constraints imposed by these indicate that HB1 must be present in the folded conformations. It is difficult to obtain information on the structure of the C- and N-termini since the peak of the S1 isotope label is not clearly isolated from the main amide I' band¹. Previous dual labeling of alanine substitutions at A1 and A10 in a mutant TZ2C have shown that the A1 and A10 residues exhibit minimal coupling, indicating fraying at the end³¹. This observation, and the high frequency of S1, suggest that the native folded state of TZ2 has on average 4 of 6 possible cross-strand H-bonds. We therefore place the free energy minimum of the initial folded state (F_0) at 4 native cross strand hydrogen bonds.

Based on qualitatively similar early time difference spectra observed for diglycine,¹⁵ the T-jump perturbs these type I' β -turn structures on the <5 ns timescale, leading to a net weakening or loss of peptide and/or solvent H-bonds. Also, the K8-2 and TT peaks shift up in frequency, which suggests loosening of the cross-strand native H-bonds, but the signs of complete loss of these H-bonds are not observed. Loss of HB4 would result in similar spectral changes, and we cannot entirely rule it out, especially because some simulations identify a partially folded state with three native H-bonds that is separated from the native state by a small barrier⁴¹. Considering it would require significant conformational change, we expect the outer W2-W11 side-chain packing to be retained. Because of these spectral changes, we believe that the folded well at the final temperature (F) is slightly shifted to lower hydrogen bond number from the initial temperature either as a result of changing interactions with the solvent or fraying, and the T-jump leads to an abrupt nanosecond relaxation within this well. The large spectral shifts are interesting because they indicate the possibility for frequency-shifting and broadening that do not result from significant conformational change of the peptide backbone. These results are pertinent for equilibrium temperature dependent studies in which shifts are normally attributed to protein backbone conformational change, but may be the result of changes in interaction with the solvent, as has been observed for the model peptide N-methylacetamide^{56, 57}.

On the $\sim 1.2 \mu\text{s}$ timescale, we believe that the native turn structure melts into a disordered state D . The transition state between D and F likely contains HBs 1 and perhaps 2, consistent with a zip-out folding mechanism for this transition. The high temperature equilibrium TZ2

spectrum displays both disordered and β -sheet characteristic peaks. Previously, we assigned this spectrum to partially folded molecules^{1, 12}. Based on the absence of appearance of frequency-shifted β -sheet peaks or blue-shifted TT or K8-2 peaks on the microsecond timescale, we think the disordered state retains few native H-bonds. Thus, the high temperature spectrum may result from a mixture of disordered and folded molecules, although we are only speculating on the high temperature states based on the data presented here, that was taken at an initial temperature of 35°C. The lack of intensity in the K8 isotope label region suggests that the disordered state is not completely extended into the solvent where multiple H-bonds to an exposed K8 carbonyl would result in a red-shifted K8 peak. Consistent with this conclusion, the disordered state observed in some simulations is also collapsed^{23, 25, 36, 37, 43, 48, 58}, and fluorescence studies indicate fully extended structures only at higher temperatures and high denaturant concentrations^{23, 24}.

The presence of a well-defined K8-1 equilibrium peak suggests that at 35°C there exist stable and likely compact conformations with solvent exposed K8 carbonyls, the M_0 state. This peak decreases in intensity with increasing temperature, which indicates that it is distinct from the disordered state, which increases in population with increasing temperature. Comparing the ratio of the intensity of the K8-2 and K8-1 peaks suggests that ~70% of the population has native turn conformations and 30% has a solvent exposed K8.

The solvent-exposed K8 conformations are lost on both the <5 ns and μ s timescale. As with the β -turn structures, we expect the loss of H-bonds to solvent due to increased kinetic energy will cause a shift in the frequency of the K8-1 peak, especially considering the solvent exposed K8 structures can accept multiple H-bonds from the solvent. The <5 ns changes reflect solvent dominated relaxation from M_0 to M . We also expect that the *disordered* structures will have a <5 ns response due to average loss of H-bonds to solvent, but this is not expected to generate a significant difference in the spectra. At our initial temperature the majority of the population is in the F or M states and the disordered state is broad enough that slight changes in solvation probably correspond to reweighting of populations within the original state.

The μ s timescale indicates that T-jump induced conformational changes to the solvent-exposed K8 structures (K8-1) proceed over a significant energetic barrier. This barrier indicates that the solvent-exposed K8 state is compact and stabilized in structures that include non-native central HB contacts and retain tryptophan packing, such as the bulged structure in Figure 8.2. Misregistered states have been observed in some long simulations^{10, 23, 35, 41, 43}. In a Markov state

analysis in which misregistered states were identified, the four longest timescales observed were due to conversion of native-to-disordered and misregistered-to-disordered structures¹⁰. This is consistent with our interpretation, in which we assign the observed μs relaxation to these two processes.

The structural dissimilarity of misregistered structures relative to the native state indicate that direct transitions between F and M states are highly improbable. These states differ in strand orientation and registry, and in the side-chain packing. A concerted inversion of strands and side-chains is needed to make this transition, and therefore, an exchange of these structures is far more likely to proceed step-wise through the D state. These structural differences may also explain why they have not been observed in some simulations, where extensive sampling would be required. High temperature simulation would be at a disadvantage since the solvent exposed K8 population is lost to a greater extent relative to the F state.

If the disordered state links M and F, and has the highest free energy barrier, this also explains why all barrier crossing kinetics are observed within the 1-2 μs window. In the most general case T-jump relaxation kinetics involving three states in series would be bi-exponential. However, both folding and unfolding rates are reflected in the relaxation kinetics, and if the fastest rate constants are D \rightarrow M and D \rightarrow F, these rates would be the dominant effects in the observed relaxation times. In fact, many hairpins exhibit observed relaxation time constants in the 1-5 μs range, supporting that the folding rates dominate the response⁵⁹. At the temperatures studied in this work, previous T-jump IR experiments have observed the convergence of frequency-dependent exponential timescales, suggesting that if this previously observed frequency dependence is due to different kinetics associated with the M and F states, then it is consistent with the convergence that we observe. Although the F is energetically more stable, it is unsurprising that the observed rates for F to D and M to D conversion may be similar.

The solvent exposed K8 M conformations and β -turn structures appear to be off-pathway states, as indicated in our proposed free energy surface.

8.4.2 Mechanism

The collapsed nature of our disordered state indicates that there is some hydrophobic collapsed state at all times, although the tryptophans are not necessarily well ordered. Thus, some hydrophobic collapse precedes formation of native cross strand H-bonding. This is consistent with fluorescence experiments, which have noted that full loss of the hydrophobic

collapsed structure only occurs at high temperatures or under high denaturant concentrations^{23, 24}. The order in which tryptophan side chains pack into native orientations and different H-bonds form is still in question.

Although the conducted T-jump fluorescence studies were done with different solvent and temperature conditions, which prevents direct comparison to our results, fluorescence relaxation rates were also in the μs time range^{24, 34}. The coincidence of rates determined with IR, which is sensitive to H-bonding structure, and rates determined with fluorescence, which is sensitive to tryptophan packing, suggests that tryptophan ordering and H-bonding are highly coupled. Geometrically, this makes sense – correct H-bonding or tryptophan packing requires nearly native contacts of sidechains and the backbone.

The present experiments cannot definitively distinguish between zipping-in or zipping-out mechanisms. The M/M_0 and F/F_0 observed shifted states are structurally similar, and provide no definitive insight into the mechanism. The M and F may be due to disordering, but they can also be explained by a solvent H-bond loosening. β -strand structure, mid-strand H-bonds, and the turn H-bond are lost simultaneously on the $\sim\mu\text{s}$ timescale. This is consistent with some simulations that identify a partially folded peptide with 2-3 H-bonds as a short lived transition state, which would be present in too small a concentration for our experiments to observe^{37, 41, 48}. All of our experiments were conducted at 35°C to maximize the difference signal and minimize the possibility for aggregation. Experiments by Keiderling and Hauser indicate frequency dependent folding rates at lower temperatures, suggesting the observation of an intermediate state³¹. The slower relaxation rates observed for the turn region versus the midstrand region might indicate unfolding from the termini towards the turn, but these rates are the convolution of multiple processes (the sum of the forward and reverse rate in the simplest two state case) and cannot be conclusively analyzed without correct identification of the underlying free energy surface. Consistent with our data, the previously observed frequency dependence of the rates disappears at higher temperatures. This convergence at higher temperatures may be due to simple convergence of separate rates which makes the intermediate state timescale indistinguishable. The convergence may also be due to tuning of the favorability of intermediate states which would change the amplitude of different pathways.

Our data does allow us to identify the misfolded state as an off-pathway intermediate. The geometry of our states indicate that conversion between folded and misregistered states

involves passage through a hub-like disordered state.

8.5 Conclusion

T-jump t-2D IR provides spectral characterization of multiple kinetic states. Here we apply isotope labeling strategies to probe specific H-bonds within the peptide TZ2. We experimentally identify 5 different kinetic states. The folded, native state is composed of at least 4 cross-strand H-bonds and a type I' β -turn. An off-pathway misfolded state consists of misregistered structures. The disordered state is compact and unstructured. Conversion between the misfolded and folded states requires loss of stabilizing native and non-native contacts and movement through the disordered state. Isotope-edited T-jump t-2D IR experiments provide sensitivity to multiple structures and increased spectral information. Experiments such as this strive to bridge the gap between experiment and atomistically-detailed MD simulations.

8.6 Acknowledgment

I would like to thank Joshua Lessing for helping to purify some of the trpzip2 samples. Chunte Sam Peng helped to collect the data, and he and Carlos Baiz provided appreciated feedback on the chapter.

8.7 References

- (1) A. W. Smith, J. Lessing, Z. Ganim, C. S. Peng, A. Tokmakoff, S. Roy, T. L. C. Jansen, and J. Knoester, "Melting of a β -Hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations," *J. Phys. Chem. B* **114**, 10913-10924 (2010).
- (2) G. R. Bowman, V. A. Voelz, and V. S. Pande, "Taming the complexity of protein folding," *Curr. Opin. Struct. Biol.* **21**, 4-11 (2011).
- (3) B. Audun, "Is it always possible to distinguish two- and three-state systems by evaluating the van't Hoff enthalpy?," *Physical Biology* **1**, 152 (2004).
- (4) A. G. Cochran, N. J. Skelton, and M. A. Starovasnik, "Tryptophan zippers: Stable, monomeric β -hairpins," *Proc. Natl. Acad. Sci. USA* **99**, 9081-9081 (2001).
- (5) B. Doug, "What have we learned from the studies of two-state folders, and what are the unanswered questions about two-state protein folding?," *Physical Biology* **6**, 015001 (2009).
- (6) J. Kubelka, J. Hofrichter, and W. A. Eaton, "The protein folding 'speed limit'," *Curr. Opin. Struct. Biol.* **14**, 76-88 (2004).
- (7) M. S. Searle and B. Ciani, "Design of β -sheet systems for understanding the thermodynamics and kinetics of protein folding," *Curr. Opin. Struct. Biol.* **14**, 458-464 (2004).
- (8) N. Singhal, C. D. Snow, and V. S. Pande, "Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin," *J. Chem. Phys.* **121**, 415-425 (2004).
- (9) J. D. Chodera, W. C. Swope, J. W. Pitera, and K. A. Dill, "Long-Time Protein Folding Dynamics from Short-Time Molecular Dynamics Simulations," *Multiscale Modeling & Simulation* **5**, 1214-1226 (2006).
- (10) J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, "Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics," *The Journal of Chemical Physics* **126**, 155101 (2007).
- (11) V. A. Voelz, G. R. Bowman, K. Beauchamp, and V. S. Pande, "Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1-39)," *J. Am. Chem. Soc.* **132**, 1526-1528 (2010).
- (12) A. W. Smith and A. Tokmakoff, "Amide I two-dimensional infrared spectroscopy of β -hairpin peptides," *The Journal of Chemical Physics* **126**, 045109 (2007).
- (13) A. Remorino, I. V. Korendovych, Y. Wu, W. F. DeGrado, and R. M. Hochstrasser, "Residue-Specific Vibrational Echoes Yield 3D Structures of a Transmembrane Helix Dimer," *Science* **332**, 1206-1209 (2011).
- (14) H. S. Chung, M. Khalil, A. W. Smith, and A. Tokmakoff, "Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics," *Rev. Sci. Instrum.* **78**, 063101 (2007).
- (15) K. C. Jones, Z. Ganim, C. S. Peng, and A. Tokmakoff, "Transient two-dimensional spectroscopy with linear absorption corrections applied to temperature-jump two-dimensional infrared," *J. Opt. Soc. Am. B* **29**, 118-129 (2012).
- (16) J. Bredenbeck, J. Helbing, R. Behrendt, C. Renner, L. Moroder, J. Wachtveitl, and P. Hamm, "Transient 2D-IR spectroscopy: Snapshots of the nonequilibrium ensemble during the picosecond conformational transition of a small peptide.," *J. Phys. Chem. B* **107**, 8654-8660 (2003).

- (17) C. Kolano, J. Helbing, M. Kozinski, W. Sander, and P. Hamm, "Watching hydrogen-bond dynamics in a β -turn by transient two-dimensional infrared spectroscopy," *Nature* **444**, 469-472 (2006).
- (18) R. Kania, A. I. Stewart, I. P. Clark, G. M. Greetham, A. W. Parker, M. Towrie, and N. T. Hunt, "Investigating the vibrational dynamics of a 17e- metalcarbonyl intermediate using ultrafast two dimensional infrared spectroscopy," *Phys. Chem. Chem. Phys.* **12**, 1051-1063 (2010).
- (19) C. R. Baiz, M. J. Nee, R. McCanne, and K. J. Kubarych, "Ultrafast nonequilibrium Fourier-transform two-dimensional infrared spectroscopy," *Opt. Lett.* **33**, 2533-2535 (2008).
- (20) W. Xiong, J. E. Laaser, P. Paoprasert, R. A. Franking, R. J. Hamers, P. Gopalan, and M. T. Zanni, "Transient 2D IR Spectroscopy of Charge Injection in Dye-Sensitized Nanocrystalline Thin Films," *J. Am. Chem. Soc.* **131**, 18040-18041 (2009).
- (21) E. R. Andresen and P. Hamm, "Site-specific difference 2D-IR spectroscopy of bacteriorhodopsin," *J. Phys. Chem. B* **113**, 6520-6527 (2009).
- (22) H. S. Chung, Z. Ganim, K. C. Jones, and A. Tokmakoff, "Transient 2D IR spectroscopy of ubiquitin unfolding dynamics," *Proc. Natl. Acad. Sci. U. S. A.*, 14237-14242 (2007).
- (23) W. Y. Yang, J. W. Pitera, W. C. Swope, and M. Gruebele, "Heterogeneous Folding of the trpzip Hairpin: Full Atom Simulation and Experiment," *J. Mol. Biol.* **336**, 241-251 (2004).
- (24) W. Y. Yang and M. Gruebele, "Detection-Dependent Kinetics as a Probe of Folding Landscape Microstructure," *J. Am. Chem. Soc.* **126**, 7758-7759 (2004).
- (25) C. D. Snow, L. Qiu, D. Du, F. Gai, S. J. Hagen, and V. S. Pande, "Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy," *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4077-4082 (2004).
- (26) T. Wang, Y. Xu, D. Du, and F. Gai, "Determining β -sheet stability by Fourier transform infrared difference spectra," *Biopolymers* **75**, 163-172 (2004).
- (27) A. W. Smith, H. S. Chung, Z. Ganim, and A. Tokmakoff, "Residual Native Structure in a Thermally Denatured β -Hairpin," *J. Phys. Chem. B* **109**, 17025-17027 (2005).
- (28) J. P. Wang, J. X. Chen, and R. M. Hochstrasser, "Local structure of β -hairpin isotopomers by FTIR, 2D IR, and ab initio theory," *J. Phys. Chem. B* **110**, 7545-7555 (2006).
- (29) T. E. Schrader, W. J. Schreier, T. Cordes, F. O. Koller, G. Babitzki, R. Denschlag, C. Renner, M. Loweneck, S.-L. Dong, L. Moroder, P. Tavan, and W. Zinth, "Light-triggered β -hairpin folding and unfolding," *Proceedings of the National Academy of Sciences*, 0707322104 (2007).
- (30) C. Krejtschi, R. Huang, T. A. Keiderling, and K. Hauser, "Time-resolved temperature-jump infrared spectroscopy of peptides with well-defined secondary structure: A Trpzip [β]-hairpin variant as an example," *Vib. Spectrosc.* **48**, 1-7 (2008).
- (31) K. Hauser, C. Krejtschi, R. Huang, L. Wu, and T. A. Keiderling, "Site-Specific Relaxation Kinetics of a Tryptophan Zipper Hairpin Peptide Using Temperature-Jump IR Spectroscopy and Isotopic Labeling," *J. Am. Chem. Soc.* **130**, 2984-2992 (2008).
- (32) R. Huang, L. Wu, D. McElheny, P. Bouř, A. Roy, and T. A. Keiderling, "Cross-Strand Coupling and Site-Specific Unfolding Thermodynamics of a Trpzip β -Hairpin Peptide Using ^{13}C Isotopic Labeling and IR Spectroscopy," *The Journal of Physical Chemistry B* **113**, 5661-5674 (2009).

- (33) K. Hauser, O. Ridderbusch, A. Roy, A. Hellerbach, R. Huang, and T. A. Keiderling, "Comparison of Isotopic Substitution Methods for Equilibrium and T-Jump Infrared Studies of β -Hairpin Peptide Conformation," *The Journal of Physical Chemistry B* **114**, 11628-11637 (2010).
- (34) R. Narayanan, L. Pelakh, and S. J. Hagen, "Solvent Friction Changes the Folding Pathway of the Tryptophan Zipper TZ2," *J. Mol. Biol.* **390**, 538-546 (2009).
- (35) J. W. H. Pitera, Imran; Swope, William, "Absence of reptation in the high-temperature folding of the trpzip2 β -hairpin peptide," *J. Chem. Phys.* **124**, 141102 (2006).
- (36) J. P. Ulmschneider, M. B. Ulmschneider, and A. Di Nola, "Monte Carlo vs Molecular Dynamics for All-Atom Polypeptide Folding Simulations," *The Journal of Physical Chemistry B* **110**, 16733-16742 (2006).
- (37) J. Zhang, M. Qin, and W. Wang, "Folding mechanism of β -hairpins studied by replica exchange molecular simulations," *Proteins: Struct., Funct., Bioinf.* **62**, 672-685 (2006).
- (38) C. Chen and Y. Xiao, "Observation of multiple folding pathways of β -hairpin trpzip2 from independent continuous folding trajectories," *Bioinformatics* **24**, 659-665 (2008).
- (39) H. Nymeyer, "Energy Landscape of the Trpzip2 Peptide," *The Journal of Physical Chemistry B* **113**, 8288-8295 (2009).
- (40) J. Kim and T. A. Keiderling, "All-Atom Molecular Dynamics Simulations of β -Hairpins Stabilized by a Tight Turn: Pronounced Heterogeneous Folding Pathways," *The Journal of Physical Chemistry B* **114**, 8494-8504 (2010).
- (41) L. Yang, Q. Shao, and Y. Q. Gao, "Thermodynamics and Folding Pathways of Trpzip2: An Accelerated Molecular Dynamics Simulation Study," *The Journal of Physical Chemistry B* **113**, 803-808 (2009).
- (42) X. Jiang, C. Chen, and Y. Xiao, "Improvements of network approach for analysis of the folding free-energy surface of peptides and proteins," *J. Comput. Chem.* **31**, 2502-2509 (2010).
- (43) W. Han, C.-K. Wan, and Y.-D. Wu, "PACE Force Field for Protein Simulations. 2. Folding Simulations of Peptides," *J. Chem. Theory Comput.* **6**, 3390-3402 (2010).
- (44) C. Zhang and J. Ma, "Enhanced sampling and applications in protein folding in explicit solvent," *The Journal of Chemical Physics* **132**, 244101-244116 (2010).
- (45) Q. Shao, H. Wei, and Y. Q. Gao, "Effects of Turn Stability and Side-Chain Hydrophobicity on the Folding of β -Structures," *J. Mol. Biol.* **402**, 595-609 (2010).
- (46) W. Zhuang, R. Z. Cui, D.-A. Silva, and X. Huang, "Simulating the T-Jump-Triggered Unfolding Dynamics of trpzip2 Peptide and Its Time-Resolved IR and Two-Dimensional IR Signals Using the Markov State Model Approach," *The Journal of Physical Chemistry B* **115**, 5415-5424 (2011).
- (47) N. R. Hayre, R. R. P. Singh, and D. L. Cox, "Evaluating force field accuracy with long-time simulations of a beta-hairpin tryptophan zipper peptide," *The Journal of Chemical Physics* **134**, 035103-035108 (2011).
- (48) R. Harada and A. Kitao, "Multi-Scale Free Energy Landscape calculation method by combination of coarse-grained and all-atom models," *Chem. Phys. Lett.* **503**, 145-152 (2011).
- (49) C. Krejtschi, R. Huang, T. Keiderling, and K. Hauser, "Time-resolved temperature-jump infrared spectroscopy of peptides with well-defined secondary structure: A Trpzip β -hairpin variant as an example," *Vib. Spectrosc.* **48**, 1-7 (2008).
- (50) K. C. Jones, Z. Ganim, and A. Tokmakoff, "Heterodyne-detected dispersed vibrational

- echo spectroscopy," *J. Phys. Chem. A* **113**, 14060-14066 (2009).
- (51) M. K. Yetzbacher, N. Belabas, K. A. Kitney, and D. M. Jonas, "Propagation, beam geometry, and detection distortions of peak shapes in two-dimensional Fourier transform spectra," *J. Chem. Phys.* **126**(2007).
- (52) E. R. Henry, J. Hofrichter, and M. L. J. Ludwig Brand, "[8] Singular value decomposition: Application to analysis of experimental data," in *Methods in Enzymology* (Academic Press, 1992), pp. 129-192.
- (53) Z. Ganim, H. S. Chung, A. W. Smith, L. P. DeFlores, K. C. Jones, and A. Tokmakoff, "Amide I Two-Dimensional Infrared Spectroscopy of Proteins," *Acc. Chem. Res.* **41**, 432-441 (2008).
- (54) J. Kubelka, E. R. Henry, T. Cellmer, J. Hofrichter, and W. A. Eaton, "Chemical, physical, and theoretical kinetics of an ultrafast folding protein," *Proceedings of the National Academy of Sciences* **105**, 18655-18662 (2008).
- (55) T. Cellmer, M. Buscaglia, E. R. Henry, J. Hofrichter, and W. A. Eaton, "Making connections between ultrafast protein folding kinetics and molecular dynamics simulations," *Proceedings of the National Academy of Sciences* **108**, 6103-6108.
- (56) K. E. Amunson and J. Kubelka, "On the Temperature Dependence of Amide I Frequencies of Peptides in Solution," *J. Phys. Chem. B* **111**, 9993-9998 (2007).
- (57) J. Kaminsky, P. Bour, and J. Kubelka, "Simulations of the Temperature Dependence of the Amide I Vibration," *J. Phys. Chem. A* **115**, 30-34 (2011).
- (58) L. Yang, Q. Shao, and Y. Q. Gao, "Thermodynamics and Folding Pathways of Trpzip2: An Accelerated Molecular Dynamics Simulation Study," *The Journal of Physical Chemistry B* **113**, 803-808 (2008).
- (59) V. Munoz, R. Ghirlando, F. J. Blanco, G. S. Jas, J. Hofrichter, and W. A. Eaton, "Folding and Aggregation Kinetics of a β -Hairpin," *Biochemistry* **45**, 7023-7035 (2006).

Appendix 8.A: Transient 2D IR spectra

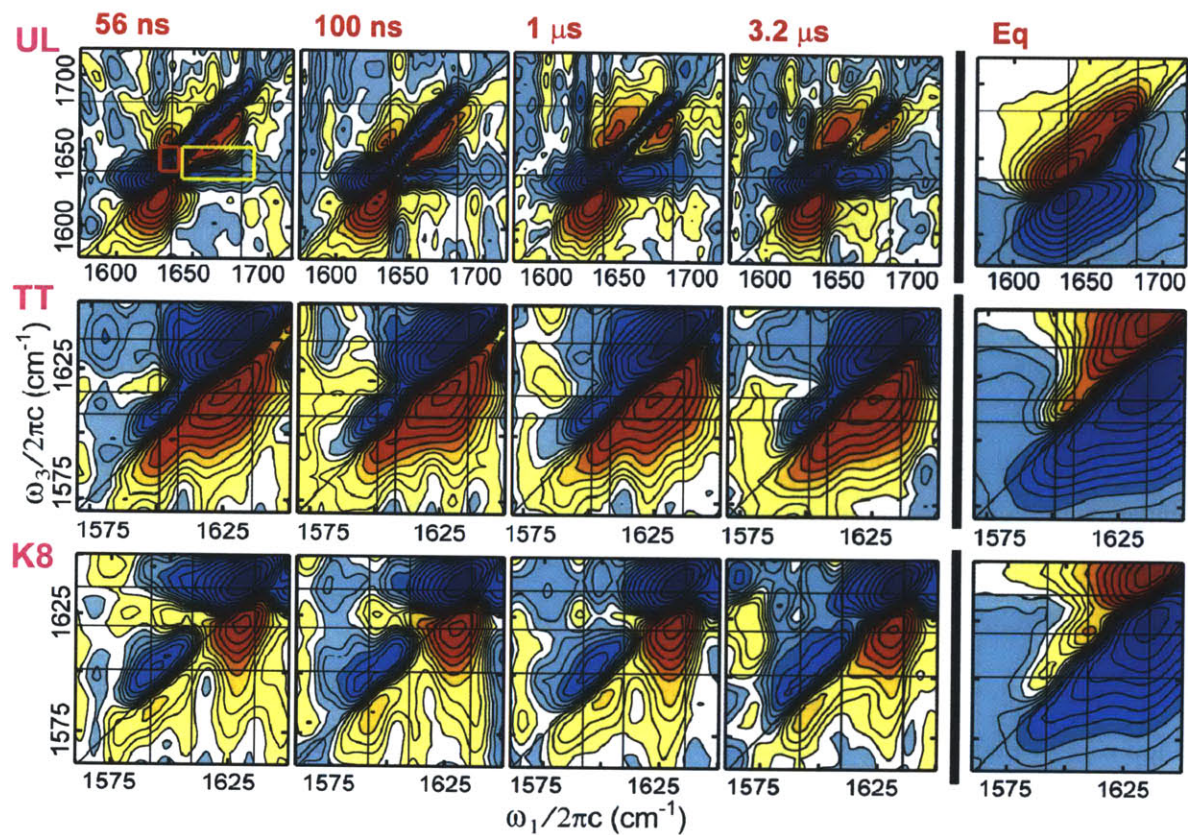


Figure 8.A.1: The t-2D IR difference data for various delays and TZ2 isotopologues. Yellow and orange boxes indicate the integrated and fit regions described in Table 8.1.

Appendix 8.B: SVD Analysis and Hypothesized Spectra

SVD provides an alternative means of representing the data. The overall process, including transformation into a physically-meaningful, model-based basis set, is diagrammed in Figure 8.B.1.

First, to reduce the full data set S to the difference data set, we subtract out the equilibrium spectrum, S_{eq} (with a constant τ -dependence of $P_{eq}(\tau)$ such that ΔS is zero at long τ , after re-equilibration):

$$\Delta S(\omega, \tau) = S(\omega, \tau) - S_{eq}(\omega)P_{eq}(\tau) \quad (8.B.1)$$

Next, as described in eq. (8.1) (step 1 in Figure 8.B.1) the data set ΔS is split through singular value decomposition into component spectra, singular value scaling factors, and component τ -dependencies:

$$\Delta S(\omega, \tau) = \mathbf{u}_1(\omega)\mathbf{s}_1\mathbf{v}_1(\tau) + \mathbf{u}_2(\omega)\mathbf{s}_2\mathbf{v}_2(\tau) + \mathbf{u}_3(\omega)\mathbf{s}_3\mathbf{v}_3(\tau) + \dots \quad (8.B.2)$$

The results from applying the SVD analysis to the UL t-DPP *difference* data are shown in the top row of Figure 8.B.2, where spectra and timetraces are ordered in descending order and labeled based on the magnitude of their associated \mathbf{S} values, which are shown for the SVD

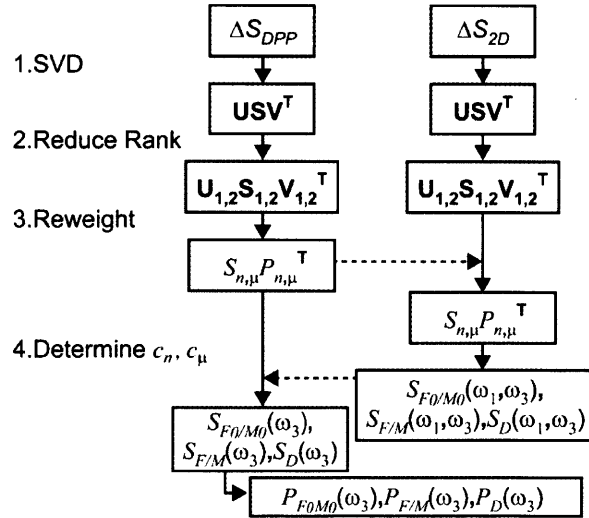


Figure 8.B.1: Flow chart depicting the SVD process and transformation. Horizontal dashed arrows indicate information feedback between the DPP and 2D IR data.

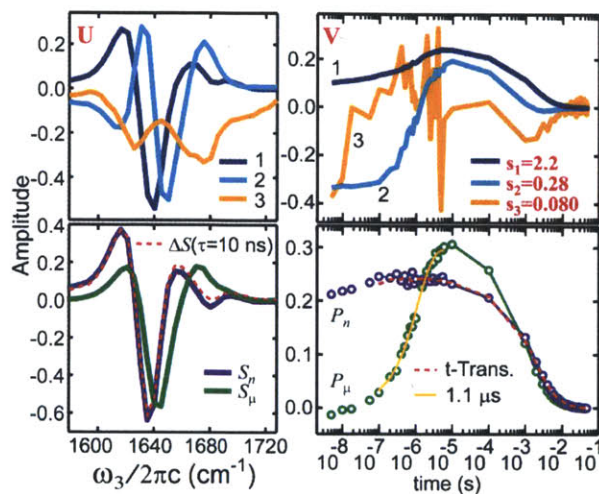


Figure 8.B.2: *Top row*– The first three UL TZ2 t-DPP \mathbf{U} , \mathbf{S} , \mathbf{V} components that result from SVD of the difference data, ΔS . *Bottom row*– The aggregate state spectra (S_n, S_μ) and τ -traces (P_n, P_μ) that result from specifying that $S_n \propto \Delta S(\tau=10 \text{ ns})$ (shown in dashed red on left) and $P_n \propto \Delta T$ (shown in dashed red on right).

decomposition of the UL, TT, and K8 t-DPP data in Figure 8.4. Ordering the spectral components by their magnitudes reveals a rapid decrease in scaling factor - for the TZ2 UL t-DPP spectra $s_1: s_2: s_3$ is 22.5:1.6:0.28 (see Figure 8.B.3). Based on the noisiness of \mathbf{V}_3 and the small magnitude of \mathbf{S}_3 relative to \mathbf{S}_1 and \mathbf{S}_2 , we reject all but the 1st and 2nd component (Figure 8.B.1, **step 2**):

$$\Delta S(\omega, \tau) \approx \mathbf{u}_1(\omega) \mathbf{s}_1 \mathbf{v}_1(\tau) + \mathbf{u}_2(\omega) \mathbf{s}_2 \mathbf{v}_2(\tau) \quad (8.B.3)$$

This decomposition into two components is unsurprising given the observation of just two distinct timescales in the data ($<5 \text{ ns}$, $\sim 1.2 \mu\text{s}$).

The spectral components from the SVD analysis are mathematical results. In order to redefine the two SVD basis spectra into physically meaningful spectra that correspond to collections of states, we separate the data based on the observed timescales. We group the conformational populations into two aggregate spectra: a spectrum that corresponds to the changes that occur in response to the T-jump in $<10 \text{ ns}$ (S_n) and a spectrum due to the collection of populations that do not respond until the μs timescale (S_μ), although merely specifying S_n does not lead to a clear S_μ . Generally, the S_n corresponds to the sum of difference spectra due to the fast, intra-well conversion, while S_μ corresponds to the sum of difference spectra due to the slower, inter-well conversion. Based on these general assignments, we also assume that the time dependence of the S_n spectrum will have a τ -dependence that is proportional to the change in

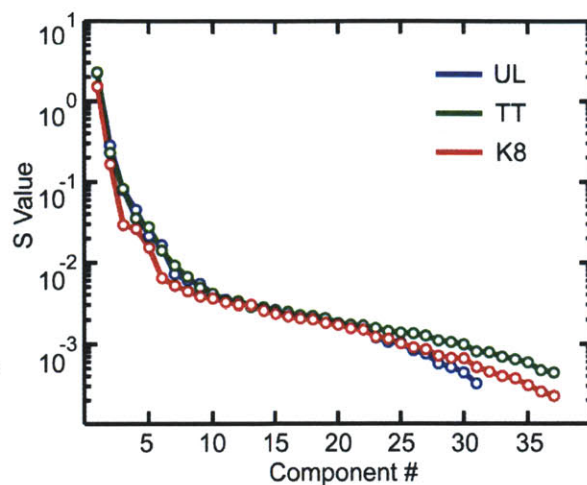


Figure 8.B.3: The singular values, S , that result from SVD applied to the three (UL, TT, K8) difference t-DPP data sets. Note that the vertical axis is logarithmic.

temperature, ΔT – the difference between the intra-well populations responds quickly enough that it is approximately proportional to the temperature change, although any inter-well conversion will cause a deviation from this strict proportionality. Based on the discussion above, we can describe the difference data using the S_n and S_μ spectra and their associated τ -dependencies:

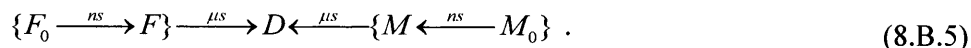
$$\begin{aligned}
 \Delta S(\omega, \tau) &\approx S_n(\omega)P_n(\tau) + S_\mu(\omega)P_\mu(\tau) \\
 &\longrightarrow S_n(\omega) \propto \Delta S(\omega, \tau = 10 \text{ ns}); \quad P_n(\tau) \propto \Delta T(\tau) \\
 &\approx c\Delta S(\omega, \tau = 10 \text{ ns})\Delta T(\tau) + S_\mu(\omega)P_\mu(\tau)
 \end{aligned} \tag{8.B.4}$$

where the second line indicates insertion of our assumptions about the nature of the S_n spectrum and its τ -dependence. The relationship between the n and μ components and the actual underlying populations is unknown, and establishing the connections requires proposing a more specific model than the description given above.

In the general case, a model can be introduced where the minimum number of states is given by the number of SVD components. Additional states may be added based on the presence of more complicated multi-exponentials or multiple peaks that can be definitively assigned, as we observe in isotope-labeled TZ2.

For our specific TZ2 system, the presence of three SVD components (two for the difference data) suggests that we have at least three states. We find that three states are inadequate to describe the data. Instead of three, we propose five states as the minimum number

of states based on the presence and response from the K8-1 and K8-2 peaks. We hypothesize the net movement of population following the T-jump to follow:



For this scheme, we would observe 3 distinguishable components if the observed timescales are the same within the noise.

Based on the proposed scheme in eq. (8.B.5), the difference spectra should be the sum of the product of the τ dependent change in population of each species i , $\Delta P_i(\tau)$, multiplied by the spectrum of each species, S_i :

$$\Delta S(\omega, \tau) = \sum_{i=F_0, F, M_0, M, D} S_i(\omega) \Delta P_i(\tau) \quad (8.B.6)$$

Based on the SVD analysis, which decomposes ΔS into just two components, multiple spectra can be lumped together because they share the same τ -dependence – we must reconcile (8.B.5), which has 5 species, with eq. (8.B.3), which has a minimum of three (two timescales translates into three states). Based on our assumptions, we first propose that states F and M and F_0 and M_0 will have proportional time-dependencies – F and M are both gained on the ns and lost on the μs timescales while F_0 and M_0 are both lost on the ns and the μs timescales, which leads us to assume that their spectra can also be lumped together into S_{F_0/M_0} and $S_{F/M}$:

$$\begin{aligned} \Delta P_{F/M}(\tau) &= \frac{1}{c_1} \Delta P_F(\tau) = \frac{1}{c_2} \Delta P_M(\tau) = \Delta P_F(\tau) + \Delta P_M(\tau) \\ \Delta P_{F_0/M_0}(\tau) &= \frac{1}{c_3} \Delta P_{F_0}(\tau) = \frac{1}{c_4} \Delta P_{M_0}(\tau) = \Delta P_{F_0}(\tau) + \Delta P_{M_0}(\tau) \\ S_{F/M}(\omega) &= c_1 S_F(\omega) + c_2 S_M(\omega) \\ S_{F_0/M_0}(\omega) &= c_3 S_{F_0}(\omega) + c_4 S_{M_0}(\omega) \end{aligned} \quad (8.B.7)$$

with $c_1+c_2 = c_3+c_4 = 1$. Based on our assumptions from eq. (8.B.4) and using the lumped states from eq. (8.B.7), we can propose the identity of the n spectrum and its population trace. For consistency with the top line of eq. (8.B.4), we can solve for the nature of the μ spectrum and population trace if the assumptions for n hold:

$$\begin{aligned} S_n(\omega) &= \frac{1}{2} [S_{F/M}(\omega) - S_{F_0/M_0}(\omega)] \propto \Delta S(\omega, \tau = 10 \text{ ns}); \\ P_n(\tau) &= \Delta P_{F/M}(\tau) - \Delta P_{F_0/M_0}(\tau) \propto \Delta T(\tau) \\ S_\mu(\omega) &= S_D(\omega) - \frac{1}{2} [S_{F/M}(\omega) + S_{F_0/M_0}(\omega)]; \quad P_\mu(\tau) = \Delta P_D(\tau) \end{aligned} \quad (8.B.8)$$

Substitution of eq. (8.B.8) into the top line of eq. (8.B.4) results in eq. (8.B.6) (with conserved population $0 = \Delta P_{F_0} + \Delta P_F + \Delta P_{M_0} + \Delta P_M + \Delta P_D$), proving that if our definitions of P_n and S_n are correct, then S_μ and P_μ must necessarily follow.

Based on eqs. (8.B.7) and (8.B.8), S_n is the difference spectrum between intra-well states (F-F₀)+(M-M₀) weighted with c_i based on the ratio of their change in populations. S_μ is then the difference spectrum due to inter-well conversion between D-(F+F₀+M+M₀), where the spectra are weighted with the same c_i as in the S_n case. The same ratios are expected at constant temperatures up to 100 μs following the T-jump – the intra-well conversion is fast and the ratio depends on the temperature. The P_n population trace is the difference in the state populations (F-F₀)+(M-M₀). The P_μ population trace follows the inter-well conversion $D = -F-F_0-M-M_0$.

The model proposed in eq. (8.B.8) is a special case in which the difference spectral changes attributed to 5 species can be described by two components (three in S). Although the coincidence of many timescales and amplitude changes seems surprising, we rationalize them within our signal to noise as follows. We expect the pulsewidth-limited <ns peptide/solvent response to occur with the same timescale and a similar amplitude for the $F_0 \rightarrow F$ and $M_0 \rightarrow M$ transitions. For the $F_0 + F \rightarrow D$ and $M_0 + M \rightarrow D$ transitions, we also expect similar timescales within the noise – we expect both inter-well conformational changes to involve crossing barrier created by possibly similar enthalpic contacts: cross-strand H-bonds and tryptophan contacts. The similarity of timescales may also be because the timescale we observe is a convolution of forward and reverse rates; even though the forward and reverse rates differ, the observed rates may still coincide. Based on these arguments, we think our proposed scheme is reasonable, especially because the proposed scheme is consistent with our spectral interpretations.

In order to calculate S_n , S_μ , P_n and P_μ , we take a linear combination of the first two SVD components weighted such that one of the new components, S_n , best fits the initial spectrum at 10 ns, and its associated P_n is proportional to the change in temperature. This is done by finding the c_a and c_b coefficients that define the linear combinations such that:

$$\begin{aligned} [\mathbf{u}_1 \mathbf{s}_1, \mathbf{u}_2 \mathbf{s}_2] \mathbf{W}_a &= [\Delta S(\tau = 10 \text{ ns}) \propto S_n, \mathbf{u}_2 \mathbf{s}_2] \\ \mathbf{W}_b^{-1} \mathbf{W}_a^{-1} [\mathbf{v}_1, \mathbf{v}_2]^T &= [\Delta T \propto P_n, P_\mu]^T \end{aligned} \quad (8.B.9)$$

with

$$W_a = \begin{pmatrix} 1 & 0 \\ c_a & 1 \end{pmatrix}, \quad W_b = \begin{pmatrix} 1 & c_b \\ 0 & 1 \end{pmatrix}. \quad (8.B.10)$$

Inserting the $W_a W_b W_b^{-1} W_a^{-1}$ identity from (8.B.10) into eq. (8.1) from the main text gives

$$\begin{aligned} \Delta S &\approx \mathbf{U}_{1,2} \mathbf{S}_{1,2} (W_a W_b W_b^{-1} W_a^{-1}) \mathbf{V}_{1,2}^T = (\mathbf{U}_{1,2} \mathbf{S} W_a W_b) (W_b^{-1} W_a^{-1} \mathbf{V}_{1,2}^T) \\ &= (\mathbf{U}_{1,2} \mathbf{S} W_a W_b) (\mathbf{V}_{1,2} (W_a^{-1} W_b^{-1})^T)^T = [S_n, S_\mu] [P_n, P_\mu]^T \end{aligned} \quad (8.B.11)$$

This process is **step 3** from Figure 8.B.1. Without using SVD analysis, we would obtain similar results choosing our own basis spectra with $S_n = \Delta S_{DPP}(\tau = 10 \text{ ns})$ and defining S_μ as the μs spectrum that remains after removing the S_n from the difference data by scaling it proportionally to the change in temperature and subtracting it from ΔS_{DPP} . SVD is a more sophisticated method of decomposition, however, because it allows us to reject noisy components.

The resulting S_n and S_μ spectra calculated from the UL t-DPP difference data using eq. (8.B.10) are shown in the bottom row of Figure 8.B.2. Based on eq. (8.B.8) S_n and S_μ are actually difference spectra. As was assumed in the proposed relations in eq. (8.B.9), the microsecond component, S_μ , starts with no population. S_μ then grows in with a 1.1 μs exponential time constant, and decays back to equilibrium with the temperature relaxation. Component S_n gains population on the $<10 \text{ ns}$ timescale, and follows the temperature as it drops back to equilibrium. As expected, the S_μ spectrum contains the peak at 1660 cm^{-1} , representing the creation of random coil, disordered structures.

To connect the better time resolution of t-HDVE with the increased spectral content of t-2D IR, the SVD analysis was also applied to the UL t-2D IR data. The t-2D IR data is noisier than the t-HDVE data, and the \mathbf{s} values are not as well separated as in the t-DPP SVD case, but the data still appears to decompose into two meaningful component spectra, \mathbf{U}_1 and \mathbf{U}_2 . As before, these spectra components are mixed spectra of the five different species. To separate S_n and S_μ spectral components, we apply the remixing shown in eq. (8.B.11) where c_a and c_b are chosen such that the t-2D IR P_n and P_μ and the ω_3 -projections of S_n and S_μ best match those from the t-DPP case. The redefinition results in the S_n and S_μ spectra shown in Figure 8.6. S_n is similar to the 56 ns t-2D IR spectrum (Figure 8.A.1), as we expect based on the assumptions that

are invoked in generating S_n and S_μ . S_μ contains the spectral features that we identified growing in on the μs timescale, in particular the intensity in the 1660 cm^{-1} region.

Hypothesizing Spectra

Although the SVD analysis and manipulation thus far has been used to identify the timescales and difference t-2D IR spectra (Figure 8.6) that appear in response to the T-jump, the S_n and S_μ spectra are still difference spectra between the equilibrium populations that are lost and the S_n and S_μ species that grow in. Isolating the actual spectrum of lumped species S_{F_0/M_0} and $S_{F/M}$ and species S_D requires further assumptions.

Inspection of eq. (8.B.8) reveals that we hypothesize that the spectrum S_n and S_μ are the sum and difference spectra, respectively of S_{F_0/M_0} and $S_{F/M}$. In order to isolate S_{F_0/M_0} and $S_{F/M}$, we assume that the equilibrium spectrum is proportional to S_{F_0/M_0} – at equilibrium we only have states M_0 and F_0 (which is not correct based on measured melting curves^{1, 4}) in the ratio that is approximately maintained following the T-jump:

$$S'_{F_0/M_0}(\omega) = S_{eq}(\omega). \quad (8.B.12)$$

where the prime indicates that S'_{F_0/M_0} is not properly normalized. With this piece of assumed information, we can vary the coefficient c_n such that the sum of S_n and S'_{F_0/M_0} gives a hypothesized $S_{F/M}$ spectrum:

$$\begin{aligned} S'_{F/M}(\omega) &= S_n(\omega) + c_n \times S'_{F_0/M_0}(\omega) \\ &= S_n(\omega) + c_n \times S_{eq}(\omega) \end{aligned} \quad (8.B.13)$$

With these two pieces of information, we can vary the coefficient c_μ such that the sum of S_μ , S'_{F_0/M_0} , and $S'_{F/M}$ gives a hypothesized S_D spectrum:

$$\begin{aligned} S'_D(\omega) &= S_\mu(\omega) + c_\mu \times \left[c_n \times S'_{F_0/M_0}(\omega) + S'_{F/M}(\omega) \right] \\ &= S_\mu(\omega) + c_\mu \times \left[c_n \times S_{eq}(\omega) + S_n(\omega) + c_n \times S_{eq}(\omega) \right]. \\ &= c_\mu \times S_n(\omega) + S_\mu(\omega) + 2c_\mu \times c_n \times S_{eq}(\omega) \end{aligned} \quad (8.B.14)$$

We are unaware of a definitive way of objectively quantifying c_i . Rather, we determine c_i by subjectively assessing the continuity and reasonableness of the hypothesized results. Although the arbitrariness of this procedure places high error bars on the results, general features can be observed. One can immediately reject extremely high or low values of c_i and reject values that give discontinuous or overly sharp features.

Because eqs. (8.B.12), (8.B.13), and (8.B.14) represent linear combinations of S_n, S_μ , and S_{eq} , we can write the transformation as the product of matrices:

$$T = \begin{pmatrix} 0 & 1 & c_\mu \\ 0 & 0 & 1 \\ 1 & c_n & 2c_\mu c_n \end{pmatrix}; \quad (8.B.15)$$

$$[S'_{F_0/M_0}, S'_{F/M}, S'_D] = [S_n, S_\mu, S_{eq}] \times T$$

Thus far, the predicted species' spectra, S'_i , are not normalized appropriately. Ideally, each S'_i will be in an intensity-per-molar unit such that the concentration dependence can be calculated in molar terms. We settle for a crude normalization of each spectrum by defining ω_{\max} , the frequency at which the equilibrium spectrum, $|S'_{F_0/M_0}{}^{2D IR}|$, is maximized. We then assume that the dipole is conserved at this ω_{\max} , which allows for normalization by multiplying S' by L :

$$L = \begin{pmatrix} 1/S'_{F_0/M_0}(\omega_{\max}) & 0 & 0 \\ 0 & 1/S'_{F/M}(\omega_{\max}) & 0 \\ 0 & 0 & 1/S'_D(\omega_{\max}) \end{pmatrix}; \quad (8.B.16)$$

$$S_{F_0/M_0, F/M, D} = S'_{F_0/M_0, F/M, D} \times L = S_{n, \mu, eq} \times T \times L$$

At this point, we have determined the hypothesized 2D IR spectra. In order to calculate the population dependence, P , we return to the DPP data, which has spectra collected at more τ values. Because we have defined our initial populations ($P_{F0}(\tau = 0) + P_{M0}(0) = 1$, $P_F(0) + P_M(0) = P_D(0) = 0$) we can convert the change in populations ΔP_i into absolute populations. We can calculate the S_i DPP counterparts (horizontal dashed line after **step 4** in Figure 8.B.1) by using the same eqs. (8.B.15) and (8.B.16) but fitting the resulting S_i^{DPP} to the projections of the S_i^{2D} .

In order to determine the population traces of the calculated hypothesized species spectra, we apply the same identity transformation analogous to eq. (8.B.11):

$$S = S_{n, \mu, eq} P_{n, \mu, eq}^T = S_{n, \mu, eq} T L L^{-1} T^{-1} P_{n, \mu, eq}^T = (S_{n, \mu, eq} T L) (L^{-1} T^{-1} P_{n, \mu, eq}^T)$$

$$= S_{F_0/M_0, F/M, D} P_{F_0/M_0, F/M, D}^T \quad (8.B.17)$$

The t-2D IR S_{F_0/M_0} , $S_{F/M}$, and S_D hypothesized spectra for the UL case are shown in Figure 8.B.4 along with the corresponding t-DPP determined P_{F_0/M_0} , $P_{F/M}$, and P_D .

Here, as described in eq. (8.B.7), the S_{F_0/M_0} and $S_{F/M}$ spectra correspond to scaled sums of

the subscript species' spectra. P_{F_0/M_0} and $P_{F/M}$ correspond to the sum of the populations of the subscripted species.

The spectra and population traces hypothesized represent a possible set of states. Summing the individual spectra scaled by their appropriate τ -dependent spectra results in the original data.

Because of all of the above assumptions, the spectra that result are hypothesized spectra that are intended to highlight the changes observed. Although the results are consistent with our proposed model, we cannot definitively prove them. In Figure 8.B.5, we show the hypothesized spectra that result from varying c_i . The hypothesized spectra become unphysical at small values of c_i and at large values they look too much like the equilibrium spectra, which suggests that the intermediate values chosen are reasonable.

Based on the calculated time traces shown in Figure 8.B.4, it appears as if the T-jump causes a 20-40% loss of the equilibrium F_0+M_0 state to create $\sim 15\%$ of $F+M$ and D each. This estimate of percentage change is highly dependent on which equilibrium frequency is used for the normalization.

The same SVD analysis procedure can be applied to the TT t-DPP and K8 t-DPP difference spectra. As with the UL case, the \mathbf{S} values and \mathbf{V}_3 time traces indicate that each data set can be well described by two components. The same redefinition of basis spectra based on separating the ns and μs spectral responses can be applied for TT t-DPP and K8 t-DPP. These t-DPP basis spectra can then be used to calculate the t-2D IR S_n and S_μ shown in Figure 8.6, both of which have the same discussed spectral features observed in the 56 ns and 3.2 μs t-2D IR spectra from Figure 8.A.1. Hypothesizing the lumped species spectra for the TT and K8 case produces the hypothesized spectra of $S_{F/M}$ and S_D shown in Figure 8.B.4. Time traces indicate the same behavior and similar percentage population shifts as shown in the UL case.

Discussion of Hypothesized spectra

Although the accuracy and integrity of the hypothesized spectra depends on making far reaching assumptions about the nature of the equilibrium population, the hypothesized spectra are consistent with the assignments made in the main text, and they provide further support for our interpretation. They are intended to exaggerate and highlight the features already identified in analysis of the raw spectra.

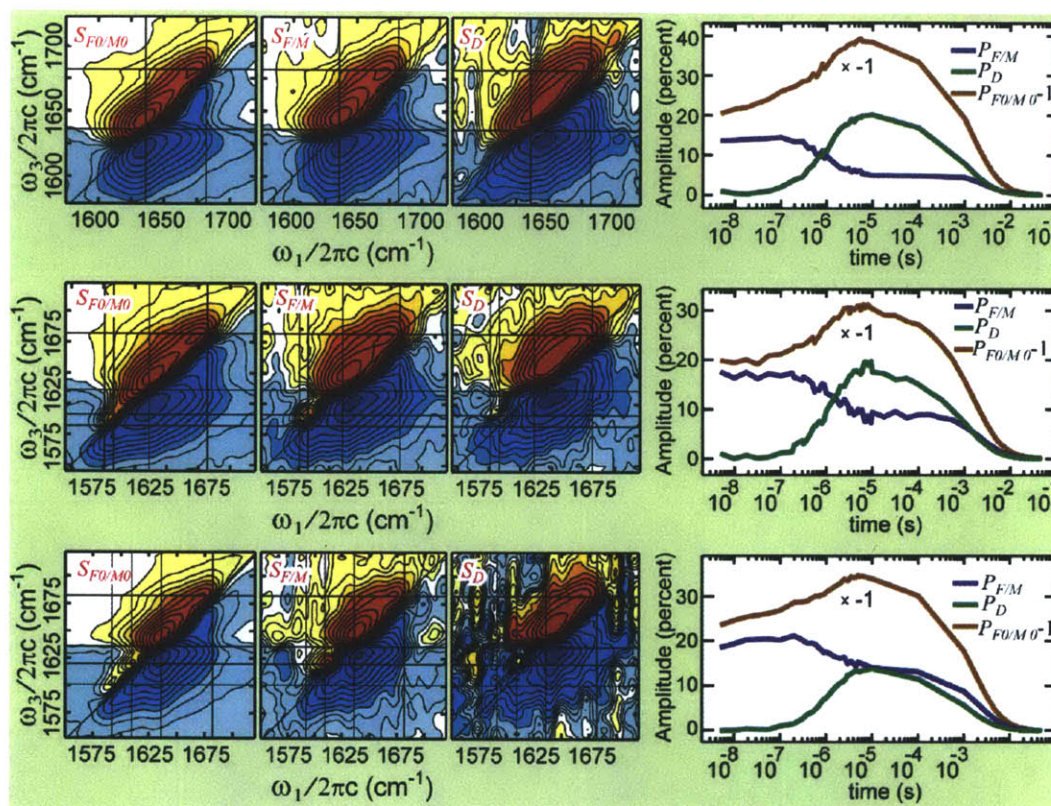


Figure 8.B.4: The hypothesized 2D IR $S_{F0/M0}$, $S_{F/M}$, and S_D spectra and $P_{F0/M0}$, $P_{F/M}$, and P_D population traces. From top to bottom, UL, TT, and K8. The contour lines are scaled as described in Figure 8.3.

The UL S_n spectrum looks similar to a homogeneously broadened β -hairpin, and the UL S_μ spectrum looks similar to a random coil spectrum, which is devoid of the dual-peak structure characteristic of an AP β -hairpin. The TT S_n spectrum contains a broadened, blue-shifted TT peak consistent with retention of mid-strand native H-bonds. The K8 S_n spectrum contains a similarly broadened and blue-shifted K8-2 peak, which is also consistent with a retained β -turn.

The lack of TT contacts in the disordered species spectrum is shown in the hypothesized TT S_μ spectrum, which contains negligible relative intensity in the isotope region. The weak peak that is present in the isotope region is further blue-shifted relative to the hypothesized S_n spectrum TT peak, and it may be due to disordered TT oscillators that are H-bonded to solvent water molecules.

Although it is noisier than the other spectra, the K8 S_μ spectrum shows negligible population in the K8 isotope spectral region, suggesting that the disordered state has a solvent

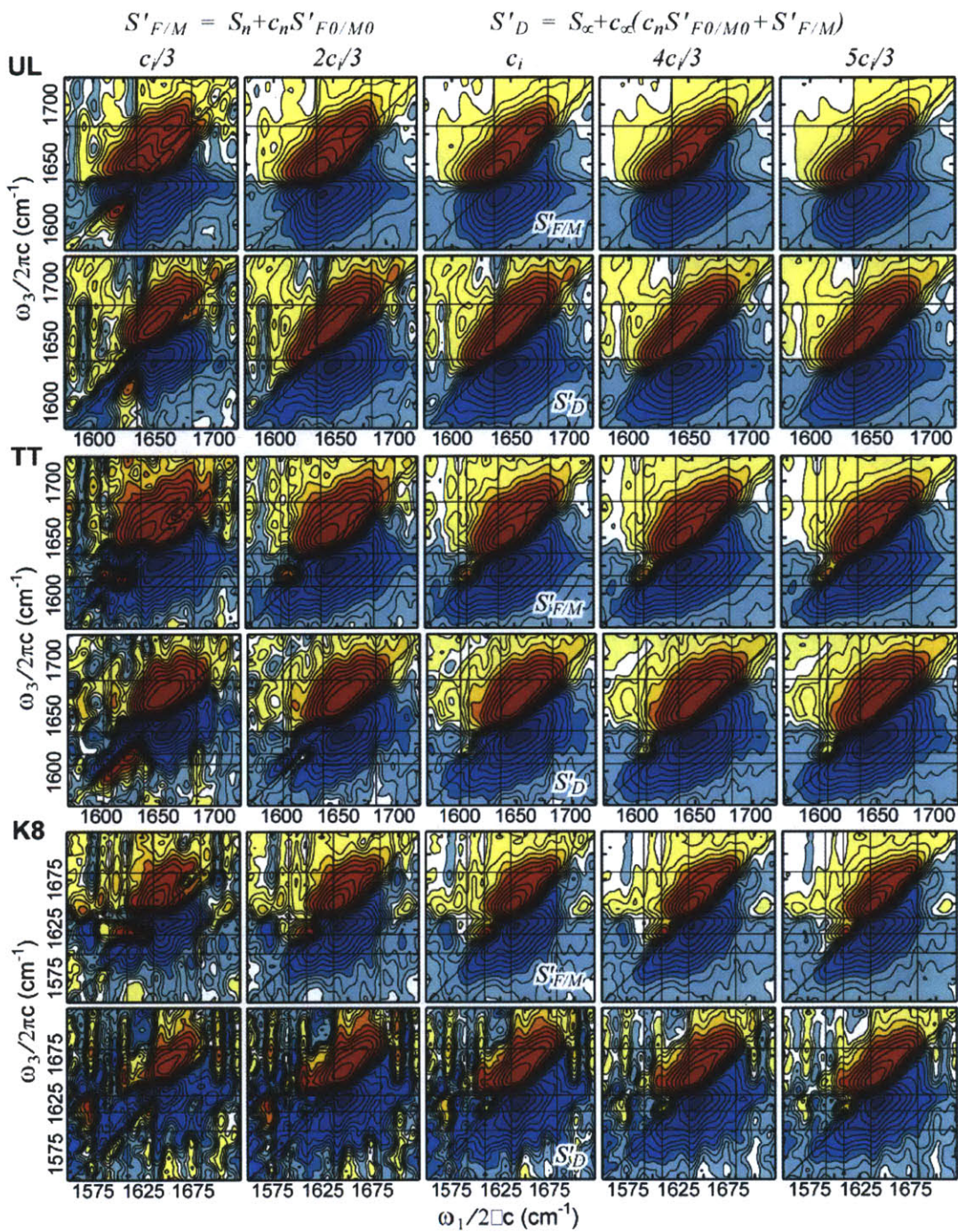


Figure 8.B.5: The effect of varying c_n and c_μ are shown in the different columns. Each row corresponds to a hypothesized S'_{F^*M} (variable c_n) or S'_D (variable c_μ) spectrum. At small or large c_n values, the S'_{F^*M} spectra tend to be discontinuous in the low frequency region or approach the equilibrium spectrum, respectively. At small or large c_μ values, the S'_D spectra tend to be discontinuous in the low frequency region or approach the S'_{F^*M} spectrum, respectively. The contour lines are scaled as described in Figure 8.3.

exposed or buried carbonyl whose peak is weak and probably inhomogeneously broadened due to the variety of electrostatic environments it can sample.

Appendix 8.C: Transient Absorption

The t-abs. of TZ2 *and* the solvent is collected along with the t-DPP spectra. In order to properly measure the t-abs. of just the peptide, the t-abs. of the solvent should be measured separately and subtracted. This essential component, the solvent t-abs., was not collected independently. Instead, we assume that SVD analysis can separate the different responses.

Although the t-abs. is noisier than the t-DPP data, SVD analysis again reveals that two components can approximately recreate the difference data. The SVD components that result from the decomposition of the UL, TT, and K8 data are shown on the left in Figure 8.C.1. Based on the continuity of the \mathbf{V} component with the 3rd largest s value and the noisiness of the \mathbf{V} trace with the 2nd largest s value, we chose to reject the 2nd and keep the 3rd components for the UL case.

We assume that the two components can again be separated as described in eqs. (8.B.4) and (8.B.8) with the addition of the solvent response. As with S_n , the solvent response should have a τ -dependence proportional to the temperature relaxation, which can be measured at frequencies above the amide I band, at $\omega=1728 \text{ cm}^{-1}$:

$$\begin{aligned}
 \Delta S^{t-abs}(\omega, \tau) &\approx S_n(\omega)P_n(\tau) + S_\mu(\omega)P_\mu(\tau) + S_{sol}(\omega)P_{sol}(\tau) \\
 &\longrightarrow S_{n/sol}(\omega) = c_5 S_n(\omega) + c_6 S_{sol}(\omega) \propto \Delta S(\omega, \tau = 10 \text{ ns}) \\
 &\longrightarrow P_{n/sol}(\tau) \propto P_n(\tau) \propto P_{sol}(\tau) \propto \Delta T(\tau) \\
 &\approx S_{n/sol}(\omega)P_{n/sol}(\tau) + S_\mu(\omega)P_\mu(\tau) \\
 &\approx c\Delta S(\omega, \tau = 10 \text{ ns})\Delta T(\tau) + S_\mu(\omega)P_\mu(\tau)
 \end{aligned} \tag{8.C.1}$$

In order to find the W_a and W_b (similar to eq. (8.B.9), but subscript n/sol instead of n) necessary to transform into these lumped spectra, we rejected the 2nd and ≥ 4 components for the UL case and the ≥ 3 components for the K8 and TT cases. For the TT case, we think that noise is contained in both the 2nd and 3rd components, and we used a linear combination of all three components – W_a and W_b are 3x3 matrices that differ from the identity matrix only by replacing the zeros in the first or second columns, respectively, with two adjustable parameters.

The $S_{n/sol}$, S_μ and $P_{n/sol}$, P_μ are shown on the center and right of Figure 8.C.1. As specified, the $P_{n/sol}$ tracks the temperature. Consistent with the t-DPP and t-2D IR traces, the P_μ starts close

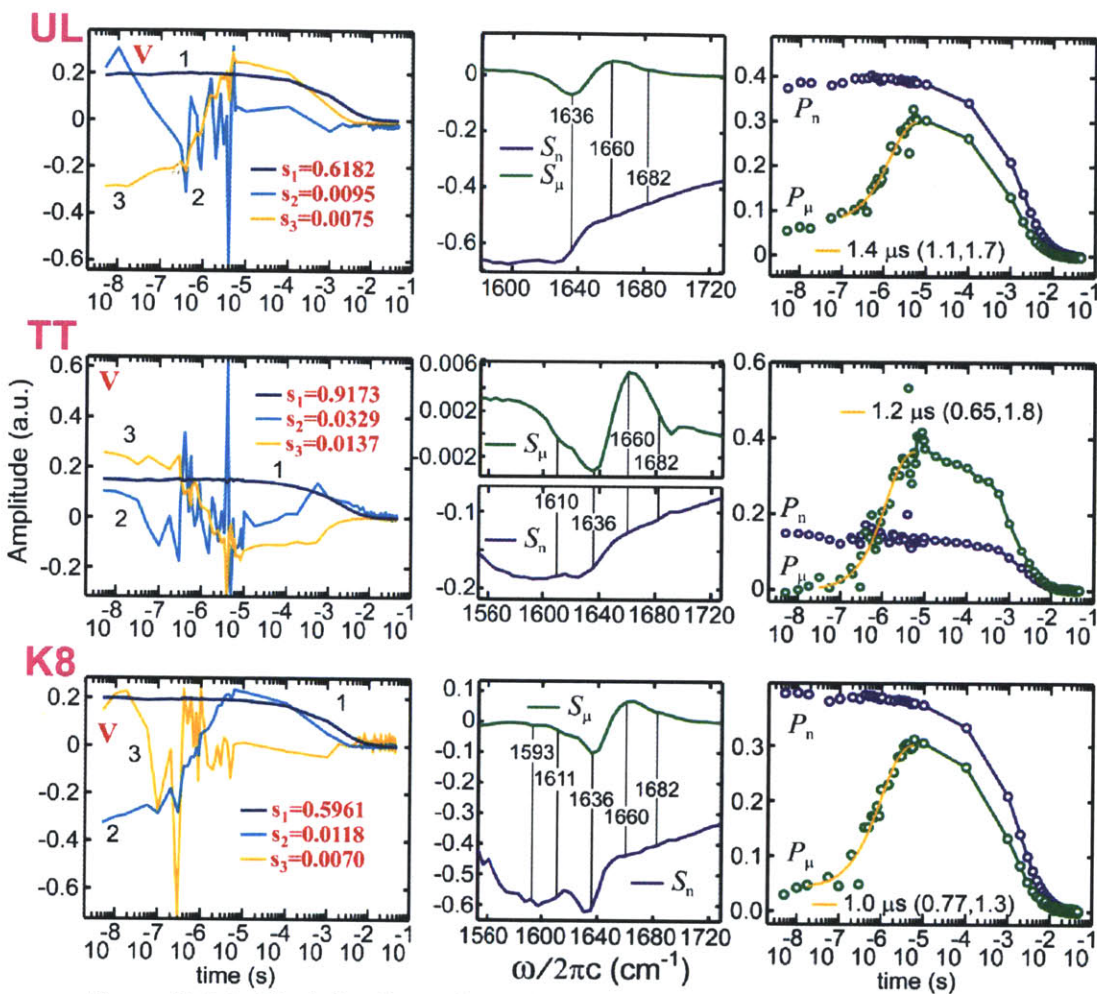


Figure 8.C.1: The left column shows the first three V and S components that result from SVD of the UL, TT, and K8 t-abs. difference data. For the UL case, the 2nd component was rejected for the S_i determination, while the 3rd component was kept because of the noisiness of the 2nd and continuity of the 3rd V component. The kept SVD components were transformed to give the aggregate S_n, S_μ spectra and P_n, P_μ population traces shown in the center and right columns.

to zero and grows in with a $\sim 1.2 \mu\text{s}$ exponential time constant. Because of the large amplitude of the solvent response, the $S_{n/sol}$ is dominated by the S_{sol} , and the S_n TZ2 features are hard to separate. As with the t-DPP and t-2D IR cases, the S_μ spectrum is expected to represent the $>1 \mu\text{s}$ inter-well conversion. Consistent with the nonlinear data, the t-abs. UL S_μ spectrum shows loss of the β -strands at 1636 and 1682 cm^{-1} and gain of disordered structures at 1660 cm^{-1} . The K8 and TT show similar main amide I loss and gain features, although the TT spectrum is noisier, and the 1682 cm^{-1} loss is not apparent. The TT, K8-1, and K8-2 isotope-label features are also lost on the $\sim 1.2 \mu\text{s}$ timescale consistent with what is observed in the transient nonlinear

spectroscopy.

Interestingly, the SVD analysis has already been applied to the temperature dependent FTIR spectra¹. Using our interpretation, we expect that the SVD analysis of the equilibrium can also separate the data into two components, one that tracks the temperature ($\sim S_n$) and one that represents the inter-well conformational changes ($\sim S_\mu$). The previously calculated SVD analysis only shows the mathematical results (\mathbf{U}_2 and \mathbf{V}_2)¹, which we expect to be a mixture of the n and μ components. As expected, the previous \mathbf{U}_2 spectra share many features with the S_μ spectra calculated here.

Chapter 9

T-Jump Studies of Insulin

9.1 Introduction

Molecular associations involving proteins are at the heart of biological function. Proteins recognize and specifically bind ligands, assemble into complexes with defined structures, and associate to form viral capsids, protein aggregates, amyloid, and fibers. Each of these processes involves questions regarding the mechanism of specific protein association. How do two molecules in solution diffuse into proximity, recognize one another, and dock to one another with specific intermolecular contacts? Despite significant theoretical and experimental efforts, predictive models that provide a molecular basis for understanding the mechanism of recognition and binding are still lacking.

Multiple paradigmatic mechanisms have been proposed to describe association processes, which differ in the resolvable states and the ordering of protein conformational change and encounter, as is shown in Figure 9.1. The oldest and simplest mechanism is the *lock-and-key*, which does not involve conformational change: the encounter of a rigid enzyme and substrate in the proper orientation will lead to association where the two species fit together due to shape and electrostatic complementarity. However, since many association processes involve initially disordered proteins³, the coupled folding and binding process must be considered in the general case. One can imagine that diffusion together leads to an encounter complex or non-specifically bound proteins⁴. The *fold-on-contact* is a general scenario in which this encounter and the accompanying necessary conformational changes for binding appear to be concerted. In this

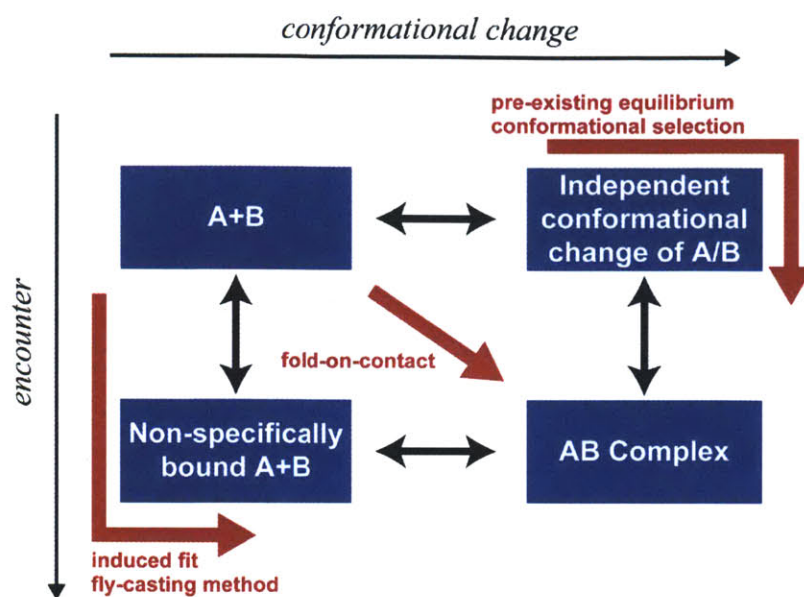


Figure 9.1: Different proposed mechanisms for association of proteins A and B. If association involves conformational change (as shown here), then the process can either occur in a concerted fashion (*fold-on-contact*) or a step-wise scenario. For step-wise association, A and B can non-specifically bind before re-arrangement into the final AB complex (*induced fit, fly-casting*) or first conformationally change followed by binding (*pre-existing equilibrium, conformational selection*).

case the encounter complex is a transition state in the association process. The *induced fit*⁶ mechanism predicts that preliminary encounter induces a conformational change that leads to binding. The *fly casting* mechanism is a variation on this process, in which the unfolded species can associate before folding into their final binding state. The expanded effective radius of encounter speeds the association process¹⁰. If conformational changes occur in the protein before binding, a *pre-existing equilibrium* exists¹². One specific case of this is if one of the species is disordered, but association only precedes when this species is folded, then a *conformation selection*¹³ will occur. Variations and combinations of these mechanisms have also been proposed^{14, 15}.

Answering general questions about the mechanism of molecular recognition and assembly in solution will be assisted by experiments that directly visualize the dynamics of the association process. In this chapter, we use 2D IR to investigate the dynamics of molecular recognition between proteins where disorder may play a role in the binding process: the monomer-dimer association of insulin. The hypothesis driving this work is that one can use transient dissociation of the dimer to prepare the same encounter species that appear in the recognition and binding process. Temperature jump (T-jump) transient 2D IR (t-2D IR) experiments of dimer dissociation will characterize conformational changes that necessarily

result from separation, which may be accompanied by disordering. These results will be used to test the validity of different conceptual pictures for molecular encounter and selective binding.

9.2 Background

9.2.1 Kinetics of molecular association

Although prediction of the rates of association from first molecular principles remains a complex problem, the kinetics of molecular associations has been developed over nearly a century. For association, partners must first diffuse into contact with one another. If diffusion is the only process determining binding, the bimolecular association rate, k_{ass} , can be predicted for spheres of uniform reactivity by:

$$k_{ass} = 4\pi DR \quad (9.1)$$

where D is the sum of the diffusion constant for each individual sphere and R is the sum of the hydrodynamic radii for the spheres. For dimeric insulin with a radius of ~ 1.5 nm, the diffusion constant is $\sim 1.5 \times 10^{-6}$ cm²/s ($15 \text{ \AA}^2/\text{ns}$)¹⁶. For small molecule chemical reactions in water, this second order rate is $1/k_{ass} \sim 100$ ps M. The observed rate of association (in s^{-1}) depends on the concentration, reflecting the increased likelihood of association with increasing concentration. Experiments often reflect much slower rates, indicating restraints on reaction imposed by the recognition and binding process¹⁷⁻¹⁹. The association rate must then be adjusted to account for these effects, which may include attraction (or repulsion) due to electric charges and/or necessary angular alignment for matching stabilizing contacts (H-bonds, hydrophobic residues, charged surfaces)^{18, 19}. For example, for diffusion with a potential for association $U(r)$, the hydrodynamic radii can be replaced with an effective reaction radius R^* defined by $(R^*)^{-1} = \int_R^\infty e^{U/RT} \frac{dr}{r^2}$ ²⁰. If

proteins bind only (and definitely) upon contact of specific reactive patches, then the k_{ass} will be slowed. For this ‘sticky patch’ scenario, modeling patches as circular areas on a sphere i which subtend a half-angle θ_i leads to a $\theta_A\theta_B(\theta_A+\theta_B)/2$ factor in eq. (9.1)²⁰. Although most observed protein association rates are slower than expected based on eq. (9.1), some are significantly faster, such as protein-DNA interactions²¹, which have in some cases been hypothesized to occur due to diffusion processes of reduced dimensionality²². Aside from accounting for attractive/repulsive forces and necessary orientational restrictions, if molecules must fold from different monomer structures into specific conformations upon binding, then the observed rate(s)

may be influenced by not only the discussed diffusion and binding rate, but also the structural transformation rate.

In order to distinguish different association mechanisms, time-resolved experiments allow for the determination of rates and the number of distinct thermodynamic states involved in the process, including the presence of intermediates. Proposing a kinetic model that is consistent with observed data allows for assessing hypothesized mechanisms. For the simplest association process, homo-dimerization, conversion of population is dictated by the association and dissociation rates:

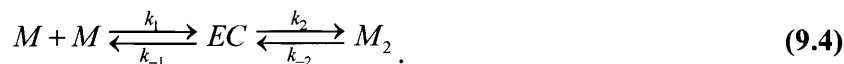


where M is the monomer and M_2 is the dimer. At equilibrium, the dissociation constant is given by the ratio of populations (inverted relative to the equilibrium constant for eq. (9.2)):

$$K_D = \frac{k_{diss}}{k_{ass}} = \frac{[M]^2}{[M_2]}. \quad (9.3)$$

For understanding the kinetics observed in time-resolved experiments, we first assume that the k_{ass} rate is independent of time. Some non-exponential behavior has been observed and explained by time-dependent rates²³⁻²⁵, although this non-exponentiality is only expected for fast (ns) nearly diffusion controlled reactions, where the spatial inhomogeneity of the monomers can affect the rate. If the reaction rates only depend on the bulk concentrations and the kinetic rates are time independent, then the kinetics for eq. (9.2) can be treated by writing out the differential equations for the time dependence of each species, and then finding solutions to the equations. For eq. (9.2), the solutions are of the form of hyperbolic sinusoidal functions²⁶, which illustrates that even simple two state models can result in complicated analytical solutions.

A more complicated scheme that may describe some of the paradigmatic mechanisms involves the following kinetics:



The dynamics require that two monomers M diffuse into proximity to form an encounter complex (EC), a loosely associated pair of molecules in contact or under the influence of each other. Then, alignment and configurational changes in the EC may be required to form the dimer M_2 . The EC lies in the gray area between two conceptual limits⁴: a stable intermediate and a fleeting transition state. These are differentiated based on the EC lifetime, population, and

barrier heights that separate it from the monomer and dimer states. For our experiment, we are generally sensitive to an *EC* that is classically viewed as an intermediate, where spectroscopically observable *EC* populations persist long enough for measurement. Absence of the experimental observation of an *EC* here does not preclude their existence. Rather, it may imply that they are fleeting complexes similar to transition states in which the rates leading away from *EC* are fast ($k_2, k_{-1} \gg k_{-2}, k_1$).

The experimental observation of the dissociation kinetics requires the preparation of a non-equilibrium state and time-resolving the induced equilibration. In the present case, T-jump experiments will be used to shift the relative population of monomers and dimers from their equilibrium values at the final temperature. If the experimentally induced changes are small the relaxation of systems described by eq. (9.2) or eq. (9.4) can be solved perturbatively (see Appendix 9.A) which results in an exponential relaxation process ($[M(t)] = [M_{eq}] + c_1 \exp[-\lambda_1 \tau] + c_2 \exp[-\lambda_2 \tau] + \dots$). In the case of eq. (9.2), following a perturbation, the system re-equilibrates with an exponential relaxation with a rate given by:

$$\lambda \approx k_{diss} + 4k_{ass}[M]_{eq}. \quad (9.5)$$

Here, the observed rate depends on the product of the association rate and the monomer equilibrium concentration, consistent with an expected faster rate at higher monomer concentrations due to increased collisions. Interestingly, for treatment of reactions of higher order or more states, as long as perturbation methods can reduce the differential equations down to effectively overall first order, then the solutions of the equations will be exponential²⁷, as is shown explicitly for two three-state kinetic schemes in Appendix 9.A. Due to the powerful simplification to exponentials provided by perturbative treatment of the kinetic rates, their use is common^{17, 26, 28}. Assessing how well exponential functions fit the data provides a point of assessment for the involved approximations: the rates are time-independent and the perturbations are small.

As described above, time-resolved spectroscopic techniques have the potential to differentiate proposed mechanisms and kinetic schemes by identifying the number of

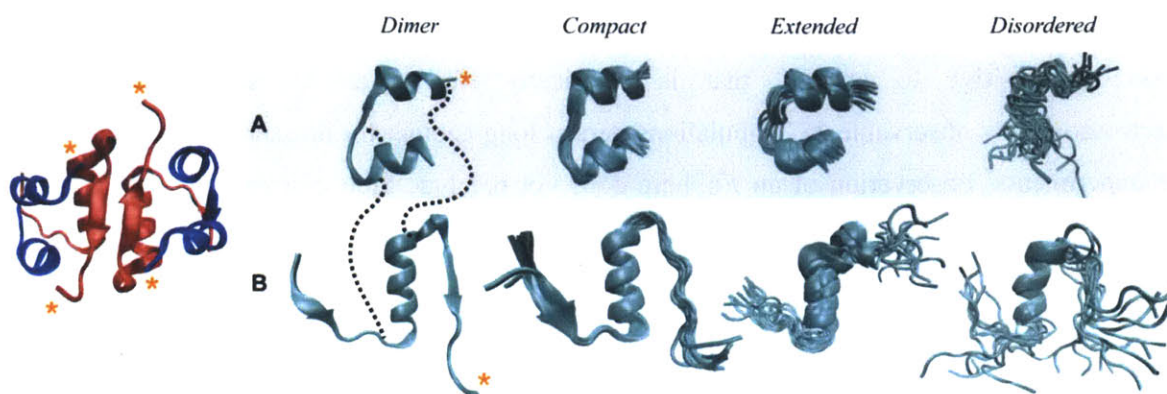


Figure 9.2: Different insulin structures determined by X-ray crystallography or NMR. At far left, the structure of the dimer (as observed in a hexamer crystal, PDB ID 4INS)². The A chain is blue while the B chain is red. C-termini are marked with an orange asterisk. The A (top row) and B (bottom row) chains are displayed separately for different determined structures. Black dashed lines indicate the three disulphide bonds, two of which connect the A and B chain. For these orientations, the A chain fits on top of the B chain to form insulin. The dimer monomer unit (4INS), *compact monomer* (2JV1⁵), *extended monomer* (1JCO⁷), and *disordered monomer* (1SF1^{8, 9}) are shown in similar orientations. For the structures assigned to monomers in solution (the three on the right), multiple possible structures are plotted from the PDB file. The structures are plotted with VMD¹¹.

intermediate states, identifying the number of timescales, and assessing the spectral changes associated with the timescales. Also, as shown in eq. (9.2) the observed rate depends on the equilibrium monomer concentration. A concentration-dependent rate is common to reactions of higher order and with more states, suggesting that the concentration-dependence can also be used to distinguish mechanisms.

9.2.2 Insulin Dimer Formation

Insulin is a model system for the study of association processes because its dissociation is coupled to a secondary structure change. It is a 51 amino acid protein that exists primarily as a monomer, dimer, and hexamer in solution. In humans, insulin is stored in the hexamer and higher oligomer form, although the protein must dissociate through the dimer into the monomer for biological activity²⁹. Despite the biological importance of the dimer-monomer transition, the complete structure of the monomer and the binding mechanism are unknown. While in the dimer and hexamer structures, the insulin monomer units have a well-defined structure²: the A chain forms a lowercase “n”-shape with α -helices towards the C- and N-terminal ends, while the B chain forms a “Z”-shape with a central α -helix and a C-terminal β -strand that forms an

interfacial anti-parallel β -sheet with the paired monomer as is shown in Figure 9.2. After dissociation, the interfacial β -sheet is lost, and the monomer β -strand is thought to disorder. The extent of disorder, however, is unknown. Through monomer-stabilizing mutations and/or the addition of co-solvents, multiple monomer structures have been observed, with varying degrees of disorder at the B chain N- and C- terminus and the A chain N-terminus, three regions that are not pinned in place by disulfide bonds^{5, 7, 8, 30-40}. These possible monomer structures can be grouped as *compact monomers* with a disordered, but compact B chain C-terminus (PDB ID 2JV1)⁵, *extended monomers* with a disordered B chain C-terminus extended into the solvent(1JCO)⁷, and *disordered monomers*, where the disordered B chain C-terminus is extended, and the B chain N-terminus and A chain N-terminus are disordered (1SF1)⁸. The accuracy of these NMR-determined structures is debatable, as increased disorder generally decreases the accuracy of structure determination. They are illustrative, however, of the regions that we may expect to disorder. The structures are shown in Figure 9.2. These monomer structures are a subset of a number of potential configurations, and they may all be present and interconverting in solution. Further evidence of multiple monomer states is indicated by an observed monomer disordering transition at high temperatures or denaturant concentrations^{8, 31, 32, 41}. A disordered or extended monomer structure is also thought to be a necessary precursor for fibrillation^{8, 42-46}. In fact, the monomer must extend beyond the *compact monomer* form to bind to insulin receptor⁴⁷.

Insulin oligomerization has been characterized through sedimentation⁴⁸⁻⁵⁰, mass spectrometry⁵¹, scattering^{52, 53}, FRET⁵⁴, pulsed field-gradient spin-echo NMR³⁹, CD⁵⁵, ultraviolet absorption^{56, 57}, kinetics⁵⁸, fluorescence depolarization⁵⁹, and FTIR^{45, 60}. While there are many methods currently used to study protein-protein interactions, every method has certain limitations. FTIR and NMR are limited to high concentrations by their low signal levels, while CD and fluorescence favor low concentration. Light scattering, CD, FTIR, and fluorescence depolarization can all be applied with adequate time resolution, but have inadequate structural sensitivity. Mass spectrometry, ultracentrifugation, and chromatography cannot be applied *in situ*^{48, 61}. Atomistic structural information is available from x-ray crystallography, solution NMR, and EPR with spin labels, but these methods are not suitable as fast transient experiments. As a system, insulin also presents its own limitations. It exists in many oligomer states, which complicates analysis, and irreversible aggregation limits the accessible temperature ranges.

The equilibrium thermodynamics of dissociation have been recently determined with two dimensional infrared spectroscopy (2D IR) with the two-state model shown in eq (9.2)⁶². In this study, the 2D IR spectroscopic signatures of the monomer and dimer were identified – the dimer exhibits a β -sheet characteristic high frequency ridge at $\omega_3 = 1690 \text{ cm}^{-1}$, while the monomer (which may be a mixture of states) is characterized by a strong peak at 1650 cm^{-1} . By monitoring these spectral features as a function of concentration, temperature, ethanol, and salt, K_D was characterized across multiple variables. Ethanol was found to stabilize the monomer state. For example, at 22°C addition of 20% EtOD caused a change in the K_D from $70 \mu\text{M}$ to 7.0 mM . The 2D IR spectra of the dimer, *compact monomer*, and *extended monomer*, were also simulated. As shown in Figure 9.3, both the monomer spectra are predicted to have a high $>1665 \text{ cm}^{-1}$ peak due to interaction between the α -helices and the B chain C-terminus region, a vibration which may lead to the weak shoulder observed in the high temperature equilibrium spectrum shown in Figure 9.7 (red arrow).

To characterize the number of states, identify the timescales separating the states, and

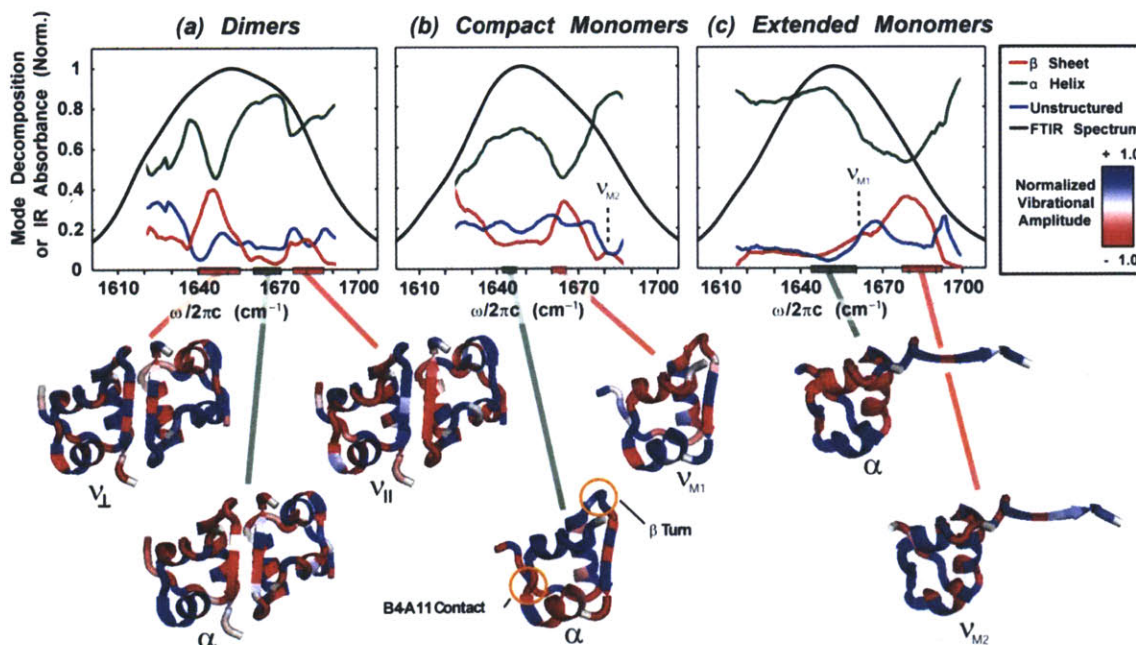


Figure 9.3: The frequency-dependent contribution of different oscillators to the underlying linear spectrum. Oscillators are grouped by the secondary structure to which they belong. Characteristic vibrations are displayed on the structures beneath, where the intensity of the color indicates the vibrational amplitude of that oscillator and the color indicates the phase. Reproduced from Ref. 1.

assess the underlying mechanism, we perform temperature-jump (T-jump) 2D IR experiments under conditions with primarily monomeric and dimeric insulin⁵¹. The T-jump causes a small perturbative change in the population of these states, and 2D IR is used as a structurally-sensitive probe to track the induced time-dependent conversion. In principle, by measuring the concentration dependence of the rate shown in eq. (9.5), the k_{diss} and k_{ass} can be calculated. At constant temperature, the spectra of different conformational states are conserved at varying concentrations – concentration-dependent spectral changes are attributed to shifting in the ratio of state populations. With varying temperature, the spectral features attributed to different states may change, but the states can be consistently assigned based on general spectral features and timescales to observe the temperature dependence of the state populations.

9.3 Methods

For all experiments performed here, insulin from bovine pancreas was purchased from Sigma-Aldrich (St. Louis, MO). Twice before use, the labile protons were exchanged for deuterons by dissolving the protein in D₂O at ~1 mg/mL, heating at 60°C for 1 h, and lyophilizing. Unless indicated otherwise, solutions contained 0.27 M DCl, 0.1 M NaCl, and 20% EtOD v/v. Some indicated samples were dissolved in 40% EtOD rather than 20%. All

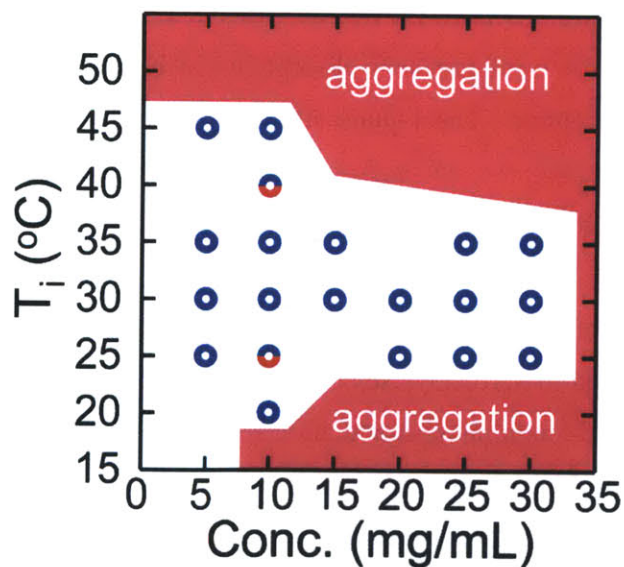


Figure 9.4: T-jump ZZZZ t-HDVE data was collected at the T_i and total concentrations indicated with circles. Blue(red) circles indicate that insulin was dissolved in D₂O with 0.1 M NaCl, 0.27 M DCl, and 20%(40%) EtOD v/v. At high and low T and high total concentrations, the parameter space is limited by aggregation that occurs in less than 8 h.

deuterated solutions were verified zinc-free and purchased from Cambridge Isotope Laboratories, Inc (Andover, MA). Based on the low pH, high ethanol content, and absence of Zn, insulin was assumed to exist almost exclusively as monomers and dimers¹. Insulin aggregation was observed after < 8 h (a typical data collection period) at low and high temperatures < 20, > 45°C for 5 mg/mL samples (Figure 9.4). Increasing concentration reduced the 8 h non-aggregation temperature range. If spectral signatures of accumulating aggregation (growth of a strong peak at ~1615-1628 cm⁻¹) were observed, the data was rejected. Concentrations of insulin were determined from the FTIR absorbance through an internal linear fit.

The transient T-jump 2D IR experiment was performed as described previously^{63, 64}. Briefly, the nonlinear signal generated by three variable-time-delayed 90 fs pulsewidth, 1660 cm⁻¹ center wavelength, 160 cm⁻¹ bandwidth, 1 kHz laser pulses was interferometrically detected on a 2x64 pixel HgCdTe array (Infrared Systems Development, Corp., Winter Park, FL). The reference beam for interferometric detection was passed through the sample. The beams were aligned in a boxcar geometry, and the resulting signal was corrected for linear absorption effects⁶⁴. The nonlinear signal was measured as a function of electronically-generated delay (τ -delay) following a 2 μ m, 6 ns, 10 mJ, 20 Hz, T-jump pulse, which excites the $\nu=2 \leftarrow 0$ O-D stretch of D₂O. The T-jump has been previously calibrated to create a $\Delta T = 10 \pm 2^\circ\text{C}$ ⁶³, as is verified by comparison of solvent transient absorption (t-Abs.) to the change in equilibrium temperature dependent absorption. The T-jump allows for time-dependent measurements over a 5 ns – 50 ms time range.

Two types of nonlinear signal, heterodyne-detected dispersed vibrational echo (HDVE) and 2DIR, were collected, both at $\tau_2 = 150$ fs. For HDVE⁶⁵, each τ -delay per data set was averaged at least 5 times (for concentrations < 15 mg, at least 12 averages). For HDVE, the reference beam was stepped to six delays separated by $\sim\pi/2$ (5 fs) intervals to allow for phase and amplitude calculation through phase-modulation spectral interferometry (5 mg/mL samples) or Fourier transform spectral interferometry (>5 mg/mL samples)⁶⁵. The real spectrum gives the dispersed pump-probe (DPP) while the amplitude gives the square root of the dispersed vibrational echo spectrum (AMP)⁶⁶. Incidentally, during t-HDVE measurements, the t-Abs. is also measured. The solvent-only reference t-Abs. is not collected. Rather, the solvent response is calculated from the 1st singular value decomposition (SVD) component, and subtracted.

For 2D IR measurements, the nonlinear signal was collected as a function of τ_1 delay. Rephasing spectra were collected at $\tau_1=0:14:2590$ fs while non-rephasing were collected at $\tau_1=0:14:1190$ fs. Fourier transform along τ_1 gives the ω_1 axis. Each τ -delay was averaged at least once, although generally three averages were taken.

Full collection of the data presented here (without considering rejected data sets) took 335 hours of 1 kHz laser time.

The T-jump is induced in the D₂O solvent surrounding insulin with a 6 ns laser pulse. The heat eventually diffuses out of the probed volume with a stretched exponential time dependence, which is shown in Figure 9.5. Due to the temperature relaxation time profile, the signals that we measure are distorted. To remove these distortions, we could de-convolve the observed signal with the temperature profile to reveal the underlying protein system response.

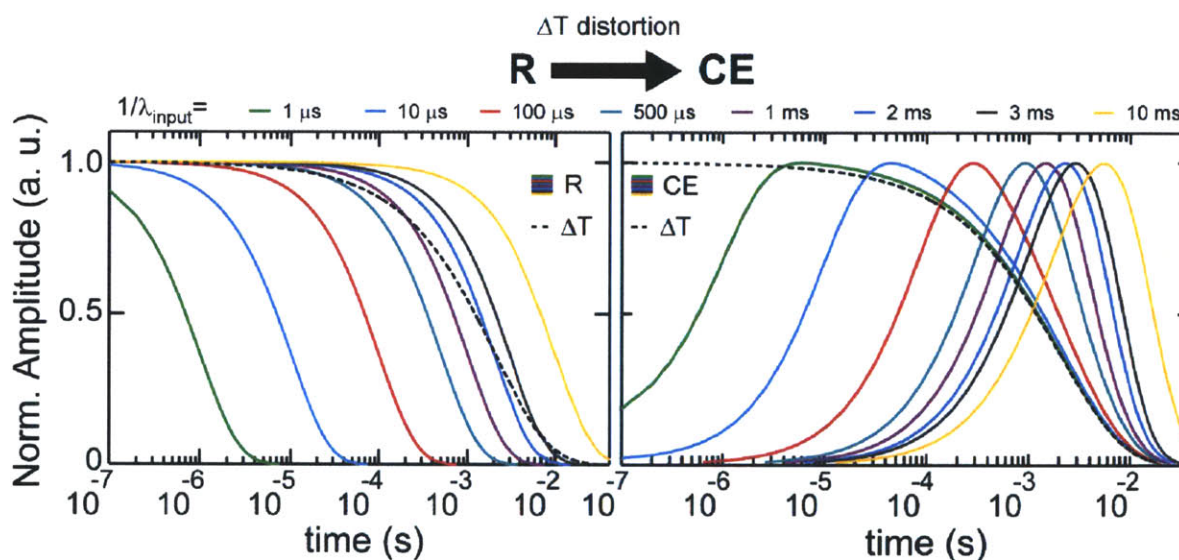


Figure 9.5: At left are exponential traces, R. Due to the stretched exponential shape of the temperature profile, ΔT , the temperature time-dependence imprints itself on the observed signals. The distortion of protein system responses can be predicted by convolving with the ΔT . Here, we have convolved exponential traces. Mono-exponential relaxation is distorted into a trace with a rise and decay. Measuring the rise and decay allows for mapping back to the underlying mono-exponential. Multiple exponential responses create a sum of CE. An instantaneous (or <5 ns) response will create a trace that matches ΔT . These convolutions ignore the temperature dependence of the rate constants. This effect is minimal at short (<100 μs) and long (>5 ms) times, where rates are dominated by the T_f and T_i rates, respectively. In the intermediate range, as the temperature drops, we expect that the extracted rate will be a sliding average between the rates at T_f and T_i . The convolutions shown above are compared to numerical kinetic models with temperature dependent rates in Chapter 3.

De-convolution, however, is easily corrupted by noise. Rather than de-convolve the signal, we have convolved exponential responses with the temperature profile. The resulting *convolved exponential* traces are shown in Figure 9.5 (and described in more detail in Chapter 3). In the following analysis, the data is fit to a linear combination of these *convolved exponential* traces. Fitting distorted signal trace data to these *convolved exponentials* is equivalent to fitting de-convolved signal traces to exponentials. If expected or observed in the data, similar convolutions can be done with arbitrary time-dependence (power law, stretched exponential) as the distortions are specific to the technique, not the proposed kinetic models. We have chosen to display the convolved exponential traces based on our general expectations, and the exponential-based analysis can be assessed based on how well the *convolved exponentials* match the data.

9.4 Results

9.4.1 Equilibrium Experiments

The monomer fraction at a given concentration and temperature can be calculated from the thermodynamic parameters determined in the previous equilibrium study assuming a two-state model¹. The dissociation constants and resulting calculated monomer fraction for the conditions used here are shown in Figure 9.6. As expected, increasing temperature and decreasing the total concentration causes an increase in the monomer fraction. For a total concentration of 10 mg/mL, we expect ~70% monomers at 20°C and ~100 % monomers at 55°C.

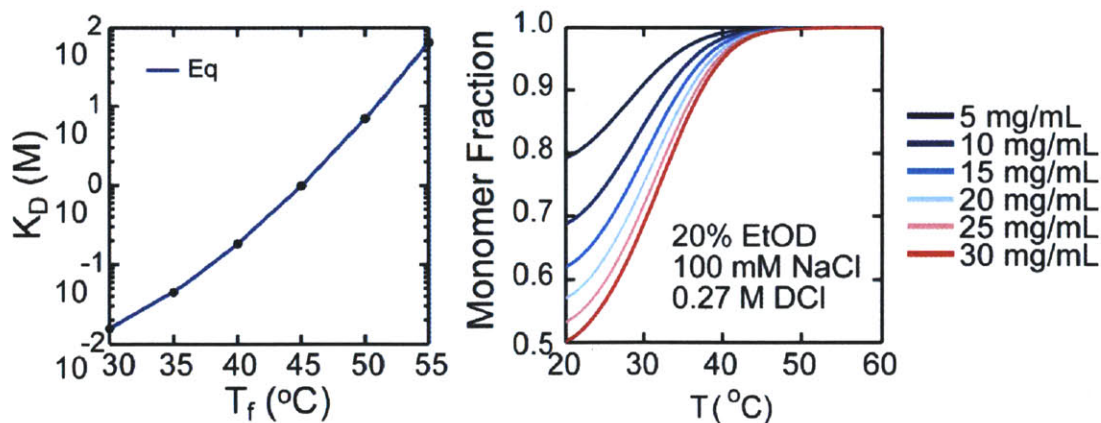


Figure 9.6: *Right* – The equilibrium K_D calculated as a function of temperature from previously determined parameters¹. Black dots indicate final temperatures probed with the transient experiments. *Left* – The monomer fraction, $[M]/([M]+2[M_2])$ calculated as a function of temperature and total concentration from the K_D at right.

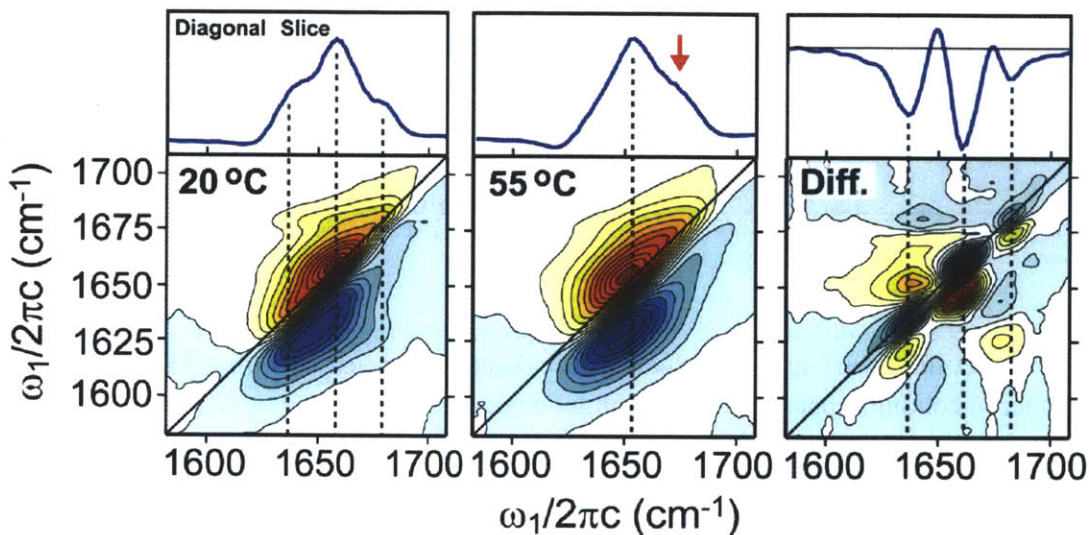


Figure 9.7: The 20°C, 55°C, and their difference ZZZY 2D IR spectra for 10 mg/mL insulin (20% EtOD, 0.1 M NaCl, 0.27 M DCl). The contours are plotted at 8.3% intervals. Reproduced from Ref. 1. The red arrow in the 55°C diagonal slice spectrum indicates a potential shoulder due to a vibrational mode that results from coupling between the monomer α -helices and B chain C-terminus, as has been predicted by simulation¹.

The ZZZY spectra at this concentration and these temperatures are shown in Figure 9.7 along with their difference. The low temperature spectrum, which has some dimer content, is characterized by a $\omega_1=1690\text{ cm}^{-1}$ ridge, which was assigned to the interfacial β -sheet. The almost exclusively monomer spectrum at 55°C is characterized by the absence of the ridge and an α -helical peak at 1650 cm^{-1} . The difference spectrum highlights these features – loss of the 1640 cm^{-1} and 1690 cm^{-1} interfacial β -sheet and shifting of intensity to the 1650 cm^{-1} α -helical region – which indicate melting of dimers into monomers. Upon T-jump, we expect to see similar spectra changes as observed in this equilibrium difference spectrum. Because the transient and equilibrium data were processed differently (the transient data have been corrected for temperature-dependent linear absorption of the solvent), the equilibrium and transient spectra are not expected to match exactly.

9.4.2 Transient Experiments

We have performed transient T-jump 2D IR experiments at a variety of concentrations (5, 10, 15, 20, 25, and 30 mg/mL; 10 mg/mL = 1.7 mM), initial temperatures (20, 25, 30, 35, 40, and 45°C), and ethanol concentrations (20 and 40% v/v) to characterize the dimer-to-monomer transition in insulin. The conditions used are tabulated in Figure 9.4.

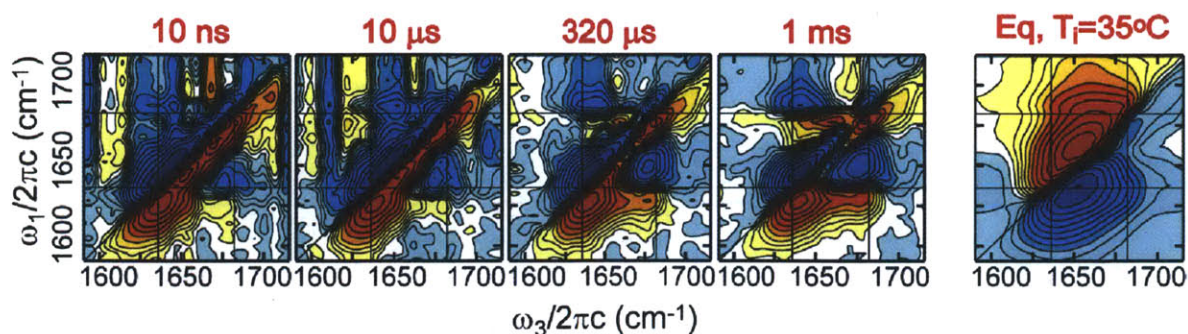


Figure 9.8: The difference ZZYY t-2D IR spectra for 20 mg/mL insulin at $T_i = 35^\circ\text{C}$. The 2D IR spectra are individually normalized and scaled such that 25 linearly-spaced contours span the entire magnitude range of $\sinh^{-1}(S_{Norm.} \times 27.29)$

Characteristic t-2D IR data sets are shown in Figure 9.8 and Figure 9.9. As observed in previous T-jump infrared experiments^{64, 67} the earliest accessible, pulse-width-limited 10 ns transient spectrum displays significant $\sim 4\%$ spectral changes. The strong diagonal loss features (blue above the diagonal, red below) are most intense on the lower frequency side of the equilibrium peaks. The other spectral feature discernible from noise is the loss cross-peak doublet in the lower right corner ($\omega_1 = 1675$, $\omega_3 = 1640$ and 1624 cm^{-1}), which can be assigned to

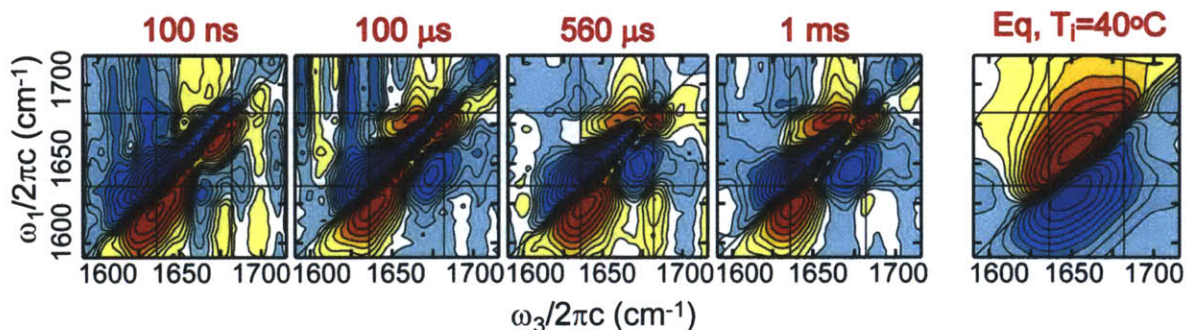


Figure 9.9: The difference ZZZZ t-2D IR spectra for 10 mg/mL insulin at $T_i = 40^\circ\text{C}$. 2D IR contours are plotted as described in Figure 9.8. The results from SVD analysis of this data are shown in Figure 9.15.

the interfacial β -sheet. These spectral features are consistent with loss of solvent/protein H-bonds: at higher temperatures, increased fluctuations within the protein cause homogeneous broadening and loss of some H-bonds, which results in diagonal loss features. The loss of some β -sheet cross-peak intensity can also be attributed to increased fluctuations and loosening of the structure. Although this may be the signature of more significant loss of structure (dissociation), the results can be explained with a generic amide I temperature response related to solvent H-

bond disruption⁶⁸ and longer timescale spectral features are significantly more convincing of dissociation.

With increasing time delays, the t-2D IR spectra show the appearance of different spectral features. The 320 μs t-2D IR spectrum shows loss features of cross-peaks (blue and red paired doublets) in the off diagonal regions (upper left: $\omega_1 = 1642$, $\omega_3 = 1687$ and 1677 cm^{-1} ; bottom right: $\omega_1 = 1677$, $\omega_3 = 1650$ and 1628 cm^{-1}). These cross-peaks are assigned to the interfacial β -sheet, and their loss indicates either disordering of the interface, dissociation, or a combination of the two.

We also observe gain features in the 320 μs spectrum. The positive (red) peak that appears in the top middle ($\omega_1 = 1660$, $\omega_3 = 1677 \text{ cm}^{-1}$) and the significant blue feature in the middle right ($\omega_1 = 1677$, $\omega_3 = 1650 \text{ cm}^{-1}$) are difficult to assign because they overlap with features expected for loss of the interfacial β -sheet. Their position and relative amplitudes, however, lead us to assign them to a monomer state, whose spectrum is expected to have increased intensity in the 1650 cm^{-1} region and a possible $\sim 1676 \text{ cm}^{-1}$ ridge¹.

A characteristic t-DPP data set is shown in Figure 9.10. The spectral features can be related to the described changes observed in the t-2D IR, whose projection onto the ω_3 axis is equivalent to the t-DPP. The major spectral changes observed following the initial pulse-width-limited 5 ns response (contrasting the 560 μs and 5 ns spectra) are the growth of the 1677 cm^{-1} positive (gain) feature and the appearance of a 1642 cm^{-1} negative feature (also attributed to gain). These two features correspond to the growth peaks identified in the 320 μs t-2D IR spectra.

The most interesting aspects of the t-DPP spectra are the clear appearance of at least four timescales. Inspection of the 1614 cm^{-1} time dependence (upper right panel of Figure 9.10) indicates behavior that approximately tracks the temperature. Comparison of the 1628 and 1682 cm^{-1} time dependencies with those from 1637 and 1647 cm^{-1} (lower left panel of Figure 9.10) reveals different timescales on the 100s of μs timescale. Inspection of longer times (lower right panel of Figure 9.10) shows 10s of ms timescale relaxation that persists long after the temperature returns back to equilibrium.

The presence of multiple time behaviors is also apparent in Figure 9.11, in which different frequencies respond on different timescales. For this t-DPP difference data, a signal corresponding to the temperature response was subtracted from each pixel (scaled to match the

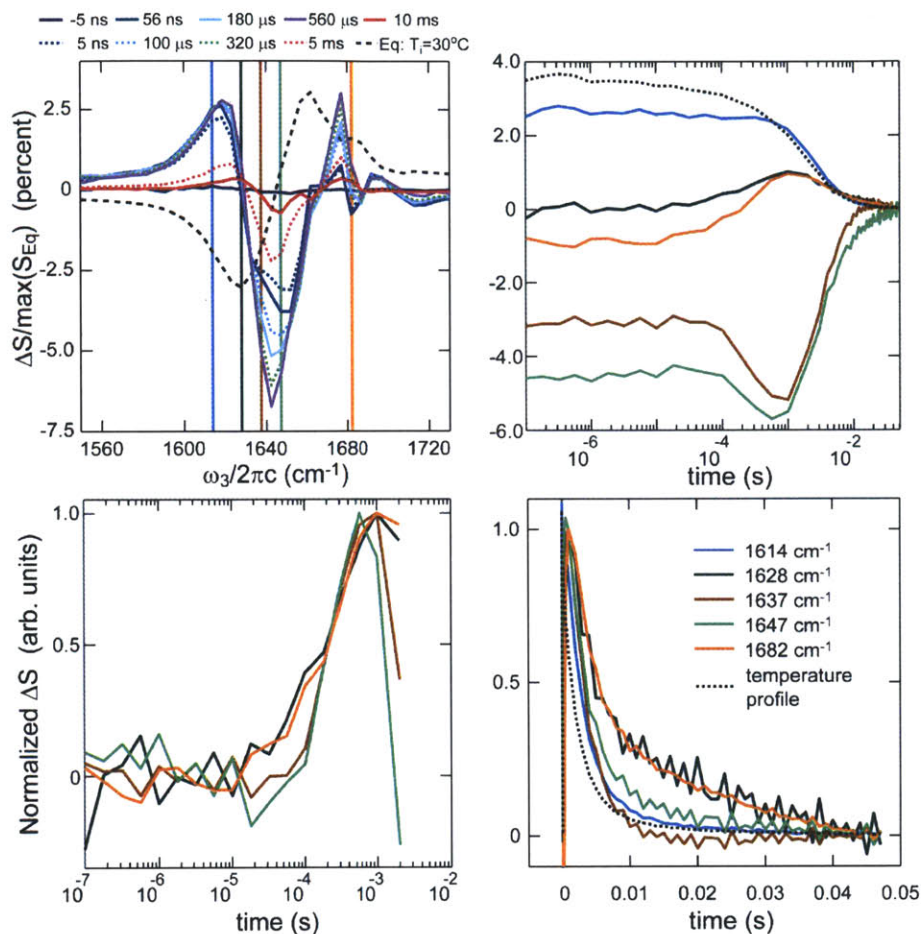


Figure 9.10: The difference ZZZZ t-DPP spectra for 10 mg/mL insulin at $T_i = 30^\circ\text{C}$. In the bottom panel, frequency traces are normalized.

early time response). This is a simple and non-invasive subtraction procedure. To assess the functional form of these traces, each data trace was fit to the convolution of the tested function with the temperature profile. Plotting different frequencies indicates that the bi-exponential form describes the data well with two sub-ms (45 and 460 μs exponential time constants) exponential timescales.

9.4.3 Separation of Spectra According to Timescale

The presence of multiple timescales and many datasets complicates fitting. Normally, individual frequency time traces are fit to potential functional forms and the simplest description that adequately describes the data is kept. For this data set, in which we have full spectral information, we strive to find a minimum number of shared time behaviors, whose linear

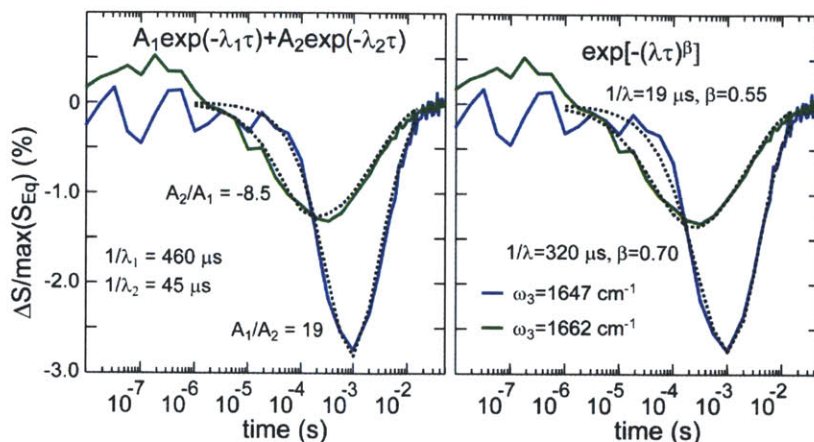


Figure 9.11: The difference ZZZZ t-DPP spectra *after* subtracting the temperature-tracking response ($\Delta S(\omega, \tau) - \Delta S(\omega, \tau = 10 \text{ ns})\Delta T(\tau)$, 10 mg/mL, $T_i=40^\circ\text{C}$). For each pixel, the amplitude at early time ($\sim 10 \text{ ns}$) indicates the amplitude of the temperature-tracking response. 1647 and 1662 cm^{-1} signals show different time behavior, which is apparent even without subtracting the temperature-tracking response (not shown). SVD is a method that can efficiently fit each frequency of the spectrum to a limited number of globally consistent exponentials. We find that 3 exponential functions and one temperature-tracking response characterize all the data. Different data sets may display a subset of the 4 responses. This data set displays two clear sub-ms exponential responses. Due to convolution with the temperature profile, each trace displays a decay and rise observed. Each panel shows the data fit to a different time dependence *after* the time dependence is convolved with the temperature profile. *Left* – Each trace is fit to the same two exponentials with varying amplitudes. Each frequency trace is dominated by one of the exponentials, as indicated by the amplitude ratios. *Right* – Each trace is fit to an independent stretched exponential.

combination reproduces the data at every frequency.

As is shown in Figure 9.11, for this specific case, both bi-exponential and stretched exponential functional forms fit the data. We choose to describe the data using multi-exponential fitting, although this choice can be assessed later. Once the exponential function form is chosen, the time traces are fit using the *convolved exponential* traces shown in Figure 9.5. Fitting the (temperature-profile distorted) data with these *convolved exponentials* is equivalent to fitting un-distorted signals with multi-exponential functions.

When exponential rates are measured and verified, it is convenient to separate the data into spectra associated with each exponential timescale. The exponential relaxation time-dependence and associated spectrum are called *kinetic phases*. Because the data has been decomposed into spectra based on their exponential time-dependence, we refer to the collection of *kinetic phases* as the *exponential basis*, which is straightforward to calculate from the data.

The information that we ultimately seek is the spectra and time-dependent concentrations of the states within the system, which we refer to as the *state basis*.

Relating the changes observed in a *kinetic phase* to the underlying states of the system is difficult. Although the amplitude of a frequency in the spectrum may be decreasing with an exponential timescale, the drop is generally due to the interconversion of many states, whose kinetics are dictated by the (exponential) solutions to their coupled differential equations. That being said, we can still relate the observed changes and timescales approximately to specific states. Each state is associated with a particular free energy well. Based on the barriers between these wells, population will interchange at a particular timescale. The fastest observed timescale corresponds to movement over the lowest barrier, suggesting that the fastest *exponential basis* spectrum corresponds to interconversion between the states with the lowest barrier. Longer timescales are due to redistribution over higher *and* lower wells (see Figure 6.9, Chapter 6). This is related to the concept of a rate-limiting step. Individual state populations respond with multi-exponential behavior. Thus, longer timescale *exponential basis* spectra are due to interconversion between many states. Hypothesizing the natural *state basis* from the raw data is difficult and requires a specific kinetic model.

9.4.4 SVD Separation

The *kinetic phases* of the *exponential basis* were identified using singular value decomposition (SVD), which linearly decomposes the data into orthogonal spectra and associated time traces. Taking a linear combination of the mathematically determined SVD time-traces allows us to calculate the *exponential basis*. That is, each data set is separated into at most four exponential time traces each with an associated spectrum. Multiplication of each spectrum, $u_i(\omega)$ by its associated *convolved exponential* signal, $v_i(\tau)$, recreates the data, $\Delta S(\omega, \tau)$:

$$\Delta S(\omega, \tau) = \sum_{i=1}^{\leq 4} u_i(\omega) v_i(\tau). \quad (9.6)$$

SVD is useful because it identifies the number of necessary exponentials, allows for faster fitting routines by treating full spectra rather than individual pixels. Fitting each pixel to a set of (at most) four exponential functions will result in the same results as the SVD method. More discussion of the process is given in Appendix 9.B.

Separating the spectra based on single exponential behavior creates a representation of the data in a particular *exponential basis* set. The explicit relationship between the *exponential*

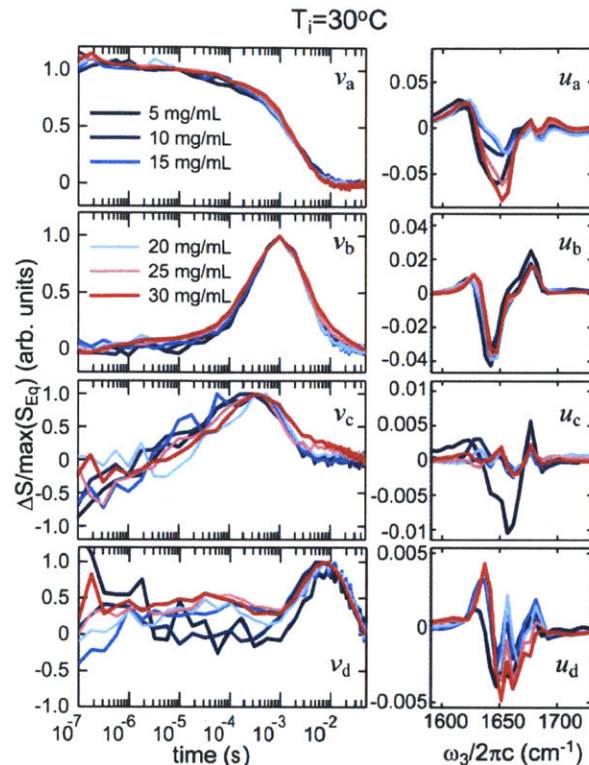


Figure 9.12: SVD analysis of the $T_i = 30^\circ\text{C}$ ZZZZ t-DPP spectra results in time traces (left) and spectra (right). Components 1, 2, 3, and 4 are mixed to produce the a, b, c, and d components shown above. Criteria for the initial stages of mixing are described in Appendix 9.B. v_c and v_d are created to match the signals predicted by convolution of the temperature profile with exponential relaxation.

basis and the *state basis* is shown in the equations in Appendix 9.A for specific two- and three-state systems. As described in the introduction, much about the time-evolution of the states in our system can be learned from the *exponential basis*. The spectra in the exponential basis (u_{a-d}) represent the combined changes in state populations on the associated exponential timescales (v_{a-d}). Thus, the u_{a-d} spectra are difference spectra. Characteristic *exponential basis* spectra and time traces are shown in Figure 9.12.

9.4.5 Separated Timescales

The separated v_{b-d} time traces are fit to bi-exponentials of the form $-C [\exp(-\lambda_{\text{rise}}\tau) - \exp(-\lambda_{\text{decay}}\tau)]$. The observed rates λ_{rise} and λ_{decay} are then mapped back to the underlying protein response based on the convolution calculations (see Figure 3.16 in Chapter 3). The mapping should produce the same results as deconvolution, although deconvolution is sensitive to noise and is more calculation intensive. The effectively deconvolved rates λ_{b-d} are shown as time

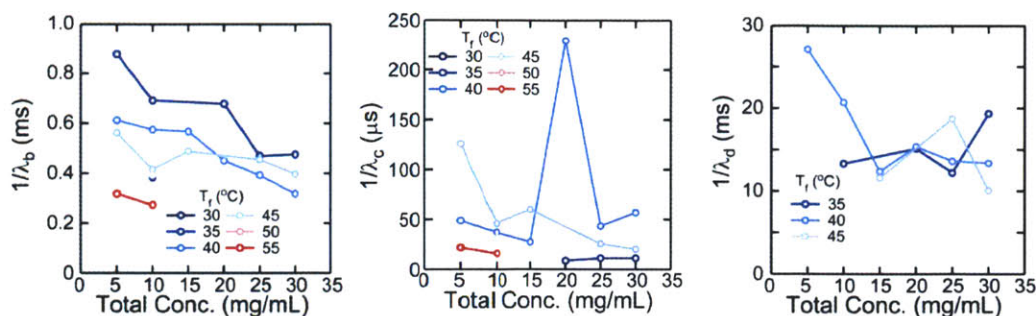


Figure 9.13: The deconvolved exponential time constants associated with components *b*, *c*, and *d*. The 95% confidence intervals (see Figure 9.C.1) calculated for fitting bi-exponentials to the v_i time traces are usually $\sim 10\%$ above and below the measured rates except for in the case of points that deviate wildly from the rest, which display higher error bars. There are no single amplitudes associated with the rates. Rather, there is a full spectrum associated with each rate. Because the v traces are normalized to one, the u spectra (as in Figure 9.12, for example) show the relative magnitude of the frequency-dependent amplitudes (spectra). The magnitudes of u_i spectra can be compared at individual frequencies as a function of temperature, as is done in Figure 9.14.

constants in Figure 9.13. Aside from the pulse-width-limited response (*a*, < 5 ns), the observed exponential rates fall into the three general categories consistent with the timescales observed in the raw data: intermediate (*b*, 250-1000 μ s), fast (*c*, 5-150 μ s), and long (*d*, 10-30 ms). As described in eq. (9.5), for a two state monomer to dimer transition (eq. (9.2)), we expect the observed rate to depend on the equilibrium monomer concentration, although the dissociation and association rate constants (k_{diss} and k_{ass}) should be independent of concentration. As we increase the total concentration, the concentration of monomers also increases even if the monomer fraction is decreasing. Thus, we expect an increase in rate with increasing total concentration. This dependence is observed in the *b* component, as is a temperature trend: as the temperature is increased, the rates increase. Neither a concentration dependence nor a temperature dependence is clear for the *c* and *d* components. Although the previously measured two-state equilibrium model thermodynamics¹ predict a potential 6-fold change in monomer concentration between our lowest and highest concentrations (0.85 mM and 5.1 mM), the slope of the concentration-dependent rate change is determined by the k_{ass} . If we assume a 10% error in our rates and a 0 to 5 mM monomer concentration range, then a rough calculation shows that if the ratio of k_{ass} / k_{dis} is < 20 (M^{-1}), then the concentration dependence will be within the noise.

Although no individual amplitude is associated with each *kinetic phase a,b,c,d* (full spectra are), the amplitude as a function of temperature can be plotted if a single frequency in the spectrum is monitored. At a particular T-jump delay, τ , the contribution of each *kinetic phase*

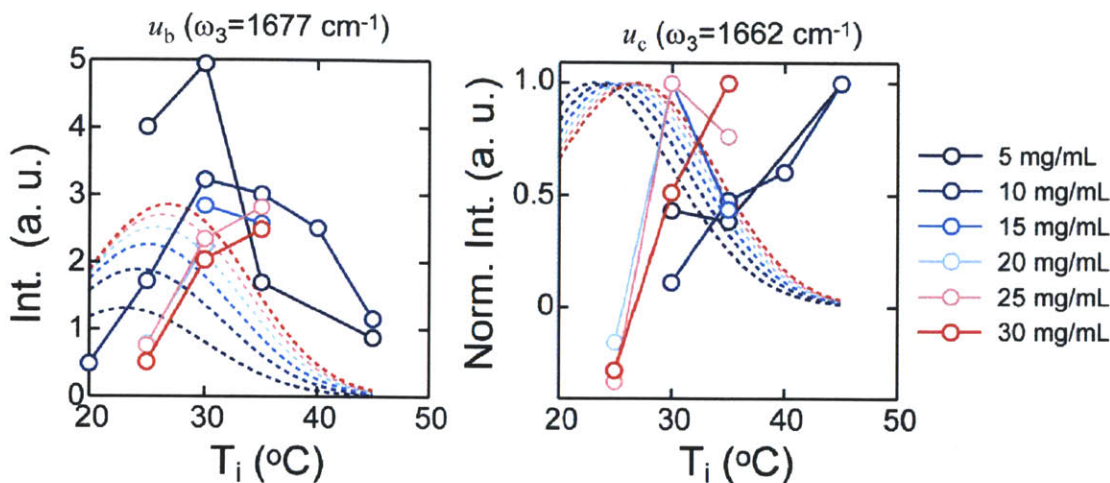


Figure 9.14: From the equilibrium thermodynamic values, the magnitude of monomer (or dimer) changes are calculated and shown with the dashed line ($[\text{monomer}(T_i)] - [\text{monomer}(T_i)]^1$). *Left* - The amplitude of the *b* component spectrum at 1677 cm^{-1} . The dashed expected curve was scaled by an arbitrary factor. The observed data was scaled by a different arbitrary curve. *Right* - The amplitude of the *c* component spectrum monitored at 1662 cm^{-1} normalized to their maximum value. The observed amplitudes depend on the response rates. The shown amplitudes have been corrected to attempt to adjust for the rate dependence. If assigned to the dissociation, the T-jump component *b* displays a trend $\sim 10^\circ\text{C}$ higher than predicted by equilibrium studies¹. Disagreement between the equilibrium-predicted and T-jump observed amplitudes may be due to the assumed two-state nature of the equilibrium studies.

contributing to the intensity observed at each frequency is equal to the product of the $u_i(\omega)$ and $v_i(\tau)$. The time traces v_i are normalized to one, such that comparison of the magnitude of the u_i gives a sense for the relative magnitudes of the different components across temperature and concentration. The $u_i(\omega)v_i(\tau)$ amplitude is normalized relative to the maximum of the equilibrium spectrum. Thus, a $u_i(\omega)v_i(\tau) = 0.04$ represents a 4% increase change relative to the maximum of the equilibrium spectrum. Thus, a larger $u_i(\omega)$ amplitude indicates an increased change in population associated with each *kinetic phase*. The temperature-dependent $u_i(\omega)$ amplitudes for *b* and *c* are shown at $\omega_3 = 1677$ and 1662 cm^{-1} , respectively, in Figure 9.14. The amplitude of the *b* spectrum tends to increase at low temperatures, reach a maximum, and decrease at high temperatures. The point of maximum amplitude increases with increasing concentration. The amplitude of the *c* spectrum tends to monotonically increase with increasing temperature. The amplitudes of the *a* component is largely independent of temperature and concentration (not shown) while the *d* component is inconclusive due to noise.

9.4.6 Separated Spectra

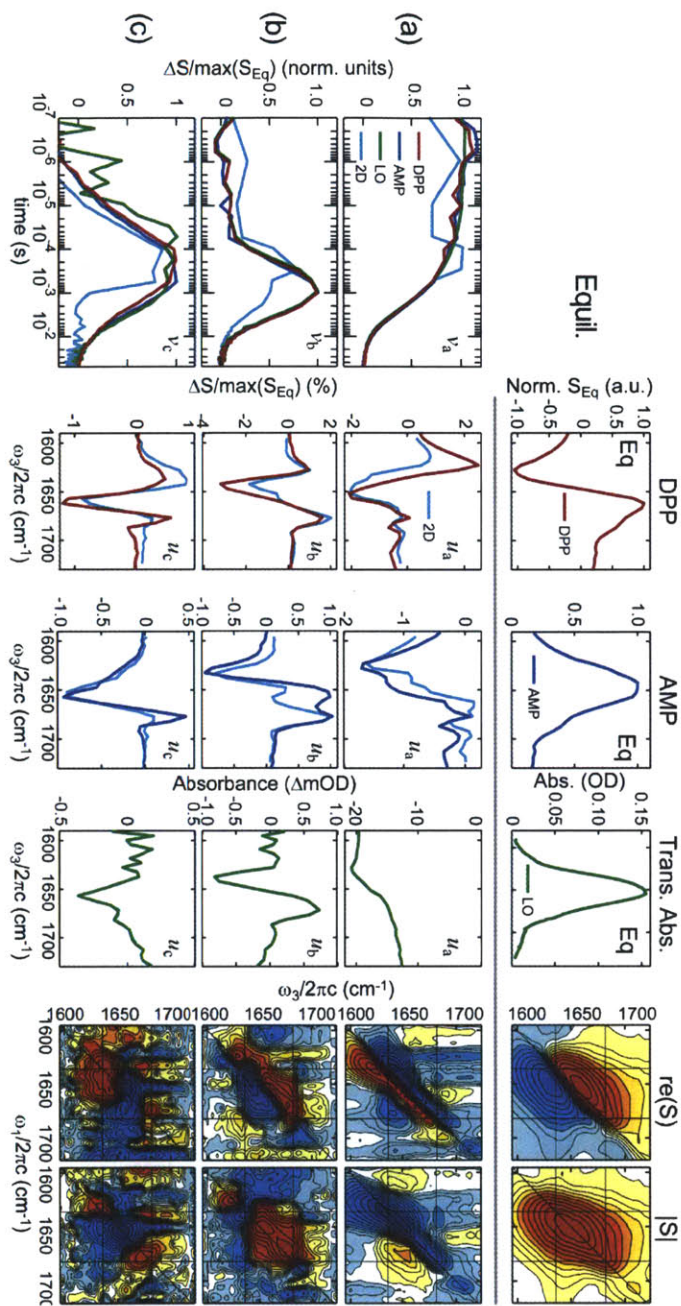


Figure 9.15: SVD analysis and subsequent remixing of the 10 mg/mL, $T_1 = 40^\circ\text{C}$ ZZZZ t-DPP and t-2D IR spectra results in time traces (left) and spectra (right). The projections of the t-2D IR surfaces are compared to the t-DPP and AMP spectra. The u_b and u_c t-2D IR spectra are baseline corrected to remove vertical stripping attributed to noise: a line is fit from the top to bottom of the spectrum at each ω_1 value and subtracted. This correction does not affect the general peak features. 2D IR contours are plotted as described in Figure 9.8.

Characteristic separated spectra and their associated time dependence are shown in Figure 9.15 (10 mg/mL, 40°C ; components a , b , and c) and Figure 9.16 (30 mg/mL, 35°C ; components a , b , and d). Assessment of the spectral features reveals the structural changes within the protein on the associated timescales. Based on their amplitudes (for example, comparing the maximum value of the DPP u spectra shown in Figure 9.12), components a and b reflect the largest changes observed in the difference spectra while c and d are on the edge of the

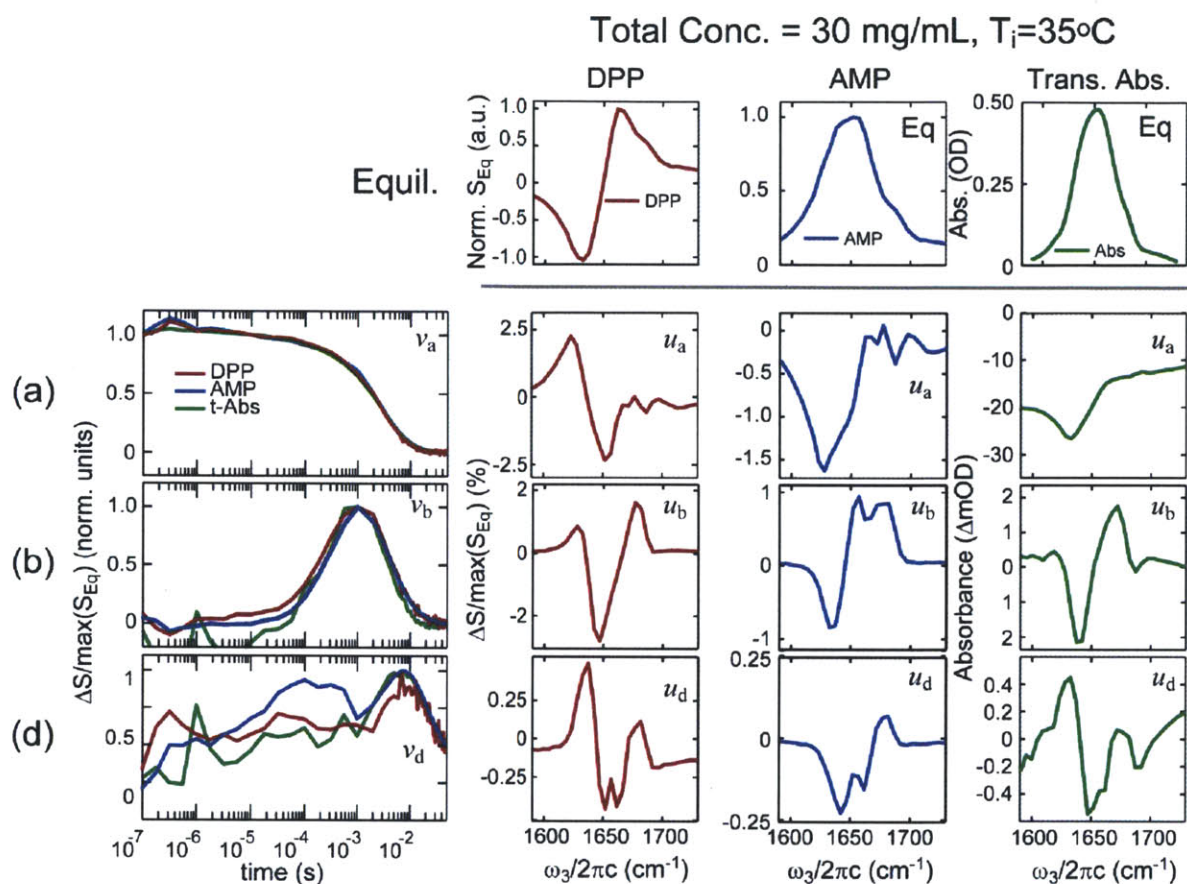


Figure 9.16: SVD analysis and subsequent remixing of the 30 mg/mL, $T_i = 35^\circ\text{C}$ ZZZZ t-DPP spectra results in time traces (left) and spectra (right).

noise level. The components reflect changes in the spectra. Thus, a large amplitude (*a* and *b*) suggests either significant spectral changes or significant population changes or both. Small amplitudes (*c* and *d*) indicate either small population changes or similarities between the spectra of the interconverting species. That is, although a large population of species may be interconverting, if the reactants and products have nearly identical spectra the difference signal observed will be small.

For component *a*, as has been observed previously^{64, 67}, the diagonal loss and off diagonal gain features in the u_a t-2D IR $\text{Re}(S)$ are consistent with homogeneous broadening (Chapter 3). The u_a t-DPP spectrum shows loss primarily on the low frequency side of equilibrium peaks. These spectral features are consistent with disruption of solvent/protein H-bonds due to increased temperatures and concomitant increased fluctuations.

Component *b* displays intermediate time behavior, and the associated spectral changes

are consistent with loss of the interfacial β -sheet. The u_b t-2D IR $\text{Re}(S)$ spectrum (Figure 9.15) shows gain and loss features. Loss is observed at the v_\perp and v_\parallel diagonal peaks ($\omega_1 = 1639$ and 1689 cm^{-1}) and the v_\perp/v_\parallel cross-peaks ($\omega_1 = 1685$, $\omega_3 = 1642$ and 1630 ; $\omega_1 = 1639$, $\omega_3 = 1692$ and 1677 cm^{-1}) attributed to loss of the interfacial β -sheet. Gain is observed at three diagonal peaks: $\omega_1 = 1623$, 1650 , 1669 cm^{-1} , which can respectively be assigned to the appearance of peaks attributed to increased inhomogeneous disorder, α -helices no longer coupled to an interfacial β -sheet, and α -helices coupled to a compact or disordered B-chain C-terminal end (M1 and M2 from Ref. 1, Figure 9.3). Simulations of insulin vibrations show coupling between the α -helix and β -sheet structures, suggesting that change will affect all the peaks to some extent¹. Consistent with the t-2D IR, the t-Abs. spectrum shows loss of the β -sheet v_\perp and v_\parallel peaks at 1637 and 1687 cm^{-1} , respectively, and gain maximized at 1672 cm^{-1} . Based on the spectral features, spectrum *b* appears to be dominated by loss of interfacial β -sheet and gain of monomer-like structure, although loss of the interfacial β -sheet does not necessarily mean dissociation, as it could lead to an associated encounter-complex.

Component *c* displays fast time behavior, and a characteristic u_c and v_c are shown in Figure 9.15. The u_c t-2D IR $\text{Re}(S)$ spectrum shows gain and loss features. Loss is observed along the diagonal at $\omega_1 = 1655 \text{ cm}^{-1}$, which corresponds to the frequency previously assigned to the α -helix. Gain features are observed at higher and lower frequencies, at $\omega_1 = 1635$ and 1674 cm^{-1} . These gain features may be attributed to a transition to a more disordered structure which can sample a wider range of environments, and thus displays increased intensity at lower and higher frequencies. Based on the loss of α -helix features, the *c* spectrum appears to be dominated by a disordering process.

Component *d* displays the slowest time behavior. A characteristic u_d and v_d are shown in Figure 9.16. The t-2D IR data are too noisy to separate the *d* behavior, but it is separable in the t-DPP, AMP, and t-Abs. spectra. The t-Abs. shows the clearest features: loss at 1647 and 1687 cm^{-1} and gain at 1633 and 1672 cm^{-1} . The 1647 loss is fairly generic, but the 1687 cm^{-1} is specific to anti-parallel β -sheet loss. The gain spectral features are consistent with oligomerization^{41, 43, 45, 60} and we observe a similar increase in the 1628 cm^{-1} region during the early stages of insulin aggregation. Although the spectra are noisy and we lack t-2D IR confirmation, the t-Abs., t-DPP, and AMP spectral features are consistent with the *d* component being dominated by oligomerization to structures larger than dimers.

9.5 Discussion

As described above, different spectral changes can be assigned to separate timescales. We have identified four timescales, *a* (pulse-width limited, strong amplitude), *b* (250-1000 μ s, strong amplitude), *c* (5-150 μ s, weak amplitude), and *d* (10-30 ms, weak amplitude). The processes occurring at these timescales can be assigned to molecular changes based on the spectral features, the concentration dependence, and other observations.

The pulse-width-limited *kinetic phase (a)* can be assigned to intrawell relaxation due to a <5 ns disruption of H-bonds. This assignment is based on the fast observed timescale, the general spectral response, and expectations based on similar behavior in previously studied systems.

The intermediate *kinetic phase (b)* can be assigned to dissociation and loss of the interfacial β -sheet based on a number of different factors. First, component *b* is the largest amplitude change we observe, suggesting that it represents the most significant conformational changes. We expect that loss of the interfacial β -sheet will introduce the largest spectral changes. Second, the spectral *b* features are dominated by loss of the interfacial β -sheet and gain of peaks associated with a monomer structure¹. Although different β -sheet structures are sometimes spectrally indistinguishable, we expect distinct spectra for the interfacial β -sheet and the *compact monomer* with any residual B-chain β -sheet structure – namely the *compact monomer* ν_{\parallel} analogous peak is expected at lower frequency relative to the interfacial β -sheet (Figure 9.3). The absence of strong β -sheet loss features in other components indicates that *b* captures the majority of the interfacial β -sheet disruption. Third, the observed *b* exponential time traces show concentration dependence. As the total concentration (and monomer concentration) increases, the rates increase. Both two- and three-state kinetic models predict that the dissociation process rate is concentration dependent, suggesting that this process is strongly linked to dissociation. Fourth, plotting the temperature dependence of the intensity of the *b* component spectrum for a particular concentration tracks the trend of increased monomer fraction expected for a dissociation process (left panel, Figure 9.14, monitored at 1677 cm^{-1}). This trend can also be observed in Figure 9.17. For dissociation, the monomer fraction is predicted to increase with a sigmoidal temperature-dependence (see Figure 9.6). However, we monitor the *change* in monomer concentration by T-jumping along this sigmoidal curve. Therefore, the expected response is a coarse derivative of the sigmoidal curve with respect to

temperature, which increases to the maximum point at temperature $T_{f,max}$ before dropping down

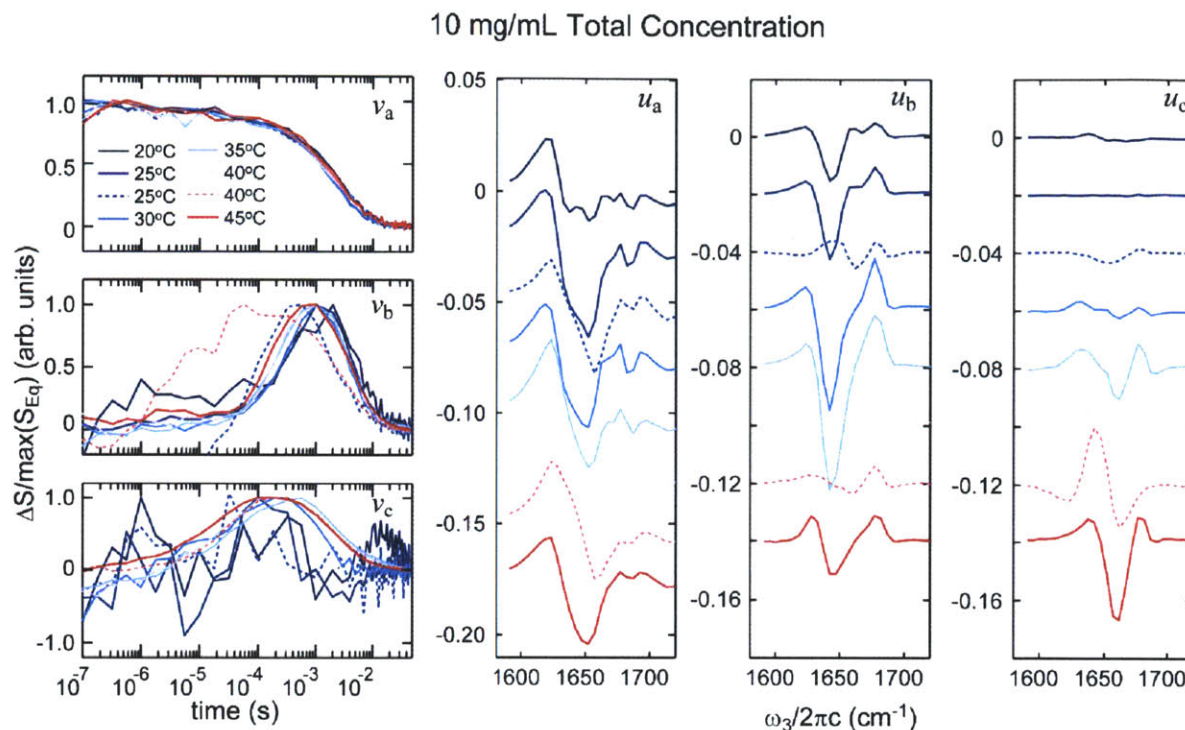


Figure 9.17: The spectra (right) and associated time-dependence (left) for 10 mg/mL samples. Each color is assigned to a temperature (T_i is reported). The dashed line refers to a 40% EtOD concentration, while the solid is 20%, the standard for this study. The $u_{a,b,c}$ spectra are offset by 0.025, 0.02, and 0.02, respectively.

to baseline when all the insulin molecules are in the monomer form. In theory, an unfolding transition (where monomer and dimer are replaced with unfolded and folded) would result in the same response. For an unfolding transition, however, we do not expect $T_{f,max}$ to be concentration dependent. The b amplitude $T_{f,max}$ shifts up with concentration supporting its assignment to dissociation.

The component c (5-150 μ s) can be assigned to a disordering process, most likely due to helical unfolding, based primarily on the spectral features. The rates are too noisy to make any statement (confirm or reject) a concentration dependence. In order to obtain the response from a predominately pure monomer disordering process, we performed T-jump experiments on insulin in 40% EtOD solvent and $T_i = 40^\circ\text{C}$, where we expect an almost purely monomer population due to the monomer-stabilization provided by ethanol. Analysis of the 40% EtOD t-DPP data shows that amplitude of the u_c spectrum is much larger than that of u_b (Figure 9.17). Therefore, we

associate the $T_i = 40^\circ\text{C}$ response from insulin in 40% EtOD with component c and assign them to correspond to monomer-specific. For the t-2D experiments under the same conditions (Figure 9.18), the observed response is dominated by spectral changes that look similar to the u_c from the 20% EtOD data, further confirming the assignment of c to monomer disordering. Figure 9.14 plots the intensity of component c (monitored at 1662 cm^{-1}). In general, the component c amplitudes are shown to increase with increasing temperature, consistent with its assignment to

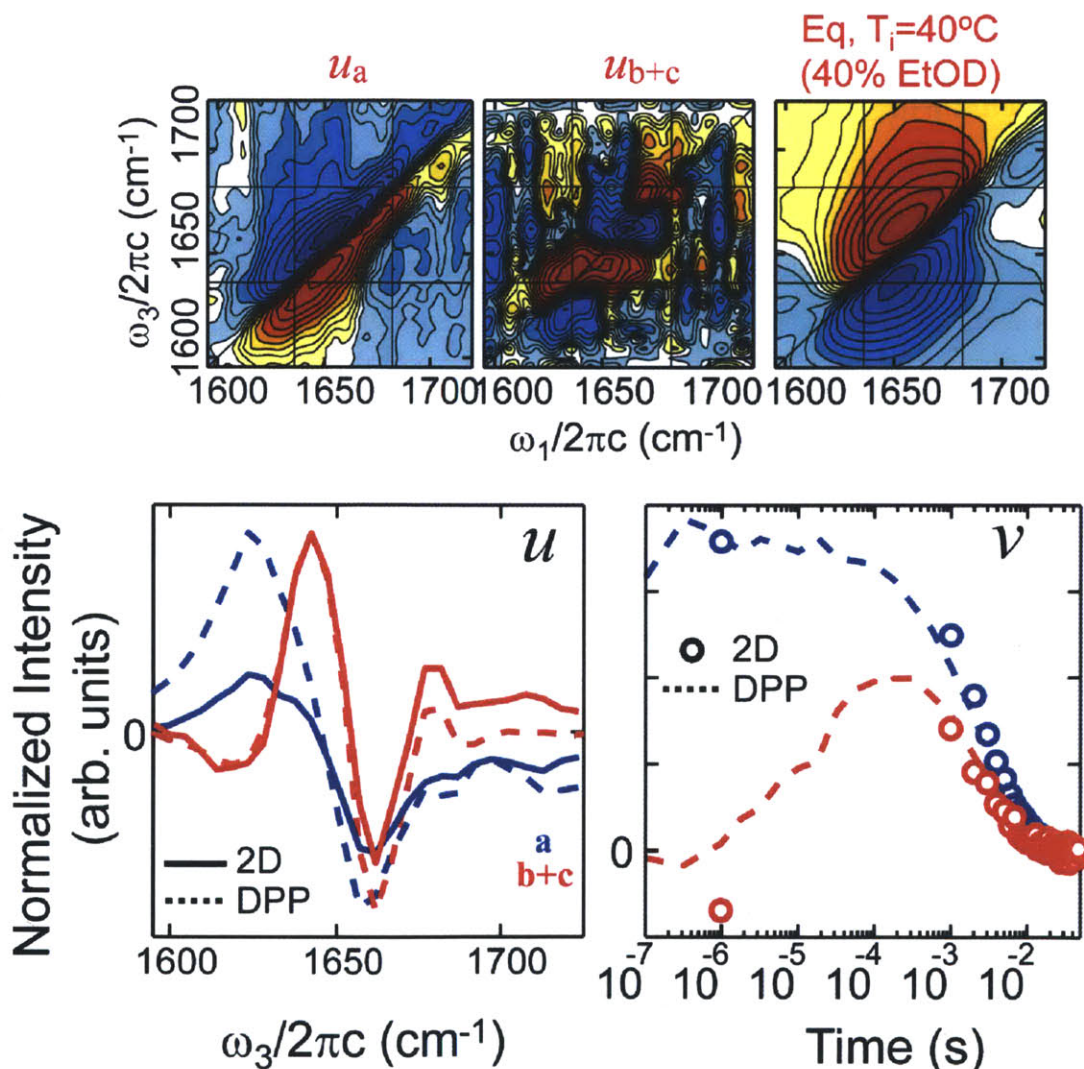


Figure 9.18: The results from SVD analysis and subsequent remixing of the 10 mg/mL, $T_i = 40^\circ\text{C}$, 40% EtOD ZZZZ t-DPP and t-2D IR spectra. Only two components are identified from the t-2D IR data, but they can be assigned to the t-DPP a and $b+c$ based on their spectral projections and time-dependence, which match the sum of the $b+c$ components from t-DPP. 2D IR contours are plotted as described in Figure 9.8.

an unfolding process that grows in with increasing temperature. Interestingly, the observed u_c spectral changes are also consistent with unfolding of parts of the dimer core without loss of the interfacial β -sheet. Therefore, if the timescale and spectral changes of unfolding of the dimer core are the same as for the monomer disordering, the component c may be due to both processes.

The component d is assigned to oligomerization based on its spectral features, but its small amplitude makes assignment less clear. The long time scale associated with it suggests that it represents the difference spectrum due to conversion between many states.

9.5.1 Mechanism

Based on the general assignments to the characteristic timescales, possible dimerization mechanisms can be assessed. We ignore the component d , whose long timescale and generic, oligomerization assignment suggest that it is a side-reaction that does not significantly affect the faster processes. Instead we focus on piecing together the b and c processes, as a is attributed to intrawell relaxation unassociated with significant conformational change. Based on equilibrium-determined K_D 's (Figure 9.6), under our conditions, the monomer fraction is always >50%. Thus, the insulin population is distributed between monomer and dimer states except for at high temperature and 40% ethanol concentrations, where it is almost exclusively monomer. Although one might be tempted to assign the fastest *kinetic phase* to the first step, it may be due to interchange from a populated intermediate. More specifically, even if the folded dimer is one of the populated states and we conceptually think of it as the beginning state for dissociation, the fast *kinetic phase* might be specific to the monomer population, which, although it may be an intermediate, is significantly populated at all conditions.

Ultimately, our ability to differentiate mechanisms is based on the sensitivity of our spectroscopic probe. Some possible mechanisms are tabulated in Figure 9.19. These represent a small, simple subset of the possible mechanisms. The mechanisms vary based on 1) the number of states (three or four) 2) the nature of the states (see Figure 9.19 caption), and 3) the processes linking each set of states (b then c or c then b). We can confidently assign the associated state to the dimer, whose structure is known from NMR⁶⁹.

Based on assessing two difference *kinetic phases*, b and c , we can identify at least three states (generally the number of observed phases is one less than the number of transitions). Other indistinguishable states may exist. When four states are identified below they indicate the

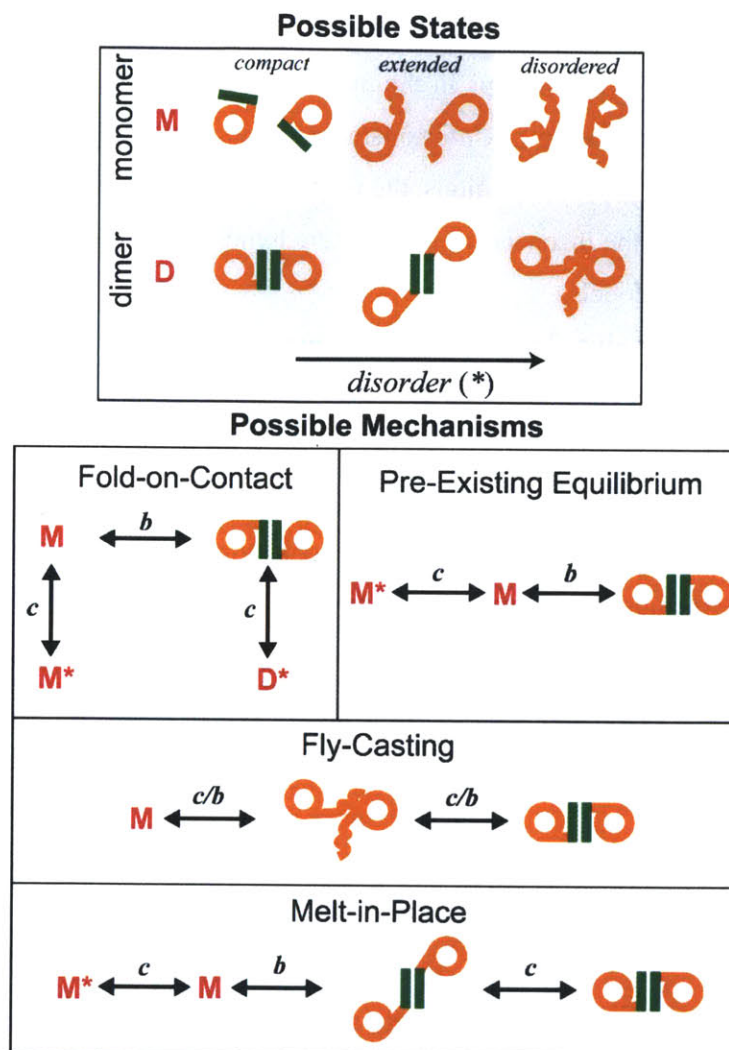


Figure 9.19: A small subset of simple possible dimerization mechanisms. *Top* – A subset of possible states. For the monomer (left to right): *compact*, *extended*, and *disordered*. For the dimer (left to right): folded, extended B-chains with intact β -sheet, non-specifically bound. *Bottom* – A subset of possible mechanisms. The asterisk indicates increased disorder relative to the **M** or **D** without the asterisk.

possibility that component c may be due to either dimer unfolding or coincidental monomer and dimer unfolding (without loss of the interfacial β -sheet),

Based on the spectral information, we may assign the b (250-1000 μ s) and c (5-150 μ s) *kinetic phases* to the dimer dissociation/interfacial β -sheet loss and disordering steps, respectively. This assignment is only valid, however, if the *observed* exponential timescales can be associated with a *single* transition between two states, rather than the mixing of different rates from different steps of the reaction. To access the validity of this assumption, quantitative

modeling was performed (Appendix 9.A) and we found that the observed rates (b or c) can be separately assigned to the sum of the forward and reverse rates (appropriately multiplied by the monomer concentration) for different steps in the mechanism with little overlap (except for the *fly-casting* scenario described below). Unfortunately, although the separation of timescales allows for assignments to individual transitions, the results were not definitive enough to assign a mechanism, as evidenced by the inconsistent temperature trends observed when fitting the b and c rates to three-state kinetic models.

Quantitative fitting produced a continuous temperature trend for one modeled scenario, the two-state *Fold-on-Contact* case, in which the c process was ignored. The results of the two-state modeling are shown in Appendix 9.C. To give a sense for the results of the fit, from $T_f = 30$ to 50°C , the k_{diss} generally increased with temperature from $1/k_{diss} = 520$ to $390 \mu\text{s}$. Conversely, the k_{ass} rate decreased significantly with temperature from $1/k_{diss} = 8 \mu\text{s}$ (30°C) to 2.7 ms (50°C). If the Arrhenius model can still be used, the decrease of rate with increasing temperature indicates a negative activation energy (Figure 9.C.1). If we believe the fit, which ignores the c process and depends heavily on the equilibrium dissociation constant to define the ratio of rates, then the negative activation energy for association may indicate the presence of a transient intermediate state⁷⁰⁻⁷².

Now that we have confidently assigned processes b and c to dimer dissociation and disordering, respectively, we discuss how these two steps manifest themselves in the following mechanisms:

Fold-on-Contact – *Fold-on-contact* is generally a two-state scenario, and we can assign the b kinetic phase to the dissociation process. There is no restraint on the monomer conformation, which is generally labeled **M** in Figure 9.19. Experimental observation of c , however, indicates more than two states (and suggests that **M** is not the fully *disordered monomer*). To reconcile this apparent contradiction, we assign the c conversion to a conformational change in monomer and/or dimer states that does not affect the dissociation process (dashed arrows, perpendicular indicates separate from dissociation). If the c transition is coupled to the dissociation process (not orthogonal), then this mechanism falls into a different classification than *fold-on-contact*.

Pre-existing Equilibrium – Here, in the b step, the dimer dissociates into monomers, although we cannot specify the degree of disorder of the monomers. In the c step, the monomer converts to a more disordered monomer state. *Disordered monomers* have been observed previously at

denaturing conditions^{8, 31, 32, 41}, and the formation of fibrils may rely on *disordered monomer* precursors^{8, 42-46}. Similar to these previous studies, we observe aggregation at high temperatures and at intermediate temperatures after long incubation times, suggesting the presence of some *disordered monomer* population. The *c* and *b* amplitude changes are consistent with population initially increasing in the center state, which then feeds the increase in a disordered monomer state at high temperatures. If the intermediate **M** state is a *compact monomer*, then this mechanism is *conformational selection*.

Fly-Casting – The *fly-casting* is an exception because there is no clear separation in the four forward and reverse rates: the *b* and *c* phases cannot be assigned to individual transitions. Thus, it is difficult to spectrally assess the mechanism. Infrared spectroscopy, however, is insensitive to the presence of a non-specifically bound dimer (encounter complex) without an interfacial β -sheet. An unfolded dimer is expected to have the same spectrum as an unfolded monomer. Given its indistinguishability, we would expect to see a single spectral change with dissociation with multiple exponential timescales indicating three states. The presence of two spectrally different *b* and *c* states suggests this mechanism is unlikely. This *fly-casting* mechanism also falls into the broader *induced fit* category.

Melt-in-place – In *melt-in-place*, the B chain β -strands extend while retaining the interfacial β -sheet *before* dissociation and disordering of the β -strands. We assign *b* to the dissociation process and *c* to the dimer disordering. We know from 40% EtOD experiments, however, that *c* is also due to monomer disordering, and we add a fourth **M*** state to reconcile this mechanism with the experimental observations. This *melt-in-place* mechanism also falls into the broader *induced fit* category, and encounter may occur through a generic *fly-casting* transition state (not the one described above).

Spectroscopically, we are only sensitive to encounter complexes with > 10 ns lifetimes and observable populations - which appear as intermediates. The only scenarios discussed here that have these long-lived intermediate encounter complexes are the *fly-casting* and *melt-in-place* mechanisms. Encounter complexes are often thought of as short-lived pairings that lie in the gray area between intermediates and transition states. Their observation may require extremely sensitive probes optimized for measuring ps – ns timescales, which are either below or on the edge of our observable time window.

From this small subset of potential scenarios, if we assess likelihood based on simplicity,

the *pre-existing equilibrium* and *fold-on-contact* mechanism (with **M**, **M***, and folded dimer) can explain the experimental observations with three states. The major difference between them is whether the association is independent of the disordering. *Fold-on-contact* assumes that the **M** and **M*** state will associate at the same rate while in *pre-existing equilibrium* only **M** will bind.

Ultimately, although we have reduced the possible mechanisms and identified the underlying states and timescales for insulin association, further experimental studies capable of supplementing our work are necessary to pinpoint the mechanism.

9.6 Conclusion

When applied to the insulin monomer-dimer transition, t-2D IR reveals four distinct timescales conserved across multiple concentrations and temperatures. These time-scales and their associated spectral changes can be generally assigned to H-bond rearrangement (< 5 ns), dissociation (250-1000 μ s), disordering (5-150 μ s), and oligomerization (10-30 ms). Based on the concentration, temperature, and ethanol dependence and the timescale assignments, we can also identify the states within our system. We have observed a folded dimer, two monomer states that differ based on their disorder, and an oligomer state. A disordered dimer state may also exist, although we have no specific evidence. The pulse-width limited response is due to H-bond loosening within these structures, although more significant changes may be masked and cannot be discounted.

These observations are based on amide I 2D IR spectroscopy's sensitivity to secondary structure. Localizing probes through isotope labeling or the use of other spectroscopic techniques will reveal other reaction coordinates. The ability to incorporate labels or dyes at specific positions in insulin through total chemical synthesis gives us access to these different reaction coordinates and expands our ability to design experiments capable of differentiating the many possible states and mechanisms⁷³. For example, equilibrium temperature-dependent experiments of insulin with isotope label pairs can reveal which segments are disordering at which temperatures. Placing a pair near the beginning of the B chain β -strand (one on B chain β -strand, one on the B chain α -helix or A chain C-terminus) creates a cross peak whose loss reveals conversion from the compact to the extended forms. If this transition occurs with a different temperature dependence than the loss of the interfacial β -sheet, then it indicates either dissociation to the *compact monomer* or a *melt-in-place* mechanism. Placing pairs on the B chain N-terminus (one on the B chain, one on the A chain disordered mid-region) or within the A

chain N-terminal α -helix will reveal possible temperature-dependent disordering of these regions. If these regions disorder at a lower temperature than the interfacial β -sheet, it may indicate the presence of disordered dimers. Once the states have been assessed through equilibrium experiments, T-jump of the isotope labeled insulin may also be performed to refine the current conclusions with the localized spectral sensitivity afforded by isotope labels. The localized probes give a better determination of the *state basis* because they can better distinguish the many disordered monomer states.

Förster resonance energy transfer (FRET) has been an attractive technique because it allows for measuring the distance between the monomers⁶². Recent attempts to measure FRET were thwarted by bad labeling efficiency⁶². Chemical synthesis of insulin may be the technique necessary for proper, efficient dye-labeling. T-jump FRET experiments are powerful because, by measuring the distance between dye-tagged monomers, they can provide the probe necessary to differentiate non-specific encounter (such as *fly-casting*) from *pre-existing equilibrium*, in which the interfacial β -sheet is lost during dissociation. 2D IR is sensitive to the β -sheet loss, but not the non-specific encounter. FRET experiments are also advantageous because they can be performed over wider concentration ranges to better resolve the concentration dependence.

Finally, without changing techniques or performing labeling, amide I T-jump 2D IR viscosity-dependent experiments may also provide a variable to distinguish different mechanisms and assess the assignment of timescales to particular processes. Diffusion-controlled reactions are expected to have a strong viscosity dependence⁷⁴. Assessing the timescale viscosity-dependence may reveal to what degree the *b*, *c*, and *d* processes are due to diffusion-controlled reactions versus conformational change, which we expect to have a weaker viscosity dependence. Increasing viscogen concentrations may also lengthen any <5 ns dissociation phenomena that might be hiding behind the H-bond loosening response. Unfortunately, viscous solutions are technically hard to work with because they slow timescales beyond the 50 ms window dictated by our T-jump repetition rate, which results in the build-up of longer-time decaying populations. This distorting accumulation of population may be overcome by flowing the sample.

9.7 Acknowledgment

I would like to thank Chunte Sam Peng and Carlos Baiz for helping to collect the data. I also thank CSP, CB, and Ziad Ganim for providing feedback on the content.

9.8 References

- (1) Z. Ganim, K. C. Jones, and A. Tokmakoff, "Insulin dimer dissociation and unfolding revealed by amide I two-dimensional infrared spectroscopy," *Phys. Chem. Chem. Phys.* **12**, 3579-3588 (2010).
- (2) E. N. Baker, T. L. Blundell, J. F. Cutfield, S. M. Cutfield, E. J. Dodson, G. G. Dodson, D. M. C. Hodgkin, R. E. Hubbard, N. W. Isaacs, C. D. Reynolds, K. Sakabe, N. Sakabe, and N. M. Vijayan, "The Structure of 2zn Pig Insulin Crystals at 1.5-Å Resolution," *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **319**, 369-& (1988).
- (3) P. E. Wright and H. J. Dyson, "Linking folding and binding," *Curr. Opin. Struct. Biol.* **19**, 31-38 (2009).
- (4) G. Schreiber, G. Haran, and H. X. Zhou, "Fundamental Aspects of Protein-Protein Association Kinetics," *Chem. Rev. (Washington, DC, U. S.)* **109**, 839-860 (2009).
- (5) W. Bocian, J. Sitkowski, E. Bednarek, A. Tarnowska, R. Kawecki, and L. Kozerski, "Structure of human insulin monomer in water/acetonitrile solution," *J. Biomol. NMR* **40**, 55-64 (2008).
- (6) D. E. Koshland, "Application of a Theory of Enzyme Specificity to Protein Synthesis," *Proc. Natl. Acad. Sci. U. S. A.* **44**, 98-104 (1958).
- (7) 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* **40**, 10732-10740 (2001).
- (8) Q.-X. Hua and M. A. Weiss, "Mechanism of insulin fibrillation: the structure of insulin under amyloidogenic conditions resembles a protein-folding intermediate," *J. Biol. Chem.* **279**, 21449-21460 (2004).
- (9) K. D. Sugase, H. J.; Wright, P. E., "Mechanism of Coupled Folding and Binding of Intrinsically Disordered Protein," *Nature* **447**, 1021-1025 (2007).
- (10) B. A. P. Shoemaker, J. J.; Wolynes, P. G., "Speeding Molecular Recognition by Using the Folding Funnel: The Fly-Casting Mechanism," *Proc. Nat. Acad. Sci.* **97**, 8868-8873 (2000).
- (11) W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal of Molecular Graphics* **14**, 33-38 (1996).
- (12) C. Tsai, S. Kumar, B. Ma, and R. Nussinov, "Folding funnels, binding funnels, and protein function," *Protein Sci.* **8**, 1181-1190 (1999).
- (13) H. R. Bosshard, "Molecular Recognition by Induced Fit: How Fit is the Concept?," *Physiology* **16**, 171-173 (2001).
- (14) G. G. Hammes, Y.-C. Chang, and T. G. Oas, "Conformational selection or induced fit: A flux description of reaction mechanism," *Proceedings of the National Academy of Sciences* **106**, 13737-13741 (2009).
- (15) R. Grunberg, J. Leckner, and M. Nilges, "Complementarity of Structure Ensembles in Protein-Protein Binding," *Structure* **12**, 2125-2136 (2004).
- (16) R. R. Walters, J. F. Graham, R. M. Moore, and D. J. Anderson, "Protein diffusion coefficient measurements by laminar flow analysis: Method and applications," *Anal. Biochem.* **140**, 190-195 (1984).
- (17) V. Munoz, R. Ghirlando, F. J. Blanco, G. S. Jas, J. Hofrichter, and W. A. Eaton, "Folding

- and Aggregation Kinetics of a β -Hairpin," *Biochemistry* **45**, 7023-7035 (2006).
- (18) M. Schlosshauer and D. Baker, "Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness," *Protein Sci.* **13**, 1660-1669 (2004).
- (19) J. Janin, "The kinetics of protein-protein recognition," *Proteins: Struct., Funct., Bioinf.* **28**, 153-161 (1997).
- (20) O. G. Berg and P. H. von Hippel, "Diffusion-controlled macromolecular interactions," *Annual Review of Biophysics and Biophysical Chemistry* **14**, 131-160 (1985).
- (21) A. D. Riggs, S. Bourgeois, and M. Cohn, "The lac repressor-operator interaction: III. Kinetic studies," *J. Mol. Biol.* **53**, 401-417 (1970).
- (22) P. H. Richter and M. Eigen, "Diffusion controlled reaction rates in spheroidal geometry: Application to repressor-operator association and membrane bound enzymes," *Biophys. Chem.* **2**, 255-263 (1974).
- (23) N. Agmon and A. Szabo, "Theory of reversible diffusion-influenced reactions," *J. Chem. Phys.* **92**, 5270-5284 (1990).
- (24) B. Cohen, D. Huppert, and N. Agmon, "Diffusion-Limited Acid-Base Nonexponential Dynamics," *J. Phys. Chem. A* **105**, 7165-7173 (2001).
- (25) D. Huppert, E. Pines, and N. Agmon, "Long-time behavior of reversible geminate recombination reactions," *J. Opt. Soc. Am. B* **7**, 1545-1550 (1990).
- (26) B. Nölting, *Protein Folding Kinetics*, 2nd Ed. ed. (Springer, Berlin, 2006).
- (27) M. N. Berberan-Santos and J. M. G. Martinho, "The integration of kinetic rate equations by matrix methods," *J. Chem. Educ.* **67**, 375 (1990).
- (28) G. G. Hammes, *Enzyme Catalysis and Regulation*, Molecular Biology (Academic Press, New York, 1982).
- (29) J. Brange, U. Ribel, J. F. Hansen, G. Dodson, M. T. Hansen, S. Havelund, S. G. Melberg, F. Norris, K. Norris, L. Snel, A. R. Sorensen, and H. O. Voigt, "Monomeric insulins obtained by protein engineering and their medical implications," *Nature* **333**, 679-682 (1988).
- (30) H. B. Olsen, S. Ludvigsen, and N. C. Kaarsholm, "Solution Structure of an Engineered Insulin Monomer at Neutral pH," *Biochemistry* **35**, 8836-8845 (1996).
- (31) M. a. Weiss, D. T. Nguyen, I. Khait, K. Inouye, B. H. Frank, M. Beckage, E. O'Shea, S. E. Shoelson, M. Karplus, and L. J. 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* **28**, 9855-9873 (1989).
- (32) Q.-x. Hua, W. Jai, and M. A. Weiss, "Conformational Dynamics of Insulin," *Frontiers in Endocrinology* **2**, 1-11 (2011).
- (33) Q. X. Hua, M. Kochoyan, and M. a. Weiss, "Structure and dynamics of des-pentapeptide-insulin in solution: the molten-globule hypothesis," *Proc. Natl. Acad. Sci. U. S. A.* **89**, 2379-2383 (1992).
- (34) V. Zoete, M. Meuwly, and M. Karplus, "A Comparison of the Dynamic Behavior of Monomeric and Dimeric Insulin Shows Structural Rearrangements in the Active Monomer," *J. Mol. Biol.* **342**, 913-929 (2004).
- (35) A. Budi, S. Legge, H. Treutlein, and I. Yarovsky, "Effect of external stresses on protein conformation: a computer modelling study," *European Biophysics Journal* **33**, 121-129 (2004).

- (36) 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. Mol. Biol.* **2**, 129-138 (1995).
- (37) Q. X. S. Hua, S. E.; Kochoyan, M.; Weiss, M.A., "Receptor Binding Redefined by a Structural Switch in a Mutant Human Insulin," *Nature* **354**, 238-241 (1991).
- (38) Y. Zhang, J. L. Whittingham, J. P. Turkenburg, E. J. Dodson, J. Brange, and G. G. Dodson, "Crystallization and preliminary crystallographic investigation of a low-pH native insulin monomer with flexible behaviour," *Acta Crystallographica Section D* **58**, 186-187 (2002).
- (39) W. Bocian, J. Sitkowski, A. Tarnowska, E. Bednarek, R. Kawecki, W. Komiski, and L. Kozerski, "Direct insight into insulin aggregation by 2D NMR complemented by PFGSE NMR," *Proteins* **71**, 1057-1065 (2008).
- (40) S. Ludvigsen, M. Roy, H. Thogersen, 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* **33**, 7998-8006 (1994).
- (41) W. Dzwolak, V. Smirnovas, R. Jansen, and R. Winter, "Insulin forms amyloid in a strain-dependent manner: An FT-IR spectroscopic study," *Protein Sci.* **13**, 1927-1932 (2004).
- (42) Q.-x. Hua, S. H. Nakagawa, W. Jia, K. Huang, N. B. Phillips, S.-q. Hu, and M. A. Weiss, "Design of an Active Ultrastable Single-chain Insulin Analog," *J. Biol. Chem.* **283**, 14703-14716 (2008).
- (43) A. Ahmad, V. N. Uversky, D. Hong, and A. L. Fink, "Early Events in the Fibrillation of Monomeric Insulin," *J. Biol. Chem.* **280**, 42669-42675 (2005).
- (44) T. N. Vinther, M. Norrman, H. M. Strauss, K. Huus, M. Schlein, T. Å. Pedersen, T. Kjeldsen, K. J. Jensen, and F. Hubálek, "Novel Covalently Linked Insulin Dimer Engineered to Investigate the Function of Insulin Dimerization," *PLoS One* **7**, e30882 (2012).
- (45) W. Dzwolak, R. Ravindra, J. Lendermann, and R. Winter, "Aggregation of Bovine Insulin Probed by DSC/PPC Calorimetry and FTIR Spectroscopy," *Biochemistry* **42**, 11347-11355 (2003).
- (46) N. B. Phillips, J. Whittaker, F. Ismail-Beigi, and M. A. Weiss, "Insulin Fibrillation and Protein Design: Topological Resistance of Single-Chain Analogs to Thermal Degradation with Application to a Pump Reservoir," *Journal of Diabetes Science and Technology* **6**, 277-288 (2012).
- (47) U. Derewenda, Z. Derewenda, E. J. Dodson, G. G. Dodson, X. Bing, and J. Markussen, "X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule: A completely inactive analogue," *J. Mol. Biol.* **220**, 425-433 (1991).
- (48) J. M. Creeth, "Sedimentation and Diffusion Studies of Insulin: the Maximum Molecular Weight," *Biochem J.* **53**, 41-47 (1953).
- (49) P. D. Jeffrey and J. H. Coates, "An Equilibrium Ultracentrifuge Study of the Effect of Ionic Strength on the Self-Association of Bovine Insulin," *Biochemistry* **5**, 3820-3824 (1966).
- (50) A. H. F. Pekar, B. H., "Conformation of Proinsulin. A Comparison of Insulin and Proinsulin Self-Association at Neutral pH," *Biochemistry* **11**, 4013-4016 (1972).
- (51) E. J. T. Nettleton, P.; Sunde, M.; Bouchard, M.; Dobson, C. M.; Robinson, C. V., "Characterization fo the Oligomeric States of Insulin in Self-Assembly and Amyloid Fibril Formation by Mass Spectrometry," *Biophys J.* **79**, 1053-1065 (2000).
- (52) C.-M. Yu, C. Y. Chin, E. I. Franses, and N.-H. L. Wang, "In situ probing of insulin

- aggregation in chromatography effluents with spectroturbidimetry," *J. Colloid Interface Sci.* **299**, 733-739 (2006).
- (53) A. K. Attri, C. Fernández, and A. P. Minton, "pH-dependent self-association of zinc-free insulin characterized by concentration-gradient static light scattering," *Biophys. Chem.* **148**, 28-33 (210).
- (54) 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," *European journal of biochemistry / FEBS* **255**, 580-587 (1998).
- (55) Y. Pocker and S. B. Biswas, "Conformational dynamics of insulin in solution. Circular dichroic studies," *Biochemistry* **19**, 5043-5049 (1980).
- (56) S. H. Strazza, R.; Walker, E.; Darnall, D. W., "The thermodynamics of bovine and porcine insulin and proinsulin association determined by concentration difference spectroscopy," *Arch. Biochem. Biophys.* **238**, 30-42 (1985).
- (57) R. S. Lord, F. Gubensek, and J. a. Rupley, "Insulin self-association. Spectrum changes and thermodynamics," *Biochemistry* **12**, 4385-4391 (1973).
- (58) R. H. Koren, G. G., "A Kinetic Study of Protein-Protein Interactions," *Biochemistry* **15**, 1165-1171 (1976).
- (59) S. Grudzielanek, R. Jansen, and R. Winter, "Solvational Tuning of the Unfolding, Aggregation and Amyloidogenesis of Insulin," *J. Mol. Biol.* **351**, 879-894 (2005).
- (60) M. Z. Bouchard, J.; Nettleton, E. J.; Dobson, C. M.; Robinson, C. V., "Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy," *Protein Sci.* **9**, 1960-1967 (2000).
- (61) P. D. C. Jeffrey, J. H., "An Equilibrium Ultracentrifuge Study of the Effect of Ionic Strength on the Self-Association of Bovine Insulin," *Biochemistry* **5**, 3820-3824 (1966).
- (62) Z. Ganim, "2D IR spectroscopy and computational modeling: application to protein folding and binding," (Massachusetts Institute of Technology, Cambridge, 2010).
- (63) H. S. Chung, M. Khalil, A. W. Smith, and A. Tokmakoff, "Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics," *Rev. Sci. Instrum.* **78**, 063101 (2007).
- (64) K. C. Jones, Z. Ganim, C. S. Peng, and A. Tokmakoff, "Transient two-dimensional spectroscopy with linear absorption corrections applied to temperature-jump two-dimensional infrared," *J. Opt. Soc. Am. B* **29**, 118-129 (2012).
- (65) K. C. Jones, Z. Ganim, and A. Tokmakoff, "Heterodyne-detected dispersed vibrational echo spectroscopy," *J. Phys. Chem. A* **113**, 14060-14066 (2009).
- (66) H. S. Chung, M. Khalil, and A. Tokmakoff, "Nonlinear infrared spectroscopy of protein conformational change during thermal unfolding," *J. Phys. Chem. B* **108**, 15332-15343 (2004).
- (67) H. S. Chung, Z. Ganim, K. C. Jones, and A. Tokmakoff, "Transient 2D IR spectroscopy of ubiquitin unfolding dynamics," *Proc. Natl. Acad. Sci. U. S. A.*, 14237-14242 (2007).
- (68) K. E. Amunson and J. Kubelka, "On the Temperature Dependence of Amide I Frequencies of Peptides in Solution," *J. Phys. Chem. B* **111**, 9993-9998 (2007).
- (69) A. M. M. Jørgensen, S. M. Kristensen, J. J. Led, and P. 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," *J. Mol. Biol.* **227**, 1146-1163 (1992).
- (70) M. Mozurkewich and S. W. Benson, "Negative activation energies and curved Arrhenius

- plots. 1. Theory of reactions over potential wells," *J. Phys. Chem.* **88**, 6429-6435 (1984).
- (71) D. L. Singleton and R. J. Cvetanovic, "Temperature dependence of the reaction of oxygen atoms with olefins," *J. Am. Chem. Soc.* **98**, 6812-6819 (1976).
- (72) N. Agmon, "Kinetics of Switchable Proton Escape from a Proton-Wire within Green Fluorescence Protein," *J. Phys. Chem. B* **111**, 7870-7878 (2007).
- (73) S. Luisier, M. Avital-Shmilovici, M. A. Weiss, and S. B. H. Kent, "Total chemical synthesis of human proinsulin," *Chem. Commun. (Cambridge, U. K.)* **46**, 8177-8179 (2010).
- (74) O. G. Berg and P. H. v. Hippel, "Diffusion-Controlled Macromolecular Interactions," *Ann. Rev. Biophys. Biophys. Chem* **14**, 131-160 (1985).

Appendix 9.A: Kinetic Models

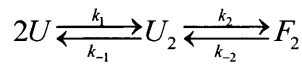
A perturbation method is applied to simplify two and three state monomer association kinetic models. In the perturbation method, the T-jump is assumed to perturb the concentration of state $[I]$ by a small amount $[i]$ such that $[I](t)=[I]_{eq}+[i](t)$. This perturbation treatment is substituted into the kinetic differential equations, and terms of order $[i]^2$ are ignored. An example is given at the end of this section and in Ref. 26. The result are tabulated below:

Model	Solution
$2F \xrightleftharpoons[k_{-1}]{k_1} F_2$	$[f](t) = [f_o] \exp(-\lambda t)$ $[f_2](t) = -0.5[f_o] \exp(-\lambda t)$ $\lambda = k_{-1} + 4k_1[F]_{eq}$
$2U \xrightleftharpoons[k_{-1}]{k_1} 2F \xrightleftharpoons[k_{-2}]{k_2} F_2$	$[u](t) = [u_o]C \exp(-\lambda_1 t) + [u_o](1-C) \exp(-\lambda_2 t)$ $[f](t) = \frac{[u_o]}{k_{-1}} \{C(k_1 - \lambda_1) \exp(-\lambda_1 t) + [u_o](1-C)(k_1 - \lambda_2) \exp(-\lambda_2 t)\}$ $[f_2](t) = -\frac{1}{2}([u](t) + [f](t))$ $p = k_1 + k_{-1} + k_{-2} + 4k_2[F]_{eq}$ $q = k_1 k_{-2} + k_{-1} k_{-2} + 4k_1 k_2 [F]_{eq}$ $\lambda_{1,2} = 0.5 p \pm (0.25 p^2 - q)^{1/2}$ $C = (k_{-1}[f_o]/[u_o] + \lambda_2 - k_1) / (\lambda_2 - \lambda_1)$
$2U \xrightleftharpoons[k_{-1}]{k_1} U_2 \xrightleftharpoons[k_{-2}]{k_2} F_2$	$[f_2](t) = [f_{2,0}]C \exp(-\lambda_1 t) + [f_{2,0}](1-C) \exp(-\lambda_2 t)$ $[u_2](t) = \left\{ \frac{k_2 - \lambda_1}{k_2} \right\} [f_{2,0}]C \exp(-\lambda_1 t) + \left\{ \frac{k_2 - \lambda_2}{k_2} \right\} [f_{2,0}](1-C) \exp(-\lambda_2 t)$ $[u](t) = -2[u_2](t) - 2[f_2](t)$ $p = k_{-1} + k_2 + k_{-2} + 4k_1[U]_{eq}$ $q = k_{-1} k_{-2} + (k_2 + k_{-2}) 4k_1 [U]_{eq}$ $\lambda_{1,2} = 0.5 p \pm (0.25 p^2 - q)^{1/2}$ $C = \frac{1}{(\lambda_2 - \lambda_1)} \left\{ k_2 \frac{[u_{2,0}]}{[f_{2,0}]} - k_{-2} + \lambda_2 \right\}$

These models must be modified slightly for our specific experiment, where the boundary conditions are different. The models are for experiments that start at $t = 0$ with concentrations that have *already* been perturbed and approach the equilibrium value at infinite time (for the

two-state case, $[F(t=0)] = [f_0] + [F]_{eq}$ while $[F(t=\infty)] = [F]_{eq}$. For the T-jump case, we perturb the system, it adjusts to the perturbation, and then relaxes back due to the temperature re-equilibration. To adjust for this, each exponential time dependence can be replaced by a bi-exponential rise and decay with equal and oppositely signed amplitudes: $\exp(-\lambda t) \rightarrow \exp(-\lambda_{rise}t) - \exp(-\lambda_{decay}t)$. This is the same as accounting for the temperature-profile distortion by replacing an exponential with a *convolved exponential*. With these sign conventions, the $[i_0]$ concentrations refer to the change in concentration expected (at equilibrium) upon changing the temperature from T_i to T_f ($[I(T_i)]_{eq} - [I(T_f)]_{eq}$).

Example:



The *equilibrium* quantities are given by the equations:

$$\frac{[U_2]}{[U]^2} = \frac{k_1}{k_{-1}} \quad (9.A.1)$$

$$\frac{[F_2]}{[U_2]} = \frac{k_2}{k_{-2}} \quad (9.A.2)$$

$$\begin{aligned} [P_{mono}] &= [U] + 2[U_2] + 2[F_2] \\ &= [U] + 2[U_2] + 2 \frac{k_2}{k_{-2}} [U_2] \\ &= [U] + 2[U]^2 \frac{k_1}{k_{-1}} + 2 \frac{k_2 k_1}{k_{-2} k_{-1}} [U]^2 \end{aligned} \quad (9.A.3)$$

where $[P_{mono}]$ is the total concentration in monomer units.

The kinetic differential equations are:

$$\frac{d[U]}{dt} = -2k_1[U]^2 + 2k_{-1}[U_2] \quad (9.A.4)$$

$$\frac{d[U_2]}{dt} = k_1[U]^2 - k_{-1}[U_2] - k_2[U_2] + k_{-2}[F_2] \quad (9.A.5)$$

$$\frac{d[F_2]}{dt} = k_2[U_2] - k_{-2}[F_2] \quad (9.A.6)$$

The perturbation definitions are given by:

$$[U] = [U]_{eq} + [u] \quad (9.A.7)$$

$$[U_2] = [U_2]_{eq} + [u_2] \quad (9.A.8)$$

$$[F_2] = [F_2]_{eq} + [f_2] \quad (9.A.9)$$

$$0 = [u] + 2[u_2] + 2[f_2] \quad (9.A.10)$$

$$[u_2] = -\frac{1}{2}[u] - [f_2]$$

For a T-jump, we assume that the concentrations are perturbed away from the equilibrium population by a small amount denoted by $[i(t)]$. It is this $[i(t)]$ that carries the time dependence of the perturbation, and it is the quantity we want to solve for. Substitution of the perturbation treatment into the kinetic equations results in:

$$\frac{d[F_2]_{eq}}{dt} + \frac{d[f_2]}{dt} = k_2([U_2]_{eq} + [u_2]) - k_{-2}([F_2]_{eq} + [f_2]) \quad (9.A.11)$$

$$\frac{d[f_2]}{dt} = k_2[u_2] - k_{-2}[f_2]$$

$$\begin{aligned} \frac{d[U]_{eq}}{dt} + \frac{d[u]}{dt} &= -2k_1([U]_{eq} + [u])^2 + 2k_{-1}([U_2]_{eq} + [u_2]) \\ &= -2k_1([U]_{eq}^2 + [u]^2 + 2[U]_{eq}[u]) + 2k_{-1}[U_2]_{eq} + 2k_{-1}[u_2] \end{aligned} \quad (9.A.12)$$

$$\begin{aligned} \frac{d[u]}{dt} &= -4k_1[U]_{eq}[u] + 2k_{-1}[u_2] - \langle 2k_1[u]^2 \rangle \\ &\approx -4k_1[U]_{eq}[u] + 2k_{-1}[u_2] \end{aligned}$$

where the triangular brackets indicate the squared term that is ignored in this approximate perturbative treatment.

Substitution of eq. (9.A.10) into eq. (9.A.12) leads to:

$$\begin{aligned} \frac{d[u]}{dt} &= -4k_1[U]_{eq}[u] - 2k_{-1}\left(\frac{1}{2}[u] + [f_2]\right) \\ &= -(4k_1[U]_{eq} + k_{-1})[u] - 2k_{-1}[f_2] \end{aligned} \quad (9.A.13)$$

Substitution of eq. (9.A.10) into eq. (9.A.11) leads to:

$$\begin{aligned} \frac{d[f_2]}{dt} &= -k_2\left(\frac{1}{2}[u] + [f_2]\right) - k_{-2}[f_2] \\ &= -k_2\frac{1}{2}[u] - (k_2 + k_{-2})[f_2] \end{aligned} \quad (9.A.14)$$

$$[u] = -2 \frac{1}{k_2} \frac{d[f_2]}{dt} - 2 \frac{1}{k_2} (k_2 + k_{-2}) [f_2]$$

Differentiation of eq. (9.A.11) and substitution of eq. (9.A.13) and eq. (9.A.14) leads to:

$$\begin{aligned} \frac{d^2[f_2]}{dt^2} &= -k_2 \frac{1}{2} \frac{d[u]}{dt} - (k_2 + k_{-2}) \frac{d[f_2]}{dt} \\ &= k_2 \frac{1}{2} \left\{ (4k_1[U]_{eq} + k_{-1})[u] + 2k_{-1}[f_2] \right\} - (k_2 + k_{-2}) \frac{d[f_2]}{dt} \\ &= k_2 \frac{1}{2} \left\{ -(4k_1[U]_{eq} + k_{-1}) \left\{ 2 \frac{1}{k_2} \frac{d[f_2]}{dt} + 2 \frac{1}{k_2} (k_2 + k_{-2}) [f_2] \right\} + 2k_{-1}[f_2] \right\} - (k_2 + k_{-2}) \frac{d[f_2]}{dt} \\ &= -(4k_1[U]_{eq} + k_{-1}) \left\{ \frac{d[f_2]}{dt} + (k_2 + k_{-2}) [f_2] \right\} + k_2 k_{-1} [f_2] - (k_2 + k_{-2}) \frac{d[f_2]}{dt} \\ &= -4k_1[U]_{eq} \frac{d[f_2]}{dt} - k_{-1} (k_2 + k_{-2}) [f_2] - 4k_1[U]_{eq} (k_2 + k_{-2}) [f_2] - \\ &\quad k_{-1} \frac{d[f_2]}{dt} + k_2 k_{-1} [f_2] - (k_2 + k_{-2}) \frac{d[f_2]}{dt} \\ &= \frac{d[f_2]}{dt} \left\{ -4k_1[U]_{eq} - k_{-1} - k_2 - k_{-2} \right\} + [f_2] \left\{ -k_{-1} k_{-2} - (k_2 + k_{-2}) 4k_1[U]_{eq} \right\} \\ 0 &= \frac{d^2[f_2]}{dt^2} + \frac{d[f_2]}{dt} \left\{ k_{-1} + k_2 + k_{-2} + 4k_1[U]_{eq} \right\} + [f_2] \left\{ k_{-1} k_{-2} + (k_2 + k_{-2}) 4k_1[U]_{eq} \right\} \end{aligned} \tag{9.A.15}$$

Which leads to the differential equation solution:

$$[f_2](t) = [f_{2,0}] C \exp(-\lambda_1 t) + [f_{2,0}] (1 - C) \exp(-\lambda_2 t) \tag{9.A.16}$$

$[u_2](t)$ is solved for by substituting eq. (9.A.14) and eq. (9.A.16) into eq. (9.A.10):

$$\begin{aligned} [u_2](t) &= \frac{1}{k_2} \left\{ \frac{d[f_2](t)}{dt} + k_{-2} [f_2](t) \right\} \\ &= -\frac{1}{k_2} \left\{ [f_{2,0}] \lambda_1 C \exp(-\lambda_1 t) + [f_{2,0}] \lambda_2 (1 - C) \exp(-\lambda_2 t) \right\} \\ &\quad + \frac{k_{-2}}{k_2} \left\{ [f_{2,0}] C \exp(-\lambda_1 t) + [f_{2,0}] (1 - C) \exp(-\lambda_2 t) \right\} \\ &= \left\{ \frac{k_{-2} - \lambda_1}{k_2} \right\} [f_{2,0}] C \exp(-\lambda_1 t) + \left\{ \frac{k_{-2} - \lambda_2}{k_2} \right\} [f_{2,0}] (1 - C) \exp(-\lambda_2 t) \end{aligned} \tag{9.A.17}$$

Now that the other two populations are known, $[u](t)$ can be re-written by eq. (9.A.10):

$$[u](t) = -2[u_2](t) - 2[f_2](t) \tag{9.A.18}$$

λ is solved for by differentiating eq. (9.A.16) and solving the resulting quadratic equation:

$$\begin{aligned}
[f_2](t) &= [f_{2,0}]C \exp(-\lambda_1 t) + [f_{2,0}](1-C) \exp(-\lambda_2 t) \\
\frac{d[f_2](t)}{dt} &= -\lambda_1 [f_{2,0}]C \exp(-\lambda_1 t) - \lambda_2 [f_{2,0}](1-C) \exp(-\lambda_2 t) \\
\frac{d^2[f_2](t)}{dt^2} &= \lambda_1^2 [f_{2,0}]C \exp(-\lambda_1 t) + \lambda_2^2 [f_{2,0}](1-C) \exp(-\lambda_2 t) \\
&= \lambda_1^2 [f_{2,0}]C \exp(-\lambda_1 t) + \lambda_2^2 [f_{2,0}](1-C) \exp(-\lambda_2 t) \\
&\quad - p \{ \lambda_1 [f_{2,0}]C \exp(-\lambda_1 t) + \lambda_2 [f_{2,0}](1-C) \exp(-\lambda_2 t) \} + \\
&\quad q \{ [f_{2,0}]C \exp(-\lambda_1 t) + [f_{2,0}](1-C) \exp(-\lambda_2 t) \} \\
0 &= \lambda_1^2 [f_{2,0}]C - p\lambda_1 [f_{2,0}]C + q [f_{2,0}]C \\
&= [f_{2,0}]C \{ \lambda_1^2 - p\lambda_1 + q \} \\
0 &= \lambda_2^2 [f_{2,0}](1-C) - p\lambda_2 [f_{2,0}](1-C) + q [f_{2,0}](1-C) \\
&= [f_{2,0}](1-C) \{ \lambda_2^2 - p\lambda_2 + q \}
\end{aligned} \tag{9.A.19}$$

C is solved by equating the $d[f_2]/dt$ from eq. (9.A.11) with the derivative at $t = 0$ given by differentiating eq. (9.A.16) (also from eq. (9.A.19)):

$$\begin{aligned}
\frac{d[f_2](t=0)}{dt} &= k_2 [u_{2,0}] - k_{-2} [f_{2,0}] \\
&= -\lambda_1 [f_{2,0}]C - \lambda_2 [f_{2,0}](1-C) \\
k_2 \frac{[u_{2,0}]}{[f_{2,0}]} - k_{-2} &= (\lambda_2 - \lambda_1)C - \lambda_2 \\
C &= \frac{1}{(\lambda_2 - \lambda_1)} \left\{ k_2 \frac{[u_{2,0}]}{[f_{2,0}]} - k_{-2} + \lambda_2 \right\}
\end{aligned} \tag{9.A.20}$$

Appendix 9.B: SVD separation of *kinetic phases*

In order to determine the *kinetic phases* that comprise the *exponential basis*, we must find the spectral amplitudes associated with different timescales. The natural procedure for doing this, which was described above, is illustrated as the *Fitting Approach* in Figure 9.B.1. Each frequency would be fit to a minimum set of exponentials drawn from the set of *convolved exponential* traces (equivalent to fitting to a multi-exponential). The same set of consistent exponentials with different amplitudes would be used for each pixel within a data-set. For global consistency, we expect trends (and the same number of exponentials) across data-sets. Stitching together the amplitudes associated with each *convolved exponential* at each frequency allows for determination of the full spectrum associated with each rise-and-fall time behavior. Unfortunately, fitting multiple exponentials for each frequency requires complicated fitting routines and convergence after multiple iterations. There is a large flexibility in fitting each pixel separately; ideally, an entire set of frequencies (a spectrum) can be fit at once rather than treating each pixel individually.

A different approach, the *SVD Approach* (Figure 9.B.1), achieves the same results as the

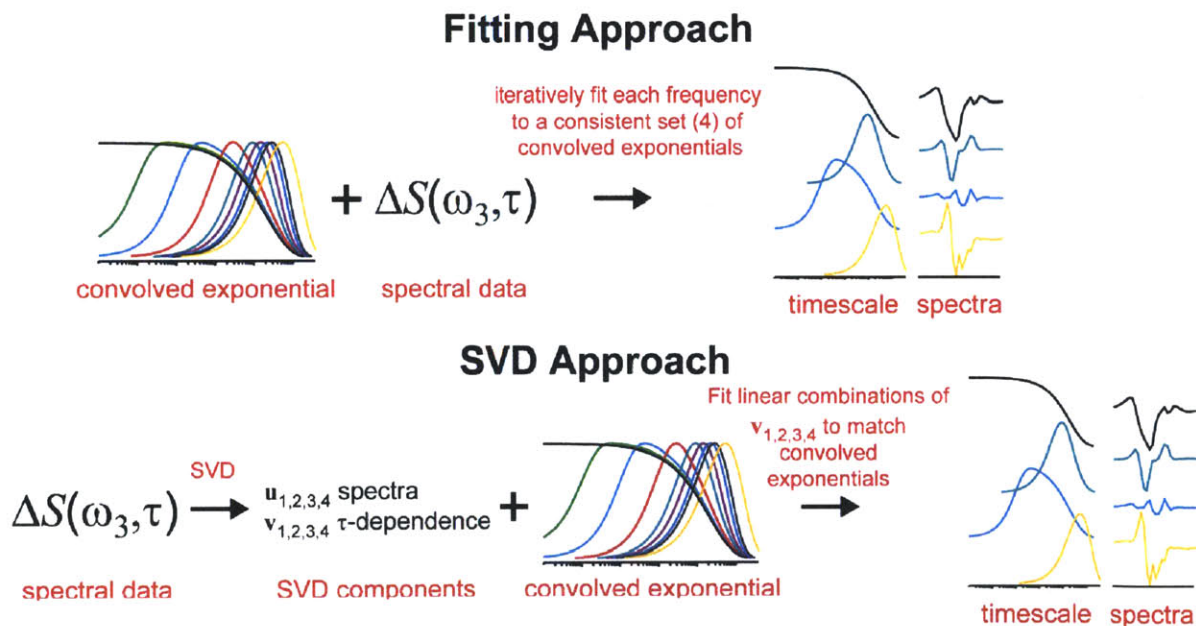


Figure 9.B.1: Illustration of two different approaches whose goal is to separate spectral changes based on their associated timescales. The *convolved exponentials* represent fitting the undistorted data to exponentials, and they are generated through convolution of mono-exponential decays with the temperature profile (Figure 9.5).

Fitting Approach (spectra separated by timescale behavior) by assessing full spectra rather than individual frequencies. Singular value decomposition is used to split the data into orthogonal spectra with associated orthogonal time-dependencies. SVD analysis results in many components due to the presence of noise. Thus, the results must be assessed and the true components must be separated from the noise (analogous to determining the number of exponentials needed in the *Fitting Approach*), sometimes using subjective criteria. As with analyzing individually frequencies, the mathematically generated SVD components may be a mixture of multiple exponential responses whose linear combination may produce what appears to be an intermediate exponential due to experimental fitting limitations, such as noise and the limited sampling of time points. In order to separate the underlying behavior, we take linear combinations of the SVD results in order to match the *convolved exponentials*, which, again, is equivalent to fitting undistorted data to simple multi-exponentials.

We use the *SVD Approach* to separate the spectral responses based on timescale because SVD works on the full spectrum rather than individual frequencies and SVD provides a better way of assessing the number of underlying timescales.

SVD was performed on each data set individually. Without considering the consistency of the associated spectra, generally, the observed exponential timescales fall into four categories: pulse-width limited (tracks the temperature profile, which exhibits a stretched exponential), fast (exponential time constant of 5-150 μ s), intermediate (250-1000 μ s), and long (10-30 ms). Individual data sets usually only require three components (three timescales) to recreate the data within the noise level, but the appearance of different timescales in other data sets (of similar concentration and/or temperature) suggests we should be flexible and open-minded about keeping or rejecting the fourth component. This fourth component is assessed based on continuity of \mathbf{v}_4 and consistency of the time trace and spectra across temperature and concentration data sets. That is to say, if independent analysis of multiple data sets results in a similar fourth component, the consistency suggests that the component is not the result of noise. Continuity can be assessed by calculating the shifted autocorrelation:

$$I_{AC} = \sum_{i=1}^{z-1} \mathbf{v}_j(\tau_i) \mathbf{v}_j(\tau_{i+1}) \quad (9.B.1)$$

where I_{AC} is the intensity of the shifted autocorrelation and z is the number of discrete τ_i points *sampled*. Components with an $I_{AC} < 0.4$ were not kept. Some components with $I_{AC} > 0.4$ were also

merged with others (the sum of two was kept while the difference of two was rejected) if the two \mathbf{v} traces and \mathbf{u} spectra were clearly similar and only differed in the sign of noise spikes at early time.

Once the rejected components are removed based on the above described criteria, we then take the linear combination of the retained components to produce a variety of time-dependencies that correspond to the *convolved exponentials* (Chapter 3 or SI). An example is shown in Figure 9.B.2, Figure 9.B.3, and Figure 9.12. The process of taking linear combinations of the SVD components is repeated independently for each data set. Most data sets only have three timescales that can be separated from the noise, although some show four and some two.

The initial, mathematical SVD components are labeled as 1-4. The amplitude of $1 > 2 > 3 \sim 4$, as shown in Figure 9.B.3. The 4 (or fewer) components were mixed to result in the *exponential basis* (exponential time dependence associated with spectra), which are labeled with a-d. a and b have similar, large intensity. c and d have similar, small intensity. Starting with \mathbf{u}_{1-4} and \mathbf{v}_{1-4} , the steps taken for transformation from 1-4 to a - d are:

- 1) a : Combine weighted sums of \mathbf{v}_{1-3} to give a v_a that matches the temperature-profile. Combine weighted sums of \mathbf{u}_{1-3} to give a u_a that matches the sum of spectra over the range 10 – 560 ns. These ensure that the *kinetic phase a* tracks the temperature and matches the early time spectra, as expected for the pulse-width limited response.
- 2) b : Based on its large intensity relative to components 3 and 4, and because it matches a *convolved exponential*, Set $v_b = \mathbf{v}_2$. Fit v_b to a bi-exponential of the form $F_b = C [\exp(-\lambda_{b1}\tau) - \exp(-\lambda_{b2}\tau)]$. This fitting is used in the next step. v_b corresponds to the intermediate time exponential *kinetic phase* (250-1000 μ s).
- 3) c : Fit \mathbf{v}_3 to the form $C [\exp(-\lambda_{b1}\tau) - \exp(-\lambda_{b2}\tau)] + F_b$. Subtract F_b to try to remove the mixed-in v_b component from \mathbf{v}_3 . Subtract a scaled \mathbf{v}_4 from the remainder to produce a v_c that matches a fast time *convolved exponential*. Thus, v_c corresponds to the fast *kinetic phase* (5-150 μ s).
- 4) d : If necessary, subtract residual scaled v_b from \mathbf{v}_4 such that the resulting v_d matches a long time (10-30 ms) *convolved exponential*.

Figure 9.B.2 shows the intermediate spectra that result from steps 1), 2), and the subtraction of v_b in 3). As is shown in Figure 9.B.2, the ' \mathbf{v}_3, v_b ' time traces vary with concentration and appear to be a mixture of multiple timescales. The \mathbf{v}_4 traces seem to have a complementary variation.

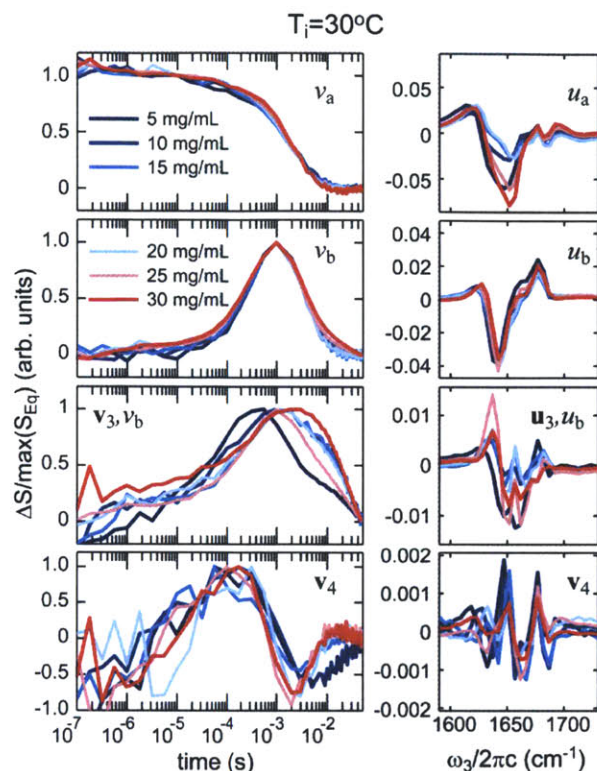


Figure 9.B.2: SVD analysis of the $T_1 = 30^\circ\text{C}$ ZZZZ t-DPP spectra results in time traces (left) and spectra (right). Components 1, 2, and 3 are mixed to create components a, b, and 3/b through the following criteria: *i*) v_a matches the temperature profile from the transient absorption data, *ii*) u_a matches the 10 ns difference spectrum, *iii*) the v_3 component is fit to the sum of v_b and a bi-exponential, and then v_b is scaled based on the fit and subtracted from v_3 to give ' v_3, v_b '. As is shown here, the ' v_3, v_b ' time traces vary with concentration and appear to be a mixture of multiple timescales. The v_4 traces seem to have a complementary variation. The traces are further mixed and the results are shown in Figure 9.12.

Applying steps 1-4 results in the *kinetic* phases shown in Figure 9.12. From the above steps, each component is sequentially the focus of attention. Aside from *a*, the transformation was done by looking at the time dependence. Defining all four time traces also transforms the associated *u* spectra. Due to the noisy nature of some of the data and difficulty in consistently automating the process for all data sets, the *c* and *d* mixing was done by hand, and is therefore more qualitative (and therefore has higher error bars) than *a* and *b*.

After making these transformations, we can assess the reasonableness of our initial selection to use exponentials. We observe that the separated timescales match the *convolved exponentials*, which indicates that data un-distorted by the temperature-relaxation profile would

fit to multi-exponentials (Figure 9.B.3). Thus, the results of the above separation of the data further justify the means. The observed timescales mirror those observed in the unseparated data, and they correspond to the signals we expect based on the convolution modeling. Although

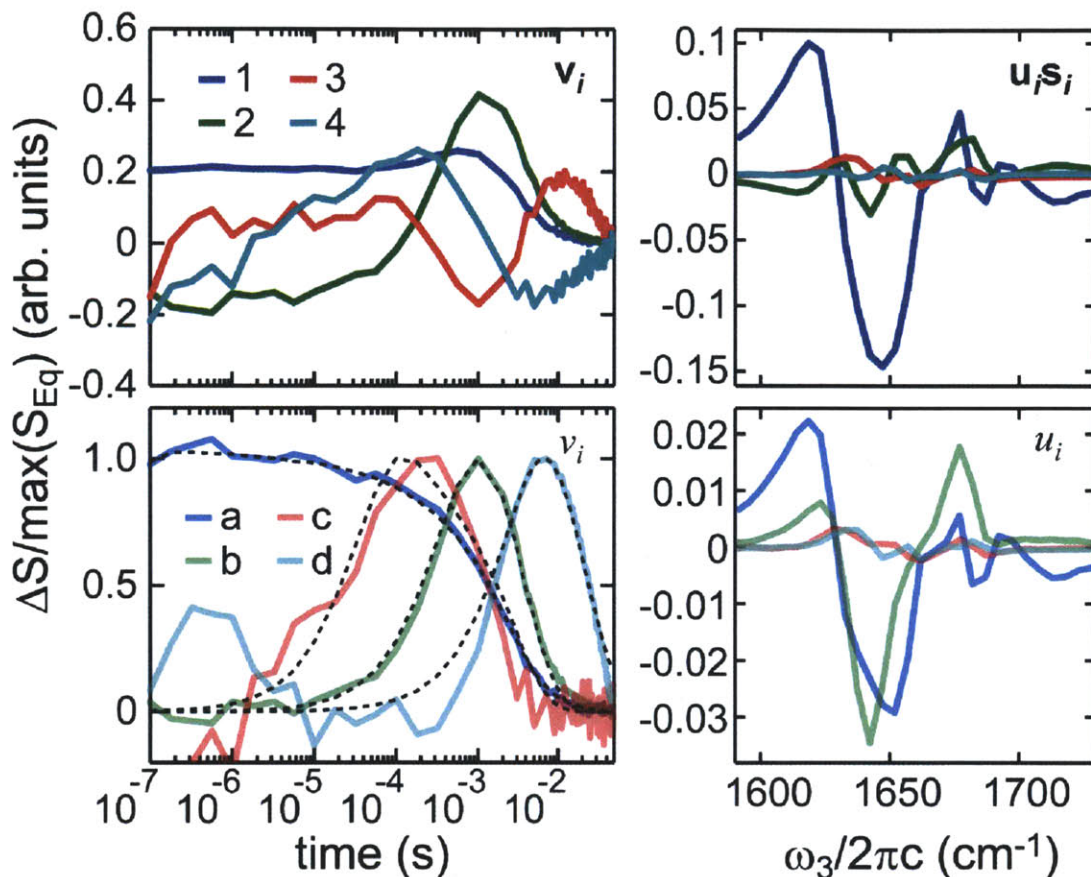


Figure 9.B.3: SVD analysis of the 10 mg/mL insulin, $T_i = 30^\circ\text{C}$ ZZZZ t-DPP spectra shown in Figure 9.10 results in time traces (top left) and spectra (top right). Linear combinations of the 1,2,3,4 components results in the traces and spectra shown at bottom. The black dashed lines indicate the signals predicted by convolution of the temperature profile with exponential responses. For this and all other representations of separated time and spectral features, the product of the time dependence and spectral amplitude result in the $\Delta S/\max(S_{\text{Eq}})$ units. That is, if $v_i(\tau)=1$ and $u_i(\omega)=0.02$, then the observed difference signal (normalized the maximum of the equilibrium spectrum) at that (τ,ω) point is $1 \times 0.02 = 0.02$, or 2% change.

each data set is processed independently, there is consistency with the results – pulse-width-limited, fast, intermediate, and slow timescales are observed and the spectra associated with each are consistent across data sets. Through SVD analysis and subsequent re-mixing, we separate the spectra associated with each exponential timescale. The u_{a-d} spectra represent the *changes* in

spectra associated with the v_{a-d} time-traces. Thus, the u_{a-d} spectra are difference spectra.

Although the exponentiality of the data was assumed in the above separation, qualitatively the results are somewhat independent of the functional form of the time-dependent relaxation. First, the temperature tracking *kinetic phase a* is universally a pulse-width-limited response that is independent of assumed functional behavior. Similarly, the *d* phase is separated from *b* and *c* by virtue of being longer than the temperature relaxation. Generally *b* and *c* are fast enough that their decay to zero is defined by the temperature relaxation profile. *d* is so long that its rise is dominated by the temperature relaxation profile. Thus, the ~ 3 ms timescale of the temperature relaxation sets a separating barrier for distinguishing *b/c* from *d*. Convolution of any reasonable (continuous) time behavior with the temperature-relaxation profile will create this separating barrier. For *b* and *c*, the raw data indicates at least two time behaviors on the 5-1000 μ s timescale (see frequency traces in Figure 9.11). Generally, these would have been separated into some early and long time behavior regardless of whether the fitting function was a stretched exponential or similar function. Thus, we believe that the data would robustly tend towards separation into generic pulse-width-limited, fast, intermediate, and slow *kinetic phases* regardless of variation in the functional form.

Appendix 9.C: Fitting a two-state model

Based on the equations given in Appendix 9.A, we can attempt to determine the k_1 (association) and k_{-1} (dissociation) rates based on the observed rates (λ) and equilibrium-measured dissociation constant. Here, we ignore the a , c , and d behavior and assume b describes the dissociation process. The results of the fitting are shown in Figure 9.C.1. Fitting to the dissociation constant has a large effect on the fits. The $T_f = 40^\circ\text{C}$ final fit is poor, but inspection of Figure 9.13 indicates that perhaps the three higher concentration points are too low, as one can expect they should lie between the $T_f = 35$ and 45°C points.

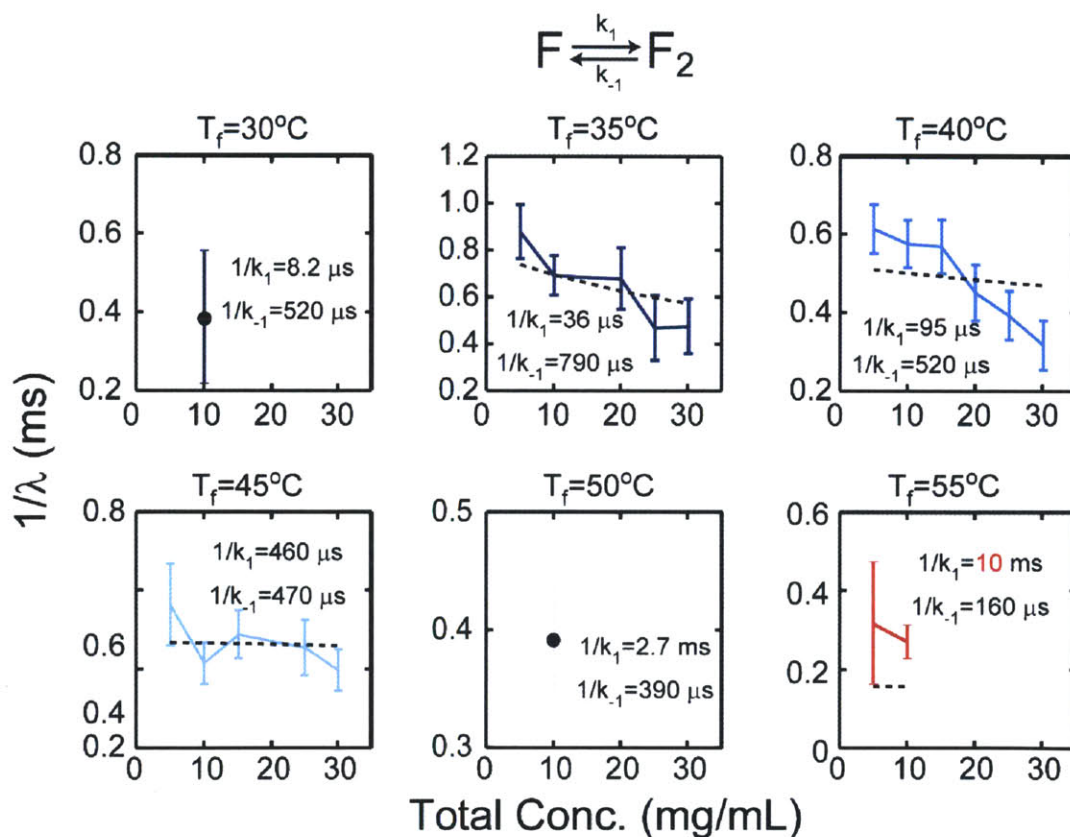


Figure 9.C.1: The de-convolved observed rates (solid colored lines) from the b process. The 95% confidence intervals are plotted as error bars. The dashed lines indicate the expected concentration dependence from the fit (inset) forward and reverse rates.

The temperature dependence of the rates is continuous. The dissociation rate increases with temperature, while the association rate decreases. Plotting the logarithm of the rates versus

inverse temperature allows us to inspect the activation barrier, as is shown in Figure 9.C.2. The negative enthalpy (decreasing rate with increasing temperature) for the association rate, k_1 , is unusual, but not necessarily unexpected. It can be rationalized given that the increase in temperature will increase the kinetic energy of the monomers such that favorable encounters that lead to association are rarer. It may also indicate a transient intermediate⁷⁰⁻⁷².

There is a relationship between the amplitude of changes expected due to a change in

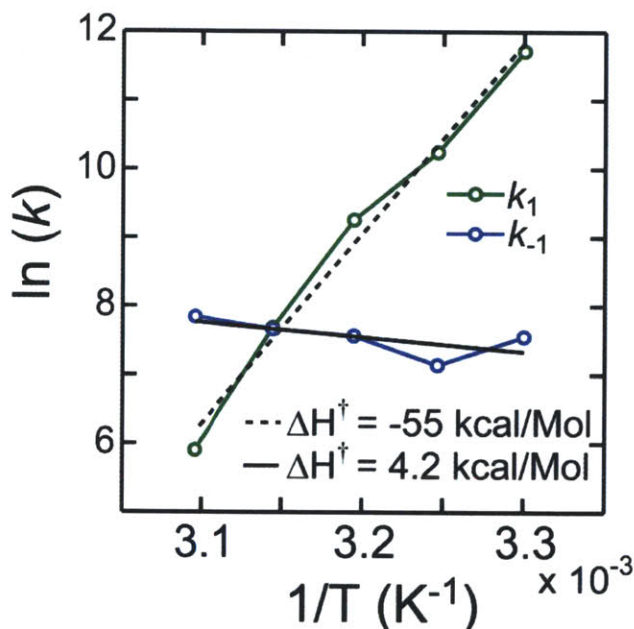


Figure 9.C.2: The logarithm of the rates determined from the two-state b timescale fit is plotted versus inverse temperature. The slope of the resulting line normally corresponds to the activation energy.

temperature and the relative temperature-dependent forward and reverse rates. That is, once the rates are calculated, we can calculate the populations at different temperatures. This allows us to predict the concentration dependence. Written out explicitly:

$$\begin{aligned} [F](t) &= [F]_{eq} + [f](t) = [F]_{eq} + [f_o] \exp(-\lambda t) \\ [F_2](t) &= [F_2]_{eq} + [f_2](t) = [F_2]_{eq} - 0.5[f_o] \exp(-\lambda t) \end{aligned} \quad (9.C.1)$$

This can be re-written in matrix form as:

$$\begin{pmatrix} [F](t_1) & \cdots & [F](t_z) \\ [F_2](t_1) & \cdots & [F_2](t_z) \end{pmatrix} = \begin{pmatrix} [f_o] & [F]_{eq} \\ -0.5[f_o] & [F_2]_{eq} \end{pmatrix} \begin{pmatrix} \exp(-\lambda t_1) & \cdots & \exp(-\lambda t_z) \\ 1 & \cdots & 1 \end{pmatrix} \quad (9.C.2)$$

where there are z discrete time points sampled in the data. Thus, Z contains columns related to the concentration dependence of the two species, X contains factors, and L contains a column

each with exponential time-dependence and constant time-dependence.

In the *state basis* the data, $S(\omega, t)$, can be written as the product of the column spectra associated with each state $\Sigma(\omega, \text{state}) = \{\sigma_{F1}, \sigma_{F2}\}$ and the concentration matrix:

$$S = \Sigma Z^T \quad (9.C.3)$$

In the *exponential basis* we separate the signal into spectra associated with exponential time dependence:

$$S = \Omega L^T \quad (9.C.4)$$

$$S = \begin{pmatrix} u_b(\omega_1) & S_{eq}(\omega_1) \\ \vdots & \vdots \\ u_b(\omega_x) & S_{eq}(\omega_x) \end{pmatrix} \begin{pmatrix} \exp(-\lambda t_1) & \cdots & \exp(-\lambda t_n) \\ 1 & \cdots & 1 \end{pmatrix}$$

where u_b is the spectrum associated with the λ exponential relaxation, S_{eq} is the equilibrium spectrum, and there are x discrete frequencies.

Equating eq. (9.C.3) and (9.C.4) followed by substitution of eq. (9.C.2) leads to:

$$\begin{aligned} \Omega L^T &= \Sigma Z^T \\ &= \Sigma X L^T \\ \Omega &= \Sigma X \end{aligned} \quad (9.C.5)$$

X is invertible, which means that the *state basis* spectra can be calculated through:

$$\Sigma = \Omega X^{-1} \quad (9.C.6)$$

while Z can be calculated through eq. (9.C.2). Thus, we can calculate the spectra associated with each state if we know 1) the *exponential basis* spectra and time-traces and 2) the X matrix. The X matrix depends on the kinetic model. For both two and three-state models, the analogous $[f_0]$ factor must be known – the population change of state $[F]$ when the temperature changes from T_i to T_f . Thus, we must know the rates (and therefore populations) at both initial and final temperatures; each T-jump measurement provides the rates at T_f . Inconsistencies across different temperatures (as well as generally noisy rate constant determination) in the three-state modeling results prevent this state spectrum calculation, as the analogous $[f_0]$ factors are unphysical and erroneous.

Because the protein response is convolved with the temperature profile, we must adjust our L matrix. First, the amplitudes are adjusted by $A(\lambda)$ a λ -dependent factor – the observed amplitudes sometimes do not reach their full values because the temperature return prematurely causes them to decay. $A(\lambda)$ is drawn from a mapping from model calculations (Chapter 3). Second, the single exponential is replaced by a rise and decay exponential, as described in

Appendix 9.A. After adjustment for our experiment, the L matrix is:

$$L = \begin{pmatrix} A(\lambda) [\exp(-\lambda_{rise} t_1) - \exp(-\lambda_{decay} t_1)] & \cdots & A(\lambda) [\exp(-\lambda_{rise} t_n) - \exp(-\lambda_{decay} t_n)] \\ 1 & \cdots & 1 \end{pmatrix} \quad (9.C.7)$$

Figure 9.C.3 shows the results of transforming from the *exponential basis* to the *state basis* using the b component spectra and timescales and the rates calculated in Figure 9.C.1. The dashed concentration dependent lines indicate the concentration changes expected based on the calculated rate constants (in the absence of the temperature re-equilibrating). The data was transformed as described in the above equations to match these calculated concentration time-dependence. The results of the transformation are shown in the solid lines. Thus, the data set (ignoring components a , c , and d) can be described with the center spectra scaled by the

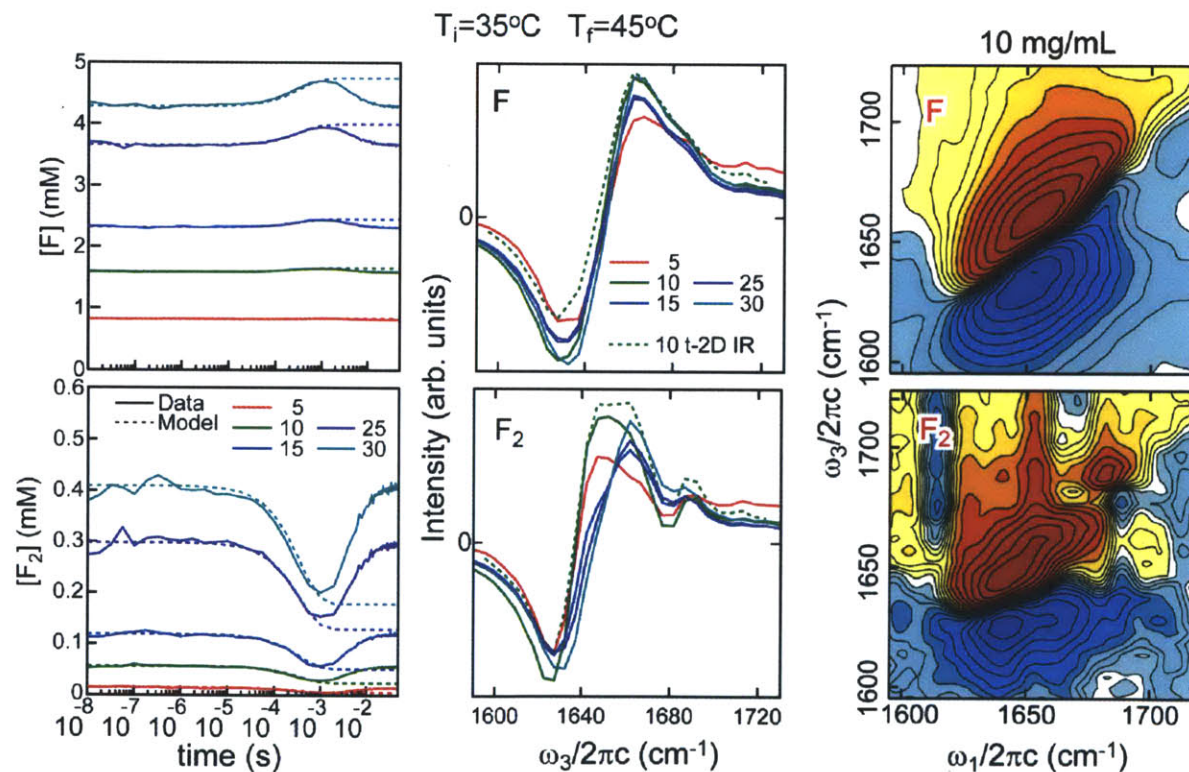


Figure 9.C.3: The calculated time-dependent concentrations (left), ZZZZ DPP spectra (center), and ZZZZ 2D IR spectra (right) for the F and F_2 states. Different colors indicate different concentration data sets (in mg/mL). The dashed lines at left indicate the calculated concentration dependence based on the calculated rates *without* the decay ($\lambda_{decay} = 0$) to indicate the concentration change expected in the absence of the temperature re-equilibration. 2D IR contours are plotted as described in Figure 9.8.

concentrations at left. The calculated spectra are fairly consistent across different concentration data sets. The amplitude of the F_2 spectrum is $\sim 2x$ the amplitude of the F spectrum. As expected, the F_2 spectrum has strong ν_{\perp} and ν_{\parallel} features characteristic of the interfacial β -sheet. The F spectrum has less pronounced features characteristic of α -helical structures.

Kevin C. Jones

275 Putnam Avenue, Apt 2
Cambridge, MA 02139

Phone: 857-998-0272
Email: kcj@mit.edu

Education

Massachusetts Institute of Technology Cambridge, MA
Ph.D., Physical Chemistry (Course 5), June 2012
Thesis: Temperature-Jump 2D IR Spectroscopy to Study Protein Conformational Dynamics

Yale University New Haven, CT
Bachelor of Science, May 2005, Magna Cum Laude
Major: Chemistry

Teaching

Teaching Assistant at MIT Cambridge, MA
5.35 *Module 1: Fundamentals of Spectroscopy*, Fall 2009
Laboratory TA - 30 students
5.33 *Advanced Chemical Experimentation and Instrumentation*, Fall 2005, 2006, 2007
Laboratory TA - 30 students
5.60 *Thermodynamics and Kinetics*, Spring 2006
Head Recitation TA - 40 students

Honors/Awards

Morse Travel Grant, MIT 2011
Magna Cum Laude, Yale University 2005

Presentations

Modern Optics and Spectroscopy Seminar (MIT), 2012 *invited talk*
Time Resolved Vibrational Spectroscopy XIV and XV, 2009, 2011 *poster, talk*
American Physical Society, March Meeting, 2011 *talk*
Vibrational Gordon Research Conference, 2006, 2008, 2010 *poster, poster, poster*
Biophysical Society, 53rd Annual Meeting, 2009 *poster*

Publications

1. C. R. Baiz, C. S. Peng, M. E. Reppert, **K. C. Jones**, and A. Tokmakoff, "Coherent two-dimensional infrared spectroscopy: Quantitative analysis of protein secondary structure in solution," *Analyst* **137**, 1793-1799 (2012).
2. **K. C. Jones**, Z. Ganim, C. S. Peng, and A. Tokmakoff, "Transient two-dimensional spectroscopy with linear absorption corrections applied to temperature-jump two-dimensional infrared," *J. Opt. Soc. Am. B* **29**, 118-129 (2012).
3. C. S. Peng, **K. C. Jones**, and A. Tokmakoff, "Anharmonic Vibrational Modes of Nucleic Acid Bases Revealed by 2D IR Spectroscopy," *J. Am. Chem. Soc.* **133**, 15650-15660 (2011).
4. Z. Ganim, **K. C. Jones**, and A. Tokmakoff, "Insulin dimer dissociation and unfolding revealed by amide I two-dimensional infrared spectroscopy," *Phys. Chem. Chem. Phys.* **12**, 3579-3588 (2010).
5. **K. C. Jones**, Z. Ganim, and A. Tokmakoff, "Heterodyne-detected dispersed vibrational echo spectroscopy," *J. Phys. Chem. A* **113**, 14060-14066 (2009).

Kevin C. Jones

275 Putnam Avenue, Apt 2
Cambridge, MA 02139

Phone: 857-998-0272
Email: kcj@mit.edu

6. Z. Ganim, H. S. Chung, A. W. Smith, L. P. DeFlores, **K. C. Jones**, and A. Tokmakoff, "Amide I Two-Dimensional Infrared Spectroscopy of Proteins," *Acc. Chem. Res.* **41**, 432-441 (2008).
7. M. F. DeCamp, L. P. DeFlores, **K. C. Jones**, and A. Tokmakoff, "Single-shot two-dimensional infrared spectroscopy," *Optics Express* **15**, 233-241 (2007).
8. H. S. Chung, Z. Ganim, **K. C. Jones**, and A. Tokmakoff, "Transient 2D IR spectroscopy of ubiquitin unfolding dynamics," *Proc. Natl. Acad. Sci. U. S. A.*, 14237-14242 (2007).

Patents

C. R. Baiz, **K. C. Jones**, A. Tokmakoff, "Vibrational Spectroscopy for Quantitative Measurement of Analytes" (Patent Pending, 2012)

Peer Counseling

Chemistry Resources for Easing Friction and Stress
Confidential Peer Counselor at MIT, 2008-2012

Cambridge, MA

Other Activities

Chemistry Department Summer Volleyball League, Captain
Propagators (2008), Spiketroscopists (2009, 2010)

Chemistry Department IM Soccer team, Captain
Buckyballers (Spring 2007, Fall 2010)